Principles of Quantum Communication Theory: A Modern Approach

Sumeet Khatri, Ludovico Lami, and Mark M. Wilde

October 6, 2024

Preface

[IN PROGRESS]

Acknowledgements

[IN PROGRESS]

We dedicate this book to the memory of Jonathan P. Dowling. Jon was generous and kind-hearted, and he always gave all of his students his full, unwavering support. His tremendous impact on the lives of everyone who met him will ensure that his memory lives on and that he will not be forgotten. We will especially remember Jon's humour and his sharp wit. We are sure that, as he had promised, this book would have made the perfect doorstop for his office.

Sumeet Khatri acknowledges support from the National Science Foundation under Grant No. 1714215 and the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship. Mark M. Wilde acknowledges support from the National Science Foundation over the past decade (specifically from Grant Nos. 1350397, 1714215, 1907615, 2014010), and is indebted and grateful to Patrick Hayden for hosting him for a sabbatical at Stanford University during calendar year 2020, with support from Stanford QFARM and AFOSR (FA9550-19-1-0369).

Table of Contents

Chapter 1

Introduction

[IN PROGRESS]

Part I

Preliminaries

Before starting our journey though quantum communication protocols, it is necessary for us to learn about and understand the various mathematical and physical concepts involved in their construction and analysis. To this end, we begin in Chapter 2 by providing an overview of the mathematics required for understanding quantum communication protocols, and quantum information more broadly. Then, in Chapters 3–6, we study the basic axioms of quantum mechanics, including quantum states and measurements (Chapter 3); followed quantum channels, with the general theory and many examples (Chapter 4); followed by fundamental quantum information processing tasks, such as teleportation, super-dense coding, and hypothesis testing (Chapter 5); and then distinguishability measures for states and channels, such as fidelity, trace distance, and diamond distance (Chapter 6). Entropies and entanglement measures are crucial in quantifying the performance of quantum communication protocols, but they are also interesting in their own right, and they have applications in other areas of mathematical physics. In Chapters 7–9, we study these quantities in detail.

Chapter 2

Mathematical Tools

In this chapter, we learn about the various mathematical concepts required for the analysis of quantum communication protocols. We mostly provide a summary of the main definitions and results needed in later chapters, and we omit several of the proofs. For further details on the concepts presented here, as well as for proofs not explicity given here, please consult the Bibliographic Notes (Section 2.6) at the end of the chapter.

Linear algebra forms the core mathematical foundation of quantum information theory for finite-dimensional quantum systems, and thus it is worthwhile for us to start by reviewing the basics of linear algebra, with an emphasis on linear operators. We then proceed to give a summary of several relevant definitions and results in real and convex analysis, probability theory, and semi-definite programming. Concepts from real analysis play an important role in quantum information theory. Indeed, as we discover later, the capacity of a quantum channel is defined as a limit, which is a core notion in real analysis. Convexity plays a prominent role as well. Not only is the set of quantum states a convex set, but also the operator Jensen inequality, a foundational statement about operator convex functions, is a fundamental inequality that leads to various quantum data-processing inequalities. The latter data-processing principle is one of the central tenets of quantum information that allows for placing limitations on the communication capacities of quantum channels. Probability theory is essential as well, due to the probabilistic nature of quantum mechanics and the inevitable and unpredictable errors that occur when communicating information over quantum channels. Finally, semi-definite programming is a remarkably useful tool, not only as an analytical tool but also for numerically calculating relevant quantities of interest. Semi-definite programming has also played a pivotal role in many of the substantive advances that have taken place in quantum information theory during the past several decades, and so it has become one of the standard tools in the quantum information theorist's toolkit.

2.1 Finite-Dimensional Hilbert Spaces

The primary mathematical object in quantum theory is the Hilbert space. We consider only finite-dimensional Hilbert spaces, denoted by $\mathcal{H}$, throughout this book, and we use $\dim(\mathcal{H})$ to denote the dimension of $\mathcal{H}$. Although we consider finite-dimensional spaces exclusively in this book, we note here that many of the statements and claims extend directly to the case of separable, infinite-dimensional Hilbert spaces, especially for operationally-defined tasks and information quantities. However, we do not delve into these details.

A d-dimensional Hilbert space ($1 \leq d < \infty$) is defined to be a complex vector space equipped with an inner product[1]. We use the notation $|\psi\rangle$ to denote a vector in $\mathcal{H}$. An inner product is a function $\langle\cdot|\cdot\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that satisfies the following properties:

- *Non-negativity*: $\langle\psi|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$, and $\langle\psi|\psi\rangle = 0$ if and only if $|\psi\rangle = 0$.
- *Conjugate bilinearity*: For all $|\psi_1\rangle, |\psi_2\rangle, |\phi_1\rangle, |\phi_2\rangle \in \mathcal{H}$ and $\alpha_1, \beta_1, \alpha_2, \beta_2 \in \mathbb{C}$,

$$\begin{aligned}\langle\alpha_1\psi_1 + \beta_1\phi_1|\alpha_2\psi_2 + \beta_2\phi_2\rangle = \overline{\alpha_1}\alpha_2\langle\psi_1|\psi_2\rangle + \overline{\alpha_1}\beta_2\langle\psi_1|\phi_2\rangle \\ + \overline{\beta_1}\alpha_2\langle\phi_1|\psi_2\rangle + \overline{\beta_1}\beta_2\langle\phi_1|\phi_2\rangle. \qquad (2.1.1)\end{aligned}$$

- *Conjugate symmetry*: $\langle\psi|\phi\rangle = \overline{\langle\phi|\psi\rangle}$ for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}$.

In the above, $\overline{\alpha}$ denotes the complex conjugate of $\alpha \in \mathbb{C}$. Throughout this book, the term "Hilbert space" always refers to a finite-dimensional Hilbert space.

All d-dimensional Hilbert spaces are isomorphic to the vector space $\mathbb{C}^d$ equipped with the Euclidean inner product. By two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ being

[1] This definition suffices in the finite-dimensional case. More generally, a Hilbert space is a complete inner product space; please consult the Bibliographic Notes (Section 2.6).

isomorphic, we mean that there is a bijective linear mapping $U : \mathcal{H} \to \mathcal{H}'$ such that

$$\langle U\varphi|U\psi\rangle = \langle\varphi|\psi\rangle, \tag{2.1.2}$$

for all $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$, and U is called an isomorphism. For the finite-dimensional case of interest for us, U is a unitary operator (discussed in more detail in Section 2.2.6). Note that $\mathbb{C}^d$ is the vector space of d-dimensional column vectors with elements in $\mathbb{C}$. We let $\{|i\rangle\}_{i=0}^{d-1}$ denote an orthonormal basis, called the *standard basis* or *computational basis*, for the Hilbert space with respect to the Euclidean inner product. The vector $|i\rangle$ is defined to be a column vector with its $(i+1)^{\text{th}}$ entry equal to one and all others equal to zero, so that

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad , \quad |d-1\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{2.1.3}$$

The inner product $\langle i|j\rangle$ evaluates to $\langle i|j\rangle = \delta_{i,j}$ for all $i, j \in \{0, \dots, d-1\}$, where the Kronecker delta function is defined as

$$\delta_{i,j} \coloneqq \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \tag{2.1.4}$$

More generally, for two vectors $|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i |i\rangle$ and $|\phi\rangle = \sum_{i=0}^{d-1} \beta_i |i\rangle$, with $\alpha_i = \langle i|\psi\rangle \in \mathbb{C}$ and $\beta_i = \langle i|\phi\rangle \in \mathbb{C}$ being the respective components of $|\psi\rangle$ and $|\phi\rangle$ in the standard basis, the inner product $\langle\psi|\phi\rangle$ is defined as

$$\langle\psi|\phi\rangle \coloneqq \sum_{i=0}^{d-1} \overline{\alpha_i}\beta_i. \tag{2.1.5}$$

The Euclidean norm of a vector $|\psi\rangle \in \mathcal{H}$, denoted by $\||\psi\rangle\|_2$, is the norm induced by the inner product, i.e.,

$$\||\psi\rangle\|_2 \coloneqq \sqrt{\langle\psi|\psi\rangle}. \tag{2.1.6}$$

The *Cauchy–Schwarz inequality* establishes an upper bound on the modulus of the inner product of two vectors $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ in terms of the product of their norms:

$$|\langle\psi|\phi\rangle|^2 \leq \langle\psi|\psi\rangle \cdot \langle\phi|\phi\rangle = \||\psi\rangle\|_2^2 \cdot \||\phi\rangle\|_2^2, \tag{2.1.7}$$

with equality if and only if $|\phi\rangle = \alpha|\psi\rangle$ for some $\alpha \in \mathbb{C}$.

Given a vector $|\psi\rangle \in \mathcal{H}$, its *dual vector*, denoted by $\langle\psi|$, is defined to be a linear functional from $\mathcal{H}$ to $\mathbb{C}$ such that $\langle\psi|(|\phi\rangle) = \langle\psi|\phi\rangle$ for all $|\phi\rangle \in \mathcal{H}$. If $|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i|i\rangle$, then $\langle\psi| = \sum_{i=0}^{d-1} \overline{\alpha_i}\langle i|$, where $\langle i|$ can be interpreted, based on (2.1.3), as a row vector with its $(i+1)^{\text{th}}$ entry equal to one and all other entries equal to zero; i.e., $\langle i| = (|i\rangle)^{\mathsf{T}}$, where $(\cdot)^{\mathsf{T}}$ denotes the matrix transpose.

The tensor product of vectors, operators, and Hilbert spaces plays an important role in quantum theory. For example, it is used to describe the state of multiple quantum systems. For two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with dimensions d_A and d_B, respectively, along with associated orthonormal bases $\{|i\rangle_A\}_{i=0}^{d_A-1}$ and $\{|j\rangle_B\}_{j=0}^{d_B-1}$, the tensor product vector $|i\rangle_A \otimes |j\rangle_B$ is a vector in a $(d_A d_B)$-dimensional Hilbert space with a one in its $(i \cdot d_B + j + 1)^{\text{th}}$ entry and zeros elsewhere. Notice here that we have employed the labels A and B in order to keep track of the Hilbert spaces of the vectors in the tensor product. Later on, when we move to the study of quantum information, we will see that the label A can be associated to a quantum system in possession of "Alice" and the label B can be associated to a quantum system in possession of "Bob." As an example of the tensor-product vector $|i\rangle_A \otimes |j\rangle_B$, if $d_A = 2$, $d_B = 3$, $i = 0$, and $j = 2$, then

$$|i\rangle \otimes |j\rangle = |0\rangle \otimes |2\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ 0 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.1.8}$$

More generally, for vectors $|\psi\rangle_A = \sum_{i=0}^{d_A-1} \alpha_i|i\rangle_A$ and $|\phi\rangle_B = \sum_{j=0}^{d_B-1} \beta_j|j\rangle_B$, the tensor-product vector $|\psi\rangle_A \otimes |\phi\rangle_B$ is given by

$$|\psi\rangle_A \otimes |\phi\rangle_B = \sum_{i=0}^{d_A-1} \alpha_i|i\rangle_A \otimes |\phi\rangle_B \tag{2.1.9}$$

$$= \sum_{i=0}^{d_A-1} \sum_{j=0}^{d_B-1} \alpha_i\beta_j|i\rangle_A \otimes |j\rangle_B. \tag{2.1.10}$$

As an example with $d_A = 2$ and $d_B = 3$, we find that $|\psi\rangle_A \otimes |\phi\rangle_B$ can be calculated by a generalization of the "stack-and-multiply" procedure used in (2.1.8):

$$|\psi\rangle_A \otimes |\phi\rangle_B = \begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} \otimes \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix}\alpha_0 \cdot \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}\\ \alpha_1 \cdot \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}\end{pmatrix} = \begin{pmatrix}\alpha_0\beta_0\\ \alpha_0\beta_1\\ \alpha_0\beta_2\\ \alpha_1\beta_0\\ \alpha_1\beta_1\\ \alpha_1\beta_2\end{pmatrix}. \quad (2.1.11)$$

The tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is defined to be the Hilbert space spanned by the vectors $|i\rangle_A \otimes |j\rangle_B$ defined above:

$$\mathcal{H}_A \otimes \mathcal{H}_B := \text{span}\{|i\rangle_A \otimes |j\rangle_B : 0 \le i \le d_A - 1, 0 \le j \le d_B - 1\}. \quad (2.1.12)$$

The inner product on $\mathcal{H}_A \otimes \mathcal{H}_B$ is given by

$$(\langle i|_A \otimes \langle j|_B)(|i'\rangle_A \otimes |j'\rangle_B) = \langle i|i'\rangle\langle j|j'\rangle = \delta_{i,i'}\delta_{j,j'} \quad (2.1.13)$$

for all i, i', j, j' satisfying $0 \le i, i' \le d_A - 1$ and $0 \le j, j' \le d_B - 1$. The Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ consequently has dimension $d_A d_B$. We often use the notation $\mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$, as well as the abbreviation $|i, j\rangle_{AB} \equiv |i\rangle_A \otimes |j\rangle_B$. We often also use the notation $\mathcal{H}_{A^n} \equiv \mathcal{H}_A^{\otimes n}$ to refer to the n-fold tensor product of $\mathcal{H}_A$.

The *direct sum* of $\mathcal{H}_A$ and $\mathcal{H}_B$, denoted by $\mathcal{H}_A \oplus \mathcal{H}_B$, is defined to be the Hilbert space of vectors of the form $|\psi\rangle_A \oplus |\phi\rangle_B$, with $|\psi\rangle_A \in \mathcal{H}_A$ and $|\phi\rangle_B \in \mathcal{H}_B$, where

$$|\psi\rangle_A \oplus |\phi\rangle_B := \begin{pmatrix}|\psi\rangle_A\\ |\phi\rangle_B\end{pmatrix}. \quad (2.1.14)$$

In other words, $\mathcal{H}_A \oplus \mathcal{H}_B$ can be viewed as the Hilbert space of column vectors formed by stacking elements of the constituent Hilbert spaces. Observe that if $\mathcal{H}_A$ has the same dimension as $\mathcal{H}_B$, then we can write

$$|\psi\rangle_A \oplus |\phi\rangle_B = |0\rangle \otimes |\psi\rangle_A + |1\rangle \otimes |\phi\rangle_B, \quad (2.1.15)$$

where $\{|0\rangle, |1\rangle\}$ is the standard basis for a two-dimensional Hilbert space.

Exercise 2.1

Verify (2.1.15).

If $\{|i\rangle_A\}_{i=0}^{d_A-1}$ and $\{|j\rangle_B\}_{j=0}^{d_B-1}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, then

$$\left\{\begin{pmatrix} |i\rangle_A \\ 0 \end{pmatrix} : 0 \le i \le d_A - 1\right\} \cup \left\{\begin{pmatrix} 0 \\ |j\rangle_B \end{pmatrix} : 0 \le j \le d_B - 1\right\} \tag{2.1.16}$$

is an orthonormal basis for $\mathcal{H}_A \oplus \mathcal{H}_B$ under the inner product

$$(\langle i|_A \oplus \langle j|_B)(|i'\rangle_A \oplus |j'\rangle_B) = \langle i|i'\rangle + \langle j|j'\rangle = \delta_{i,i'} + \delta_{j,j'}. \tag{2.1.17}$$

for all $0 \le i, i' \le d_A - 1$ and $0 \le j, j' \le d_B - 1$. Consequently, $\mathcal{H}_A \oplus \mathcal{H}_B$ has dimension $d_A + d_B$. One of the simplest examples of a direct-sum Hilbert space is $\mathbb{C} \oplus \mathbb{C} = \mathbb{C}^2$. More generally, the d-fold direct sum $\mathbb{C}^{\oplus d}$ is equal to $\mathbb{C}^d$.

If $\mathcal{H}$ is a d-dimensional Hilbert space, then the k-fold direct sum $\mathcal{H}^{\oplus k}$ is a kd-dimensional Hilbert space. Consequently, it is isomorphic to $\mathbb{C}^k \otimes \mathcal{H}$, and the isomorphism is a generalization of the simple example presented in (2.1.15). Indeed, let $\mathcal{H}_X \equiv \mathbb{C}^k$, with orthonormal basis $\{|i\rangle_X\}_{i=0}^{k-1}$, and let $\mathcal{H}_A \equiv \mathcal{H}$, with orthonormal basis $\{|j\rangle_A\}_{j=0}^{d-1}$. We then have the correspondence

$$|i\rangle_X \otimes |j\rangle_A \leftrightarrow \begin{pmatrix} 0 \\ \vdots \\ 0 \\ |j\rangle_A \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.1.18}$$

holding for all $0 \le i \le k - 1$ and all $0 \le j \le d - 1$, where on the right-hand side there is a one in the $(i \cdot d + j + 1)^{\text{th}}$ entry of the column vector and zeros elsewhere. Then, for an element $|\psi_0\rangle_A \oplus |\psi_1\rangle_A \oplus \cdots \oplus |\psi_{k-1}\rangle_A \in \mathcal{H}_A^{\oplus k}$, we have

$$\begin{pmatrix} |\psi_0\rangle_A \\ |\psi_1\rangle_A \\ \vdots \\ |\psi_{k-1}\rangle_A \end{pmatrix} \leftrightarrow \sum_{i=0}^{k-1} |i\rangle_X \otimes |\psi_i\rangle_A. \tag{2.1.19}$$

The isomorphism between $\mathcal{H}^{\oplus k}$ and $\mathbb{C}^k \otimes \mathcal{H}$ given by (2.1.18) and (2.1.19) is relevant in the context of superpositions of quantum states and entanglement.

2.2 Linear Operators

Linear operators are relevant in quantum theory for describing states of quantum systems, as well as physical evolutions of the states, including measurements and unitary evolutions as special cases of general physical evolutions. Given a Hilbert space $\mathcal{H}_A$ with dimension d_A and a Hilbert space $\mathcal{H}_B$ with dimension d_B, a linear operator $X : \mathcal{H}_A \to \mathcal{H}_B$ is defined to be a function such that

$$X(\alpha|\psi\rangle_A + \beta|\phi\rangle_A) = \alpha X|\psi\rangle_A + \beta X|\phi\rangle_A \tag{2.2.1}$$

for all $\alpha, \beta \in \mathbb{C}$ and $|\psi\rangle_A, |\phi\rangle_A \in \mathcal{H}_A$. For clarity, we sometimes write $X_{A\to B}$ to explicitly indicate the input and output Hilbert spaces of the linear operator X.

Given two vectors $|\psi\rangle_A \in \mathcal{H}_A$ and $|\phi\rangle_B \in \mathcal{H}_B$, one can construct the associated linear operator from A to B given by $|\phi\rangle_B\langle\psi|_A$, which acts on an arbitrary input $|\chi\rangle_A \in \mathcal{H}_A$ as

$$(|\phi\rangle_B\langle\psi|_A)\,|\chi\rangle_A = |\phi\rangle_B\,\langle\psi|\chi\rangle. \tag{2.2.2}$$

We use $\mathbb{1}$ to denote the identity operator, which is defined as the unique linear operator such that $\mathbb{1}|\psi\rangle = |\psi\rangle$ for every vector $|\psi\rangle$. For clarity, when needed, we write $\mathbb{1}_d$ to indicate the identity operator acting on a d-dimensional Hilbert space.

Exercise 2.2

Given an orthonormal basis $\{|e_k\rangle\}_{k=1}^d$ for a d-dimensional Hilbert space, prove that

$$\mathbb{1}_d = \sum_{k=1}^{d} |e_k\rangle\!\langle e_k|. \tag{2.2.3}$$

We denote the set of all linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$ by $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$. If $\mathcal{H}_A = \mathcal{H}_B$, then $\mathrm{L}(\mathcal{H}_A) := \mathrm{L}(\mathcal{H}_A, \mathcal{H}_A)$, and we sometimes indicate the input Hilbert space $\mathcal{H}_A$ of $X \in \mathrm{L}(\mathcal{H}_A)$ by writing X_A. In particular, we often write X_{AB} when referring to linear operators in $\mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, i.e., when referring to linear operators acting on a tensor-product Hilbert space.

The set $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is itself a $d_A d_B$-dimensional vector space. The standard basis for $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is defined to be

$$\{|i\rangle_B\langle j|_A : 0 \le i \le d_B - 1,\ 0 \le j \le d_A - 1\}. \tag{2.2.4}$$

By applying (2.1.3), we see that the operator $|i\rangle_B\langle j|_A$ has a matrix representation as a $d_B \times d_A$ matrix with the $(i+1, j+1)^{\text{th}}$ entry equal to one and all other entries equal to zero, i.e.,

$$\begin{aligned} |0\rangle_B\langle 0|_A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad |0\rangle_B\langle 1|_A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \dots, \\ |d_B - 1\rangle_B\langle d_A - 1|_A = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \end{aligned} \tag{2.2.5}$$

Using this basis, we can write a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ as

$$X_{A\to B} = \sum_{i=0}^{d_B-1}\sum_{j=0}^{d_A-1} X_{i,j}|i\rangle_B\langle j|_A, \tag{2.2.6}$$

where $X_{i,j} \coloneqq \langle i|_B X|j\rangle_A$. This follows because

$$\begin{aligned} X_{A\to B} &= \mathbb{1}_B X_{A\to B}\mathbb{1}_A && (2.2.7) \\ &= \left(\sum_{i=0}^{d_B-1} |i\rangle\!\langle i|_B\right) X_{A\to B} \left(\sum_{j=0}^{d_A-1} |j\rangle\!\langle j|_A\right) && (2.2.8) \\ &= \sum_{i=0}^{d_B-1}\sum_{j=0}^{d_A-1} \langle i|_B X_{A\to B}|j\rangle_A |i\rangle_B\langle j|_A && (2.2.9) \\ &= \sum_{i=0}^{d_B-1}\sum_{j=0}^{d_A-1} X_{i,j}|i\rangle_B\langle j|_A. && (2.2.10) \end{aligned}$$

We can thus interpret a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ as a $d_B \times d_A$ matrix with the $(i+1, j+1)^{\text{th}}$ element equal to $X_{i,j} = \langle i|_B X|j\rangle_A$, where $0 \leq i \leq d_B - 1$ and $0 \leq j \leq d_A - 1$. For example, if $d_A = 2$ and $d_B = 3$, then

$$X = \begin{pmatrix} X_{0,0} & X_{0,1} \\ X_{1,0} & X_{1,1} \\ X_{2,0} & X_{2,1} \end{pmatrix}. \tag{2.2.11}$$

Exercise 2.3

Show that every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, expressed as in (2.2.6), can be written as

$$X_{A\to B} = \sum_{i=0}^{d_B-1} |i\rangle_B \langle\psi_i|_A = \sum_{j=0}^{d_A-1} |\phi_j\rangle_B \langle j|_A, \tag{2.2.12}$$

where $\{\langle\psi_i|_A\}_{i=0}^{d_B-1}$ and $\{|\phi_j\rangle_B\}_{j=0}^{d_A-1}$ are the rows and columns, respectively, of X.

2.2.1 Tensor Product

Given two linear operators $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ and $Y \in \mathrm{L}(\mathcal{H}_{A'}, \mathcal{H}_{B'})$, their tensor product $X \otimes Y$ is the unique linear operator in $\mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_{A'}, \mathcal{H}_B \otimes \mathcal{H}_{B'})$ such that

$$(X \otimes Y)(|\psi\rangle_A \otimes |\phi\rangle_{A'}) = X|\psi\rangle_A \otimes Y|\phi\rangle_{A'} \tag{2.2.13}$$

for all $|\psi\rangle_A \in \mathcal{H}_A$ and $|\phi\rangle_{A'} \in \mathcal{H}_{A'}$. The matrix representation of $X \otimes Y$ is the Kronecker product of the matrix representations of X and Y, which is a matrix generalization of the "stack-and-multiply" procedure from (2.1.11). For example, if $d_A = d_B = 2$ and $d_{A'} = d_{B'} = 3$, then

$$X \otimes Y = \begin{pmatrix} X_{0,0} & X_{0,1} \\ X_{1,0} & X_{1,1} \end{pmatrix} \otimes \begin{pmatrix} Y_{0,0} & Y_{0,1} & Y_{0,2} \\ Y_{1,0} & Y_{1,1} & Y_{1,2} \\ Y_{2,0} & Y_{2,1} & Y_{2,2} \end{pmatrix} \tag{2.2.14}$$

$$= \begin{pmatrix} X_{0,0} \cdot \begin{pmatrix} Y_{0,0} & Y_{0,1} & Y_{0,2} \\ Y_{1,0} & Y_{1,1} & Y_{1,2} \\ Y_{2,0} & Y_{2,1} & Y_{2,2} \end{pmatrix} & X_{0,1} \cdot \begin{pmatrix} Y_{0,0} & Y_{0,1} & Y_{0,2} \\ Y_{1,0} & Y_{1,1} & Y_{1,2} \\ Y_{2,0} & Y_{2,1} & Y_{2,2} \end{pmatrix} \\ X_{1,0} \cdot \begin{pmatrix} Y_{0,0} & Y_{0,1} & Y_{0,2} \\ Y_{1,0} & Y_{1,1} & Y_{1,2} \\ Y_{2,0} & Y_{2,1} & Y_{2,2} \end{pmatrix} & X_{1,1} \cdot \begin{pmatrix} Y_{0,0} & Y_{0,1} & Y_{0,2} \\ Y_{1,0} & Y_{1,1} & Y_{1,2} \\ Y_{2,0} & Y_{2,1} & Y_{2,2} \end{pmatrix} \end{pmatrix} \tag{2.2.15}$$

$$= \begin{pmatrix} X_{0,0}Y_{0,0} & X_{0,0}Y_{0,1} & X_{0,0}Y_{0,2} & X_{0,1}Y_{0,0} & X_{0,1}Y_{0,1} & X_{0,1}Y_{0,2} \\ X_{0,0}Y_{1,0} & X_{0,0}Y_{1,1} & X_{0,0}Y_{1,2} & X_{0,1}Y_{1,0} & X_{0,1}Y_{1,1} & X_{0,1}Y_{1,2} \\ X_{0,0}Y_{2,0} & X_{0,0}Y_{2,1} & X_{0,0}Y_{2,2} & X_{0,1}Y_{2,0} & X_{0,1}Y_{2,1} & X_{0,1}Y_{2,2} \\ X_{1,0}Y_{0,0} & X_{1,0}Y_{0,1} & X_{1,0}Y_{0,2} & X_{1,1}Y_{0,0} & X_{1,1}Y_{0,1} & X_{1,1}Y_{0,2} \\ X_{1,0}Y_{1,0} & X_{1,0}Y_{1,1} & X_{1,0}Y_{1,2} & X_{1,1}Y_{1,0} & X_{1,1}Y_{1,1} & X_{1,1}Y_{1,2} \\ X_{1,0}Y_{2,0} & X_{1,0}Y_{2,1} & X_{1,0}Y_{2,2} & X_{1,1}Y_{2,0} & X_{1,1}Y_{2,1} & X_{1,1}Y_{2,2} \end{pmatrix}. \tag{2.2.16}$$

2.2.2 Image, Kernel, and Support

The *image* of a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, denoted by $\mathrm{im}(X)$, is the set defined as

$$\mathrm{im}(X) := \{|\phi\rangle_B \in \mathcal{H}_B : |\phi\rangle_B = X|\psi\rangle_A,\ |\psi\rangle_A \in \mathcal{H}_A\}. \tag{2.2.17}$$

It is also known as the column space or range of X. The image of X is a subspace of $\mathcal{H}_B$. The *rank* of X, denoted by $\mathrm{rank}(X)$, is defined[2] to be the dimension of $\mathrm{im}(X)$. Note that $\mathrm{rank}(X) \le \min\{d_A, d_B\}$ for all $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$.

The *kernel* of a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, denoted by $\ker(X)$, is defined to be the set of vectors in the input space $\mathcal{H}_A$ of X for which the output is the zero vector; i.e.,

$$\ker(X) := \{|\psi\rangle_A \in \mathcal{H}_A : X|\psi\rangle_A = 0\}. \tag{2.2.18}$$

It is also known as the null space of X. The following dimension formula holds:

$$d_A = \mathrm{rank}(X) + \dim(\ker(X)), \tag{2.2.19}$$

and it is known as the *rank-nullity theorem* (the quantity $\dim(\ker(X))$ is called the nullity of X).

The *support* of a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, denoted by $\mathrm{supp}(X)$, is defined to be the orthogonal complement of its kernel:

$$\mathrm{supp}(X) := \ker(X)^\perp := \{|\psi\rangle \in \mathcal{H}_A : \langle\psi|\phi\rangle = 0\ \ \forall |\phi\rangle \in \ker(X)\}. \tag{2.2.20}$$

It is also known as the row space or coimage of X.

See Figure 2.1 for a visual representation of the subspaces $\mathrm{im}(X)$, $\ker(X)$, and $\mathrm{supp}(X)$ corresponding to a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$. We use the notions of support and kernel extensively in Chapter 7, when proving properties of quantum relative entropy and its variants, which are core distinguishability measures in quantum information.

A linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is called *injective* (or *one-to-one*) if, for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A$, $X|\psi\rangle = X|\phi\rangle$ implies $|\psi\rangle = |\phi\rangle$. A necessary and sufficient condition for X to be injective is that the kernel of X contains only the zero vector

[2] The rank of a linear operator can also be equivalently defined as the number of its singular values; please see Theorem 2.1.

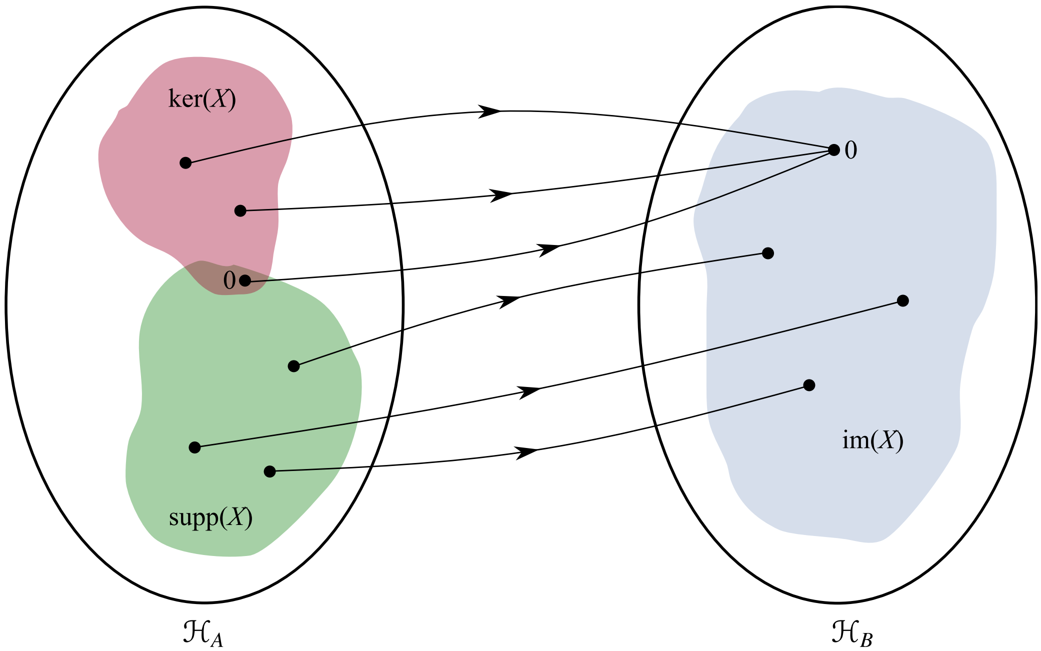


FIGURE 2.1: Visual representation of the subspaces $\text{im}(X)$, $\ker(X)$, and $\text{supp}(X)$ corresponding to a linear operator $X \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$. Note that only the zero vector is contained in both $\ker(X)$ and $\text{supp}(X)$.

(i.e., the column vector in which all of the elements are equal to zero), which implies that $\dim(\ker(X)) = 0$.

A linear operator $X \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$ is called *surjective* (or *onto*) if, for all $|\phi\rangle \in \mathcal{H}_B$, there exists $|\psi\rangle \in \mathcal{H}_A$ such that $X|\psi\rangle = |\phi\rangle$. A necessary and sufficient condition for X to be surjective is that $\text{rank}(X) = d_B$.

Exercise 2.4

Prove that a linear operator $X \in \text{L}(\mathcal{H})$ with the same, finite-dimensional input and output Hilbert space $\mathcal{H}$ is injective if and only if it is surjective. (*Hint*: use the rank-nullity theorem in (2.2.19).)

A linear operator $X \in \text{L}(\mathcal{H})$ that is both injective and surjective is known as a *bijection*. By definition, every bijection is *invertible*, meaning that there exists a unique linear operator, denoted by X^{-1}, such that $XX^{-1} = X^{-1}X = \mathbb{1}$.

2.2.3 Trace

The *trace* of a linear operator $X \in \mathrm{L}(\mathcal{H})$ acting on a d-dimensional Hilbert space $\mathcal{H}$ is defined as

$$\mathrm{Tr}[X] \coloneqq \sum_{i=0}^{d-1} \langle i|X|i\rangle, \tag{2.2.21}$$

which can be interpreted as the sum of the diagonal elements of the matrix corresponding to X in the standard basis.

Exercise 2.5

Prove that the trace of a linear operator is independent of the choice of basis used in (2.2.21). In other words, prove that $\sum_{i=0}^{d-1}\langle i|X|i\rangle = \sum_{k=1}^{d}\langle e_k|X|e_k\rangle$ for every orthonormal basis $\{|e_k\rangle\}_{k=1}^{d}$. (*Hint*: use (2.2.3).)

The trace satisfies the *cyclicity* property: for all pairs of operators $X, Y \in \mathrm{L}(\mathcal{H})$, it holds that

$$\mathrm{Tr}[XY] = \mathrm{Tr}[YX]. \tag{2.2.22}$$

Similarly, for $X, Y, Z \in \mathrm{L}(\mathcal{H})$,

$$\mathrm{Tr}[XYZ] = \mathrm{Tr}[YZX] = \mathrm{Tr}[ZXY]. \tag{2.2.23}$$

More generally, the cyclicity property holds for linear operators with different input and output Hilbert spaces: for $Z_{A\to B} \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, $Y_{B\to C} \in \mathrm{L}(\mathcal{H}_B, \mathcal{H}_C)$, and $X_{C\to A} \in \mathrm{L}(\mathcal{H}_C, \mathcal{H}_A)$,

$$\mathrm{Tr}[X_{C\to A}Y_{B\to C}Z_{A\to B}] = \mathrm{Tr}[Y_{B\to C}Z_{A\to B}X_{C\to A}] \tag{2.2.24}$$
$$= \mathrm{Tr}[Z_{A\to B}X_{C\to A}Y_{B\to C}]. \tag{2.2.25}$$

Exercise 2.6

1. Prove the equalities in (2.2.24) and (2.2.25).
2. Prove that $\mathrm{Tr}[X \otimes Y] = \mathrm{Tr}[X]\mathrm{Tr}[Y]$ for all $X \in \mathrm{L}(\mathcal{H}_A)$ and $Y \in \mathrm{L}(\mathcal{H}_B)$.

2.2.4 Transpose and Conjugate Transpose

Consider $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ as written in (2.2.6). The *transpose* of X is denoted by X^T or alternatively by $\mathrm{T}(X)$, and it is defined as

$$X^\mathsf{T} \equiv \mathrm{T}(X) := \sum_{i=0}^{d_B-1} \sum_{j=0}^{d_A-1} X_{i,j} |j\rangle_A \langle i|_B. \tag{2.2.26}$$

Note that the transpose is basis dependent, in the sense that it is defined with respect to a particular basis (in the case above, we have defined it with respect to the standard bases of $\mathcal{H}_A$ and $\mathcal{H}_B$). Furthermore, taking the transpose with respect to one orthonormal basis can lead to an operator different from that found by taking the transpose with respect to a different orthonormal basis. In this sense, we could more precisely refer to the operation in (2.2.26) as the "standard transpose." The standard transpose can also be understood as a linear superoperator (an operator on operators) with the following representation:

$$\mathrm{T}(X) = \sum_{i=0}^{d_B-1} \sum_{j=0}^{d_A-1} (|j\rangle_A \langle i|_B)\, X\, (|j\rangle_A \langle i|_B) = \sum_{i=0}^{d_B-1} \sum_{j=0}^{d_A-1} X_{i,j} |j\rangle_A \langle i|_B. \tag{2.2.27}$$

Superoperators are discussed in more detail in Section 2.2.11.

The *conjugate transpose* of $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, also known as the *Hermitian conjugate* or the *adjoint* of X, is the linear operator $X^\dagger \in \mathrm{L}(\mathcal{H}_B, \mathcal{H}_A)$ defined as

$$X^\dagger := \sum_{i=0}^{d_B-1} \sum_{j=0}^{d_A-1} \overline{X_{i,j}} |j\rangle_A \langle i|_B. \tag{2.2.28}$$

The adjoint of X is the unique linear operator that satisfies

$$\langle \phi | X \psi \rangle = \langle X^\dagger \phi | \psi \rangle \tag{2.2.29}$$

for all $|\psi\rangle \in \mathcal{H}_A$ and $|\phi\rangle \in \mathcal{H}_B$.

Exercise 2.7

Prove that the conjugate transpose is a basis-independent operation, i.e., that one does not need to specify a basis in order to take the conjugate transpose of a linear operator.

2.2.5 Hilbert–Schmidt Inner Product, Vectorization, and Transpose Trick

On the vector space $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, we define the *Hilbert–Schmidt inner product* as follows:

$$\langle X, Y \rangle \coloneqq \mathrm{Tr}[X^\dagger Y], \quad X, Y \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B). \tag{2.2.30}$$

A key application of the Hilbert–Schmidt inner product in quantum mechanics is the Born rule: In Section 3.3, we will learn that the probability of obtaining an outcome in an experiment is equal to the Hilbert–Schmidt inner product of the state in which the system is prepared and a measurement operator corresponding to the measurement outcome. With this inner product, the vector space $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ becomes an inner product space, and thus a Hilbert space. The basis defined in (2.2.4) is an orthonormal basis for $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ under this inner product. The *Hilbert–Schmidt norm* (or *Schatten 2-norm*) of an operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is defined from the Hilbert–Schmidt inner product as follows:

$$\|X\|_2 \coloneqq \sqrt{\langle X, X \rangle}, \tag{2.2.31}$$

which is a generalization of the Euclidean norm in (2.1.6). The Cauchy–Schwarz inequality for the Hilbert–Schmidt inner product is as follows:

$$|\langle X, Y \rangle|^2 \leq \langle X, X \rangle \cdot \langle Y, Y \rangle = \|X\|_2^2 \cdot \|Y\|_2^2, \tag{2.2.32}$$

for all $X, Y \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$. Observe that the inequality in (2.2.32) is a generalization of that in (2.1.7).

The Hilbert space $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is isomorphic to the tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, where the isomorphism is defined by

$$|i\rangle_B\langle j|_A \leftrightarrow |j\rangle_A \otimes |i\rangle_B \eqqcolon \mathrm{vec}(|i\rangle_B\langle j|_A) \tag{2.2.33}$$

for all $i \in \{0, \ldots, d_B - 1\}$ and $j \in \{0, \ldots, d_A - 1\}$. The operation vec on the right in (2.2.33) is the "vectorize" operation, which transposes the rows of a $d_B \times d_A$ matrix with respect to the standard basis and then stacks the resulting columns in order one on top of the next in order to construct a $d_A d_B$-dimensional column vector. Specifically, for a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ written as in (2.2.6),

$$\mathrm{vec}(X) = \sum_{i=0}^{d_B-1} \sum_{j=0}^{d_A-1} X_{i,j} |j\rangle_A \otimes |i\rangle_B. \tag{2.2.34}$$

The following are useful identities involving the vec operation that we call upon repeatedly throughout this book.

1. For every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$,

$$\mathrm{vec}(X) = (\mathbb{1}_A \otimes X_{A\to B})|\Gamma\rangle_{AA}, \tag{2.2.35}$$

where

$$|\Gamma\rangle_{AA} := \sum_{i=0}^{d_A-1} |i,i\rangle_{AA}. \tag{2.2.36}$$

For reasons that will become clear later, we refer to $|\Gamma\rangle_{AA}$ as the "maximally entangled vector." Note that $|\Gamma\rangle_{AA} = \mathrm{vec}(\mathbb{1}_A)$. For clarity, when needed, we write $|\Gamma_d\rangle$ to refer to the vector defined in (2.2.36) when each Hilbert space has dimension d.

We also note that, for two vectors $|\psi\rangle_B \in \mathcal{H}_B$ and $|\phi\rangle_A \in \mathcal{H}_A$,

$$\mathrm{vec}(|\psi\rangle_B\langle\phi|_A) = \overline{|\phi\rangle}_A \otimes |\psi\rangle_B. \tag{2.2.37}$$

Exercise 2.8

1. Prove (2.2.35).
2. Prove the equality in (2.2.37) by writing both $|\psi\rangle_B$ and $|\phi\rangle_A$ in terms of the orthonormal bases $\{|i\rangle_B\}_{i=0}^{d_B-1}$ and $\{|j\rangle_A\}_{j=0}^{d_A-1}$, respectively, and using (2.2.33).

2. For every vector $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, there exists a unique linear operator $X^{\psi}_{A\to B} \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ such that

$$|\psi\rangle_{AB} = (\mathbb{1}_A \otimes X^{\psi}_{A\to B})|\Gamma\rangle_{AA} = \mathrm{vec}(X^{\psi}_{A\to B}). \tag{2.2.38}$$

In particular, if $|\psi\rangle_{AB} = \sum_{i=0}^{d_A-1}\sum_{j=0}^{d_B-1}\alpha_{i,j}|i,j\rangle_{AB}$, then

$$X^{\psi}_{A\to B} = \sum_{j=0}^{d_B-1}\sum_{i=0}^{d_A-1}\alpha_{i,j}|j\rangle_B\langle i|_A. \tag{2.2.39}$$

Alternatively, there exists a linear operator $Y^{\psi}_{B\to A} \in \mathrm{L}(\mathcal{H}_B, \mathcal{H}_A)$ such that

$$|\psi\rangle_{AB} = (Y^{\psi}_{B\to A} \otimes \mathbb{1}_B)|\Gamma\rangle_{BB}. \tag{2.2.40}$$

This operator is given by

$$Y^{\psi}_{B\to A} = \sum_{j=0}^{d_B-1}\sum_{i=0}^{d_A-1}\alpha_{i,j}|i\rangle_A\langle j|_B, \tag{2.2.41}$$

and one finds by inspection of (2.2.39)–(2.2.41) that $X^{\psi}_{A\to B} = \mathrm{T}(Y^{\psi}_{B\to A})$.

3. *Transpose trick*: For every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, the following equality holds:

$$(\mathbb{1}_A \otimes X_{A\to B})|\Gamma\rangle_{AA} = ((X^{\mathsf{T}})_{B\to A} \otimes \mathbb{1}_B)|\Gamma\rangle_{BB}. \tag{2.2.42}$$

4. For every linear operator $X \in \mathrm{L}(\mathbb{C}^d)$,

$$\mathrm{Tr}[X] = \langle\Gamma|(X \otimes \mathbb{1})|\Gamma\rangle = \langle\Gamma|(\mathbb{1} \otimes X)|\Gamma\rangle = \langle\Gamma|\mathrm{vec}(X). \tag{2.2.43}$$

Exercise 2.9

1. Prove (2.2.42).
2. Prove (2.2.43).
3. Let $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, $Y \in \mathrm{L}(\mathcal{H}_C, \mathcal{H}_A)$, and $Z \in \mathrm{L}(\mathcal{H}_C, \mathcal{H}_D)$. Prove that

$$\mathrm{vec}(XYZ^{\dagger}) = (\overline{Z} \otimes X)\mathrm{vec}(Y). \tag{2.2.44}$$

4. Prove that $\|X\|_2 = \|\mathrm{vec}(X)\|_2$ for every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$.

2.2.6 Notable Classes of Linear Operators

The following classes of linear operators are particularly notable.

- *Hermitian operators*: Also known as *self-adjoint* operators, they are linear operators that are equal to their conjugate transpose. That is, $X \in \mathrm{L}(\mathcal{H})$ is Hermitian if $X^\dagger = X$. The set of all Hermitian operators is a real vector space with dimension d^2, where $d = \dim(\mathcal{H})$. This means that every Hermitian operator X can be expanded in terms of an orthonormal basis $\{B_k\}_{k=1}^{d^2}$ of Hermitian operators such that

$$X = \sum_{k=1}^{d^2} x_k B_k, \tag{2.2.45}$$

where x_k are *real numbers*. Orthonormality is with respect to the Hilbert–Schmidt inner product, which means that $\langle B_k, B_\ell \rangle = \mathrm{Tr}[B_k B_\ell] = \delta_{k,l}$ and $x_k = \langle B_k, X \rangle = \mathrm{Tr}[B_k X]$ for all $k, \ell \in \{1, 2, \ldots, d^2\}$.

An example of an orthonormal basis of d^2 Hermitian operators is the following:

$$\sigma_{0,0}^{(d)} := \frac{\mathbb{1}_d}{\sqrt{d}}, \tag{2.2.46}$$

$$\sigma_{k,\ell}^{(d;+)} := \frac{1}{\sqrt{2}} \left(|k\rangle\langle \ell| + |\ell\rangle\langle k| \right), \quad 0 \leq k < \ell \leq d-1, \tag{2.2.47}$$

$$\sigma_{k,\ell}^{(d;\mathrm{i})} := \frac{1}{\sqrt{2}} \left(-\mathrm{i}|k\rangle\langle \ell| + \mathrm{i}|\ell\rangle\langle k| \right), \quad 0 \leq k < \ell \leq d-1, \tag{2.2.48}$$

$$\sigma_{k,k}^{(d)} := \frac{1}{\sqrt{k(k+1)}} \left(\left(\sum_{j=0}^{k-1} |j\rangle\langle j| \right) - k|k\rangle\langle k| \right), \quad 1 \leq k \leq d-1, \tag{2.2.49}$$

Observe that all of the above operators, with the exception of $\sigma_{0,0}^{(d)}$, are traceless. If we scale each of them by $\sqrt{d}$, then they are called the *generalized Gell-Mann matrices*.

When $d = 2$, the generalized Gell-Mann matrices reduce to the *Pauli matrices*:

$$\mathbb{1} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \sqrt{2}\sigma_{0,0}^{(2)}, \quad X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sqrt{2}\sigma_{0,1}^{(2;+)}, \tag{2.2.50}$$

$$Y := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} = \sqrt{2}\sigma_{0,1}^{(2;\mathrm{i})}, \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sqrt{2}\sigma_{1,1}^{(2)}. \tag{2.2.51}$$

The Pauli matrices are important in the context of quantum mechanics, and quantum information more generally, as they can be used to describe the quantum states of two-dimensional quantum systems, as well as their evolution. They are also involved in fundamental quantum information processing protocols such as quantum teleportation. We elaborate upon these points in Chapters 3–5.

Exercise 2.10

Prove that the operators in (2.2.46)–(2.2.49) do indeed form an orthonormal basis for the vector space of Hermitian operators. More generally, prove that they form an orthonormal basis for the vector space $\mathrm{L}(\mathbb{C}^d)$ of *all* linear operators.

- *Positive semi-definite operators*: A Hermitian operator $X \in \mathrm{L}(\mathcal{H})$ is positive semi-definite if $\langle\psi|X|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$. For every positive semi-definite operator X, there exists a linear operator Y such that $X = Y^\dagger Y$. X is called *positive definite* if $\langle\psi|X|\psi\rangle > 0$ for all $|\psi\rangle \in \mathcal{H}$ such that $|\psi\rangle \neq 0$. We write $X \geq 0$ if X is positive semi-definite, and $X > 0$ if X is positive definite. If $X - Z$ is positive semi-definite for Hermitian operators X and Z, then we write $X \geq Z$, and if $X - Z$ is positive definite, then we write $X > Z$. The ordering $X \geq Z$ on Hermitian operators is a partial order called the Löwner order and is discussed more in Definition 2.12.

Exercise 2.11

Let $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ be a linear operator, with $\mathcal{H}_A$ and $\mathcal{H}_B$ arbitrary. Prove that $X^\dagger X$ is positive semi-definite.

- *Density operators*: These are Hermitian operators that are positive semi-definite and have unit trace. Density operators are generalizations of probability distributions from classical information theory and describe the states of a quantum system, as detailed in Chapter 3.

- *Unitary operators*: These are linear operators whose inverses are equal to their adjoints. That is, $U \in \mathrm{L}(\mathcal{H})$ is unitary if $U^\dagger U = UU^\dagger = \mathbb{1}$. Unitary operators generalize invertible maps or permutations from classical information theory and describe the noiseless evolution of the state of a quantum system, as detailed in Chapters 3 and 4.

Exercise 2.12

Let $U \in \mathrm{L}(\mathcal{H})$ be a unitary operator acting on a d-dimensional Hilbert space $\mathcal{H}$.

1. Given an orthonormal basis $\{|e_k\rangle\}_{k=1}^{d}$ for $\mathcal{H}$, prove that the set $\{|f_k\rangle\}_{k=1}^{d}$, with $|f_k\rangle := U|e_k\rangle$ for all $1 \leq k \leq d$, is another orthonormal basis for $\mathcal{H}$.

2. Using the transpose trick identity in (2.2.42), prove that

$$|\Gamma\rangle = (\overline{U} \otimes U)|\Gamma\rangle. \tag{2.2.52}$$

3. Using 1. and 2., conclude that the following identity holds for every orthonormal basis $\{|e_k\rangle\}_{k=1}^{d}$ for $\mathcal{H}$:

$$|\Gamma\rangle = \sum_{k=1}^{d} \overline{|e_k\rangle} \otimes |e_k\rangle. \tag{2.2.53}$$

- *Isometric operators* or *isometries*: A linear operator $V \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ is isometric if $V^\dagger V = \mathbb{1}_A$ (we also say that V is an isometry). Isometries also describe the noiseless evolution of the state of a quantum system, as detailed in Chapters 3 and 4.

Exercise 2.13

Let $V \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ be an isometry.

1. Prove that $\langle V\psi|V\phi\rangle = \langle\psi|\phi\rangle$ for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A$.

2. Using 1., prove that V is injective.

3. Using 2., prove that $d_B \geq d_A$. (*Hint*: use the rank-nullity theorem in (2.2.19).)

4. Prove that, if $d_A = d_B$, then V is a unitary. (*Hint*: use the result of Exercise 2.4.)

- *Projection operators*: A linear operator $P \in \mathrm{L}(\mathcal{H})$ is a projection if it is Hermitian and satisfies $P^2 = P$. Such operators are also sometimes called *orthogonal projection operators*.

Note that every linear operator X can be decomposed as

$$X = \mathrm{Re}[X] + \mathrm{i}\,\mathrm{Im}[X], \tag{2.2.54}$$

where $\mathrm{Re}[X] := \frac{1}{2}(X + X^\dagger)$ and $\mathrm{Im}[X] := \frac{1}{2\mathrm{i}}(X - X^\dagger)$ are both Hermitian operators, generalizing how complex numbers can be decomposed into real and imaginary parts.

2.2.7 Singular Value, Schmidt, and Polar Decompositions

An important fact that we make use of throughout this book is the *singular value decomposition theorem*.

Theorem 2.1 Singular Value Decomposition

For every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ with $r := \mathrm{rank}(X)$, there exist strictly positive real numbers $\{s_k > 0 : 1 \leq k \leq r\}$, called the *singular values of* X, and orthonormal vectors $\{|e_k\rangle_B : 1 \leq k \leq r\}$ and $\{|f_k\rangle_A : 1 \leq k \leq r\}$, such that

$$X = \sum_{k=1}^{r} s_k |e_k\rangle_B \langle f_k|_A. \tag{2.2.55}$$

This is called the *singular value decomposition of* X.

Remark: From (2.2.55), we see that the rank of a linear operator X, which we defined earlier as the dimension of the image of X, is equal to the number of singular values of X.

The singular value decomposition theorem can be written in the following matrix form that is familiar from elementary linear algebra. We first extend the orthonormal vectors $\{|e_k\rangle_B : 1 \leq k \leq r\}$ in $\mathcal{H}_B$ to an orthonormal basis $\{|e_k\rangle_B : 1 \leq k \leq d_B\}$ for $\mathcal{H}_B$, and similarly, we extend the set $\{|f_k\rangle_A : 1 \leq k \leq r\}$ of orthonormal vectors in $\mathcal{H}_A$ to an orthonormal basis $\{|f_k\rangle_A : 1 \leq k \leq d_A\}$ for $\mathcal{H}_A$. Then, by defining the unitaries

$$W := \sum_{k=1}^{d_B} |e_k\rangle_B\langle k-1|_B, \qquad V := \sum_{k=1}^{d_A} |f_k\rangle_A\langle k-1|_A, \tag{2.2.56}$$

we can write (2.2.55) as

$$X = WSV^\dagger, \tag{2.2.57}$$

where

$$S := \sum_{k=1}^{r} s_k|k-1\rangle_B\langle k-1|_A \tag{2.2.58}$$

is the matrix of singular values. Note that the matrix S is a $d_B \times d_A$ rectangular matrix and (2.2.58) is only specifying the non-zero entries of S along the diagonal.

The singular value decomposition can be used to prove the following useful theorem for vectors in a tensor-product Hilbert space.

Theorem 2.2 Schmidt Decomposition

Let $|\psi\rangle_{AB}$ be a vector in the tensor-product Hilbert space $\mathcal{H}_{AB}$. Let $X_{A\to B}$ be the linear operator with matrix elements $\langle j|_B X|i\rangle_A = \langle i,j|\psi\rangle_{AB}$, and let $r = \text{rank}(X)$. Then, there exist strictly positive *Schmidt coefficients* $\{\lambda_k\}_{k=1}^r$, and orthonormal vectors $\{|e_k\rangle_A\}_{k=1}^r$ and $\{|f_k\rangle_B\}_{k=1}^r$, such that

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}\, |e_k\rangle_A \otimes |f_k\rangle_B. \tag{2.2.59}$$

The quantity r is called the *Schmidt rank*, satisfying $r \leq \min\{d_A, d_B\}$.

Proof: This is a direct application of the singular value decomposition (Theorem 2.1). Consider the operator X defined in the statement of the theorem, having matrix elements $\langle j|_B X|i\rangle_A = \langle i,j|\psi\rangle_{AB}$. Observe then that $\text{vec}(X) = |\psi\rangle_{AB}$ (see (2.2.33)).

Now, by the singular-value decomposition (Theorem 2.1), we can write X as $X = \sum_{k=1}^r s_k|f_k\rangle_B\langle g_k|_A$, where s_k are strictly positive numbers, $r = \text{rank}(X) \leq \min\{d_A, d_B\}$, and $\{|f_k\rangle_B : 1 \leq k \leq r\}$ and $\{|g_k\rangle_A : 1 \leq k \leq r\}$ are sets of orthonormal vectors in $\mathcal{H}_B$ and $\mathcal{H}_A$, respectively. Letting $\lambda_k = s_k^2$, and upon vectorizing X as written in this form (see (2.2.33)), we find that

$$|\psi\rangle_{AB} = \text{vec}(X) = \sum_{k=1}^{r} \sqrt{\lambda_k}\, \overline{|g_k\rangle}_A \otimes |f_k\rangle_B. \tag{2.2.60}$$

Letting $|e_k\rangle_A := \overline{|g_k\rangle}_A$, the result follows. ■

A simple but important consequence of the Schmidt decomposition theorem is that every vector $|\psi\rangle_{AB}$ in a tensor-product Hilbert space $\mathcal{H}_{AB}$ can be written as

$$|\psi\rangle_{AB} = \sum_{k=1}^{\min\{d_A,d_B\}} \sqrt{p_k}|u_k\rangle_A \otimes |v_k\rangle_B, \tag{2.2.61}$$

where $p_k = s_k^2$ for $1 \leq k \leq r$ and $p_k = 0$ for $r < k \leq \min\{d_A, d_B\}$. The vectors $|u_k\rangle_A$ and $|v_k\rangle_B$ are such that $|u_k\rangle_A = |e_k\rangle_A$ and $|v_k\rangle_B = |f_k\rangle_B$ for $1 \leq k \leq r$, and the remaining vectors combine to form orthonormal bases for a $\min\{d_A, d_B\}$-dimensional subspace of $\mathcal{H}_A \otimes \mathcal{H}_B$.

Exercise 2.14

Using arguments similar to the proof of Theorem 2.2, prove that every linear operator $X_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ can be written as

$$X_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}\, E_A^k \otimes F_B^k, \tag{2.2.62}$$

where the set $\{\lambda_k\}_{k=1}^r$ consists of strictly positive reals, $\{E_A^k\}_{k=1}^r$ and $\{F_B^k\}_{k=1}^r$ are orthonormal sets of linear operators acting on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and $r = \text{rank}(M)$, where $M \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_B)$ is defined by $\langle j, \ell|_{BB} M|i, k\rangle_{AA} = \langle i, j|_{AB} X|k, \ell\rangle_{AB}$ for all $0 \leq i, k \leq d_A - 1$ and $0 \leq j, \ell \leq d_B - 1$.

Another important decomposition of a linear operator is the *polar decomposition*.

Theorem 2.3 Polar Decomposition

Every linear operator $X \in \mathrm{L}(\mathcal{H})$ can be written as $X = UP$, where U is a unitary operator and P is a positive semi-definite operator. In particular,

$P = |X| := \sqrt{X^\dagger X}$.

Proof: If $X = WSV^\dagger$ is the singular value decomposition of X, then we can take $P = VSV^\dagger$ and $U = WV^\dagger$. ∎

The polar decomposition generalizes the polar form $z = re^{i\theta}$ of a complex number z. Indeed, the fact that $r \geq 0$ is in correspondence with P being a positive semi-definite operator, and the phase $e^{i\theta}$ is in correspondence with the unitary U, the latter having eigenvalues on the unit circle (i.e., of the form $e^{i\theta}$).

2.2.8 Spectral Theorem

Given a linear operator $X \in L(\mathcal{H})$ acting on some Hilbert space $\mathcal{H}$, if there exists a vector $|\psi\rangle \in \mathcal{H}$ such that

$$X|\psi\rangle = \lambda|\psi\rangle, \tag{2.2.63}$$

then $|\psi\rangle$ is said to be an *eigenvector of* X with associated *eigenvalue* λ. The set of all eigenvectors associated with an eigenvalue λ is a subspace of $\mathcal{H}$ called the *eigenspace of* X *associated with* λ, and the *multiplicity* of λ is the number of linearly independent eigenvectors of X that are associated with λ (in other words, it is the dimension of the eigenspace of X associated with λ). The eigenspace of X associated with λ is equal to $\ker(X - \lambda I)$.

The spectral theorem, which we state below, allows us to decompose every *normal operator* X, i.e., an operator that commutes with its adjoint, so that $XX^\dagger = X^\dagger X$, in terms of its eigenvalues and projections onto its corresponding eigenspaces. We employ it most often when analyzing quantum states and observables.

Theorem 2.4 Spectral Theorem

For every normal operator $X \in L(\mathcal{H})$, there exists $n \in \mathbb{N}$ such that

$$X = \sum_{j=1}^{n} \lambda_j \Pi_j, \tag{2.2.64}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ are the distinct eigenvalues of X and $\Pi_1, \Pi_2, \ldots, \Pi_n$

are the *spectral projections* onto the corresponding eigenspaces, which satisfy $\Pi_1 + \Pi_2 + \cdots + \Pi_n = \mathbb{1}_{\mathcal{H}}$ and $\Pi_i \Pi_j = \delta_{i,j}\Pi_i$. The decomposition in (2.2.64) is unique and is called the *spectral decomposition* of X. The spectrum of X is denoted by $\text{spec}(X) \coloneqq \{\lambda_1, \lambda_2, \dots, \lambda_n\}$.

REMARK: The multiplicity of an eigenvalue λ_i is equal to the rank of the corresponding projection Π_i.

The property $\Pi_i\Pi_j = \delta_{i,j}\Pi_i$ of the spectral projections indicates that the eigenspaces of a linear operator are orthogonal to each other. In other words, for every eigenvector $|\psi_{\lambda_1}\rangle$ associated with the eigenvalue λ_1 and every eigenvector $|\psi_{\lambda_2}\rangle$ associated with the eigenvector λ_2, where $\lambda_1 \neq \lambda_2$, we have that $\langle\psi_{\lambda_1}|\psi_{\lambda_2}\rangle = 0$.

If we take the spectral decomposition of a normal operator $X \in \mathrm{L}(\mathcal{H})$ in (2.2.64) and find orthonormal bases for the corresponding eigenspaces, then X can be written as

$$X = \sum_{k=1}^{d} \lambda_k |\psi_k\rangle\langle\psi_k|, \tag{2.2.65}$$

where $d = \dim(\mathcal{H})$ and $\{|\psi_k\rangle\}_{k=1}^{d}$ is a set of orthonormal vectors such that

$$|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2| + \cdots + |\psi_d\rangle\langle\psi_d| = \mathbb{1}_{\mathcal{H}}. \tag{2.2.66}$$

From this decomposition, we see clearly that $|\psi_k\rangle$ is an eigenvector of X with associated eigenvalue λ_k.

The spectral theorem is a statement of a fact familiar from elementary linear algebra, that every normal operator can be diagonalized by a unitary. Indeed, if we consider the spectral decomposition in (2.2.65), and we define the unitary operator $U \coloneqq \sum_{k=1}^{d} |\psi_k\rangle\langle k-1|$, then (2.2.65) can be written as $X = UDU^\dagger$, where $D \coloneqq \sum_{k=1}^{d} \lambda_k |k-1\rangle\langle k-1|$ is a diagonal matrix.

Note that for the decomposition in (2.2.65) the numbers $\lambda_k \in \mathbb{C}$ are not all distinct because the eigenspace associated with each eigenvalue can have dimension greater than one. Also, note that the decomposition in (2.2.65) is generally not unique because the decomposition of the spectral projections into orthonormal vectors is not unique.

From (2.2.65), it is evident that the rank of a normal operator X (recall the discussion around (2.2.17)) is equal to the number of non-zero eigenvalues of X

(including their multiplicities). Furthermore, the support of X (recall (2.2.20)) is equal to the span of all eigenvectors of X associated with the non-zero eigenvalues of X. In particular,

$$\Pi_X := \sum_{k:\lambda_k \neq 0} |\psi_k\rangle\langle\psi_k| \tag{2.2.67}$$

is the projection onto the support of X. It is also evident that the trace of X is equal to the sum of its eigenvalues, i.e., $\mathrm{Tr}[X] = \sum_{k=1}^{d} \lambda_k = \sum_{k:\lambda_k \neq 0} \lambda_k$.

Exercise 2.15

Let P be a projection operator.

1. Prove that the eigenvalues of P are either 0 or 1. Prove that $\mathrm{Tr}[P] = \mathrm{rank}(P)$.
2. Using 1., conclude that $\mathrm{rank}(X) = \mathrm{Tr}[\Pi_X]$ for every linear operator X, where Π_X is the projection onto the support of X, as defined in (2.2.67).

The singular values of a linear operator X (not necessarily normal) are related to the eigenvalues of $X^\dagger X$ and $XX^\dagger$ in the following way. Let $\{s_k\}_{k=1}^{\mathrm{rank}(X)}$ be the set of singular values of X, and let $\{\lambda_k\}_{k=1}^{\mathrm{rank}(X)}$ be the non-zero eigenvalues of $X^\dagger X$, which are the same as the non-zero eigenvalues of $XX^\dagger$. (Note that both $X^\dagger X$ and $XX^\dagger$ are normal operators, so that the spectral theorem applies to them.) Then, $s_k = \sqrt{\lambda_k}$ for all $1 \leq k \leq \mathrm{rank}(X)$. In particular, if X is a Hermitian operator with non-zero eigenvalues $\{\omega_k\}_{k=1}^{\mathrm{rank}(X)}$, then $s_k = \sqrt{\omega_k^2} = |\omega_k|$ for all $1 \leq k \leq \mathrm{rank}(X)$.

Exercise 2.16

1. Prove that every Hermitian operator has real eigenvalues.
2. Prove that every unitary operator has eigenvalues with unit modulus; i.e., if λ is an eigenvalue of a given unitary operator, then $|\lambda| = 1$.

If X is a Hermitian operator, then we can split it into a *positive part*, denoted by X_+, and a *negative part*, denoted by X_-, so that

$$X = X_+ - X_- . \tag{2.2.68}$$

This follows because X has real eigenvalues (see Exercise 2.16). In particular, if

$X = \sum_{k=1}^{d} \lambda_k |\psi_k\rangle\langle\psi_k|$ is a spectral decomposition of X, then

$$X_+ := \sum_{k:\lambda_k \geq 0} \lambda_k |\psi_k\rangle\langle\psi_k|, \quad X_- := \sum_{k:\lambda_k < 0} |\lambda_k| \, |\psi_k\rangle\langle\psi_k|. \tag{2.2.69}$$

Note that both X_+ and X_- are positive semi-definite operators, and they are supported on orthogonal subspaces, meaning that $X_+ X_- = 0$. Such a decomposition of a Hermitian operator X into positive and negative parts is called the *Jordan–Hahn decomposition* of X.

Exercise 2.17

Using the Jordan–Hahn decomposition, prove that there exists a basis for $\mathrm{L}(\mathbb{C}^d)$ consisting entirely of positive semi-definite operators, for all $d \geq 2$.

Another important fact is that a Hermitian operator X is positive semi-definite if and only if all of its eigenvalues are non-negative, and positive definite if and only if all of its eigenvalues are strictly positive. The latter implies that every positive definite operator $X \in \mathrm{L}(\mathcal{H})$ has full support, i.e., $\mathrm{supp}(X) = \mathcal{H}$ and $\mathrm{rank}(X) = \dim(\mathcal{H})$. Thus, positive definite operators are invertible.

Exercise 2.18

For every positive semi-definite operator X, prove that the operator $X + \varepsilon \mathbb{1}$ is positive definite for all $\varepsilon > 0$.

2.2.8.1 Functions of Hermitian Operators

Using the spectral decomposition as in (2.2.65), we can define a function of a Hermitian operator. Basic functions of density operators that we employ extensively are $x \to -x \log_2 x$ for the quantum entropy and $x \to x^\alpha$ for the Rényi entropy.

Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function with domain $\mathrm{dom}(f)$. For every Hermitian operator $X \in \mathrm{L}(\mathcal{H})$ with spectrum $\mathrm{spec}(X) \subseteq \mathrm{dom}(f)$ and spectral decomposition $X = \sum_{j=1}^{n} \lambda_j \Pi_j$ as in (2.2.64), we define $f(X)$ as the operator

$$f(X) := \sum_{j=1}^{n} f(\lambda_j) \Pi_j. \tag{2.2.70}$$

Note that using the alternative form of the spectral decomposition given in (2.2.65) we could also write

$$f(X) \coloneqq \sum_{k=1}^{d} f(\lambda_k)|\psi_k\rangle\langle\psi_k|, \tag{2.2.71}$$

where $d = \dim(\mathcal{H})$. An immediate consequence of these definitions is that if X is a Hermitian operator and U is a unitary operator, then

$$f(UXU^\dagger) = Uf(X)U^\dagger. \tag{2.2.72}$$

This is due to the fact that X and $UXU^\dagger$ have the same eigenvalues and multiplicities.

Functions that arise frequently throughout this book are as follows:

- *Power functions*: for some real number α, the function $f_\alpha(x) = x^\alpha$ extends to Hermitian operators via (2.2.70) as

  $$X^\alpha \coloneqq \sum_{k=1}^{d} \lambda_k^\alpha |\psi_k\rangle\langle\psi_k|. \tag{2.2.73}$$

  However, in order to apply f_α we need to make sure that the spectrum of X is contained in the domain of f_α. We distinguish four cases:

  - If $\alpha \in \mathbb{N}$ is a non-negative integer, then the domain of f_α is the entire real line $\mathbb{R}$, and hence (2.2.73) is well defined for all Hermitian operators X. Here we follow the convention that $0^0 = 0$.
  - If $\alpha \in \mathbb{Z} \setminus \mathbb{N}$ is a negative integer, then the domain of f_α is $\mathbb{R} \setminus \{0\}$, and therefore (2.2.73) applies to all full-rank Hermitian operators X, i.e., to all X such that $\text{rank}(X) = d$.
  - If $\alpha > 0$ but $\alpha \notin \mathbb{N}$, then the domain of f_α is $[0, \infty)$, and therefore (2.2.73) applies to all positive semi-definite operators $X \geq 0$.
  - If $\alpha < 0$ but $\alpha \notin \mathbb{Z}$, then the domain of f_α is $(0, \infty)$, and therefore (2.2.73) applies to all positive definite operators $X > 0$.

  When $\alpha = \frac{1}{2}$ and $X \geq 0$, we typically use the notation $\sqrt{X}$ to refer to $X^{\frac{1}{2}}$. In particular, $\sqrt{X}$ is the unique positive semi-definite operator such that $\sqrt{X}\sqrt{X} = X$.

  For $\alpha = 0$ and full-rank X, the following equality holds:

  $$X^0 = \sum_{k:\lambda_k \neq 0} |\psi_k\rangle\langle\psi_k| = \Pi_X, \tag{2.2.74}$$

  where Π_X is the projection onto the support of X; recall (2.2.67).

- *Logarithm functions*: For the function $\log_b : (0, \infty) \to \mathbb{R}$ with base $b > 0$, we define

$$\log_b(X) := \sum_{k:\lambda_k>0} \log_b(\lambda_k)|\psi_k\rangle\langle\psi_k|. \tag{2.2.75}$$

  We deal throughout this book exclusively with the base-2 logarithm $\log_2$ and the base-e logarithm $\log_e \equiv \ln$.

We end this section with a lemma that is used several times in Chapter 7.

Lemma 2.5

Let $X \in \mathrm{L}(\mathcal{H})$, and let f be a function such that the squares of the singular values of X are in the domain of f. Then

$$Xf(X^\dagger X) = f(XX^\dagger)X. \tag{2.2.76}$$

Proof: This is a direct consequence of the singular value decomposition theorem (Theorem 2.1). Let $X = WSV^\dagger$ be a singular value decomposition of L, where W and V are unitary operators and S is a diagonal, positive semi-definite operator. Then

$$Xf(X^\dagger X) = WSV^\dagger f\left(\left(WSV^\dagger\right)^\dagger WSV^\dagger\right) \tag{2.2.77}$$
$$= WSV^\dagger f(VSW^\dagger WSV^\dagger) \tag{2.2.78}$$
$$= WSV^\dagger f(VS^2V^\dagger). \tag{2.2.79}$$

Now, we use the fact that $f(VS^2V^\dagger) = Vf(S^2)V^\dagger$, which holds because V is unitary. Using as well the fact that $Sf(S^2) = f(S^2)S$, we obtain

$$Xf(X^\dagger X) = WSf(S^2)V^\dagger \tag{2.2.80}$$
$$= Wf(S^2)SV^\dagger \tag{2.2.81}$$
$$= Wf(SVV^\dagger S)W^\dagger WSV^\dagger \tag{2.2.82}$$
$$= f(WSV^\dagger VSW^\dagger)WSV^\dagger \tag{2.2.83}$$
$$= f(XX^\dagger)X. \tag{2.2.84}$$

This concludes the proof. ■

2.2.9 Norms

A *norm* on a Hilbert space $\mathcal{H}$ (more generally a vector space) is a function $\|\cdot\| : \mathcal{H} \to \mathbb{R}$ that satisfies the following properties:

- *Non-negativity*: $\||\psi\rangle\| \geq 0$ for all $|\psi\rangle \in \mathcal{H}$, and $\||\psi\rangle\| = 0$ if and only if $|\psi\rangle = 0$.
- *Scaling* (also called *homogeneity*): For every $\alpha \in \mathbb{C}$ and $|\psi\rangle \in \mathcal{H}$, $\|\alpha|\psi\rangle\| = |\alpha| \, \||\psi\rangle\|$.
- *Triangle inequality*: For all $|\psi\rangle, |\phi\rangle \in \mathcal{H}$, $\||\psi\rangle + |\phi\rangle\| \leq \||\psi\rangle\| + \||\phi\rangle\|$.

An immediate consequence of the scaling property and the triangle inequality is that a norm is convex:

$$\|\lambda|\psi\rangle + (1-\lambda)|\phi\rangle\| \leq \lambda \, \||\psi\rangle\| + (1-\lambda) \, \||\phi\rangle\| \tag{2.2.85}$$

for all vectors $|\psi\rangle$ and $|\phi\rangle$ and all $\lambda \in [0, 1]$.

In this section, we are primarily interested in the Hilbert space $\mathrm{L}(\mathcal{H})$ of linear operators $X : \mathcal{H} \to \mathcal{H}$ for some Hilbert space $\mathcal{H}$. The following norm for linear operators is used extensively in this book.

Definition 2.6 Schatten Norm

For every linear operator $X \in \mathrm{L}(\mathcal{H})$ acting on a Hilbert space $\mathcal{H}$, we define its *Schatten α-norm* as

$$\|X\|_\alpha := (\mathrm{Tr}[|X|^\alpha])^{\frac{1}{\alpha}}, \tag{2.2.86}$$

for all $\alpha \in [1, \infty)$, where $|X| := \sqrt{X^\dagger X}$. We let

$$\|X\|_\infty := \lim_{\alpha \to \infty} \|X\|_\alpha . \tag{2.2.87}$$

Throughout this book, we extend the function $\|\cdot\|_\alpha$ to include $\alpha \in (0, 1)$ (with the definition exactly as in (2.2.86)), although in this case it is not a norm because it does not satisfy the triangle inequality.

Norms are typically employed in pure mathematics to measure the lengths of vectors or operators, and different norms give different ways of measuring length. In quantum information, we employ norms to measure entropy and information of quantum states and channels (see Chapter 7). The parameter α for the Schatten norm then becomes the Rényi parameter for the Rényi entropy.

Exercise 2.19

Let X be a linear operator, and let $\{s_k\}_{k=1}^r$ be the set of singular values of X, where $r := \text{rank}(X)$. Prove that

$$\|X\|_\alpha = \left(\sum_{k=1}^r s_k^\alpha\right)^{\frac{1}{\alpha}} \tag{2.2.88}$$

for all $\alpha \in (0, \infty)$.

If X is a Hermitian operator with non-zero eigenvalues $\{\lambda_k\}_{k=1}^r$, then its singular values are $s_k = |\lambda_k|$, where $k \in \{1, \ldots, r\}$ (see Section 2.2.8). Therefore, for all $\alpha \in (0, \infty)$,

$$\|X\|_\alpha = \left(\sum_{k=1}^r |\lambda_k|^\alpha\right)^{\frac{1}{\alpha}} \quad (X \text{ Hermitian}). \tag{2.2.89}$$

Exercise 2.20

Let X be a linear operator and $\alpha \in (0, \infty)$. Prove that

$$\left\|X^\dagger X\right\|_\alpha = \left\|XX^\dagger\right\|_\alpha = \|X\|_{2\alpha}^2 . \tag{2.2.90}$$

We now state several important properties of the Schatten norm.

Proposition 2.7 Properties of Schatten Norm

For all $\alpha \in [1, \infty]$, the Schatten norm $\|\cdot\|_\alpha$ has the following properties.

1. *Monotonicity*: For every linear operator X, the function $X \mapsto \|X\|_\alpha$ is monotonically non-increasing with α; i.e., for $\alpha \geq \beta > 0$, the following holds:

$$\|X\|_\alpha \leq \|X\|_\beta . \tag{2.2.91}$$

In particular, we then have that $\|X\|_\infty \leq \|X\|_\alpha \leq \|X\|_1$ for all $\alpha \in [1, \infty]$

and every linear operator X.

2. *Isometric invariance*: For all isometries U and V,

$$\|X\|_\alpha = \left\|UXV^\dagger\right\|_\alpha . \tag{2.2.92}$$

3. *Submultiplicativity*: For all linear operators X, Y, and Z,

$$\|XYZ\|_\alpha \leq \|X\|_\infty \|Y\|_\alpha \|Z\|_\infty . \tag{2.2.93}$$

Consequently, for all linear operators X and Y,

$$\|XY\|_\alpha \leq \|X\|_\alpha \|Y\|_\alpha . \tag{2.2.94}$$

4. *Multiplicativity with respect to tensor product*: For all linear operators X and Y,

$$\|X \otimes Y\|_\alpha = \|X\|_\alpha \|Y\|_\alpha . \tag{2.2.95}$$

5. *Direct-sum property*: Given a collection $\{X_x\}_{x\in\mathcal{X}}$ of linear operators indexed by a finite alphabet $\mathcal{X}$, the following equality holds for every orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$:

$$\left\|\sum_{x\in\mathcal{X}} |x\rangle\langle x| \otimes X_x\right\|_\alpha^\alpha = \sum_{x\in\mathcal{X}} \|X_x\|_\alpha^\alpha . \tag{2.2.96}$$

6. *Duality*: For every linear operator X,

$$\|X\|_\alpha = \sup_Y \left\{\left|\mathrm{Tr}[Y^\dagger X]\right| : \|Y\|_\beta \leq 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1\right\}. \tag{2.2.97}$$

The equality above implies *Hölder's inequality*:

$$\left|\mathrm{Tr}[Z^\dagger X]\right| \leq \|X\|_\alpha \|Z\|_\beta \tag{2.2.98}$$

which holds for all linear operators X and Z, where $\alpha, \beta \in [1, \infty]$ satisfy $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. In this sense, the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, with $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, are said to be *dual* to each other.

Proof:

1. By the scaling property, up to dividing by $\|X\|_\beta$ it suffices to prove (2.2.91) for the case where $\|X\|_\beta = 1$. Let $\{s_k\}_{k=1}^r$ denote the singular values of X, where $r := \text{rank}(X)$. From
$$1 = \|X\|_\beta^\beta = \sum_{k=1}^r s_k^\beta \tag{2.2.99}$$
we deduce that $s_k \leq 1$ for all k. Since $\beta \leq \alpha$, this entails that $s_k^\beta \geq s_k^\alpha$ for all k. Summing over k yields precisely
$$1 = \sum_{k=1}^r s_k^\beta \geq \sum_{k=1}^r s_k^\alpha = \|X\|_\alpha^\alpha , \tag{2.2.100}$$
i.e., $\|X\|_\alpha \leq 1 = \|X\|_\beta$, as claimed. In fact, this proof shows that $\|X\|_\alpha$ is monotone non-increasing for all $\alpha > 0$, not only for $\alpha \geq 1$.

2. Isometric invariance holds because the singular values of a linear operator X do not change under the action $X \mapsto UXV^\dagger$, for isometries U and V.

3. For a proof of (2.2.93), see the Bibliographic Notes (Section 2.6). Submultiplicativity in (2.2.94) follows immediately from (2.2.93) by taking $Z = \mathbb{1}$, using the fact that $\|\mathbb{1}\|_\infty = 1$ (see Section 2.2.9.1 below), and using monotonicity, which implies that $\|X\|_\infty \leq \|X\|_\alpha$.

4.-5. Multiplicativity with respect to the tensor product and the direct sum property follow immediately from the definition of $\|\cdot\|_\alpha$.

6. We provide a proof of (2.2.97) in the special case $\alpha = 1, \beta = \infty$ in Proposition 2.10 below. For all other values of α and β, please consult the Bibliographic Notes (Section 2.6). Given (2.2.97), for all linear operators X and Z, let $Y = \frac{Z}{\|Z\|_\beta}$. Then, $\|Y\|_\beta \leq 1$, which means that
$$\frac{1}{\|Z\|_\beta} \left|\text{Tr}[Z^\dagger X]\right| = \left|\text{Tr}[Y^\dagger X]\right| \leq \|X\|_\alpha \Rightarrow \left|\text{Tr}[Z^\dagger X]\right| \leq \|X\|_\alpha \|Z\|_\beta , \tag{2.2.101}$$
which is the inequality in (2.2.98). ■

In addition to the variational characterization of the Schatten norm $\|\cdot\|_\alpha$ given in (2.2.97), we have the following variational characterization, which extends to $\alpha \in (0, 1)$.

Proposition 2.8

Let $\alpha \in (0,1) \cup (1,\infty]$. Then, for every β such that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, and every positive semi-definite operator X,

$$\|X\|_\alpha = \begin{cases} \inf\{\mathrm{Tr}\big[XY^{\frac{1}{\beta}}\big] : Y > 0,\ \mathrm{Tr}[Y] = 1\} & \text{if } \alpha \in [0,1), \\ \sup\{\mathrm{Tr}\big[XY^{\frac{1}{\beta}}\big] : Y \geq 0,\ \mathrm{Tr}[Y] = 1\} & \text{if } \alpha \in [1,\infty). \end{cases} \tag{2.2.102}$$

Proof: Please consult the Bibliographic Notes (Section 2.6). ■

2.2.9.1 Schatten ∞-Norm (Spectral/Operator Norm)

An important case of the Schatten norms is the Schatten ∞-norm, which we recall from (2.2.87) is defined as

$$\|X\|_\infty := \lim_{\alpha\to\infty} \|X\|_\alpha\,, \tag{2.2.103}$$

Proposition 2.9 Schatten ∞-Norm is Largest Singular Value

For every linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, $\|X\|_\infty$ is equal to the largest singular value of X, which we denote by $s_{\max}$, i.e.,

$$\|X\|_\infty = s_{\max}. \tag{2.2.104}$$

Proof: Let $\vec{s} := (s_k)_{k=1}^r$ denote the vector of singular values of X, where $r := \mathrm{rank}(X)$. Then for all $\alpha \geq 1$ and $k \in \{1,\dots,r\}$, the inequalities $s_{\max}^\alpha \leq \sum_{k=1}^r s_k^\alpha \leq r s_{\max}^\alpha$ hold. Raising everything to the power of $1/\alpha$ and taking the limit $\alpha \to \infty$ concludes the proof, because $\lim_{\alpha\to\infty} r^{1/\alpha} = 1$, given that r is fixed. ■

Due to Proposition 2.9, the term *spectral norm* is often used to refer to the Schatten ∞-norm. It is also referred to as the *operator norm*, because it is the norm induced by the Euclidean norm on the underlying Hilbert space on which the operator X acts, i.e.,

$$\|X\|_\infty = \sup_{|\psi\rangle\neq 0} \frac{\|X|\psi\rangle\|_2}{\||\psi\rangle\|_2} = \sup_{|\psi\rangle:\||\psi\rangle\|_2=1} \|X|\psi\rangle\|_2\,. \tag{2.2.105}$$

In the equation above, we have employed the shorthand sup, which stands for supremum. We also often employ inf for infimum. These concepts are reviewed in Section 2.3.1.

Exercise 2.21

Using the fact that X has a singular value decomposition of the form $X = \sum_{k=1}^{\text{rank}(X)} s_k |e_k\rangle\langle f_k|$ (see Theorem 2.1), prove (2.2.105). Similarly, prove that

$$\|X\|_\infty = \sup\{|\langle\psi|X|\phi\rangle| : \||\psi\rangle\|_2 = \||\phi\rangle\|_2 = 1\}. \tag{2.2.106}$$

If X is Hermitian and positive semi-definite, then $\|X\|_\infty$ is equal to the largest eigenvalue of X, and we can write

$$\|X\|_\infty = \sup_{|\psi\rangle : \||\psi\rangle\|_2 = 1} \langle\psi|X|\psi\rangle \tag{2.2.107}$$

$$= \sup_{\rho \geq 0}\{\text{Tr}[X\rho] : \text{Tr}[\rho] = 1\} \quad (X \text{ positive semi-definite}). \tag{2.2.108}$$

More generally, if X is Hermitian and $\{\lambda_k\}_{k=1}^{\text{rank}(X)}$ is the set of its eigenvalues, then

$$\|X\|_\infty = \sup_{|\psi\rangle : \||\psi\rangle\|_2 = 1} |\langle\psi|X|\psi\rangle| \tag{2.2.109}$$

$$= \max_{1 \leq k \leq \text{rank}(X)} |\lambda_k| \quad (X \text{ Hermitian}). \tag{2.2.110}$$

Exercise 2.22

Let U be a unitary operator. Prove that $\|U\|_\infty = 1$. More generally, prove that $\|V\|_\infty = 1$ for every isometry V.

2.2.9.2 Schatten 1-Norm (Trace Norm)

Another important special case of the Schatten α-norm is $\alpha = 1$. In this case, we refer to it as the *trace norm*, and by applying (2.2.88), it is equal to the sum of the singular values of X:

$$\|X\|_1 = \sum_{k=1}^{\text{rank}(X)} s_k. \tag{2.2.111}$$

If X is Hermitian and positive semi-definite, then $\|X\|_1$ is equal to the sum of the eigenvalues of X, i.e., to the trace of X:

$$\|X\|_1 = \mathrm{Tr}[X] \quad \text{(X positive semi-definite)}. \tag{2.2.112}$$

More generally, if X is Hermitian and $\{\lambda_k\}_{k=1}^{\mathrm{rank}(X)}$ is the set of its eigenvalues, then

$$\|X\|_1 = \sum_{k=1}^{\mathrm{rank}(X)} |\lambda_k| \quad \text{(X Hermitian)}. \tag{2.2.113}$$

Exercise 2.23

Consider two vectors $|\psi\rangle, |\phi\rangle \in \mathbb{C}^d$, with $d \geq 2$. Show that

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1^2 = (\langle\psi|\psi\rangle + \langle\phi|\phi\rangle)^2 - 4\,|\langle\psi|\phi\rangle|^2 . \tag{2.2.114}$$

We now provide a proof of the variational characterization of the Schatten norm in (2.2.97) for the special case of $\alpha = 1$ and $\beta = \infty$.

Proposition 2.10 Variational Characterization of Trace Norm

For all $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$, the trace norm of X has the following variational characterization:

$$\|X\|_1 = \sup_{Y:\|Y\|_\infty \leq 1} \left|\mathrm{Tr}[Y^\dagger X]\right|, \tag{2.2.115}$$

where the optimization is with respect to all $Y \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ with spectral norm bounded from above by one.

PROOF: Let $X = \sum_{k=1}^r s_k |e_k\rangle_B\langle f_k|_A$ be the singular value decomposition of X, where $r := \mathrm{rank}(X)$. Let $Y \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ be such that $\|Y\|_\infty \leq 1$. Then,

$$\left|\mathrm{Tr}[Y^\dagger X]\right| = \left|\mathrm{Tr}\left[Y^\dagger \left(\sum_{k=1}^r s_k |e_k\rangle_B\langle f_k|_A\right)\right]\right| \tag{2.2.116}$$

$$= \left|\sum_{k=1}^r s_k \langle e_k|_B Y |f_k\rangle_A\right| \tag{2.2.117}$$

$$\leq \sum_{k=1}^r s_k \,|\langle e_k|_B Y |f_k\rangle_A| , \tag{2.2.118}$$

where the last line is due to the triangle inequality. Now, using (2.2.106), we have

$$|\langle e_k|_B Y|f_k\rangle_A| \leq \|Y\|_\infty \leq 1, \tag{2.2.119}$$

for every $k \in \{1, 2, \ldots, r\}$. Therefore,

$$\left|\text{Tr}[Y^\dagger X]\right| \leq \sum_{k=1}^{r} s_k = \|X\|_1 \,, \tag{2.2.120}$$

which holds for every $Y \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$ satisfying $\|Y\|_\infty \leq 1$, so that the inequality

$$\sup_{Y:\|Y\|_\infty \leq 1} \left|\text{Tr}[Y^\dagger X]\right| \leq \|X\|_1 \tag{2.2.121}$$

holds. The opposite inequality holds by making a particular choice for Y. We pick Y to be the following linear operator defined from the singular value decomposition of X: $Y = \sum_{k=1}^{r} |e_k\rangle_B\langle f_k|_A$. Observe that $\|Y\|_\infty = 1$. Thus,

$$\sup_{Y:\|Y\|_\infty \leq 1} \left|\text{Tr}[Y^\dagger X]\right| \geq \left|\text{Tr}\left[\left(\sum_{k'=1}^{r} |f_{k'}\rangle_A\langle e_{k'}|_B\right)\left(\sum_{k=1}^{r} s_k|e_k\rangle_B\langle f_k|_A\right)\right]\right| \tag{2.2.122}$$

$$= \left|\sum_{k=1}^{r} s_k\right| \tag{2.2.123}$$

$$= \|X\|_1 \,. \tag{2.2.124}$$

This completes the proof. ■

Remark: Observe that Proposition 2.10 can be generalized as follows for every linear operator $X_{A\to B} \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$:

$$\|X\|_1 = \sup_{Y:\|Y\|_\infty \leq 1} \text{Re}\left(\text{Tr}[Y^\dagger X]\right), \tag{2.2.125}$$

where, as before, the optimization is with respect to every operator $Y \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$ with spectral norm bounded from above by one. Indeed, for every complex number $z \in \mathbb{C}$, the inequality $\text{Re}(z) \leq |\text{Re}(z)| \leq |z|$ holds, which means that

$$\sup_{Y:\|Y\|_\infty \leq 1} \text{Re}\left(\text{Tr}[Y^\dagger X]\right) \leq \sup_{Y:\|Y\|_\infty \leq 1} \left|\text{Tr}[Y^\dagger X]\right| = \|X\|_1 \,. \tag{2.2.126}$$

Then, to obtain the opposite inequality, the same choice for Y as in the the proof of Proposition 2.10 can be made, because for that choice of Y we have $\text{Tr}[Y^\dagger X] = \|X\|_1$, which is real, so that $\text{Re}(\text{Tr}[Y^\dagger X]) = \|X\|_1$. We can thus conclude (2.2.125).

We also remark that in both (2.2.115) and (2.2.125), it suffices to optimize with respect to isometries. In particular, because $\|U\|_\infty = 1$ for every isometry U (see Exercise 2.22), using similar techniques as in the proof of Proposition 2.10, it is straightforward to prove that for all $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$,

$$\|X\|_1 = \sup_{\substack{U_{B\to A} \\ \text{isometry}}} |\mathrm{Tr}[U_{B\to A} X_{A\to B}]| = \sup_{\substack{U_{B\to A} \\ \text{isometry}}} \mathrm{Re}\,(\mathrm{Tr}[U_{B\to A} X_{A\to B}])\,, \quad d_A \geq d_B, \tag{2.2.127}$$

$$\|X\|_1 = \sup_{\substack{V_{A\to B} \\ \text{isometry}}} \left|\mathrm{Tr}[V_{A\to B}(X_{A\to B})^\dagger]\right| = \sup_{\substack{V_{A\to B} \\ \text{isometry}}} \mathrm{Re}\left(\mathrm{Tr}[V_{A\to B}(X_{A\to B})^\dagger]\right), \quad d_A \leq d_B. \tag{2.2.128}$$

In particular, if $d_A = d_B = d$, then the optimization in (2.2.127) and (2.2.128) is with respect to unitary operators, so that for all $X \in \mathrm{L}(\mathbb{C}^d)$,

$$\|X\|_1 = \sup_{\substack{U \in \mathrm{L}(\mathbb{C}^d) \\ \text{unitary}}} |\mathrm{Tr}[UX]| = \sup_{\substack{U \in \mathrm{L}(\mathbb{C}^d) \\ \text{unitary}}} \mathrm{Re}\,(\mathrm{Tr}[UX])\,. \tag{2.2.129}$$

The monotonicity result in Proposition 2.7 implies that

$$\|X\|_\infty \leq \|X\|_1\,. \tag{2.2.130}$$

for every linear operator X. The following proposition gives a tighter bound than the one in (2.2.130) for the special case when X is a traceless Hermitian operator.

Lemma 2.11

Let X be a Hermitian operator satisfying $\mathrm{Tr}[X] = 0$. Then,

$$\|X\|_\infty \leq \frac{1}{2}\|X\|_1\,. \tag{2.2.131}$$

Proof: Let the Jordan–Hahn decomposition of X be given by

$$X = X_+ - X_-, \tag{2.2.132}$$

where $X_+, X_- \geq 0$ and $X_+X_- = 0$. Then,

$$\|X\|_1 = \mathrm{Tr}[X_+] + \mathrm{Tr}[X_-]. \tag{2.2.133}$$

Since $\mathrm{Tr}[X] = 0$, we have that $\mathrm{Tr}[X_+] = \mathrm{Tr}[X_-]$, which means that

$$\|X\|_1 = 2\mathrm{Tr}[X_+]. \tag{2.2.134}$$

We also have that

$$\|X\|_\infty = \max\{\|X_+\|_\infty, \|X_-\|_\infty\}, \tag{2.2.135}$$

because $X_+X_- = 0$. Then, since $\|X_\pm\|_\infty \leq \|X_\pm\|_1 = \mathrm{Tr}X_\pm$ by (2.2.130) and (2.2.112), we deduce that

$$\|X\|_\infty = \max\{\|X_+\|_\infty, \|X_-\|_\infty\} \leq \mathrm{Tr}[X_+] = \frac{1}{2}\|X\|_1, \tag{2.2.136}$$

thus concluding the proof. ■

We remark that the monotonicity inequality $\|X\|_\infty \leq \|X\|_1$ in (2.2.130) can be reversed to give

$$\|X\|_1 \leq d\,\|X\|_\infty \tag{2.2.137}$$

for every linear operator X acting on a d-dimensional Hilbert space. This holds because

$$\|X\|_1 = \sum_{k=1}^{r} s_k \leq \sum_{k=1}^{r} \max_{k\in\{1,\dots,r\}} s_k = r \max_{k\in\{1,\dots,r\}} s_k = r\,\|X\|_\infty \leq d\,\|X\|_\infty, \tag{2.2.138}$$

where $\{s_k\}_{k=1}^r$ is the set of singular values of X and $r := \mathrm{rank}(X)$. (We used this very same reasoning in the proof of Proposition 2.9.)

Using Proposition 2.10, we can establish the following slight strengthening of the Hölder inequality in (2.2.98):

$$\left\|Z^\dagger X\right\|_1 \leq \|X\|_\alpha \|Z\|_\beta, \tag{2.2.139}$$

which holds for all linear operators X and Z and $\alpha, \beta \in [1,\infty]$ satisfying $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. This actually follows by a direct application of the Hölder inequality itself:

$$\left|\mathrm{Tr}[UZ^\dagger X]\right| \leq \|X\|_\alpha \left\|ZU^\dagger\right\|_\beta = \|X\|_\alpha \|Z\|_\beta, \tag{2.2.140}$$

which holds for every isometry U. Therefore, it follows from (2.2.115) that

$$\left\|Z^\dagger X\right\|_1 = \sup_U \left|\mathrm{Tr}[UZ^\dagger X]\right| \leq \|X\|_\alpha \|Z\|_\beta. \tag{2.2.141}$$

2.2.10 Operator Inequalities

Throughout this book, we make use of the *Löwner partial order* for Hermitian operators. It is useful as a way of comparing two Hermitian operators in $\mathrm{L}(\mathcal{H})$, generalizing the way in which we compare two real numbers.

Definition 2.12 Löwner Partial Order for Hermitian Operators

For two Hermitian operators X and Y, the expression $X \geq Y$ is an *operator inequality* and means that $X - Y \geq 0$, i.e., that $X - Y$ is positive semi-definite. We also write $X \leq Y$ to mean $Y - X \geq 0$. The expressions $X > Y$ and $Y < X$ mean that $X - Y$ is positive definite.

The relations "$\geq$" and "$\leq$" satisfy the following expected properties: $X \leq Y$ and $X \geq Y$ imply that $X = Y$, and $X \leq Y$ and $Y \leq Z$ imply that $X \leq Z$. The term "partial order" is used because not every pair (X, Y) of Hermitian operators satisfies either $X \geq Y$ or $X \leq Y$.

Definition 2.13 Operator Convex, Concave, Monotone Functions

Let X and Y be Hermitian operators, and let $f : \mathbb{R} \to \mathbb{R}$ be a function extended to Hermitian operators as in (2.2.70).

1. The function f is called *operator convex* if for all $\lambda \in [0, 1]$ and Hermitian operators X and Y, the following inequality holds:

$$f(\lambda X + (1 - \lambda)Y) \leq \lambda f(X) + (1 - \lambda) f(Y). \tag{2.2.142}$$

   We call f *operator concave* if $-f$ is operator convex.

2. The function f is called *operator monotone* if $X \leq Y$ implies $f(X) \leq f(Y)$ for all Hermitian operators X and Y. We call f *operator anti-monotone* if $-f$ is operator monotone.

The functions considered in Section 2.2.8.1 have the following properties with respect to Definition 2.13:

- The function $x \mapsto x^\alpha$ is operator monotone for $\alpha \in [0, 1]$ and $x \in [0, \infty)$, operator anti-monotone for $\alpha \in [-1, 0)$ and $x \in (0, \infty)$, operator convex for $\alpha \in [-1, 0)$ and $x \in (0, \infty)$, operator convex for $[1, 2]$ and $x \in [0, \infty)$, and operator concave for $\alpha \in (0, 1]$ and $x \in [0, \infty)$. Note that the function $x \mapsto x^\alpha$ is neither operator monotone, operator convex, nor operator concave for $\alpha < -1$ and $\alpha > 2$.

- The function $x \mapsto \log_b(x)$, for every base $b > 0$ and $x \in (0, \infty)$, is operator monotone and operator concave.

- The function $x \mapsto x \log_b(x)$, for every base $b > 0$ and $x \in [0, \infty)$, is operator convex[3].

For proofs of these properties, please see the Bibliographic Notes (Section 2.6). We note here that these properties are critical for understanding quantum entropies, as detailed in Chapter 7. Especially, the data-processing inequality for quantum relative entropy, which is at the heart of understanding quantum communication limits, is intimately related to operator convexity.

We now state some basic operator inequalities that we use repeatedly throughout the book.

Lemma 2.14 Basic Operator Inequalities

Let $X, Y \in \mathrm{L}(\mathcal{H})$ be Hermitian operators acting on a Hilbert space $\mathcal{H}$.

1. $X \geq 0 \Rightarrow ZXZ^\dagger \geq 0$ for all $Z \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$. In particular, $X \geq Y \Rightarrow ZXZ^\dagger \geq ZYZ^\dagger$ for all $Z \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$.
2. $X \geq Y \Rightarrow \mathrm{Tr}[X] \geq \mathrm{Tr}[Y]$. More generally, $X \geq Y \Rightarrow \mathrm{Tr}[WX] \geq \mathrm{Tr}[WY]$ for all $W \in \mathrm{L}(\mathcal{H})$ satisfying $W \geq 0$.
3. For every Hermitian operator X with maximum and minimum eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$, respectively, $\lambda_{\min}\mathbb{1} \leq X \leq \lambda_{\max}\mathbb{1}$.
4. Let X and Y have their spectrum in some interval $I \subset \mathbb{R}$, and let $f : I \to \mathbb{R}$ be a monotone increasing function. If $X \leq Y$, then $\mathrm{Tr}[f(X)] \leq \mathrm{Tr}[f(Y)]$. In particular, if X and Y are positive semi-definite, then

$$0 \leq X \leq Y \Rightarrow \mathrm{Tr}[X^\alpha] \leq \mathrm{Tr}[Y^\alpha] \quad \forall\, \alpha > 0. \tag{2.2.143}$$

(Note that we are *not* requiring f to be *operator* monotone.)

PROOF:

1. $X \geq 0$ implies that $\langle\psi|X|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$. Then, for every vector $|\phi\rangle \in \mathcal{H}'$, we have $\langle\phi|ZXZ^\dagger|\phi\rangle \geq 0$ because $Z^\dagger|\phi\rangle \equiv |\psi\rangle$ is some vector in $\mathcal{H}$. Therefore, $ZXZ^\dagger \geq 0$.

   Now, $X \geq Y$ is equivalent to $X - Y \geq 0$. Let $W = X - Y$. Then, from the

[3] Note that, because $\lim_{x\to 0} x \log_b(x) = 0$, we take the convention that $0 \log_b(0) = 0$ throughout this book.

arguments in the previous paragraph, we have $ZWZ^\dagger \geq 0$ for all Z, which implies that $ZXZ^\dagger - ZYZ^\dagger \geq 0$, i.e., $ZXZ^\dagger \geq ZYZ^\dagger$, as required.

2. $X \geq Y$ implies that $X - Y \geq 0$. The trace of a positive semi-definite operator is non-negative, since positive semi-definite operators have non-negative eigenvalues and the trace of every normal operator is equal to the sum of its eigenvalues. Thus, $X - Y \geq 0$ implies $\text{Tr}[X - Y] \geq 0$, which implies that $\text{Tr}[X] \geq \text{Tr}[Y]$, as required.

   Next, let W be a positive semi-definite operator. Using 1. above, $X \geq Y$ implies that $\sqrt{W}X\sqrt{W} \geq \sqrt{W}Y\sqrt{W}$. Then, using the result of the previous paragraph, we obtain $\text{Tr}[\sqrt{W}X\sqrt{W}] \geq \text{Tr}[\sqrt{W}Y\sqrt{W}]$. Finally, by cyclicity of the trace (recall (2.2.23)), we find that $\text{Tr}[WX] \geq \text{Tr}[WY]$, as required.

3. This result follows from the fact that, for every Hermitian operator $X \in \mathrm{L}(\mathcal{H})$ with eigenvalues $\{\lambda_k\}_{k=1}^{\dim(\mathcal{H})}$, the eigenvalues of $X + t\mathbb{1}$ are equal to $\{\lambda_k + t\}_{k=1}^{\dim(\mathcal{H})}$ for every $t \in \mathbb{R}$. In particular, then, by definition of the minimum eigenvalue, $X - \lambda_{\min}\mathbb{1} \geq 0$, because all of the eigenvalues of $X - \lambda_{\min}\mathbb{1}$ are non-negative. Similarly, by definition of the maximum eigenvalue, $X - \lambda_{\max}\mathbb{1} \leq 0$, because all of the eigenvalues of $X - \lambda_{\max}\mathbb{1}$ are non-positive.

4. Let $\lambda_i^\downarrow(X)$ denote the sequence of decreasingly ordered eigenvalues of X. Then the inequalities $\lambda_i^\downarrow(X) \leq \lambda_i^\downarrow(Y)$ hold for all $i \in \{1, \dots, \dim(\mathcal{H})\}$. These inequalities are a consequence of the Courant–Fischer–Weyl minimax principle (please consult the Bibliographic Notes in Section 2.6 for a reference to this principle). Then, the desired inequality follows directly from the fact that $\text{Tr}[f(X)] = \sum_{i=1}^{\dim(\mathcal{H})} f(\lambda_i^\downarrow(X))$, as well as the monotonicity of f. The inequality in (2.2.143) follows because the function x^α with domain $x \geq 0$ is monotone for all $\alpha > 0$. ■

Lemma 2.15 Araki–Lieb–Thirring Inequality

Let X and Y be positive semi-definite operators acting on a finite-dimensional Hilbert space. For all $q \geq 0$:

1. $\text{Tr}\left[\left(Y^{\frac{1}{2}}XY^{\frac{1}{2}}\right)^{rq}\right] \geq \text{Tr}\left[\left(Y^{\frac{r}{2}}X^rY^{\frac{r}{2}}\right)^{q}\right]$ for all $0 \leq r < 1$.

2. $\text{Tr}\left[\left(Y^{\frac{1}{2}}XY^{\frac{1}{2}}\right)^{rq}\right] \leq \text{Tr}\left[\left(Y^{\frac{r}{2}}X^rY^{\frac{r}{2}}\right)^{q}\right]$ for all $r \geq 1$.

Proof: Please consult the Bibliographic Notes (Section 2.6). ■

The operator Jensen inequality below is the linchpin of several quantum data-processing inequalities presented later on in Chapter 7. These in turn are repeatedly used in Parts II and III to place fundamental limits on quantum communication protocols. As such, the operator Jensen inequality is a significant bridge that connects convexity to information processing.

Theorem 2.16 Operator Jensen Inequality

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function with $\text{dom}(f) = I \subset \mathbb{R}$ (where I is an interval). Then, the following are equivalent:

1. f is operator convex.
2. For all $n \in \mathbb{N}$, the inequality
$$f\left(\sum_{k=1}^{n} A_k^\dagger X_k A_k\right) \leq \sum_{k=1}^{n} A_k^\dagger f(X_k) A_k \tag{2.2.144}$$
holds for every collection $\{X_k\}_{k=1}^n$ of Hermitian operators acting on a Hilbert space $\mathcal{H}$ with spectrum contained in I and every collection $\{A_k\}_{k=1}^n$ of linear operators in $\mathrm{L}(\mathcal{H}', \mathcal{H})$ satisfying $\sum_{k=1}^n A_k^\dagger A_k = \mathbb{1}_{\mathcal{H}'}$.
3. For every Hermitian operator $X \in \mathrm{L}(\mathcal{H})$ with spectrum in I and every isometry $V \in \mathrm{L}(\mathcal{H}', \mathcal{H})$, the following inequality holds:
$$f(V^\dagger X V) \leq V^\dagger f(X) V. \tag{2.2.145}$$

Proof: We first prove that 2. ⇒ 1. Let X and Y be Hermitian operators with their eigenvalues in I. Let $\lambda \in [0, 1]$. We can take $n = 2$, $A_1 = \sqrt{\lambda}\mathbb{1}$, $X_1 = X$, $A_2 = \sqrt{1-\lambda}\mathbb{1}$, $X_2 = Y$, and the following operator inequality is an immediate consequence of (2.2.144):

$$f(\lambda X + (1-\lambda)Y) \leq \lambda f(X) + (1-\lambda) f(Y). \tag{2.2.146}$$

Since X, Y, and λ are arbitrary, it follows that f is operator convex.

3. is actually a special case of 2., found by setting $n = 1$ and taking $A_1 = V$ and $X_k = X$, with V an isometry and X Hermitian with eigenvalues in I.

Now we prove that 3. ⇒ 2. Fix $n \in \mathbb{N}$ and the sets $\{A_k\}_{k=1}^n$ and $\{X_k\}_{k=1}^n$ of operators such that they satisfy the conditions specified in 2. Define the following Hermitian operator:

$$X := \sum_{k=1}^{n} X_k \otimes |k\rangle\langle k|, \tag{2.2.147}$$

as well as the isometry

$$V := \sum_{k=1}^{n} A_k \otimes |k\rangle, \tag{2.2.148}$$

where $\{|k\rangle\}_{k=1}^n$ is an orthonormal basis. The condition $\sum_{k=1}^n A_k^\dagger A_k = \mathbb{1}$ and a calculation imply that V is an isometry (satisfying $V^\dagger V = \mathbb{1}$). Another calculation implies that

$$V^\dagger X V = \sum_{k=1}^{n} A_k^\dagger X_k A_k. \tag{2.2.149}$$

Since

$$f(X) = f\left(\sum_{k=1}^{n} X_k \otimes |k\rangle\langle k|\right) = \sum_{k=1}^{n} f(X_k) \otimes |k\rangle\langle k|, \tag{2.2.150}$$

which follows as a consequence of (2.2.70), a similar calculation implies that

$$V^\dagger f(X) V = \sum_{k=1}^{n} A_k^\dagger f(X_k) A_k. \tag{2.2.151}$$

Then the desired inequality in (2.2.144) follows from (2.2.149), (2.2.151), and (2.2.145).

We finally prove that 1. ⇒ 3. Fix the operator X and isometry V, as specified in 3. Let M be a Hermitian operator in $\mathrm{L}(\mathcal{H}')$ with spectrum in I. Let $P := \mathbb{1}_{\mathcal{H}} - VV^\dagger$, and observe that P is a projection (i.e., $P^2 = P$), $V^\dagger P = 0$, and $PV = 0$. Set

$$Z := \begin{pmatrix} X & 0 \\ 0 & M \end{pmatrix}, \qquad U := \begin{pmatrix} V & P \\ 0 & -V^\dagger \end{pmatrix}, \qquad W := \begin{pmatrix} V & -P \\ 0 & V^\dagger \end{pmatrix}. \tag{2.2.152}$$

Observe that U and W are unitary operators (these are called unitary dilations of the isometry V). By direct calculation, we then find that

$$U^\dagger Z U = \begin{pmatrix} V^\dagger X V & V^\dagger X P \\ PXV & PXP + VMV^\dagger \end{pmatrix}, \tag{2.2.153}$$

$$W^\dagger Z W = \begin{pmatrix} V^\dagger X V & -V^\dagger X P \\ -PXV & PXP + VMV^\dagger \end{pmatrix}, \tag{2.2.154}$$

so that

$$\frac{1}{2}\left(U^\dagger Z U + W^\dagger Z W\right) = \begin{pmatrix} V^\dagger X V & 0 \\ 0 & PXP + VMV^\dagger \end{pmatrix}. \tag{2.2.155}$$

From the same reasoning that leads to (2.2.150), and using (2.2.155), we find that

$$\begin{aligned} &\begin{pmatrix} f(V^\dagger X V) & 0 \\ 0 & f\left(PXP + VBV^\dagger\right) \end{pmatrix} \\ &\quad = f\left(\begin{pmatrix} V^\dagger X V & 0 \\ 0 & PXP + VBV^\dagger \end{pmatrix}\right) &(2.2.156)\\ &\quad = f\left(\frac{1}{2}\left(U^\dagger Z U + W^\dagger Z W\right)\right) &(2.2.157)\\ &\quad \leq \frac{1}{2} f\left(U^\dagger Z U\right) + \frac{1}{2} f\left(W^\dagger Z W\right) &(2.2.158)\\ &\quad = \frac{1}{2} U^\dagger f(Z)\, U + \frac{1}{2} W^\dagger f(Z)\, W &(2.2.159)\\ &\quad = \begin{pmatrix} V^\dagger f(X) V & 0 \\ 0 & P f(X) P + V f(B) V^\dagger \end{pmatrix}. &(2.2.160) \end{aligned}$$

The inequality follows from the assumption that f is operator convex. The third equality follows from (2.2.70). The final equality follows because

$$f(Z) = \begin{pmatrix} f(X) & 0 \\ 0 & f(M) \end{pmatrix}, \tag{2.2.161}$$

and by applying (2.2.155) again, with the substitutions $Z \to f(Z)$, $X \to f(X)$, and $M \to f(M)$. It follows that

$$\begin{pmatrix} f\left(V^\dagger X V\right) & 0 \\ 0 & f\left(PXP + VBV^\dagger\right) \end{pmatrix} \leq \begin{pmatrix} V^\dagger f(X) V & 0 \\ 0 & P f(X) P + V f(B) V^\dagger \end{pmatrix}, \tag{2.2.162}$$

and we finally conclude that $f(V^\dagger X V) \leq V^\dagger f(X) V$ by examining the upper left blocks in the operator inequality in (2.2.162). ■

2.2.11 Superoperators

Just as we have been considering linear operators of the form $X : \mathcal{H}_A \to \mathcal{H}_B$, with input and output Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, we can consider linear operators with input Hilbert space $\mathrm{L}(\mathcal{H}_A)$ and output Hilbert space $\mathrm{L}(\mathcal{H}_B)$. We use the term *superoperator* to refer to a linear operator acting on the Hilbert space of linear operators. Specifically, a superoperator is a function $\mathcal{N} : \mathrm{L}(\mathcal{H}_A) \to \mathrm{L}(\mathcal{H}_B)$ such that

$$\mathcal{N}(\alpha X + \beta Y) = \alpha\mathcal{N}(X) + \beta\mathcal{N}(Y) \tag{2.2.163}$$

for all $\alpha, \beta \in \mathbb{C}$ and $X, Y \in \mathrm{L}(\mathcal{H}_A)$. It is often helpful to indicate explicitly the input and output Hilbert spaces of a superoperator $\mathcal{N} : \mathrm{L}(\mathcal{H}_A) \to \mathrm{L}(\mathcal{H}_B)$ by writing $\mathcal{N}_{A\to B}$. We make use of this notation throughout the book.

For every superoperator $\mathcal{N}_{A\to B}$, there exists $n \in \mathbb{N}$, and sets $\{K_i\}_{i=1}^n$ and $\{L_i\}_{i=1}^n$ of operators in $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ such that

$$\mathcal{N}_{A\to B}(X_A) = \sum_{i=1}^{n} K_i X_A L_i^\dagger, \tag{2.2.164}$$

for all $X_A \in \mathrm{L}(\mathcal{H}_A)$. This follows as a consequence of the requirement that $\mathcal{N}_{A\to B}$ has a linear action on X_A and the isomorphism in (2.2.33). The transpose operation discussed previously in (2.2.27) is an example of a superoperator. In Chapter 4, we see that quantum physical evolutions of quantum states, known as quantum channels, are other examples of superoperators with additional constraints on the sets $\{K_i\}_{i=1}^n$ and $\{L_i\}_{i=1}^n$.

We denote the identity superoperator by $\mathrm{id} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H})$, and by definition it satisfies $\mathrm{id}(X) = X$ for all $X \in \mathrm{L}(\mathcal{H})$.

Given two superoperators $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$, their tensor product $\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'}$ is the unique superoperator with input Hilbert space $\mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and output Hilbert space $\mathrm{L}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$, such that

$$(\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(X_A \otimes Y_B) = \mathcal{N}_{A\to A'}(X_A) \otimes \mathcal{M}_{B\to B'}(Y_B) \tag{2.2.165}$$

for all $X_A \in \mathrm{L}(\mathcal{H}_A)$ and $Y_B \in \mathrm{L}(\mathcal{H}_B)$. We use the abbreviation

$$\mathcal{N}_{A\to A'} \otimes \mathrm{id}_{B\to B'} \equiv \mathcal{N}_{A\to A'}, \quad \mathrm{id}_{A\to A'} \otimes \mathcal{M}_{B\to B'} \equiv \mathcal{M}_{B\to B'} \tag{2.2.166}$$

throughout this book whenever a superoperator acts only on one of the tensor factors of the underlying Hilbert space of linear operators.

Definition 2.17 Hermiticity-Preserving Superoperator

A superoperator $\mathcal{N}$ is called *Hermiticity preserving* if $\mathcal{N}(X)$ is Hermitian for every Hermitian input X.

Exercise 2.24

Using (2.2.54), prove that a superoperator $\mathcal{N}$ is Hermiticity preserving if and only if $\mathcal{N}(X^\dagger) = \mathcal{N}(X)^\dagger$ for every linear operator X.

Definition 2.18 Adjoint of a Superoperator

The *adjoint* of a superoperator $\mathcal{N} : \mathrm{L}(\mathcal{H}_A) \to \mathrm{L}(\mathcal{H}_{A'})$ is the unique superoperator $\mathcal{N}^\dagger : \mathrm{L}(\mathcal{H}_{A'}) \to \mathrm{L}(\mathcal{H}_A)$ that satisfies

$$\langle Y, \mathcal{N}(X)\rangle = \langle \mathcal{N}^\dagger(Y), X\rangle \tag{2.2.167}$$

for all $X \in \mathrm{L}(\mathcal{H}_A)$ and $Y \in \mathrm{L}(\mathcal{H}_{A'})$, where $\langle \cdot, \cdot \rangle$ is the Hilbert–Schmidt inner product defined in (2.2.30).

Exercise 2.25

Let $\mathcal{N}_{A\to B}$ be a superoperator represented as in (2.2.164).

1. Prove that the adjoint $\mathcal{N}^\dagger$ is given by $\mathcal{N}^\dagger(Y) = \sum_{i=1}^n K_i^\dagger Y L_i$ for every linear operator Y.

2. If $\mathcal{N}$ is Hermiticity preserving, then prove that an alternate operator-sum representation of $\mathcal{N}$ is $\mathcal{N}(X) = \sum_{i=1}^n L_i X K_i^\dagger$ for all $X \in \mathrm{L}(\mathcal{H}_A)$.

3. Using 1. and 2., prove that if $\mathcal{N}$ is Hermiticity preserving, then so is its adjoint $\mathcal{N}^\dagger$.

Definition 2.19 Trace-Preserving and Unital Superoperator

Let $\mathcal{N}_{A\to B}$ be a superoperator.

1. $\mathcal{N}$ is called *trace preserving* if $\text{Tr}[\mathcal{N}(X)] = \text{Tr}[X]$ for all $X \in \text{L}(\mathcal{H}_A)$.

2. $\mathcal{N}$ is called *unital* if $\mathcal{N}(\mathbb{1}_A) = \mathbb{1}_B$.

REMARK: Observe that if $\mathcal{N}$ is trace preserving and unital, and if $\mathcal{H}_A$ and $\mathcal{H}_B$ have finite dimensions, then we find that $d_A = d_B$, by taking the trace on both sides of $\mathcal{N}(\mathbb{1}_A) = \mathbb{1}_B$. This means that, in the finite-dimensional case, it is necessary for trace-preserving and unital superoperators to have the same input and output dimensions.

Exercise 2.26

Let $\mathcal{N}_{A\to B}$ be a trace-preserving superoperator represented as in (2.2.164).

1. Prove that $\sum_{i=1}^{n} K_i^\dagger L_i = \mathbb{1}_A$.

2. Using 1., show that the adjoint $\mathcal{N}^\dagger$ is unital. Thus, the adjoint of every trace-preserving superoperator is unital.

For every superoperator $\mathcal{N}_{A\to B} : \text{L}(\mathcal{H}_A) \to \text{L}(\mathcal{H}_B)$, we define its *induced trace norm* $\|\mathcal{N}\|_1$ as

$$\|\mathcal{N}\|_1 := \sup\left\{\frac{\|\mathcal{N}(X)\|_1}{\|X\|_1} : X \in \text{L}(\mathcal{H}_A), X \neq 0\right\} \tag{2.2.168}$$

$$= \sup\{\|\mathcal{N}(X)\|_1 : X \in \text{L}(\mathcal{H}_A), \|X\|_1 \leq 1\}. \tag{2.2.169}$$

Then, for all $X \in \text{L}(\mathcal{H}_A)$, it immediately follows that

$$\|\mathcal{N}(X)\|_1 \leq \|\mathcal{N}\|_1 \|X\|_1 . \tag{2.2.170}$$

Exercise 2.27

Prove that

$$\|\mathcal{N}\|_1 = \sup_{\substack{U\in \text{L}(\mathcal{H}_B)\\ \text{unitary}}} \left\|\mathcal{N}^\dagger(U)\right\|_\infty \tag{2.2.171}$$

for every superoperator $\mathcal{N}_{A\to B}$, where the optimization is with respect to every unitary operator U acting on $\mathcal{H}_B$.

Definition 2.20 Diamond Norm

Let $\mathcal{N}_{A\to B}$ be a superoperator. The quantity

$$\|\mathcal{N}\|_\diamond := \sup\{\|(\mathrm{id}_R \otimes \mathcal{N}_{A\to B})(X_{RA})\|_1 : X_{RA} \in \mathrm{L}(\mathcal{H}_{RA}),\ \|X_{RA}\|_1 \leq 1\} \tag{2.2.172}$$

is known as the *diamond norm* of $\mathcal{N}$, where the optimization is with respect to every linear operator X_{RA}, and there is an implicit optimization over Hilbert spaces $\mathcal{H}_R$ of dimension $d_R \geq 1$.

Theorem 2.21

For every superoperator $\mathcal{N}_{A\to B}$,

$$\|\mathcal{N}\|_\diamond = \|\mathrm{id}_A \otimes \mathcal{N}_{A\to B}\|_1 \tag{2.2.173}$$

$$= \sup\{\|(\mathrm{id}_A \otimes \mathcal{N}_{A\to B})(|\psi\rangle\langle\phi|_{AA})\|_1 : \||\psi\rangle_{AA}\|_2 = \||\phi\rangle_{AA}\|_2 = 1\}. \tag{2.2.174}$$

Furthermore, if $\mathcal{N}$ is Hermiticity preserving, then

$$\|\mathcal{N}\|_\diamond = \sup\{\|(\mathrm{id}_A \otimes \mathcal{N}_{A\to B})(|\psi\rangle\langle\psi|_{AA})\|_1 : \||\psi\rangle_{AA}\|_2 = 1\}. \tag{2.2.175}$$

PROOF: Please see the Bibliographic Notes in Section 2.6. ■

We study the diamond norm in detail in Chapter 6 in the context of quantum channels.

2.3 Analysis and Probability

In this section, we briefly review some essential concepts from mathematical analysis, in particular the concepts of the limit, supremum, and infimum, as well as the continuity of real-valued functions. We also discuss compact sets, convex sets, and functions, as well as the basic notions of probability distributions.

2.3.1 Limits, Infimum, Supremum, and Continuity

Limit of a sequence

We start with the definition of the limit of a sequence of real numbers. A sequence $\{s_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ of real numbers is said to have the *limit* ℓ, written $\lim_{n\to\infty} s_n = \ell$, if for all $\varepsilon > 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that, for all $n \geq n_\varepsilon$, the inequality $|s_n - \ell| < \varepsilon$ holds.

One can think of the concept of a limit intuitively as a game between two players, an antagonist and a protagonist. The antagonist goes first, and gets to pick an arbitrary $\varepsilon > 0$. The protagonist wins if he reports back an entry in the sequence $\{s_n\}_n$ such that $|s_n - \ell| < \varepsilon$. If the protagonist reports back an entry s_n such that $|s_n - \ell| \geq \varepsilon$, then the antagonist wins. If the limit exists and is equal to ℓ, then the protagonist always wins by taking n sufficiently large (i.e., larger than n_ε) and then reporting back s_n. If the limit does not exist or if the limit is not equal to ℓ, then the protagonist cannot necessarily win with the strategy of taking n sufficiently large; in this case, there exists a choice of $\varepsilon > 0$, such that for all $n_\varepsilon \in \mathbb{N}$, there exists $n \geq n_\varepsilon$ such that $|s_n - \ell| \geq \varepsilon$. In this latter case, the choice of $\varepsilon > 0$ can again be understood as a strategy of the antagonist.

Infimum and supremum

Let us now recall the concepts of the infimum and supremum of subsets of the real numbers. Roughly speaking, they are generalizations of the concepts of the minimum and maximum, respectively, of a set. Formally, let $E \subset \mathbb{R}$.

- A point $x \in \mathbb{R}$ is a *lower bound* of E if $y \geq x$ for all $y \in E$. If x is the greatest such lower bound, then x is called the *infimum* of E, and we write $x = \inf E$.

- A point $x \in \mathbb{R}$ is an *upper bound* of E if $y \leq x$ for all $y \in E$. If x is the least such upper bound, then x is called the *supremum* of E, and we write $x = \sup E$.

The supremum and infinum may or may not be contained in the subset E. For example, let $E = \{\frac{1}{n}\}_{n\in\mathbb{N}}$. Then, $\sup E = 1 \in E$, but $\inf E = 0 \notin E$. As another example, let $E = [0, 1)$. Then $\sup E = 1 \notin E$ and $\inf E = 0 \in E$. If the supremum is contained in E, then it is equal to the maximum element of E. Similarly, if the infimum is contained in E, then it is equal to the minimum element of E.

When considering a function $F : S \to \mathbb{R}$ defined on a subset S of $\mathrm{L}(\mathcal{H})$, its infimum and supremum are defined for the set $E = \{F(X) : X \in S\}$. Specifically,

$$\inf_{X \in S} F(X) \coloneqq \inf\{F(X) : X \in S\}, \tag{2.3.1}$$

and

$$\sup_{X \in S} F(X) \coloneqq \sup\{F(X) : X \in S\}. \tag{2.3.2}$$

Limit inferior and limit superior

Turning back to limits, the limit of a sequence need not always exist. A particularly illuminating example is the sequence $\{r^n\}_{n \in \mathbb{N}}$ for $r \in \mathbb{R}$. If $-1 < r < 1$, then the limit exists and is equal to zero. If $r > 1$, then the sequence never converges to a finite value and so the limit does not exist. We say that the sequence diverges to $+\infty$ in this case. If $r < -1$, then the sequence oscillates and diverges (but it does not specifically diverge to either $+\infty$ or $-\infty$). If $r = -1$, then the sequence oscillates back and forth between -1 and $+1$ and so the limit does not exist.

Given that the limit of a sequence need not always exist, it can be helpful to have a reasonable substitute for this asymptotic concept that does always exist. Such a substitute is provided by two quantities: the limit inferior and limit superior of a sequence. We now define the limit inferior and limit superior, noting that they can be understood as asymptotic versions of the infimum and supremum just discussed.

- We say that s is an *asymptotic lower bound* on the sequence $\{s_n\}_{n \in \mathbb{N}}$ if for all $\varepsilon > 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that, for all $n \geq n_\varepsilon$, the inequality $s_n > s - \varepsilon$ holds. The *limit inferior* is the greatest asymptotic lower bound and is denoted by
$$\liminf_{n \to \infty} s_n. \tag{2.3.3}$$
- The definition of the limit superior is essentially opposite to that of the limit inferior. We say that s is an *asymptotic upper bound* on the sequence $\{s_n\}_{n \in \mathbb{N}}$ if for all $\varepsilon > 0$, there exists $n_\varepsilon \in \mathbb{N}$, such that for all $n \geq n_\varepsilon$, the inequality $s_n < s + \varepsilon$ holds. The *limit superior* is the least asymptotic upper bound and is denoted by
$$\limsup_{n \to \infty} s_n. \tag{2.3.4}$$

The limit inferior and limit superior always exist by extending the real line $\mathbb{R}$ to include $-\infty$ and $+\infty$. Furthermore, every asymptotic lower bound on the sequence cannot exceed an asymptotic upper bound, implying that the following inequality holds for every sequence $\{s_n\}_{n\in\mathbb{N}}$:

$$\liminf_{n\to\infty} s_n \leq \limsup_{n\to\infty} s_n. \tag{2.3.5}$$

If the opposite inequality holds for a sequence $\{s_n\}_{n\in\mathbb{N}}$, then the limit of the sequence exists and we can write

$$\liminf_{n\to\infty} s_n = \limsup_{n\to\infty} s_n = \lim_{n\to\infty} s_n. \tag{2.3.6}$$

This collapse is a direct consequence of the definitions of limit, limit inferior, and limit superior.

Limits and continuity

We now consider the limit and continuity of a function. Specifically, we consider real-valued functions $F : \mathrm{L}(\mathcal{H}) \to \mathbb{R}$ defined on the space of linear operators acting on a Hilbert space $\mathcal{H}$. We view this space as a normed vector space with either the trace norm $\|\cdot\|_1$ or the spectral norm $\|\cdot\|_\infty$. In the definitions that follow, we use $\|\cdot\|$ to denote either one of these norms.

- *Limit*: We write $\lim_{X\to X_0} F(X) = y$ to the mean the following: for all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that $|F(X) - y| < \varepsilon$ for all $X \in \mathrm{L}(\mathcal{H})$ for which $\|X - X_0\| < \delta_\varepsilon$.

- *Continuity at a point*: We say that F is *continuous at* X_0 if for all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that $|F(X) - F(X_0)| < \varepsilon$ for every point X for which $\|X - X_0\| < \delta_\varepsilon$. F is said to be *continuous* if F is continuous at X_0 for all $X_0 \in \mathrm{L}(\mathcal{H})$.

- *Uniform continuity*: We say that f is *uniformly continuous* if for all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that $|F(X) - F(X')| < \varepsilon$ for all $X, X' \in \mathrm{L}(\mathcal{H})$ for which $\|X - X'\| < \delta_\varepsilon$.

- *Upper and lower semi-continuity*: We say that F is *upper semi-continuous at* X_0 if for all $\varepsilon > 0$, there exists a neighborhood $N_{X_0,\varepsilon}$ of X_0 such that: if $F(X_0) > -\infty$, then $F(X) \leq F(X_0) + \varepsilon$ for all $X \in N_{X_0,\varepsilon}$; if $F(X_0) = -\infty$, then $\lim_{X\to X_0} F(X) = -\infty$. F called *upper semi-continuous* if it is upper semi-continuous at X_0 for all $X_0 \in \mathrm{L}(\mathcal{H})$.

We say that f is *lower semi-continuous at* X_0 if for all $\varepsilon > 0$, there exists a neighborhood $N_{X_0,\varepsilon}$ of X_0 such that: if $F(X_0) < +\infty$, then $F(X) \geq F(X_0) - \varepsilon$ for all $X \in N_{X_0,\varepsilon}$; if $F(X_0) = +\infty$, then $\lim_{X\to X_0} F(X) = +\infty$.

2.3.2 Compact Sets

A subset S of a finite-dimensional topological vector space is called *compact* if every sequence of elements in S has a subsequence that converges to an element in S. For finite-dimensional vector spaces, a subset S is compact if and only if it is closed and bounded. An important fact about compact sets is that the infimum and supremum of continuous functions defined on compact sets are always achieved; in other words, the minimum and maximum exist and are thus attained by points in the set. In practice, for optimization problems over compact sets, the infimum can be replaced by a minimum and the supremum can be replaced by a maximum. More formally, if $F : S \to \mathbb{R}$ is a continuous function defined on a compact subset S of $\mathrm{L}(\mathcal{H})$, then

$$\inf_{X\in S} F(X) = \min_{X\in S} F(X) \quad \text{and} \quad \sup_{X\in S} F(X) = \max_{X\in S} F(X). \tag{2.3.7}$$

An important example of a compact set is the set $\{X \in \mathrm{L}(\mathcal{H}) : X \geq 0, \mathrm{Tr}[X] \leq 1\}$ of positive semi-definite operators with trace bounded from above by one. This set contains the set of density operators acting on $\mathcal{H}$.

2.3.3 Convex Sets and Functions

A subset C of a vector space is called *convex* if, for all elements $u, v \in C$ and for all $\lambda \in [0, 1]$, we have $\lambda u + (1 - \lambda)v \in C$. We often call $\lambda u + (1 - \lambda)v$ a *convex combination* of u and v. More generally, for every set $S = \{v_x\}_{x\in\mathcal{X}}$ of elements of a real vector space indexed by an alphabet $\mathcal{X}$, and every function $p : \mathcal{X} \to [0, 1]$ with $p(x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x\in\mathcal{X}} p(x) = 1$, the sum $\sum_{x\in\mathcal{X}} p(x)v_x$ is called a convex combination of the vectors in S. The *convex hull* of S is the convex set of all possible convex combinations of the vectors in S.

Throughout this book, in the context of convex sets and functions, we consider the real vector space of Hermitian operators acting on some Hilbert space. Then, an important example of a convex subset is the set of all positive semi-definite operators. Indeed, if X and Y are positive semi-definite operators, then $\lambda X+(1-\lambda)Y$ is a positive semi-definite operator for all $\lambda \in [0,1]$. From now on, we assume C to be a convex subset of the set of Hermitian operators, and we use X, Y, and Z to denote arbitrary elements of C.

An element $Z \in C$ is called an *extreme point of* C if Z cannot be written as a non-trivial convex combination of other vectors in C. Formally, Z is an extreme point if every decomposition of Z as the convex combination $Z = \lambda X + (1-\lambda)Y$, such that $\lambda \in (0,1)$ (so that the decomposition is non-trivial), implies that $X = Y = Z$. An important fact is that every convex set is equal to the convex hull of its extreme points.

We now define convex and concave functions.

Definition 2.22 Convex and Concave Functions

A function $F : C \to \mathbb{R}$ defined on a convex subset $C \subseteq \mathrm{L}(\mathcal{H})$ is a *convex function* if, for all $X, Y \in C$ and $\lambda \in [0,1]$, the following inequality holds:

$$F(\lambda X + (1-\lambda)Y) \leq \lambda F(X) + (1-\lambda)F(Y). \tag{2.3.8}$$

A function $F : C \to \mathbb{R}$ is a *concave function* if $-F$ is a convex function.

It follows from an inductive argument using Definition 2.22 that a convex function $F : C \to \mathbb{R}$ on a convex subset $C \subset \mathrm{L}(\mathcal{H})$ satisfies

$$F\left(\sum_{x\in\mathcal{X}} p(x)X_x\right) \leq \sum_{x\in\mathcal{X}} p(x)F(X_x) \tag{2.3.9}$$

for every set $\{X_x\}_{x\in\mathcal{X}}$ of elements in C and every function $p : \mathcal{X} \to [0,1]$ defined on $\mathcal{X}$, such that $p(x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x\in\mathcal{X}} p(x) = 1$.

- A function $F : C \to \mathbb{R}$ defined on a convex subset $C \subseteq \mathrm{L}(\mathcal{H})$ is called *quasi-convex* if for all $X, Y \in C$ and all $\lambda \in [0,1]$ the following inequality holds:
$$F(\lambda X + (1-\lambda)Y) \leq \max\{F(X), F(Y)\}. \tag{2.3.10}$$
- A function $F : C \to \mathbb{R}$ defined on a convex subset $C \subseteq \mathrm{L}(\mathcal{H})$ is called *quasi-concave* if $-F$ is quasi-convex. Specifically, F is called quasi-concave if

for all $X, Y \in C$ and all $\lambda \in [0, 1]$, the following inequality holds:

$$F(\lambda X + (1 - \lambda)Y) \geq \min\{F(X), F(Y)\}. \tag{2.3.11}$$

2.3.4 Fenchel–Eggleston–Carathéodory Theorem

We mentioned above that the convex hull of a subset S of a real vector space is the convex set consisting of all convex combinations of the vectors in S. A fundamental result is that if the underlying vector space has dimension d, then, in order to obtain an element in the convex hull of S, it suffices to take a convex combination of no more than $d + 1$ elements of S. If S is connected and compact, then no more than d elements are required. We state this formally as follows.

Theorem 2.23 Fenchel–Eggleston–Carathéodory Theorem

Let S be a set of vectors in a real d-dimensional vector space ($d < \infty$), and let $\text{conv}(S)$ denote the convex hull of S. Then, an arbitrary element $v \in \text{conv}(S)$ can be expressed as a convex combination of at most $m \leq d + 1$ elements in S. If S is connected and compact, then the same statement holds with $m \leq d$.

Proof: Please consult the Bibliographic Notes (Section 2.6). ■

2.3.5 Minimax Theorems

We often encounter expressions of the following form in quantum information theory:

$$\inf_{X \in S} \sup_{Y \in S'} F(X, Y). \tag{2.3.12}$$

The expression above contains both an infimum and a supremum over subsets $S, S' \subseteq \text{L}(\mathcal{H})$ of some real-valued function $F : S \times S' \to \mathbb{R}$.

Expressions such as the one in (2.3.12) arise in the context of two-player zero-sum games. In such a game, the function $F(X, Y)$ represents the reward of a protagonist, who chooses elements $Y \in S'$ in order to maximize F. The antagonist chooses elements $X \in S$ in order to minimize F, i.e., to minimize the reward to the protagonist[4]. The worst-case scenario for the antagonist is that, no matter what element $X \in S$ it chooses, the protagonist chooses the "best" possible element in S' for their benefit, so that the reward is $G(X) \coloneqq \sup_{Y \in S'} F(X,Y)$. The optimal reward of the antagonist in this scenario is thus given by $\inf_{X \in S} G(X)$, which is the quantity in (2.3.12).

On the other hand, the worst-case scenario for the protagonist is that, no matter what element $Y \in S'$ they choose, the antagonist chooses the "best" possible element in S for their benefit, so that the reward is $\widetilde{G}(Y) \coloneqq \inf_{X \in S} F(X,Y)$. The optimal reward of the protagonist in this scenario is thus given by $\sup_{Y \in S'} \widetilde{G}(Y)$, i.e.,

$$\sup_{Y \in S'} \inf_{X \in S} F(X,Y). \tag{2.3.13}$$

Intuitively, it is advantageous for the protagonist to achieve a higher reward when playing second (in reaction to the antagonist's choice). This intuition is captured by the following "max-min inequality":

$$\sup_{Y \in S'} \inf_{X \in S} F(X,Y) \leq \inf_{X \in S} \sup_{Y \in S'} F(X,Y), \tag{2.3.14}$$

which always holds. A mnemonic trick to remember it is that the internal optimization (inf on the left-hand side and sup on the right-hand side) is the one that determines the inequality, because it represents the player who plays second. We now prove this formally. Observe that for all $X \in S, Y \in S'$, we have that $\widetilde{G}(Y) \leq F(X,Y)$. It then follows that $\sup_{Y \in S'} \widetilde{G}(Y) \leq \sup_{Y \in S'} F(X,Y)$ by applying the definition of supremum. Since this latter inequality holds for all $X \in S$, the definition of infimum implies that $\sup_{Y \in S'} \widetilde{G}(Y) \leq \inf_{X \in S} \sup_{Y \in S'} F(X,Y)$, which is precisely the inequality in (2.3.14).

Many proofs that we present in this book require determining when the inequality opposite to the one in (2.3.14) holds, i.e.,

$$\inf_{X \in S} \sup_{Y \in S'} F(X,Y) \overset{?}{\leq} \sup_{Y \in S'} \inf_{X \in S} F(X,Y), \tag{2.3.15}$$

which is known as the "min-max inequality." If it holds, then the inequality in (2.3.14) is saturated and becomes an equality. The game-theoretic interpretation of the situation when the reverse inequality holds is that, for the sets S and S' and

[4]This is due the fact that the reward of the antagonist is equal to $-F(X,Y)$ as a consequence of the zero-sum property of the game.

the reward function F, there is no advantage to playing first; i.e., the reward is the same regardless of who goes first, as long as the protagonist and antagonist play optimal strategies. It is thus important to know under what conditions this reverse inequality holds.

We now present theorems for two classes of functions that tell us when the inequality (and thus the equality) in (2.3.15) holds.

Theorem 2.24 Sion Minimax

Let S be a compact and convex subset of a normed vector space and let S' be a convex subset of a normed vector space. Let $F : S \times S' \to \mathbb{R}$ be a real-valued function such that

1. The function $F(\cdot, Y) : S \to \mathbb{R}$ is lower semi-continuous and quasi-convex on S for every $Y \in S'$.

2. The function $F(X, \cdot) : S' \to \mathbb{R}$ is upper semi-continuous and quasi-concave on S' for every $X \in S$.

Then,

$$\inf_{X \in S} \sup_{Y \in S'} F(X, Y) = \sup_{Y \in S'} \inf_{X \in S} F(X, Y). \tag{2.3.16}$$

Furthermore, on both sides of the above equation, the infimum can be replaced with a minimum.

Proof: Please consult the Bibliographic Notes (Section 2.6). ■

Theorem 2.25 Mosonyi–Hiai Minimax

Let S be a compact set in a normed vector space, let $S' \subseteq \mathbb{R}$, and let $F : S \times S' \to \mathbb{R} \cup \{+\infty, -\infty\}$. Suppose that

1. The function $F(\cdot, y) : S \to \mathbb{R} \cup \{+\infty, -\infty\}$ is lower semi-continuous for every $y \in S'$.

2. The function $F(X, \cdot) : S' \to \mathbb{R} \cup \{+\infty, -\infty\}$ is either monotonically increasing or monotonically decreasing for every $X \in S$.

Then,

$$\inf_{X \in S} \sup_{y \in S'} F(X, y) = \sup_{y \in S'} \inf_{X \in S} F(X, y). \tag{2.3.17}$$

Furthermore, the infimum can be replaced with a minimum.

PROOF: Please consult the Bibliographic Notes (Section 2.6). ■

2.3.6 Probability Distributions

Throughout this book, we are concerned for the most part with discrete probability distributions, and the following definitions suffice for our needs. A *discrete probability distribution* is a function $p : \mathcal{X} \to [0, 1]$ defined on an finite alphabet $\mathcal{X}$ such that $p(x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(x) = 1$. Formally, we can consider the alphabet $\mathcal{X}$ to be the set of realizations of a discrete random variable $X : \Omega \to \mathcal{X}$ from the space Ω of experimental outcomes, called the sample space, to the set $\mathcal{X}$. We then write p_X to denote the probability distribution of the random variable X, i.e., $p_X(x) \equiv \Pr[X = x]$.

The *expected value* or *mean* $\mathbb{E}[X]$ of a random variable X taking values in $\mathcal{X} \subset \mathbb{R}$ is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \, p_X(x). \tag{2.3.18}$$

For every function $g : \mathcal{X} \to \mathbb{R}$, we define $g(X)$ to be the random variable $g \circ X : \Omega \to \mathbb{R}$ with image $\{g(x) : x \in \mathcal{X}\}$. Then,

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_X(x). \tag{2.3.19}$$

A useful fact is *Markov's inequality*: if X is a non-negative random variable, then for all $a > 0$ we have

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}. \tag{2.3.20}$$

Intuitively, Markov's inequality tells us that the probability that X takes on values that are much larger than its mean must be small.

Jensen's inequality is the following: if X is a random variable with finite mean, and f is a real-valued convex function acting on the output of X, then

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]. \tag{2.3.21}$$

This is a very special case of the more elaborate operator Jensen inequality presented previously in Theorem 2.16.

As we explain in Chapter 3, observables O in quantum mechanics (which are merely Hermitian operators) generalize random variables, such that their expectation is given by $\mathbb{E}[O] \equiv \langle O \rangle_\rho := \operatorname{Tr}[O\rho]$ for a density operator ρ. In this case, an application of the operator Jensen inequality from Theorem 2.16 leads to the following: for a Hermitian operator O, a density operator ρ, and an operator convex function f,

$$f(\operatorname{Tr}[O\rho]) \leq \operatorname{Tr}[f(O)\rho], \tag{2.3.22}$$

which we can alternatively write as

$$f(\langle O \rangle_\rho) \leq \langle f(O) \rangle_\rho . \tag{2.3.23}$$

2.4 Semi-Definite Programming

Semi-definite programs (SDPs) constitute an important class of optimization problems that arise frequently in quantum information theory. An SDP is a constrained optimization problem in which the optimization variable is a positive semi-definite operator X, the objective function is linear in X, and the constraint is an operator inequality featuring a linear function of X. Not only are SDPs useful as an analytical tool, but there also exist a number of numerical solvers that can be used for evaluating these optimization problems (one can use the `CVX` package for MATLAB or the `CVXPY` package for Python).

Semi-definite programs play an important role in quantum information because, for a number of operational tasks of interest, we are trying to maximize a linear objective function over the sets of quantum states or measurements, which are specified by semi-definite constraints. Furthermore, many of the different communication capacities of quantum channels are difficult to characterize or compute, and it can helpful to find semi-definite relaxations of them that are efficiently computable. These are two common ways in which semi-definite programs arise in this book.

Definition 2.26 Semi-Definite Program

Given a Hermiticity-preserving superoperator Φ and Hermitian operators A and B, a *semi-definite program* (SDP) corresponds to two optimization problems. The first is the *primal SDP*, which is defined as

$$\begin{aligned} \text{maximize} \quad & \text{Tr}[AX] \\ \text{subject to} \quad & \Phi(X) \leq B, \\ & X \geq 0. \end{aligned} \tag{2.4.1}$$

The second optimization problem is the *dual SDP*, which is defined as

$$\begin{aligned} \text{minimize} \quad & \text{Tr}[BY] \\ \text{subject to} \quad & \Phi^\dagger(Y) \geq A, \\ & Y \geq 0. \end{aligned} \tag{2.4.2}$$

We let

$$S(\Phi, A, B) := \sup_{X \geq 0} \{\text{Tr}[AX] : \Phi(X) \leq B\}, \tag{2.4.3}$$

$$\widehat{S}(\Phi, A, B) := \inf_{Y \geq 0} \{\text{Tr}[BY] : \Phi^\dagger(Y) \geq A\} \tag{2.4.4}$$

denote the optimal values of the primal and dual SDPs, respectively.

Remark: Equality constraints of the form $\text{Tr}[CX] = c$, where C is a Hermitian matrix and $c \in \mathbb{R}$ is a real number, can be incorporated into (2.4.1) by rephrasing them as pairs of inequalities $\text{Tr}[CX] \leq c$ and $\text{Tr}[CX] \geq c$.

A variable X for the primal SDP in (2.4.3) is called a *feasible point* if it is positive semi-definite ($X \geq 0$) and satisfies the constraint $\Phi(X) \leq B$, and it is called a *strictly feasible point* if X is positive definite ($X > 0$) and the constraint is satisfied with a strict inequality, i.e., if $\Phi(X) < B$. The same definitions apply to the dual SDP in (2.4.4): a variable Y is a feasible point if $Y \geq 0$ and $\Phi^\dagger(Y) \geq A$, and it is strictly feasible if $Y > 0$ and $\Phi^\dagger(Y) > A$. By convention, if there is no primal feasible operator X, then $S(\Phi, A, B) = -\infty$, and if there is no dual feasible operator Y, then $\widehat{S}(\Phi, A, B) = +\infty$. It is also possible for $S(\Phi, A, B) = +\infty$ or $\widehat{S}(\Phi, A, B) = -\infty$. A simple example of $S(\Phi, A, B) = +\infty$ is when $A = 1$ is a scalar, $\Phi(X) = 0$, and $B = 1$ (so that the constraint in (2.4.3) is trivially satisfied), and a simple example of $\widehat{S}(\Phi, A, B) = -\infty$ is when $B = -1$, $\Phi(Y) = 0$, and $A = -1$.

REMARK: As discussed in Remark , an SDP can incorporate equality constraints by turning each one into a pair of inequality constraints. However, when determining whether a given variable X is strictly feasible, one should group back all such inequality pairs into equality constraints and consider strict feasibility only with respect to the remaining inequality constraints.

Proposition 2.27 Weak Duality

For every SDP corresponding to Φ, A, and B, the following weak duality inequality holds:

$$S(\Phi, A, B) \leq \widehat{S}(\Phi, A, B). \tag{2.4.5}$$

PROOF: Let $X \geq 0$ be primal feasible, and let $Y \geq 0$ be dual feasible. Then the following holds

$$\mathrm{Tr}[AX] \leq \mathrm{Tr}[\Phi^\dagger(Y)X] = \mathrm{Tr}[Y\Phi(X)] \leq \mathrm{Tr}[YB]. \tag{2.4.6}$$

The first inequality follows from the assumption that Y is dual feasible, so that we have $A \leq \Phi^\dagger(Y)$, and by applying 2. of Lemma 2.14. The equality holds by definition of the adjoint map $\Phi^\dagger$; see (2.2.167). The last inequality follows from the assumption that X is primal feasible, so that we have $\Phi(X) \leq B$, and by applying 2. of Lemma 2.14. Since the inequality holds for all primal feasible X and for all dual feasible Y, we can take a supremum over the left-hand side of (2.4.6) and an infimum over the right-hand side of (2.4.6), and we thus arrive at the weak duality inequality in (2.4.5). ■

There is a deep connection between the weak duality inequality in Proposition 2.27 and the max-min inequality in (2.3.14). This is realized by introducing the Lagrangian $\mathcal{L}(\Phi, A, B, X, Y)$ for the SDP as follows:

$$\mathcal{L}(\Phi, A, B, X, Y) := \mathrm{Tr}[AX] + \mathrm{Tr}[BY] - \mathrm{Tr}[\Phi(X)Y]. \tag{2.4.7}$$

Note that the following equalities hold, which are helpful in the discussion below:

$$\mathcal{L}(\Phi, A, B, X, Y) = \mathrm{Tr}[AX] + \mathrm{Tr}[(B - \Phi(X))Y] \tag{2.4.8}$$

$$= \mathrm{Tr}[BY] + \mathrm{Tr}[(A - \Phi^\dagger(Y))X]. \tag{2.4.9}$$

By first taking an infimum over $Y \geq 0$ and then a supremum over $X \geq 0$, we find that

$$\sup_{X\geq 0} \inf_{Y\geq 0} \mathcal{L}(\Phi, A, B, X, Y) = S(\Phi, A, B). \tag{2.4.10}$$

This equality follows because

$$\sup_{X\geq 0}\inf_{Y\geq 0}\mathcal{L}(\Phi, A, B, X, Y) = \sup_{X\geq 0}\left\{\operatorname{Tr}[AX] + \inf_{Y\geq 0}\operatorname{Tr}[(B-\Phi(X))Y]\right\}. \tag{2.4.11}$$

The inner infimum with respect to $Y \geq 0$ forces the outer optimization to be with respect to every feasible point X for the primal SDP in (2.4.3). In this sense, the variable Y can be thought as a "Lagrange multiplier", analogous to Lagrange multipliers that are used in constrained optimization problems in elementary calculus. Indeed, suppose that an infeasible $X \geq 0$ is chosen, meaning that the constraint $\Phi(X) \leq B$ is violated. This means that there exists a non-trivial negative eigenspace of $B - \Phi(X)$. Let $|\varphi\rangle$ be a unit vector in this negative eigenspace. We can then pick $Y = c|\varphi\rangle\langle\varphi|$ for $c > 0$ and take the limit $c \to \infty$, so that $\inf_{Y\geq 0}\operatorname{Tr}[(B-\Phi(X))Y] = -\infty$. So a violation of the constraint $\Phi(X) \leq B$ incurs an infinite cost for the outer optimization with respect to $X \geq 0$, meaning that the corresponding point X can be effectively discarded. The constraint $\Phi(X) \leq B$ is therefore forced to be satisfied, leading to the equality in (2.4.10).

If we instead take a supremum over $X \geq 0$ first and then take an infimum over $Y \geq 0$, it follows that

$$\inf_{Y\geq 0}\sup_{X\geq 0}\mathcal{L}(\Phi, A, B, X, Y) = \widehat{S}(\Phi, A, B). \tag{2.4.12}$$

This time, the equality follows because

$$\inf_{Y\geq 0}\sup_{X\geq 0}\mathcal{L}(\Phi, A, B, X, Y) = \inf_{Y\geq 0}\left\{\operatorname{Tr}[BY] + \sup_{X\geq 0}\operatorname{Tr}[(A-\Phi^\dagger(Y))X]\right\} \tag{2.4.13}$$

Similar to what was argued previously, the inner optimization variable X is a Lagrange multiplier that forces the outer optimization to be restricted to dual feasible points only; the constraint $\Phi^\dagger(Y) \geq A$ is thus forced to be satisfied, leading to the equality in (2.4.12).

Now, by examining (2.3.14), (2.4.10), and (2.4.12), we see that the weak duality inequality in Proposition 2.27 can be understood as a consequence of the max-min inequality in (2.3.14).

The inequality opposite to the one in (2.4.5) does not hold in general; if it does, it implies that $S(\Phi, A, B) = \widehat{S}(\Phi, A, B)$. We then say that the SDP corresponding to Φ, A, and B has the *strong duality* property, or that it satisfies strong duality. Considering the discussion above in terms of the Lagrangian of the SDP, we also can understand strong duality as being equivalent to a minimax theorem holding.

Theorem 2.28 Slater's Condition

Slater's condition is a sufficient condition for strong duality to hold, and it is given as follows:

1. If there exists $X \geq 0$ such that $\Phi(X) \leq B$ and there exists $Y > 0$ such that $\Phi^\dagger(Y) > A$, then $S(\Phi, A, B) = \widehat{S}(\Phi, A, B)$. Furthermore, there exists a primal feasible operator X for which $\text{Tr}[AX] = S(\Phi, A, B)$.

2. If there exists $Y \geq 0$ such that $\Phi^\dagger(Y) \geq A$ and there exists $X > 0$ such that $\Phi(X) < B$, then $S(\Phi, A, B) = \widehat{S}(\Phi, A, B)$. Furthermore, there exists a dual feasible operator Y for which $\text{Tr}[BY] = \widehat{S}(\Phi, A, B)$.

Remark: The nomenclature Slater's "condition" (rather than "conditions") is commonly used, but note that one can check either one of the two sufficient conditions above to determine if strong duality holds.

For many SDPs of interest, it is straightforward to determine if Slater's condition holds. We provide an example in Section 2.4.1.

Complementary slackness for SDPs is useful for understanding further constraints on an optimal primal operator X and an optimal dual operator Y.

Proposition 2.29 Complementary Slackness of SDPs

Consider an arbitrary SDP corresponding to Φ, A, and B, and suppose that strong duality holds. Then the following complementary slackness conditions hold for feasible X and Y if and only if they are optimal:

$$YB = Y\Phi(X), \tag{2.4.14}$$

$$\Phi^\dagger(Y)X = AX. \tag{2.4.15}$$

Proof: On the one hand, suppose that X is primal feasible, that Y is dual feasible, and that they satisfy (2.4.14)–(2.4.15). Then it is clear by inspecting (2.4.6) that the inequalities are saturated, thus implying that X is primal optimal and Y is dual optimal.

On the other hand, suppose that X is primal optimal and that Y is dual optimal. Then, by this assumption, it follows that $\text{Tr}[AX] = \text{Tr}[BY]$ so that the inequalities in (2.4.6) are saturated. This means that

$$\mathrm{Tr}[(\Phi^\dagger(Y) - A)X] = 0 \tag{2.4.16}$$

$$\mathrm{Tr}[Y(B - \Phi(X))] = 0. \tag{2.4.17}$$

Since $\Phi^\dagger(Y) - A$ and X are positive semi-definite, the equality in (2.4.16) implies that $(\Phi^\dagger(Y) - A)X = 0$, which is equivalent to (2.4.14). Similarly, since $B - \Phi(X)$ and Y are positive semi-definite, the equality in (2.4.17) implies that $Y(B - \Phi(X)) = 0$, which is equivalent to (2.4.15). ■

If the matrices A and B and the map Φ involved in an SDP are of reasonable size, then the SDP can be computed efficiently using numerical solvers (specifically, the time required is polynomial in the size of these objects and polynomial in the logarithm of the inverse of the numerical accuracy desired). As mentioned earlier, SDPs arise frequently in quantum information, with some examples appearing in Chapter 6. Furthermore, SDPs appear in some of the upper bounds for rates of quantum communication protocols that we consider in Parts II and III.

Exercise 2.28

Consider the following pair of primal and dual optimization problems:

$$\begin{aligned} &\text{maximize} && \mathrm{Tr}[CZ] \\ &\text{subject to} && \Psi(Z) = D, \\ &&& Z \geq 0 \end{aligned} \qquad \begin{aligned} &\text{minimize} && \mathrm{Tr}[DW] \\ &\text{subject to} && \Psi^\dagger(W) \geq C, \\ &&& W \text{ Hermitian}, \end{aligned} \tag{2.4.18}$$

where C and D are Hermitian operators and Ψ is a Hermiticity-preserving superoperator. Let us show that these problems are SDPs, i.e., that they are equivalent to the optimization problems in (2.4.1) and (2.4.2).

1. Given C, D, and Ψ for the optimization problems in (2.4.18), find A, B, and Φ such that these optimization problems can be expressed in the forms presented in (2.4.1) and (2.4.2). (*Hint*: Start by using the fact that $\Psi(Z) = D$ if and only if $\Psi(Z) \leq D$ and $-\Psi(Z) \leq -D$.)

2. Conversely, given A, B, and Φ for the SDPs in (2.4.1) and (2.4.2), find C, D, and Ψ such that those SDPs can be expressed in the forms in (2.4.18). (*Hint*: Start by using the fact that $\Phi(X) \leq B$ if and only if there exists $S \geq 0$ such that $\Phi(X) + S = B$.)

Reasoning analogous to that in Exercise 2.28 can be used to show that the following pair of optimization problems are also SDPs, equivalent to the ones in (2.4.1) and (2.4.2):

$$\begin{array}{llll} \text{minimize} & \text{Tr}[CZ] & \text{maximize} & \text{Tr}[DW] \\ \text{subject to} & \Psi(Z) = D, & \text{subject to} & \Psi^\dagger(W) \leq C, \\ & Z \geq 0 & & W \text{ Hermitian} \end{array} \tag{2.4.19}$$

Exercise 2.29

1. Consider the following SDP in primal form:

$$\sup_{X\geq 0} \{\text{Tr}[AX] : \Phi_1(X) \leq B_1,\ \Phi_2(X) = B_2\}, \tag{2.4.20}$$

where A, B_1, B_2 are Hermitian operators and Φ_1, Φ_2 are Hermiticity-preserving superoperators. Show that the dual SDP is given by

$$\inf_{\substack{Y_1\geq 0,\\ Y_2 \text{ Hermitian}}} \left\{\text{Tr}[B_1Y_1] + \text{Tr}[B_2Y_2] : \Phi_1^\dagger(Y_1) + \Phi_2^\dagger(Y_2) \geq A\right\}. \tag{2.4.21}$$

Furthermore, evaluate Slater's conditions for strong duality, as well as the conditions for complementary slackness.

2. Now suppose that the primal SDP has the form

$$\inf_{Y\geq 0} \{\text{Tr}[BY] : \Phi_1(Y) \geq A_1,\ \Phi_2(Y) = A_2\}, \tag{2.4.22}$$

where A_1, A_2, B are Hermitian operators and Φ_1, Φ_2 are Hermiticity-preserving superoperators. Show that the dual SDP is given by

$$\sup_{\substack{X_1\geq 0,\\ X_2 \text{ Hermitian}}} \left\{\text{Tr}[A_1X_1] + \text{Tr}[A_2X_2] : \Phi_1^\dagger(X_1) + \Phi_2^\dagger(X_2) \leq B\right\}. \tag{2.4.23}$$

Furthermore, evaluate Slater's conditions for strong duality, as well as the conditions for complementary slackness.

2.4.1 SDPs for Spectral and Trace Norm, Maximum and Minimum Eigenvalue

In this section, we provide semi-definite programs for calculating the spectral and trace norms of Hermitian operators, as well as their largest and smallest eigenvalues.

Theorem 2.30 SDPs for the Spectral Norm of Hermitian Operators

Let H be a Hermitian operator, and consider the following functions:

$$f(H) := \sup_{X_1,X_2\geq 0} \{\mathrm{Tr}[H(X_1 - X_2)] : \mathrm{Tr}[X_1 + X_2] \leq 1\}, \tag{2.4.24}$$

$$\widehat{f}(H) := \inf_{t\geq 0} \{t : -t\mathbb{1} \leq H \leq t\mathbb{1}\}. \tag{2.4.25}$$

The quantities above can be computed via SDPs, and in fact, the following equality holds

$$f(H) = \widehat{f}(H) = \|H\|_\infty. \tag{2.4.26}$$

That is, $f(H)$ is equal to the largest singular value of the Hermitian operator H.

Proof: Given that the optimization in (2.4.24) is a maximization, let us first show that (2.4.24) can be written in the form of $S(\Phi, A, B)$ in (2.4.3). Indeed if we let

$$X = \begin{pmatrix} X_1 & Z^\dagger \\ Z & X_2 \end{pmatrix}, \qquad A = \begin{pmatrix} H & 0 \\ 0 & -H \end{pmatrix}, \tag{2.4.27}$$

$$\Phi(X) = \mathrm{Tr}[X_1 + X_2], \qquad B = 1, \tag{2.4.28}$$

then we have that

$$f(H) = \sup_{X\geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\}. \tag{2.4.29}$$

The constraint $X \geq 0$ implies that $X_1, X_2 \geq 0$. Furthermore, notice that the operator Z appears neither in the objective function $\mathrm{Tr}[H(X_1 - X_2)]$ nor in the constraint $\mathrm{Tr}[X_1 + X_2] \leq 1$. Thus, the operator Z plays no role in the optimization, and so we can simply set $Z = 0$, so that

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}. \tag{2.4.30}$$

Thus, (2.4.24) is indeed an SDP in primal form.

Now, recall from (2.2.109) that the spectral norm of H is given by the maximum of the absolute values of the eigenvalues of H. In particular, we can write

$$\|H\|_\infty = \max\{|\lambda_{\max}|, |\lambda_{\min}|\}, \tag{2.4.31}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues, respectively, of H. Note that we always have $\lambda_{\max} \geq \lambda_{\min}$. Let $|\phi_{\max}\rangle$ be an eigenvector of H satisfying $H|\phi_{\max}\rangle = \lambda_{\max}|\phi_{\max}\rangle$, and let $|\phi_{\min}\rangle$ be an eigenvector of H satisfying $H|\phi_{\min}\rangle = \lambda_{\min}|\phi_{\min}\rangle$. Let us suppose at first that $\lambda_{\max} \geq 0$. Then one feasible choice of X_1 and X_2 in (2.4.24) is $X_1 = |\phi_{\max}\rangle\langle\phi_{\max}|$ and $X_2 = 0$, and for this choice, we find that $f(H) \geq \lambda_{\max} = |\lambda_{\max}|$. If $\lambda_{\max} \leq 0$, then another feasible choice of X_1 and X_2 in (2.4.24) is $X_1 = 0$ and $X_2 = |\phi_{\max}\rangle\langle\phi_{\max}|$, and for this choice, we find that $f(H) \geq -\lambda_{\max} = |\lambda_{\max}|$. Therefore, we conclude that

$$f(H) \geq |\lambda_{\max}|. \tag{2.4.32}$$

Now, suppose that $\lambda_{\min} \geq 0$. Then a feasible choice of X_1 and X_2 in (2.4.24) is $X_1 = |\phi_{\min}\rangle\langle\phi_{\min}|$ and $X_2 = 0$, and for this choice, we find that $f(H) \geq \lambda_{\min} = |\lambda_{\min}|$. If $\lambda_{\min} \leq 0$, then another feasible choice of X_1 and X_2 in (2.4.24) is $X_1 = 0$ and $X_2 = |\phi_{\min}\rangle\langle\phi_{\min}|$, and for this choice, we find that $f(H) \geq -\lambda_{\min} = |\lambda_{\min}|$. Therefore, we conclude that

$$f(H) \geq \max\{|\lambda_{\max}|, |\lambda_{\min}|\} = \|H\|_\infty. \tag{2.4.33}$$

It now remains to prove the reverse inequality, namely, the inequality $f(H) \leq \|H\|_\infty$. To prove this, let us show that $\widehat{f}(H)$, as defined in (2.4.25), is given by the SDP dual to the one that defines $f(H)$. In order to do this, we should determine the map $\Phi^\dagger$, which is the adjoint of Φ. Since $B = 1$ and $\Phi(X) = \operatorname{Tr}[X_1 + X_2]$ are scalars, we take $Y = t$ to be a scalar also. Then, we find that

$$\operatorname{Tr}[Y\Phi(X)] = t\operatorname{Tr}[X_1 + X_2] \tag{2.4.34}$$

$$= \operatorname{Tr}\left[\begin{pmatrix} t\mathbb{1} & 0 \\ 0 & t\mathbb{1}\end{pmatrix}\begin{pmatrix} X_1 & 0 \\ 0 & X_2\end{pmatrix}\right] \tag{2.4.35}$$

$$= \operatorname{Tr}[\Phi^\dagger(Y)X], \tag{2.4.36}$$

from which we conclude that

$$\Phi^\dagger(Y) = \Phi^\dagger(t) = \begin{pmatrix} t\mathbb{1} & 0 \\ 0 & t\mathbb{1}\end{pmatrix}. \tag{2.4.37}$$

Plugging this into the standard form of the dual in (2.4.4), we find that

$$\inf_{Y\geq 0}\left\{\text{Tr}[BY] : \Phi^\dagger(Y)\geq A\right\} = \inf_{t\geq 0}\left\{t : \begin{pmatrix} t\mathbb{1} & 0 \\ 0 & t\mathbb{1}\end{pmatrix} \geq \begin{pmatrix} H & 0 \\ 0 & -H\end{pmatrix}\right\} \tag{2.4.38}$$

$$= \inf_{t\geq 0}\{t : t\mathbb{1}\geq H,\ t\mathbb{1}\geq -H\} \tag{2.4.39}$$

$$= \inf_{t\geq 0}\{t : -t\mathbb{1}\leq H\leq t\mathbb{1}\} \tag{2.4.40}$$

$$= \widehat{f}(H). \tag{2.4.41}$$

Let us now recall property 3. of Lemma 2.14, which states that $\lambda_{\min}\mathbb{1} \leq H \leq \lambda_{\max}\mathbb{1}$. By combining with (2.4.33), we find that $\lambda_{\max}\mathbb{1} \leq \|H\|_\infty \mathbb{1}$ and $\lambda_{\min}\mathbb{1} \geq -\|H\|_\infty \mathbb{1}$, which implies that

$$-\|H\|_\infty \mathbb{1} \leq H \leq \|H\|_\infty \mathbb{1}. \tag{2.4.42}$$

Thus, we see that $\|H\|_\infty$ is a feasible choice for t in (2.4.40), which implies that

$$\widehat{f}(H) \leq \|H\|_\infty . \tag{2.4.43}$$

Now, combining the inequalities in (2.4.33) and (2.4.43) gives us $f(H) \geq \|H\|_\infty \geq \widehat{f}(H)$. Then, using the weak duality inequality from Proposition 2.27, which for our case implies that $f(H) \leq \widehat{f}(H)$, we conclude that the primal and dual optimal values are equal to each other and equal to the spectral norm of H: $f(H) = \widehat{f}(H) = \|H\|_\infty$. ■

We proved (2.4.26) by employing clever guesses for primal feasible and dual feasible points. Doing so is possible in this case because the problem is simple enough to begin with, and we could apply knowledge from linear algebra to make these clever guesses. Although it is sometimes possible to make clever guesses and arrive at analytical solutions like we did above, in many cases it is not possible. In such cases, it can be helpful to check Slater's condition in Theorem 2.28 explicitly in order to see if strong duality holds. So let us do so for the SDPs corresponding to $f(H)$ and $\widehat{f}(H)$. For the primal SDP in (2.4.24), a strictly feasible point consists of the choice $X_1 = \alpha\frac{\mathbb{1}}{d}$ and $X_2 = \beta\frac{\mathbb{1}}{d}$ such that $\alpha, \beta > 0$ and $\alpha + \beta < 1$, where d is the dimension of $\mathbb{1}$. Then we clearly have $X_1 > 0$, $X_2 > 0$, and $\text{Tr}[X_1 + X_2] < 1$, so that X_1 and X_2 are strictly feasible, as claimed. A feasible point for the dual consists of the choice $\gamma \geq \|H\|_\infty$. Thus, strong duality holds, further confirming that $f(H) = \widehat{f}(H)$, as shown above.

We now remark about the complementary slackness conditions from Proposition 2.29 for the SDPs corresponding to $f(H)$ and $\widehat{f}(H)$, which apply to optimal primal X and optimal dual Y. In this case, the conditions reduce to

$$t = t\,\mathrm{Tr}[X_1 + X_2], \tag{2.4.44}$$

$$\begin{pmatrix} t\mathbb{1} & 0 \\ 0 & t\mathbb{1} \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} = \begin{pmatrix} H & 0 \\ 0 & -H \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}, \tag{2.4.45}$$

and the latter is the same as the following two separate conditions:

$$tX_1 = HX_1, \qquad -tX_2 = HX_2. \tag{2.4.46}$$

If we have prior knowledge about the operator H (e.g., that it is non-zero), then we conclude that the optimal $t \neq 0$ and the condition in (2.4.44) implies that $\mathrm{Tr}[X_1 + X_2] = 1$. In this case, we can conclude that the inequality constraint in (2.4.24) is loose and it suffices to optimize over X_1 and X_2 satisfying the constraint with equality. The conditions in (2.4.46) indicate that the image of the optimal X_1 should be in the eigenspace of H with optimal eigenvalue t, and the image of the optimal X_2 should be in the eigenspace of H with optimal eigenvalue $-t$. Observe that these complementary slackness conditions are consistent with the choices that we made above.

As a final remark, if H is actually positive semi-definite, then the lower bound constraint in (2.4.25) is unnecessary. Letting P be a positive semi-definite operator, we thus find that

$$f(P) = \|P\|_\infty = \inf_{t \geq 0} \{t : P \leq t\mathbb{1}\}. \tag{2.4.47}$$

Note that, in this case, $\|P\|_\infty$ is the largest eigenvalue of P.

Exercise 2.30 SDPs for the Trace Norm of Hermitian Operators

1. Let H be a Hermitian operator. Like the spectral norm of H, as shown above, prove that the trace norm of H can also be computed using an SDP. Specifically, prove that

$$\|H\|_1 = \sup_{\Lambda_1, \Lambda_2 \geq 0} \{\mathrm{Tr}[H(\Lambda_1 - \Lambda_2)] : \Lambda_1, \Lambda_2 \leq \mathbb{1}\}. \tag{2.4.48}$$

(*Hint*: Use (2.2.69) and (2.2.113).)

2. Show that an alternate SDP formulation for $\|H\|_1$ is

$$\|H\|_1 = \inf_{Y_1,Y_2\geq 0}\{\text{Tr}[Y_1 + Y_2] : Y_1 \geq H,\, Y_2 \geq -H\}. \tag{2.4.49}$$

(*Hint*: Show that the SDP in (2.4.49) is dual to the one in (2.4.48), and then prove strong duality.)

Exercise 2.31 SDPs for the Maximum and Minimum Eigenvalue of Hermitian Operators

Let H be a Hermitian operator. Prove that the maximum and minimum eigenvalues of H, denoted by $\lambda_{\max}(H)$ and $\lambda_{\min}(H)$, respectively, have the following SDP characterizations:

$$\lambda_{\min}(H) = \inf_{\rho\geq 0}\{\text{Tr}[H\rho] : \text{Tr}[\rho] = 1\} \tag{2.4.50}$$

$$= \sup_{t\in\mathbb{R}}\{t : H \geq t\mathbb{1}\} \tag{2.4.51}$$

and

$$\lambda_{\max}(H) = \sup_{\rho\geq 0}\{\text{Tr}[H\rho] : \text{Tr}[\rho] = 1\} \tag{2.4.52}$$

$$= \inf_{t\in\mathbb{R}}\{t : t\mathbb{1} \geq H\}. \tag{2.4.53}$$

(*Hint*: Use the spectral theorem (Theorem 2.4) and the duality of SDPs.)

2.5 Symmetric Subspace

Given a d-dimensional Hilbert space $\mathcal{H}$ and an n-fold tensor product $\mathcal{H}^{\otimes n}$ of $\mathcal{H}$, for $n \geq 2$, it is often important to consider a permutation of the individual Hilbert spaces in the n-fold tensor product. This is especially the case in quantum information theory, because we often assume or it is often the case that the resources involved have permutation symmetry. These permutations can be implemented using a unitary representation of the symmetric group on n elements.

The symmetric group on n elements, denoted by $\mathcal{S}_n$, is defined to be the set of permutations of the set $\{1, 2, \ldots, n\}$. A permutation in $\mathcal{S}_n$ is an invertible function $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ that describes how each element in the set $\{1, 2, \ldots, n\}$ should be rearranged, or permuted. An example of a permutation in $\mathcal{S}_3$ is the function π such that $\pi(1) = 3$, $\pi(2) = 1$, and $\pi(3) = 2$. Since there are $n!$ ways to permute n distinct elements, it follows that the set $\mathcal{S}_n$ contains $n!$ elements.

Given a permutation $\pi \in \mathcal{S}_n$ and an orthonormal basis $\{|i\rangle\}_{i=0}^{d-1}$ for $\mathcal{H}$, we define the unitary permutation operators W^π acting on $\mathcal{H}^{\otimes n}$ by

$$W^\pi |i_1, i_2, \ldots, i_n\rangle = |i_{\pi^{-1}(1)}, i_{\pi^{-1}(2)}, \ldots, i_{\pi^{-1}(n)}\rangle, \quad 0 \le i_1, i_2, \ldots, i_n \le d-1, \tag{2.5.1}$$

so that an alternative expression for W^π is

$$W^\pi = \sum_{i_1, i_2, \ldots, i_n} |i_{\pi^{-1}(1)}, i_{\pi^{-1}(2)}, \ldots, i_{\pi^{-1}(n)}\rangle\langle i_1, i_2, \ldots, i_n| \tag{2.5.2}$$

$$= \sum_{i_1, i_2, \ldots, i_n} |i_1, i_2, \ldots, i_n\rangle\langle i_{\pi(1)}, i_{\pi(2)}, \ldots, i_{\pi(n)}| . \tag{2.5.3}$$

Since the set $\{|i_1, i_2, \ldots, i_n\rangle : 0 \le i_1, i_2, \ldots, i_n \le d-1\}$ is an orthonormal basis for $\mathcal{H}^{\otimes n}$, the definition in (2.5.1) extends to every vector in $\mathcal{H}^{\otimes n}$ by linearity. The operators in the set $\{W^\pi\}_{\pi \in \mathcal{S}_n}$ constitute a unitary representation of $\mathcal{S}_n$, in the sense that

$$(W^\pi)^\dagger = W^{\pi^{-1}}, \quad W^{\pi_1} W^{\pi_2} = W^{\pi_1 \circ \pi_2} \tag{2.5.4}$$

for all $\pi, \pi_1, \pi_2 \in \mathcal{S}_n$.

Exercise 2.32

Prove the equality in (2.5.3), and prove both equalities in (2.5.4).

Given a d-dimensional Hilbert space $\mathcal{H}$ and the unitary representation $\{W^\pi\}_{\pi \in \mathcal{S}_n}$ of $\mathcal{S}_n$ defined in (2.5.1), we are interested in the subspace of vectors $|\psi\rangle \in \mathcal{H}^{\otimes n}$ that are *invariant* under permutations, i.e., $W^\pi|\psi\rangle = |\psi\rangle$ for all $\pi \in \mathcal{S}_n$. We call this subspace the *symmetric subspace of* $\mathcal{H}^{\otimes n}$, and it is formally defined as

$$\mathrm{Sym}_n(\mathcal{H}) := \mathrm{span}\{|\psi\rangle \in \mathcal{H}^{\otimes n} : W^\pi|\psi\rangle = |\psi\rangle \text{ for all } \pi \in \mathcal{S}_n\}. \tag{2.5.5}$$

A vector $|\psi\rangle \in \mathrm{Sym}_n(\mathcal{H})$ is sometimes called *symmetric*. The subspace

$$\mathrm{ASym}_n(\mathcal{H}) := \mathrm{span}\{|\psi\rangle \in \mathcal{H}^{\otimes n} : W^\pi|\psi\rangle = \mathrm{sgn}(\pi)|\psi\rangle \text{ for all } \pi \in \mathcal{S}_n\} \tag{2.5.6}$$

is called the *anti-symmetric subspace of* $\mathcal{H}^{\otimes n}$, where $\text{sgn}(\pi)$ is the *sign* of the permutation π, defined as $\text{sgn}(\pi) = (-1)^{T(\pi)}$ where $T(\pi)$ is the number of transpositions into which π can be decomposed[5].

The operator

$$\Pi_{\text{Sym}_n(\mathcal{H})} := \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} W^\pi \tag{2.5.7}$$

is the orthogonal projection onto the symmetric subspace of $\mathcal{H}^{\otimes n}$, while

$$\Pi_{\text{ASym}_n(\mathcal{H})} := \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \text{sgn}(\pi) W^\pi \tag{2.5.8}$$

is the orthogonal projection onto the anti-symmetric subspace of $\mathcal{H}^{\otimes n}$.

Exercise 2.33

1. Prove that $\Pi_{\text{Sym}_n(\mathcal{H})}$ and $\Pi_{\text{ASym}_n(\mathcal{H})}$ are projections, as claimed above.

2. Prove that $\text{Sym}_n(\mathcal{H})$ and $\text{ASym}_n(\mathcal{H})$ are orthogonal subspaces of $\mathcal{H}^{\otimes n}$ by showing that

$$\Pi_{\text{Sym}_n(\mathcal{H})} \Pi_{\text{ASym}_n(\mathcal{H})} = 0. \tag{2.5.9}$$

This implies that $\langle \psi_s | \psi_a \rangle = 0$ for all $|\psi_s\rangle \in \text{Sym}_n(\mathcal{H})$ and $|\psi_a\rangle \in \text{ASym}_n(\mathcal{H})$.

Exercise 2.34

Let $\mathcal{H}$ be a d-dimensional Hilbert space, $d \geq 2$. Show that, for $n = 2$,

$$\Pi_{\text{Sym}_2(\mathcal{H})} = \frac{1}{2}(\mathbb{1}_d \otimes \mathbb{1}_d + F), \tag{2.5.10}$$

$$\Pi_{\text{ASym}_2(\mathcal{H})} = \frac{1}{2}(\mathbb{1}_d \otimes \mathbb{1}_d - F), \tag{2.5.11}$$

[5] A transposition is a permutation that permutes only two elements of the set $\{1, 2, \dots, n\}$. Any permutation $\pi \in \mathcal{S}_n$ can be decomposed into a product of transpositions. Although this decomposition is in general not unique, the parity of the number $T(\pi)$ of transpositions into which π can be decomposed is unique, so that $\text{sgn}(\pi)$ is well defined.

where $F \coloneqq W^\pi$ is the representation of the permutation $\pi = (1\,2)$, i.e.,

$$F = \sum_{k,k'=0}^{d-1} |k\rangle\langle k'| \otimes |k'\rangle\langle k|. \tag{2.5.12}$$

In quantum information theory, F is referred to as the *swap operator*.

We focus primarily on the symmetric subspace of $\mathcal{H}^{\otimes n}$ in this book, and so we now provide some additional facts about it.

The following set of vectors constitutes an orthonormal basis for the symmetric subspace $\mathrm{Sym}_n(\mathcal{H})$ corresponding to the d-dimensional Hilbert space $\mathcal{H}$:

$$|n_1, n_2, \dots, n_d\rangle \coloneqq \frac{1}{\sqrt{n!\left(\prod_{j=1}^d n_j!\right)}} \sum_{\pi \in \mathcal{S}_n} W^\pi(|0\rangle^{\otimes n_1} \otimes |1\rangle^{\otimes n_2} \otimes \cdots \otimes |d-1\rangle^{\otimes n_d}), \tag{2.5.13}$$

where $n_1, n_2, \dots, n_d \geq 0$ are such that $\sum_{j=1}^d n_j = n$. We often call this the *occupation number basis* for $\mathrm{Sym}_n(\mathcal{H})$. The reason for this name is that, physically, each of the n Hilbert spaces $\mathcal{H}$ corresponds to a quantum system, and each n_j tells us how many of the n quantum systems are in the state given by $|j-1\rangle$. (We formally draw the correspondence between Hilbert spaces and quantum systems in Chapter 3.) The number of elements in this basis is equal to the number of ways of selecting n elements, with repetition, from a set of d distinct elements. This number is equal to $\binom{d+n-1}{n}$. Consequently, the dimension of $\mathrm{Sym}_n(\mathcal{H})$ is

$$\dim(\mathrm{Sym}_n(\mathcal{H})) = \binom{d+n-1}{n} = \binom{d+n-1}{d-1}. \tag{2.5.14}$$

Exercise 2.35

Let $d \geq 2$ and $n = 2$. Show that the basis elements $|n_1, n_2, \dots, n_d\rangle$ of $\mathrm{Sym}_2(\mathbb{C}^d)$ are given as follows:

$$|n_1, n_2, \dots, n_d\rangle = |j-1, j-1\rangle, \tag{2.5.15}$$

if $n_j = 2$, $n_\ell = 0\ \forall \ell \neq j$, and

$$|n_1, n_2, \dots, n_d\rangle = \frac{1}{\sqrt{2}}(|j-1, k-1\rangle + |k-1, j-1\rangle), \tag{2.5.16}$$

if $n_j = n_k = 1$, $k \neq j$ and $n_\ell = 0\ \forall \ell \neq j, k$.

REMARK: The direct sum vector space

$$\mathcal{F}_B(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathrm{Sym}_n(\mathcal{H}) \tag{2.5.17}$$

is called the *bosonic Fock space*. (Note that $\mathrm{Sym}_0(\mathcal{H})$ is the set of complex scalars, i.e., $\mathrm{Sym}_0(\mathcal{H}) = \mathbb{C}$.) It is an infinite-dimensional Hilbert space that is relevant for the study of quantum optical and other continuous-variable quantum systems.

An important fact that we state without proof (please consult the Bibliographic Notes in Section 2.6) is that for every d-dimensional Hilbert space $\mathcal{H}$,

$$\Pi_{\mathrm{Sym}_n(\mathcal{H})} = \binom{d+n-1}{n} \int \psi^{\otimes n}\, \mathrm{d}\psi, \tag{2.5.18}$$

where the integral on the right-hand side is taken with respect to the Haar measure over all unit vectors.

REMARK: The measure $\mathrm{d}\psi$ is also called the *Fubini-Study measure*. A concrete coordinate representation of the measure can be obtained by using the following parameterization of every unit vector $|\psi\rangle$ in a d-dimensional Hilbert space $\mathcal{H}$:

$$|\psi\rangle = \sum_{k=0}^{d-1} r_k \mathrm{e}^{\mathrm{i}\varphi_k} |k\rangle, \tag{2.5.19}$$

where $0 \leq \varphi_k \leq 2\pi$ and $r_k \geq 0$ for all $0 \leq k \leq d-1$. Furthermore, since $|\psi\rangle$ is a unit vector, we require that $\sum_{k=0}^{d-1} r_k^2 = 1$. The conditions on the coefficients r_k imply that they parameterize the positive octant of a sphere in d dimensions. As such, each r_k can be written as

$$r_0 = \prod_{k=1}^{d-1} \sin\frac{\theta_k}{2}, \tag{2.5.20}$$

$$r_m = \cos\frac{\theta_m}{2} \prod_{k=m+1}^{d-1} \sin\frac{\theta_k}{2}, \quad 1 \leq m \leq d-2, \tag{2.5.21}$$

$$r_{d-1} = \cos\frac{\theta_{d-1}}{2}, \tag{2.5.22}$$

where $0 \leq \theta_i \leq \pi$. Similarly, the angles φ_k parameterize a torus in d dimensions. The Fubini-Study measure $\mathrm{d}\psi$ is then the volume element of the coordinate system formed from the r_k and the coordinate system formed by the φ_k:

$$\mathrm{d}\psi = \frac{(d-1)!}{(2\pi)^{d-1}} \prod_{i=1}^{d-1} \cos\frac{\theta_i}{2} \sin^{2i-1}\frac{\theta_i}{2}\, \mathrm{d}\theta_i\, \mathrm{d}\varphi_i. \tag{2.5.23}$$

(Please consult the Bibliographic Notes in Section 2.6 for details.) In the case $d = 2$, we have that

$$\mathrm{d}\psi = \frac{1}{2\pi}\cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}\,\mathrm{d}\theta_1\,\mathrm{d}\varphi_1 = \frac{1}{4\pi}\sin(\theta_1)\,\mathrm{d}\theta_1\mathrm{d}\varphi_1 \quad (d=2), \tag{2.5.24}$$

recovering the familiar surface element of a 2-sphere (i.e., the surface of a 3-dimensional ball).

We often consider the case that the Hilbert space $\mathcal{H}$ is a tensor product of two Hilbert spaces, i.e., $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \equiv \mathcal{H}_{AB}$, with $\mathcal{H}_A$ a d_A-dimensional Hilbert space and $\mathcal{H}_B$ a d_B-dimensional Hilbert space. As we have seen above, if $\{|i\rangle_A\}_{i=0}^{d_A-1}$ is an orthonormal basis for $\mathcal{H}_A$ and $\{|j\rangle_B\}_{j=0}^{d_B-1}$ is an orthonormal basis for $\mathcal{H}_B$, then $\{|i,j\rangle_{AB} \equiv |i\rangle_A \otimes |j\rangle_B : 0 \leq i \leq d_A - 1, 0 \leq j \leq d_B - 1\}$ is an orthonormal basis for $\mathcal{H}_{AB}$. In this case, if we consider the n-fold tensor product $\mathcal{H}_{AB}^{\otimes n}$, then the unitary representation $\{W^\pi_{(AB)^n}\}_{\pi\in\mathcal{S}_n}$ defined in (2.5.1) acts as follows:

$$\begin{aligned} &W^\pi_{(AB)^n}(|i_1,j_1\rangle_{A_1B_1} \otimes |i_2,j_2\rangle_{A_2B_2} \otimes \cdots \otimes |i_n,j_n\rangle_{A_nB_n}) \\ &\qquad = |i_{\pi(1)},j_{\pi(1)}\rangle_{A_1B_1} \otimes |i_{\pi(2)},j_{\pi(2)}\rangle_{A_2B_2} \otimes \cdots \otimes |i_{\pi(n)},j_{\pi(n)}\rangle_{A_nB_n}, \end{aligned} \tag{2.5.25}$$

for all $0 \leq i_1, i_2, \dots, i_n \leq d_A - 1$ and all $0 \leq j_1, j_2, \dots, j_n \leq d_B - 1$. However, by rearranging the tensor factors, we find that the right-hand side of the above equation can be written as

$$W^\pi_{A^n}|i_1,i_2,\dots,i_n\rangle_{A_1\cdots A_n} \otimes W^\pi_{B^n}|j_1,j_2,\dots,j_n\rangle_{B_1B_2\cdots B_n}, \tag{2.5.26}$$

where $\{W^\pi_{A^n}\}_{\pi\in\mathcal{S}_n}$ and $\{W^\pi_{B^n}\}_{\pi\in\mathcal{S}_n}$ are the unitary representations of $\mathcal{S}_n$ acting on $\mathcal{H}_A^{\otimes n}$ and $\mathcal{H}_B^{\otimes n}$, respectively. We can thus write the projection onto $\mathrm{Sym}_n(\mathcal{H}_A \otimes \mathcal{H}_B)$ as

$$\Pi_{\mathrm{Sym}_n(\mathcal{H}_A\otimes\mathcal{H}_B)} = \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} W^\pi_{(AB)^n} \equiv \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} W^\pi_{A^n} \otimes W^\pi_{B^n}. \tag{2.5.27}$$

2.6 Bibliographic Notes

The study of inner product spaces, including Hilbert spaces, is the primary focus of functional analysis, for which we refer to the following books: (Reed and Simon, 1981; Kreyszig, 1989; Hall, 2013). In the case of finite-dimensional Hilbert spaces, which is what we consider throughout this book, many of the concepts studied in functional analysis reduce to those studied in linear algebra and matrix analysis. For these topics, we refer to Bhatia (1997); Horn and Johnson (2013); Strang (2016).

The generalized Gell-Mann matrices discussed after (2.2.49) were presented by Hioe and Eberly (1981); Bertlmann and Krammer (2008).

A review of operator monotone, operator concave, and operator convex functions is given by Bhatia (1997). The short course of Carlen (2010) is also helpful. For proofs of the properties listed immediately after Definition 2.13, see (Bhatia, 1997, Chapter V).

The proof of (2.2.93) follows immediately from (Bhatia, 1997, Problem III.6.2). A proof of (2.2.97), and therefore, of the Hölder inequality in (2.2.98), can be found in (Bhatia, 1997, Section IV & Exercise IV.2.12).

Lemma 2.11 can be found in (Audenaert and Eisert, 2005, Lemma 4). A proof of Proposition 2.8 can be found in Müller-Lennert et al. (2013, Lemma 12). Lemma 2.15 was proved by Lieb and Thirring (1976); Araki (1990). The Courant–Fischer–Weyl minimax principle, which is invoked in the proof of property 4 of Lemma 2.14, is presented in (Bhatia, 1997, Corollary III.1.2).

A proof of the operator Jensen inequality (Theorem 2.16) was given by Hansen and Pedersen (2003). In presenting the implication 1. ⇒ 3. of Theorem 2.16, we followed the proof given by Fujii et al. (2004, Theorem 3).

The notation $\|\cdot\|_\diamond$ for the quantity on the right-hand side of (2.2.172) was introduced by Kitaev (1997), and it is known as a *completely bounded trace norm* in the mathematics literature; see, for example, (Paulsen, 2003). The result in (2.2.173) is due to Smith (1983) (see Theorem 2.10 therein), but it can also be found in (Kitaev, 1997; Aharonov et al., 1998). For a proof of (2.2.175), see Theorem 3.51 in (Watrous, 2018), which also contains several more properties of the diamond norm.

For an introduction to real analysis, see (Rudin, 1976).

For an introduction to convex analysis, see (Rockafellar, 1970; Boyd and Vandenberghe, 2004), and for a proof of the Fenchel–Eggleston–Carathéodory theorem, see (Eggleston, 1958; Rockafellar, 1970).

Sion's minimax theorem (Theorem 2.24) is due to Sion (1958), and it is a generalization of a minimax theorem of von Neumann (1928). A short proof of Sion's minimax theorem can be found in (Komiya, 1988). The minimax theorem in Theorem 2.25 was presented by Mosonyi and Hiai (2011).

For an introduction to probability theory, see (Feller, 1968; Ross, 2019). Proofs of Markov's inequality (2.3.20) and Jensen's inequality (2.3.21) can be found in, e.g., (Fristedt and Gray, 1997).

For further details on semi-definite programming, see Vandenberghe and Boyd (1996); Watrous (2018). Various polynomial-time algorithms for solving semi-definite programs were developed by Khachiyan (1980); Arora et al. (2005); Arora and Kale (2007); Arora et al. (2012); Lee et al. (2015). A proof of Slater's Theorem (Theorem 2.28) can be found in (Boyd and Vandenberghe, 2004, Section 5.3.2).

For further details about the symmetric subspace of a tensor product of finite-dimensional Hilbert spaces, as well as for a proof of (2.5.18), see (Harrow, 2013) (see also Bengtsson and Zyczkowski (2017, Section 12.7)). Further details about the Fubini-Study measure $\mathrm{d}\psi$ introduced in (2.5.18) and elaborated upon in the remark immediately below it may be found in (Bengtsson and Zyczkowski, 2017, Chapter 4).

2.7 Problems

1. Prove that a linear operator $X \in \mathrm{L}(\mathcal{H})$ is positive semi-definite if and only if it can be written as $X = Y^\dagger Y$ for some $Y \in \mathrm{L}(\mathcal{H}, \mathcal{H}')$.

2. Prove that the columns of every isometry form an orthonormal set of vectors. Similarly, prove that the rows and columns of every unitary operator form orthonormal sets of vectors. (*Hint*: Consider using the expressions in (2.2.12).)

3. Let $X \in \mathrm{L}(\mathcal{H}_A)$ and $Y \in \mathrm{L}(\mathcal{H}_B)$ be normal operators, and consider their so-called

Kronecker sum:

$$X \oplus_{\mathrm{K}} Y := X \otimes \mathbb{1}_B + \mathbb{1}_A \otimes Y. \tag{2.7.1}$$

Prove that $\mathrm{spec}(X \oplus_{\mathrm{K}} Y) = \{\lambda + \mu : \lambda \in \mathrm{spec}(X),\ \mu \in \mathrm{spec}(Y)\}$. Also prove that the associated eigenvectors are of the form $|\psi\rangle \otimes |\phi\rangle$, where $|\psi\rangle$ is an eigenvector of X and $|\phi\rangle$ is an eigenvector of Y.

4. The *Hadamard product*, also known as the Schur product, of two linear operators $X, Y \in \mathrm{L}(\mathbb{C}^d)$, with $d \geq 2$, is defined to be the element-wise product of X and Y: if $X = \sum_{i,j=0}^{d-1} X_{i,j}|i\rangle\langle j|$ and $Y = \sum_{i,j=0}^{d-1} Y_{i,j}|i\rangle\langle j|$, then

$$X * Y := \sum_{i,j=0}^{d-1} X_{i,j}Y_{i,j}|i\rangle\langle j|. \tag{2.7.2}$$

   (a) Verify that for all $|\psi\rangle, |\phi\rangle \in \mathbb{C}^d$,

$$\langle\psi|X * Y|\phi\rangle = \mathrm{Tr}\left[X^{\mathsf{T}}\mathrm{diag}(\langle\psi|)Y\mathrm{diag}(|\phi\rangle)\right], \tag{2.7.3}$$

   where for $|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i|i\rangle$ and $|\phi\rangle = \sum_{j=0}^{d-1} \beta_j|j\rangle$,

$$\mathrm{diag}(\langle\psi|) := \sum_{i=0}^{d-1} \overline{\alpha_i}|i\rangle\langle i|, \quad \mathrm{diag}(|\phi\rangle) := \sum_{j=0}^{d-1} \beta_j|j\rangle\langle j|. \tag{2.7.4}$$

   (b) Prove that for all $|\psi\rangle, |\phi\rangle \in \mathbb{C}^d$,

$$|\psi\rangle\langle\psi| * |\phi\rangle\langle\phi| = (|\psi\rangle * |\phi\rangle)(\langle\psi| * \langle\phi|). \tag{2.7.5}$$

   (c) Prove that the Hadamard product of two positive semi-definite operators is positive semi-definite.

5. Let $\{|\psi_j\rangle\}_{j=1}^d$ be a set of d linearly independent vectors in $\mathbb{C}^d$, with $d \geq 2$. By definition, this means that, for all $c_1, c_2, \dots, c_d \in \mathbb{C}$, the equation $c_1|\psi_1\rangle + c_2|\psi_2\rangle + \cdots + c_d|\psi_d\rangle = 0$ implies $c_1 = c_2 = \cdots = c_d = 0$.

   (a) Let

$$T := \sum_{j=1}^{d} |\psi_j\rangle\langle j-1|. \tag{2.7.6}$$

   The operator T can be thought of as a $d \times d$ matrix whose columns are given by the vectors $|\psi_j\rangle$. Prove that T is invertible. (*Hint*: First prove that T is injective, by showing that its kernel contains only the zero vector. Then use the result of Exercise 2.4.)

(b) Using (a), prove that $\{|\psi_j\rangle\}_{j=1}^d$ is a basis for $\mathbb{C}^d$. In other words, prove that every vector $|\phi\rangle \in \mathbb{C}^d$ can be written as a unique linear combination of the vectors $|\psi_j\rangle$. We thus have that every set of d linearly independent vectors in $\mathbb{C}^d$ is a basis for $\mathbb{C}^d$.

(c) Prove that if $\sum_{j=1}^d |\psi_j\rangle\langle\psi_j| = \mathbb{1}_d$, then $\{|\psi\rangle_j\}_{j=1}^d$ is an orthonormal basis for $\mathbb{C}^d$. By combining this result with the result of Exercise 2.2, we have that a linearly independent set $\{|\psi_j\rangle\}_{j=1}^d$ of vectors in $\mathbb{C}^d$ is an orthonormal basis if and only if $\sum_{j=1}^d |\psi_j\rangle\langle\psi_j| = \mathbb{1}_d$.

6. Let $\{B_j\}_{j=1}^{d^2}$ be an orthonormal basis for $\mathrm{L}(\mathbb{C}^d)$, with $d \geq 2$.

(a) Prove that

$$\sum_{j=1}^{d^2} \overline{B_j} \otimes B_j = \Gamma_d, \tag{2.7.7}$$

where we recall that $\Gamma_d = |\Gamma_d\rangle\langle\Gamma_d| = \sum_{i,j=0}^{d-1} |i,i\rangle\langle j,j|$; see (2.2.36). Similarly, prove that

$$\sum_{j=1}^{d^2} B_j^\dagger \otimes B_j = F, \tag{2.7.8}$$

where we recall that $F = \sum_{i,j=0}^{d-1} |i,j\rangle\langle j,i|$; see (2.5.12).

(*Hint*: Start by verifying that $\{\overline{B_j}\}_{j=1}^{d^2}$ is an orthonormal basis for $\mathrm{L}(\mathbb{C}^d)$. Then, use the fact that every linear operator $Z \in \mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ can be written as $Z = \sum_{j,k=1}^{d^2} c_{j,k} \overline{B_j} \otimes B_k$ for some coefficients $c_{j,k} \in \mathbb{C}$.)

(b) Prove that for all $X \in \mathrm{L}(\mathbb{C}^d)$,

$$\sum_{j=1}^{d^2} B_j X B_j^\dagger = \mathrm{Tr}[X]\mathbb{1}_d. \tag{2.7.9}$$

(*Hint*: Use (2.7.7), along with the identities in (2.2.42)–(2.2.44).)

(c) Prove that $\{\mathrm{vec}(B_j)\}_{j=1}^{d^2}$ and $\{(B_j \otimes \mathbb{1}_d)|\Gamma_d\rangle\}_{j=1}^{d^2}$ are orthonormal bases for $\mathbb{C}^d \otimes \mathbb{C}^d$.

7. For all $d \geq 2$, construct a basis for $\mathrm{L}(\mathbb{C}^d)$ that consists entirely of density operators. (*Hint*: Consider using the eigenvectors of the orthonormal basis of Hermitian operators defined in (2.2.47)–(2.2.49).)

8. Let $\{|\psi_j\rangle\}_{j=1}^d$ be a set of linearly independent, normalized, but non-orthogonal vectors in $\mathbb{C}^d$, with $d \geq 2$. We would like to transform these vectors into a new set $\{|\phi_j\rangle\}_{j=1}^d$ of orthonormal vectors via an invertible linear operator X, such that $|\phi_j\rangle = X|\psi_j\rangle$ for all $j \in \{1, 2, \dots, d\}$.

   (a) Prove that the operator S defined as

$$S := \sum_{j=1}^{d} |\psi_j\rangle\langle\psi_j| \tag{2.7.10}$$

   is invertible and positive definite. (*Hint*: Write S in terms of the operator T defined in (2.7.6).)

   (b) Let

$$|\phi_j\rangle := S^{-\frac{1}{2}}|\psi_j\rangle \tag{2.7.11}$$

   for all $j \in \{1, 2, \dots, d\}$. Prove that $\{|\phi_j\rangle\}_{j=1}^d$ is an orthonormal basis for $\mathbb{C}^d$. (*Hint*: See problem 5.(c).) Also, prove that $\langle\phi_i|\psi_j\rangle = \langle i-1|G^{\frac{1}{2}}|j-1\rangle$ for all $i, j \in \{1, 2, \dots, d\}$, where $G := T^\dagger T$ and $T := \sum_{j=1}^d |\psi_j\rangle\langle j-1|$.

   (c) Let us now show that the vectors defined in (2.7.11) are optimal with respect to the Euclidean norm, in the following sense:

$$\inf_X \left\{ \sum_{j=1}^{d} \left\| |\psi_j\rangle - |\phi_j\rangle \right\|_2^2 : |\phi_j\rangle = X|\psi_j\rangle,\ \langle\phi_i|\phi_j\rangle = \delta_{i,j}\ \forall\, 1 \leq j \leq d \right\} \tag{2.7.12}$$

$$= \sum_{j=1}^{d} \left\| |\psi_j\rangle - S^{-\frac{1}{2}}|\psi_j\rangle \right\|_2^2, \tag{2.7.13}$$

   where the optimization in (2.7.12) is with respect to invertible linear operators X.

   i. Prove that solving the optimization problem given by (2.7.12) can be reduced to solving the optimization problem given by

$$\sup_X \left\{ \mathrm{Tr}[(X + X^\dagger)S] : XSX^\dagger = \mathbb{1}_d \right\}. \tag{2.7.14}$$

   ii. Prove that the constraint $XSX^\dagger = \mathbb{1}_d$ in (2.7.14) implies $X = US^{-\frac{1}{2}}$, where U is a unitary operator. (*Hint*: Consider a polar decomposition of X; see Theorem 2.3.) Hence, show that the optimization problem given by (2.7.14) is equivalent to

$$\sup_U \mathrm{Re}\left(\mathrm{Tr}[US^{\frac{1}{2}}]\right), \tag{2.7.15}$$

   where the optimization is with respect to unitary operators U acting on $\mathbb{C}^d$.

iii. Prove that the solution to the optimization problem given by (2.7.15) is $U = \mathbb{1}_d$, implying that the optimal X in (2.7.12) is indeed $S^{-\frac{1}{2}}$. (*Hint*: Use Proposition 2.10.)

(*Bibliographic Note*: The vectors $|\phi_j\rangle$ defined in (2.7.11) are known as the *symmetric orthogonalization* of the original vectors $|\psi_j\rangle$, and this construction is attributed to Löwdin (1950); see also (Löwdin, 1970). An alternate proof of the optimality of this construction, as worked out in part (c) of this problem, can be found in (Mayer, 2002).)

9. For the case $d = 2$ and $n = 2$, verify the equalities given by (2.5.7) and (2.5.18) by making use of (2.5.24).

10. Prove that the right-hand side of (2.5.7) is indeed the projection onto $\text{Sym}_n(\mathcal{H})$ by showing that

$$\sum_{\substack{n_1,n_2,\dots,n_d\geq 0,\\ \sum_{j=1}^d n_j = n}} |n_1, n_2, \dots, n_d\rangle\langle n_1, n_2, \dots, n_d| = \Pi_{\text{Sym}_n(\mathcal{H})}. \tag{2.7.16}$$

Chapter 3

Quantum States and Measurements

In the previous chapter, we studied several important topics in mathematics that collectively form one foundational piece for the study of quantum information processing. Another foundational piece is quantum mechanics, and in this and the following chapter, we provide an overview of it, placing particular emphasis on those aspects of it that are useful for the communication protocols that we discuss in later chapters. Many aspects of quantum mechanics cannot be explained by classical reasoning. For example, there is no strong classical analogue for pure quantum states or entanglement, and this leads to stark differences between what is possible in the classical and quantum worlds. However, at the same time, it is important to emphasize that all of classical information theory is subsumed by quantum information theory, so that whatever is possible with classical information processing is also possible with quantum information processing. Interestingly as well, quantum information processing allows for richer possibilities, with protocols such as quantum teleportation and super-dense coding.

3.1 Axioms of Quantum Mechanics

The mathematical description of quantum systems can be summarized by the following axioms. Each of these axioms is elaborated upon in the section indicated.

1. *Quantum systems*: A quantum system A is associated with a Hilbert space $\mathcal{H}_A$.

The state of the system A is described by a *density operator*, which is a unit-trace, positive semi-definite linear operator acting on $\mathcal{H}_A$. (See Section 3.2.)

2. *Bipartite quantum systems*: For distinct quantum systems A and B with associated Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the composite system AB is associated with the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. (See Section 3.2.1.)

3. *Measurement*: The measurement of a quantum system A is described by a *positive operator-valued measure (POVM)* $\{M_x\}_{x\in\mathcal{X}}$, which is defined to be a collection of positive semi-definite operators satisfying $\sum_{x\in\mathcal{X}} M_x = \mathbb{1}_{\mathcal{H}_A}$, where $\mathcal{X}$ is a finite alphabet[1]. If the system is in the state ρ and the measurement outcome is described by a random variable X, then the probability $\Pr[X = x]$ of obtaining the outcome x is given by the *Born rule* as

$$\Pr[X = x] = \mathrm{Tr}[M_x\rho]. \tag{3.1.1}$$

Furthermore, a physical *observable* O corresponds to a Hermitian operator acting on the underlying Hilbert space. Recall from the spectral theorem (Theorem 2.4) that O has a spectral decomposition as follows:

$$O = \sum_{\lambda\in\mathrm{spec}(O)} \lambda\, \Pi_\lambda, \tag{3.1.2}$$

where $\mathrm{spec}(O)$ is the set of distinct eigenvalues of O and Π_λ is a spectral projection. A measurement of O is described by the POVM $\{\Pi_\lambda\}_\lambda$, which is indexed by the distinct eigenvalues λ of O. The expected value $\langle O\rangle_\rho$ of the observable O when the state is ρ is given by

$$\langle O\rangle_\rho \coloneqq \mathrm{Tr}[O\rho] = \sum_{\lambda\in\mathrm{spec}(O)} \lambda\, \mathrm{Tr}[\Pi_\lambda\rho]. \tag{3.1.3}$$

(See Section 3.3.)

4. *Evolution*: The evolution of the state of a quantum system is described by a *quantum channel*, which is a linear, completely positive, and trace-preserving map acting on the state of the system. (See Chapter 4.)

Note that the second axiom for the description of bipartite quantum systems is sufficient to conclude that the multipartite quantum system $A_1A_2\cdots A_k$, comprising k distinct quantum systems $A_1, A_2, \ldots, A_k$, is associated with the Hilbert space $\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \otimes \cdots \otimes \mathcal{H}_{A_k}$.

[1]POVMs need not contain a finite number of elements, but we consider POVMs with a finite number of elements exclusively throughout this book.

3.2 Quantum Systems and States

Each quantum system is associated with a Hilbert space. In this book, we consider only finite-dimensional quantum systems, that is, quantum systems described by finite-dimensional Hilbert spaces. In the following, we provide a mathematical description of several finite-dimensional quantum systems, along with examples of how these systems can be physically realized.

1. *Qubit systems*: The qubit is perhaps the most fundamental quantum system and is the quantum analogue of the (classical) bit. Every physical system with two distinct states obeying the laws of quantum mechanics can be considered a qubit system. The Hilbert space associated with a qubit system is $\mathbb{C}^2$, whose standard orthonormal basis is denoted by $\{|0\rangle, |1\rangle\}$. Three common ways of physically realizing qubit systems are as follows:
   (a) The two spin states of a spin-$\frac{1}{2}$ particle.
   (b) Two distinct energy levels of an atom, such as the ground state and one of the excited states.
   (c) Clockwise and counter-clockwise directions of current flow in a superconducting electronic circuit.

2. *Qutrit systems*: A qutrit system is a quantum system consisting of three distinct physical degrees of freedom. The Hilbert space of a qutrit is $\mathbb{C}^3$, with the standard orthonormal basis denoted by $\{|0\rangle, |1\rangle, |2\rangle\}$. Qutrit systems are less commonly considered than qubit systems for implementations, although one important example of an implementation of a qutrit system occurs in quantum optical systems, which we discuss below. Like qubit systems, qutrit systems can also be physically realized using, for example, the spin states of a spin-1 atom or three distinct energy levels of an atom.

3. *Qudit systems*: A qudit system is a quantum system with d distinct degrees of freedom and is described by the Hilbert space $\mathbb{C}^d$, with the standard orthonormal basis denoted by $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$. The spin states of every spin-j atom can be used to realize a qudit system with $d = 2j + 1$. Another physical realization of a qudit system is with the d distinct energy levels of an atom.

4. *Quantum optical systems*: An important quantum system, particularly for the implementation of many quantum communication protocols, is a quantum

optical system. By a quantum optical system, we mean a physical system, such as an optical cavity or a fiber-optic cable, in which modes of light, with photons as information carriers, propagate. A mode of light has a well defined momentum, frequency, polarization, and spatial direction.

Formally, a quantum optical system with d distinct modes is described by the Fock space $\mathcal{F}_{\mathrm{B}}(\mathbb{C}^d)$, which is a Hilbert space equipped with the orthonormal occupation number basis $\{|n_1, \dots, n_d\rangle : n_1, \dots, n_d \geq 0\}$, where n_j, for $j \in \{1, \dots, d\}$, indicates the number of photons occupied in mode j. See (2.5.13) and the surrounding discussion for a brief review of the occupation number basis and Fock space.

The Fock space is infinite dimensional, but by restricting to particular subspaces, it is possible use photons to physically realize finite-dimensional quantum systems. The following two realizations of a qubit system are particularly important:

(a) A single-mode optical system, with Hilbert space $\mathcal{F}_{\mathrm{B}}(\mathbb{C})$, restricted to the subspace spanned by the orthonormal vectors $\{|0\rangle, |1\rangle\}$, interpreted as either zero or one photon occupied in the mode. The vector $|0\rangle$ corresponding to no photons is commonly called the *vacuum state vector* of the mode.

(b) A two-mode optical system, with Hilbert space $\mathcal{F}_{\mathrm{B}}(\mathbb{C}^2)$, restricted to the subspace spanned by the orthonormal vectors $\{|0,1\rangle, |1,0\rangle\}$, consisting of only one photon in total occupying either one of the two modes. This realization of a qubit system is commonly called the *dual-rail encoding* because it makes use of two modes of light. Two distinct polarization degrees of freedom of photons, such as horizontal and vertical polarizations, are commonly used as the two modes in dual-rail encodings of a qubit. One then usually lets $|H\rangle \equiv |0,1\rangle$ and $|V\rangle \equiv |1,0\rangle$ denote a horizontally- and vertically-polarized photon, respectively.

    By considering the three-dimensional subspace spanned by $\{|0,0\rangle, |0,1\rangle, |1,0\rangle\}$, that is, the dual-rail qubit system with the additional orthogonal vacuum state vector $|0,0\rangle$ of the two modes, we obtain a physical realization of a qutrit system. This particular realization of a qutrit system is relevant for communication protocols in the context of the erasure channel, which is discussed in Section 4.5.2.

Having discussed how a quantum system is mathematically described, let us now move on to the mathematical description of the state of a quantum system.

Definition 3.1 Quantum State

The state of a quantum system is described by a density operator acting on the underlying Hilbert space of the quantum system. A density operator is a unit-trace, positive semi-definite linear operator. Throughout the book, we identify a state with its corresponding density operator. We denote the set of density operators on a Hilbert space $\mathcal{H}$ as $\mathrm{D}(\mathcal{H})$.

We typically use the Greek letters ρ, σ, τ, or ω to denote quantum states.

Exercise 3.1

Prove that the set of quantum states is a *convex set*. (Recall the definition of a convex set from Section 2.3.3.) In other words, prove that for every alphabet $\mathcal{X}$ and set $\{\rho^x\}_{x\in\mathcal{X}}$ of quantum states, along with every probability distribution $p : \mathcal{X} \to [0, 1]$, the following convex combination is a quantum state:

$$\rho = \sum_{x\in\mathcal{X}} p(x)\rho^x. \tag{3.2.1}$$

The extremal points in the convex set of quantum states are called *pure states*. A pure state is a rank-one projection onto a unit vector in the Hilbert space. Concretely, pure states are of the form $|\psi\rangle\langle\psi|$ where $|\psi\rangle \in \mathcal{H}$ is a normalized vector. For convenience, we sometimes denote $|\psi\rangle\langle\psi|$ simply as ψ, and refer to the unit vector $|\psi\rangle$ as a *state vector*. Since every element of a convex set can be written as a convex combination of the extremal points in the set, every quantum state ρ that is not a pure state can be written as

$$\rho = \sum_{x\in\mathcal{X}} p(x)|\psi_x\rangle\langle\psi_x| \tag{3.2.2}$$

for some set $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ of state vectors defined with respect to a finite alphabet $\mathcal{X}$, where $p : \mathcal{X} \to [0, 1]$ is a probability distribution.

Exercise 3.2

Prove that a quantum state ρ is pure if and only if $\rho^2 = \rho$. More generally, prove that ρ is pure if and only if $\mathrm{Tr}[\rho^2] = 1$. The quantity $\mathrm{Tr}[\rho^2]$ is known as the *purity* of ρ.

A state ρ that is not pure is called a *mixed state*, because it can be thought of as arising from the lack of knowledge of which pure state from the set $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ in (3.2.2) the system has been prepared. Note that the decomposition in (3.2.2), of a quantum state into pure states, is generally not unique.

A state ρ is called *maximally mixed* if the set $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ in (3.2.2) consists of d orthonormal state vectors and the probability distribution $\{p(x)\}_{x\in\mathcal{X}}$ is uniform (i.e., $p(x) = \frac{1}{d}$ for all $x \in \mathcal{X}$). In this case, it follows that

$$\rho = \frac{\mathbb{1}_d}{d} =: \pi_d, \tag{3.2.3}$$

as a consequence of Exercise 2.2. The state π_d is called maximally mixed because it corresponds to having the most uncertainty about which state from the set $\{|\psi_k\rangle\}_{k=1}^{d}$ the system is in. This uncertainty can be quantified by using quantum entropy, and in Chapter 7, we find that the maximally mixed state π_d has the largest entropy among all states of a finite-dimensional system of dimension d, thus justifying the term "maximally mixed."

Now, let us recall the orthonormal basis of Hermitian operators defined in (2.2.46)–(2.2.49). In quantum information, it is common to scale these operators by $\sqrt{d}$, where d is the dimension, so that we have an orthogonal basis $\{S_k^{(d)}\}_{k=0}^{d^2-1}$ of Hermitian operators, with $S_0^{(d)} = \mathbb{1}_d$ and $S_k^{(d)}$, $k \in \{1, 2, \dots, d^2-1\}$, equal to the traceless operators in (2.2.47)–(2.2.49) multiplied by $\sqrt{d}$. Note here that we have also relabeled the indices of the set of operators defined in (2.2.46)–(2.2.49). These operators satisfy $\text{Tr}[(S_k^{(d)})^2] = d$ and $\text{Tr}[S_k^{(d)} S_\ell^{(d)}] = d\delta_{k,\ell}$ for all $k, \ell \in \{0, 1, \dots, d^2-1\}$. We often suppress the dimension and write $S_k \equiv S_k^{(d)}$ if the dimension is unimportant or clear from the context. Using these operators, we can write every density operator $\rho \in \text{D}(\mathbb{C}^d)$ in the following form:

$$\rho = \frac{1}{d}\left(\mathbb{1} + \sum_{k=1}^{d^2-1} r_k S_k\right), \tag{3.2.4}$$

where $r_k = \langle S_k, \rho\rangle = \text{Tr}[S_k\rho] \in \mathbb{R}$ for all $k \in \{1, 2, \dots, d^2-1\}$. The vector $\vec{r}_\rho := (r_1, r_2, \dots, r_{d^2-1}) \in \mathbb{R}^{d^2-1}$ is sometimes called the *Bloch vector*, or the *coherence vector*, of ρ; please see the Bibliographic Notes (Section 3.4) for more information on this terminology.

Exercise 3.3

Let ρ be the quantum state represented as in (3.2.4).

1. Verify that $\mathrm{Tr}[\rho] = 1$.

2. Prove that ρ is pure if and only if $\sum_{k=1}^{d^2-1} r_k^2 = d - 1$.

At this point, it is instructive to look at an example. Let us consider quantum systems with $d = 2$, i.e., qubits. The representation of an arbitrary quantum state ρ in (3.2.4) becomes

$$\rho = \frac{1}{2}(\mathbb{1} + r_1 X + r_2 Y + r_3 Z) \quad \text{(qubit state)}, \tag{3.2.5}$$

where X, Y, and Z are the Pauli operators, which we defined in (2.2.50) and (2.2.51):

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.2.6}$$

The fact that $\mathrm{Tr}[\rho] = 1$ follows from the fact that X, Y, and Z are traceless operators, while $\mathrm{Tr}[\mathbb{1}] = 2$. The condition for ρ to be positive semi-definite is left to the following exercise.

Exercise 3.4

Show that the positive semi-definiteness of every qubit state ρ, as represented in (3.2.5), is equivalent to $r_1^2 + r_2^2 + r_3^2 \leq 1$.

The condition $r_1^2 + r_2^2 + r_3^2 \leq 1$ for every qubit state ρ represented as in (3.2.5), along with the fact that $r_1, r_2, r_3 \in \mathbb{R}$, implies that the vector $\vec{r}_\rho = (r_1, r_2, r_3)$ lies on or inside the unit sphere in three dimensions, for every qubit state ρ. Furthermore, the condition in Exercise 3.3 for ρ to be pure implies that a qubit state is pure if and only if the vector $\vec{r}_\rho$ lies on the surface of the unit sphere. In quantum mechanics, this unit sphere is known as the *Bloch sphere*, and if we include all mixed states corresponding to the interior of the sphere, then we use the term *Bloch ball* to refer to the set of all qubit states. See Figure 3.1 for a visual representation of the Bloch ball.

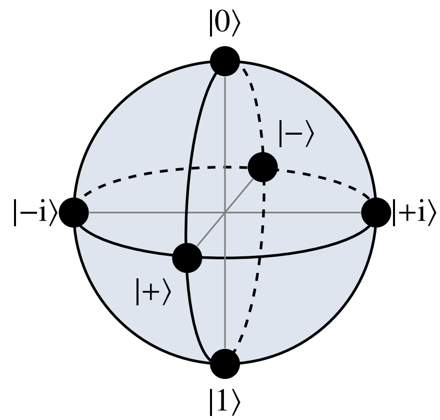


FIGURE 3.1: The quantum states in $\mathrm{D}(\mathbb{C}^2)$ of every qubit system can be represented as a point in the so-called *Bloch ball*. All pure states lie on the surface of the Bloch ball, which is known as the *Bloch sphere*. Shown are the basis state vectors $|0\rangle$ and $|1\rangle$, corresponding to the Bloch vectors $(0, 0, 1)$ and $(0, 0, -1)$, respectively. The superposition state vectors $|\pm\rangle \coloneqq \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ correspond to the Bloch vectors $(\pm 1, 0, 0)$, and the superposition state vectors $|\pm \mathrm{i}\rangle \coloneqq \frac{1}{\sqrt{2}}(|0\rangle \pm \mathrm{i}|1\rangle)$ correspond to the Bloch vectors $(0, \pm 1, 0)$.

3.2.1 Bipartite States and Schmidt Decomposition

The joint state of two distinct quantum systems A and B is described by a bipartite quantum state $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. For brevity, the joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ of the composite system AB is denoted by $\mathcal{H}_{AB}$.

Let $\{|i\rangle_A\}_{i=0}^{d_A-1}$ and $\{|j\rangle_B\}_{j=0}^{d_B-1}$ be orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then,

$$\{|i\rangle_A \otimes |j\rangle_B : 0 \leq i \leq d_A - 1,\ 0 \leq j \leq d_B - 1\} \tag{3.2.7}$$

is an orthonormal basis for $\mathcal{H}_{AB}$. For brevity, we often write $|i, j\rangle_{AB}$ instead of $|i\rangle_A \otimes |j\rangle_B$. Every state vector $|\psi\rangle_{AB} \in \mathcal{H}_{AB}$ can thus be written as

$$|\psi\rangle_{AB} = \sum_{i=0}^{d_A-1} \sum_{j=0}^{d_B-1} \alpha_{i,j} |i, j\rangle_{AB}, \tag{3.2.8}$$

where $\alpha_{i,j} = \langle i, j|\psi\rangle \in \mathbb{C}$ and $\sum_{i=0}^{d_A-1} \sum_{j=0}^{d_B-1} \left|\alpha_{i,j}\right|^2 = 1$. By the Schmidt decomposition theorem (Theorem 2.2), we can alternatively write $|\psi\rangle_{AB}$ as

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k} |e_k\rangle_A \otimes |f_k\rangle_B, \tag{3.2.9}$$

where each Schmidt coefficient λ_k is strictly positive and they all satisfy $\sum_{k=1}^{r}\lambda_k = 1$, $\{|e_k\rangle_A\}_{k=1}^{r}$ and $\{|f_k\rangle_B\}_{k=1}^{r}$ are orthonormal sets of vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and $r = \text{rank}(X)$, where $X \in \text{L}(\mathcal{H}_A, \mathcal{H}_B)$ is defined as $\langle j|_B X|i\rangle_A = \langle i, j|\psi\rangle_{AB}$ for all $0 \le i \le d_A - 1$ and $0 \le j \le d_B - 1$.

More generally, recall from Chapter 2 that we can define the orthonormal bases

$$\{|i\rangle\langle i'|_A : 0 \le i, i' \le d_A - 1\}, \quad \{|j\rangle\langle j'|_B : 0 \le j, j' \le d_B - 1\}, \tag{3.2.10}$$

for $\text{L}(\mathcal{H}_A)$ and $\text{L}(\mathcal{H}_B)$, respectively. Then, the set

$$\{|i,j\rangle\langle i',j'|_{AB} \equiv |i\rangle\langle i'|_A \otimes |j\rangle\langle j'|_B : 0 \le i, i' \le d_A - 1,\ 0 \le j, j' \le d_B - 1\} \tag{3.2.11}$$

is an orthonormal basis for $\text{L}(\mathcal{H}_{AB})$. It follows that every mixed state $\rho_{AB} \in \text{D}(\mathcal{H}_{AB})$ can be written as

$$\rho_{AB} = \sum_{i,i'=0}^{d_A-1} \sum_{j,j'=0}^{d_B-1} \beta_{i,j;i',j'} |i,j\rangle\langle i',j'|_{AB}, \tag{3.2.12}$$

where $\beta_{i,j;i',j'} = \langle |i,j\rangle\langle i',j'|, \rho_{AB}\rangle = \langle i,j|\rho_{AB}|i',j'\rangle \in \mathbb{C}$. Similarly, consider orthogonal bases $\{S_A^k\}_{k=0}^{d_A^2-1}$ and $\{S_B^\ell\}_{\ell=0}^{d_B^2-1}$ for $\text{L}(\mathcal{H}_A)$ and $\text{L}(\mathcal{H}_B)$, respectively, as defined in the paragraph above (3.2.4). Then, $\{S_A^k \otimes S_B^\ell : 0 \le k \le d_A^2 - 1,\ 0 \le \ell \le d_B^2 - 1\}$ is an orthogonal basis for $\text{L}(\mathcal{H}_{AB})$, so that every quantum state $\rho_{AB} \in \text{D}(\mathcal{H}_{AB})$ can be written as

$$\rho_{AB} = \frac{1}{d_A d_B} \sum_{k=0}^{d_A^2-1} \sum_{\ell=0}^{d_B^2-1} r_{k,\ell} S_A^k \otimes S_B^\ell, \tag{3.2.13}$$

where

$$r_{k,\ell} = \langle S_A^k \otimes S_B^\ell, \rho_{AB}\rangle = \text{Tr}[(S_A^k \otimes S_B^\ell)\rho_{AB}] \tag{3.2.14}$$

for all $k \in \{0, 1, \dots, d_A^2 - 1\}$ and $\ell \in \{0, 1, \dots, d_B^2 - 1\}$. Also, as with the Schmidt decomposition of state vectors in (3.2.9), from Exercise 2.14 we have that every mixed state ρ_{AB} can be written as

$$\rho_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}\, E_A^k \otimes F_B^k, \tag{3.2.15}$$

where each coefficient λ_k is strictly positive, $\{E_A^k\}_{k=1}^{r}$ and $\{F_B^k\}_{k=1}^{r}$ are orthonormal sets of linear operators acting on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and $r = \text{rank}(M)$, where $M \in \text{L}(\mathcal{H}_A \otimes \mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_B)$ is defined by $\langle j, \ell|_{BB} M|i, k\rangle_{AA} = \langle i, j|\rho_{AB}|k, \ell\rangle$ for all $0 \le i, j \le d_A - 1$ and $0 \le j, \ell \le d_B - 1$.

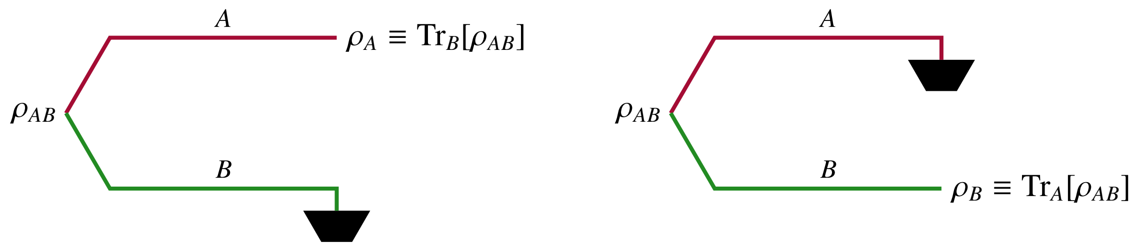


FIGURE 3.2: The partial trace superoperator (see Definition 3.2) is the mathematical representation of physically discarding a subsystem of a composite quantum system. In other words, given a bipartite state ρ_{AB} for the two quantum systems A and B, the partial trace Tr_B allows us to determine the quantum state of system A when we do not have access to system B (left), and Tr_A allows us to determine the quantum state of system B when we do not have access to system A (right).

3.2.2 Partial Trace

Recall from Section 2.2 that the trace of a linear operator X acting on a d-dimensional Hilbert space can be written as

$$\mathrm{Tr}[X] = \sum_{i=0}^{d-1} \langle i|X|i\rangle, \tag{3.2.16}$$

where $\{|i\rangle\}_{i=0}^{d-1}$ is the standard orthonormal basis. We can interpret the trace as the sum of the diagonal elements of the matrix corresponding to X written in the standard basis. From Exercise 2.5, however, we have that the trace is independent of the choice of basis used to evaluate it.

The trace is physically relevant, especially when it acts on one part of a bipartite quantum state, in which case we call it the *partial trace*. To be specific, given a state ρ_{AB} for the bipartite system AB, we are often interested in determining the state of only one of its subsystems. The partial trace Tr_B, which we define formally below, takes a state ρ_{AB} acting on the space $\mathcal{H}_{AB}$ and returns a state $\rho_A \equiv \mathrm{Tr}_B[\rho_{AB}]$ acting on the space $\mathcal{H}_A$. The partial trace is therefore the mathematical operation used to determine the state of one of the subsystems given the state of a composite system comprising two or more subsystems, and it can be thought of as the action of "discarding" one of the subsystems; see Figure 3.2. The partial trace generalizes the notion of marginalizing a joint probability distribution. Later, in Chapter 4, we see that partial trace is a particular kind of quantum channel corresponding to this discarding action.

Definition 3.2 Partial Trace

Given quantum systems A and B, the *partial trace over* B is denoted by $\mathrm{Tr}_B \equiv \mathrm{id}_A \otimes \mathrm{Tr}_B$, and it is defined as

$$\begin{aligned}\mathrm{Tr}_B[X_{AB}] &= (\mathrm{id}_A \otimes \mathrm{Tr})(X_{AB}) \\ &= \sum_{j=0}^{d_B-1} (\mathbb{1}_A \otimes \langle j|_B) X_{AB} (\mathbb{1}_A \otimes |j\rangle_B)\end{aligned} \tag{3.2.17}$$

for every linear operator $X_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The partial trace over A is defined similarly.

Remark: For every linear operator X_{AB} acting on $\mathcal{H}_{AB}$, we can define the partial trace $\mathrm{Tr}_B[X_{AB}]$ more abstractly as the unique linear operator X_A acting on $\mathcal{H}_A$ such that

$$\mathrm{Tr}[(M_A \otimes \mathbb{1}_B)X_{AB}] = \mathrm{Tr}[M_A X_A] \tag{3.2.18}$$

for every operator $M_A \in \mathrm{L}(\mathcal{H}_A)$. If we let X_{AB} be the state ρ_{AB} and M_A be a Hermitian operator, then we can interpret this equation physically in the following way: in order to determine the expectation value of an observable M_A acting on only one of the subsystems (in this case, the A subsystem), it suffices to know the reduced state ρ_A of the subsystem A rather than the joint state ρ_{AB} of the total system.

It is clear from Definition 3.2 that the partial trace is a linear superoperator. In particular, the expression in (3.2.17) defines the partial trace in precisely the operator-sum form for superoperators shown in (2.2.164).

Now, in order to explicitly determine the partial trace of a given linear operator $X_{AB} \in \mathrm{L}(\mathcal{H}_{AB})$, it suffices to know the action of the partial trace on basis elements of $\mathrm{L}(\mathcal{H}_{AB})$ because the action of every linear superoperator is completely defined by its action on basis elements. Using the orthonormal basis for $\mathrm{L}(\mathcal{H}_{AB})$ given in (3.2.11), it is straightforward to see that the action of the partial trace Tr_B on this basis is

$$\mathrm{Tr}_B[|i\rangle\langle i'|_A \otimes |j\rangle\langle j'|_B] = |i\rangle\langle i'|_A \delta_{j,j'} \tag{3.2.19}$$

for all $0 \le i, i' \le d_A - 1$, $0 \le j, j' \le d_B - 1$. Similarly, for Tr_A, we obtain

$$\mathrm{Tr}_A[|i\rangle\langle i'|_A \otimes |j\rangle\langle j'|_B] = \delta_{i,i'} |j\rangle\langle j'|_B \tag{3.2.20}$$

for all $0 \le i, i' \le d_A - 1$, $0 \le j, j' \le d_B - 1$. Then, by decomposing every linear operator X_{AB} as

$$\begin{aligned} X_{AB} &= \sum_{i,i'=0}^{d_A-1} \sum_{j,j'=0}^{d_B-1} X_{i,j;i',j'} |i\rangle\!\langle i'|_A \otimes |j\rangle\!\langle j'|_B \\ &= \sum_{i,i'=0}^{d_A-1} \sum_{j,j'=0}^{d_B-1} X_{i,j;i',j'} |i,j\rangle\!\langle i',j'|_{AB}, \end{aligned} \tag{3.2.21}$$

where $X_{i,j;i',j'} \coloneqq \langle i,j|X_{AB}|i',j'\rangle$, we find that

$$\mathrm{Tr}_B[X_{AB}] = \sum_{i,i'=0}^{d_A-1} \left(\sum_{j=0}^{d_B-1} X_{i,j;i',j} \right) |i\rangle\!\langle i'|_A, \tag{3.2.22}$$

$$\mathrm{Tr}_A[X_{AB}] = \sum_{j,j'=0}^{d_B-1} \left(\sum_{i=0}^{d_A-1} X_{i,j;i,j'} \right) |j\rangle\!\langle j'|_B. \tag{3.2.23}$$

For every bipartite linear operator X_{AB}, we let

$$X_A \equiv \mathrm{Tr}_B[X_{AB}] \text{ and } X_B \equiv \mathrm{Tr}_A[X_{AB}] \tag{3.2.24}$$

denote its partial traces. For states, we also use the terms *marginal states* or *reduced states* to refer to their partial traces.

An immediate consequence of the Schmidt decomposition theorem is that the marginal states $\rho_A \coloneqq \mathrm{Tr}_B[|\psi\rangle\!\langle\psi|_{AB}]$ and $\rho_B \coloneqq \mathrm{Tr}_A[|\psi\rangle\!\langle\psi|_{AB}]$ of every pure state $|\psi\rangle\!\langle\psi|_{AB}$ have the same non-zero eigenvalues. Indeed, using (3.2.9), we find that

$$\rho_A = \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} \mathrm{Tr}_B[|e_k\rangle\!\langle e_{k'}|_A \otimes |f_k\rangle\!\langle f_{k'}|_B] \tag{3.2.25}$$

$$= \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} |e_k\rangle\!\langle e_{k'}|_A \delta_{k,k'} \tag{3.2.26}$$

$$= \sum_{k=1}^{r} \lambda_k |e_k\rangle\!\langle e_k|_A, \tag{3.2.27}$$

$$\text{and} \quad \rho_B = \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} \mathrm{Tr}_A[|e_k\rangle\!\langle e_{k'}|_A \otimes |f_k\rangle\!\langle f_{k'}|_B] \tag{3.2.28}$$

$$= \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} \delta_{k,k'} |f_k\rangle\langle f_{k'}|_B \tag{3.2.29}$$

$$= \sum_{k=1}^{r} \lambda_k |f_k\rangle\langle f_k|_B, \tag{3.2.30}$$

in which the equalities in (3.2.27) and (3.2.30) contain spectral decompositions of ρ_A and ρ_B.

Exercise 3.5

Consider two quantum systems A and B, with $d_A = d_B = d$.

1. Calculate $\mathrm{Tr}_A[|\Gamma\rangle\langle\Gamma|_{AB}]$ and $\mathrm{Tr}_B[|\Gamma\rangle\langle\Gamma|_{AB}]$, where we recall from (2.2.36) that $|\Gamma\rangle_{AB} = \sum_{j=0}^{d-1} |j, j\rangle_{AB}$.
2. Calculate $\mathrm{Tr}_A[F_{AB}]$ and $\mathrm{Tr}_B[F_{AB}]$, where we recall from (2.5.12) that $F_{AB} = \sum_{k,k'=0}^{d-1} |k, k'\rangle\langle k', k|_{AB}$.

Below are two useful lemmas about how the support of a bipartite linear operator (recall the definition of support from Section 2.2) relates to the support of its partial traces. Their proofs are somewhat technical, and so we provide them in Appendices 3.A and 3.B.

Lemma 3.3

Let $X_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be positive semi-definite, and let $X_A := \mathrm{Tr}_B[X_{AB}]$ and $X_B := \mathrm{Tr}_A[X_{AB}]$. Then $\mathrm{supp}(X_{AB}) \subseteq \mathrm{supp}(X_A) \otimes \mathrm{supp}(X_B)$.

Lemma 3.4

Let $X_{AB}, Y_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be positive semi-definite, and suppose that $\mathrm{supp}(X_{AB}) \subseteq \mathrm{supp}(Y_{AB})$. Then $\mathrm{supp}(X_A) \subseteq \mathrm{supp}(Y_A)$, where $X_A := \mathrm{Tr}_B[X_{AB}]$ and $Y_A := \mathrm{Tr}_B[Y_{AB}]$.

3.2.3 Separable and Entangled States

The concepts of separable and entangled states are at the heart of virtually all of the communication protocols that we consider in this book. More generally, entanglement is a key distinction between the classical and quantum theories of information; it simply is not present and therefore does not play a role in classical information theory. Entanglement, in particular, is a key element of private communication and secure key distillation, and the successful distribution of entangled states among several spatially separated parties is a crucial ingredient in the implementation of such protocols over the future quantum internet. If the parties share only separable, unentangled states, then it is not possible for them to distill a key that is secure against a general quantum adversary.

We begin this section by defining separable and entangled states.

Definition 3.5 Separable and Entangled States

A bipartite state σ_{AB} is called *separable* if there exists a finite alphabet $\mathcal{X}$, a probability distribution $p : \mathcal{X} \to [0,1]$ on $\mathcal{X}$, and sets $\{\omega_A^x\}_{x\in\mathcal{X}}$ and $\{\tau_B^x\}_{x\in\mathcal{X}}$ of states for A and B, respectively, such that

$$\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x. \tag{3.2.31}$$

In other words, a state is called separable if it can be written as a convex combination of *product states*, each of which has the form $\omega_A \otimes \tau_B$. The set of separable states on $\mathcal{H}_{AB}$ is denoted by $\mathrm{SEP}(A\!:\!B)$.

A state that is not separable is called *entangled*.

Remark: Note that a separable state can always be written in the form

$$\sigma_{AB} = \sum_{x'\in\mathcal{X}'} q(x')|\psi_{x'}\rangle\langle\psi_{x'}|_A \otimes |\phi_{x'}\rangle\langle\phi_{x'}|_B \tag{3.2.32}$$

for some probability distribution $q : \mathcal{X}' \to [0,1]$ on a finite alphabet $\mathcal{X}'$ and sets of pure states $\{|\psi_{x'}\rangle\langle\psi_{x'}|_A : x' \in \mathcal{X}'\}$, $\{|\phi_{x'}\rangle\langle\phi_{x'}|_B : x' \in \mathcal{X}'\}$. In other words, separable states can always be written as a convex combination of pure product states. Indeed, from (3.2.31), we can take spectral decompositions of ω_A^x and τ_B^x,

$$\omega_A^x = \sum_{k=1}^{r_A^x} t_k^x |e_k^x\rangle\langle e_k^x|_A, \quad \tau_B^x = \sum_{\ell=1}^{r_B^x} s_\ell^x |f_\ell^x\rangle\langle f_\ell^x|_B, \tag{3.2.33}$$

where $r_A^x = \text{rank}(\omega_A^x)$ and $r_B^x = \text{rank}(\tau_B^x)$, so that

$$\rho_{AB} = \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_A^x} \sum_{\ell=1}^{r_B^x} p(x) t_k^x s_\ell^x |e_k^x\rangle\langle e_k^x|_A \otimes |f_\ell^x\rangle\langle f_\ell^x|_B. \tag{3.2.34}$$

Then, define the alphabet $\mathcal{X}' = \{x' \coloneqq (x,k,\ell) : x \in \mathcal{X},\ 1 \le k \le r_A^x,\ 1 \le \ell \le r_B^x\}$, so that x' is a superindex, and the unit vectors

$$|\psi_{x'}\rangle_A \coloneqq |e_k^x\rangle_A, \quad |\phi_{x'}\rangle_B \coloneqq |f_\ell^x\rangle_B. \tag{3.2.35}$$

Also, define the probability distribution $q : \mathcal{X}' \to [0,1]$ by

$$q(x,k,\ell) = p(x) t_k^x s_\ell^x. \tag{3.2.36}$$

Therefore, (3.2.34) can be written as

$$\sigma_{AB} = \sum_{x'\in\mathcal{X}'} q(x') |\psi_{x'}\rangle\langle\psi_{x'}|_A \otimes |\phi_{x'}\rangle\langle\phi_{x'}|_B. \tag{3.2.37}$$

From the development above, it follows that the set of separable states is the convex hull of the set of pure product states. By an application of the Fenchel–Eggleston–Carathéodory Theorem (Theorem 2.23), it follows that σ_{AB} can be written as a convex combination of no more than $\text{rank}(\sigma_{AB})^2$ pure product states. Indeed, in Theorem 2.23, it suffices to take as S the set of pure product states of the form $|\psi\rangle\langle\psi|_A \otimes |\phi\rangle\langle\phi|_B$, where $|\psi\rangle_A \otimes |\phi\rangle_B \in \text{supp}(\sigma_{AB})$. The relevant real (affine) space is the space of bipartite Hermitian operators with support contained in $\text{supp}(\sigma_{AB})$ and with trace one, which has real dimension $\text{rank}(\sigma_{AB})^2 - 1$.

In the sense that follows, bipartite separable states can be thought of as exhibiting purely classical correlations between the two parties, Alice and Bob. Suppose that Alice draws x with probability $p(x)$, prepares her system in the state ω_A^x, sends x to Bob over a classical channel, who then prepares his system in the state τ_B^x, where $x \in \mathcal{X}$ and $\mathcal{X}$ is a finite alphabet. This procedure corresponds to preparing the ensemble $\{(p(x), \omega_A^x \otimes \tau_B^x)\}_{x\in\mathcal{X}}$, and if Alice and Bob discard the label x, then their shared joint state is the separable state $\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x$.

On the other hand, no such procedure consisting of only local operations by Alice and Bob, supplemented by classical communication between them, can ever be used to generate an entangled state between them (without them already sharing some entanglement beforehand). In Section 4.6.2, we introduce local operations and classical communication (LOCC) channels and explain this point in more detail. Essentially, two entangled quantum systems are intrinsically linked in such a way that it is insufficient to describe each one individually.

Observe that a pure state ψ_{AB} is separable if and only if it is a product state, i.e., if and only if there exist pure states ϕ_A and φ_B such that $\psi_{AB} = \phi_A \otimes \varphi_B$. Recalling that every pure state has a Schmidt decomposition (see Theorem 2.2), we obtain the following result:

A pure state is entangled if and only if its Schmidt rank is strictly greater than one.

An important example of an entangled pure state is the state Φ_{AB} on two d-dimensional systems A and B, defined as $\Phi_{AB} \coloneqq |\Phi\rangle\langle\Phi|_{AB}$, where

$$|\Phi\rangle_{AB} \coloneqq \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_A \otimes |i\rangle_B = \frac{1}{\sqrt{d}} |\Gamma\rangle_{AB}, \tag{3.2.38}$$

and $|\Gamma\rangle_{AB} = \sum_{i=0}^{d-1} |i\rangle_A \otimes |i\rangle_B$ is the vector defined in (2.2.36).

Exercise 3.6

1. Show that $\mathrm{Tr}_A[\Phi_{AB}] = \mathrm{Tr}_B[\Phi_{AB}] = \frac{\mathbb{1}_d}{d}$.
2. Define $\Phi^U_{AB} \coloneqq (U_A \otimes \mathbb{1}_B)\Phi_{AB}(U_A \otimes \mathbb{1}_B)^\dagger$, for U_A a unitary acting on system A. Show that $\mathrm{Tr}_A[\Phi^U_{AB}] = \mathrm{Tr}_B[\Phi^U_{AB}] = \frac{\mathbb{1}_d}{d}$.

The state Φ_{AB} is an example of a maximally entangled state.

Definition 3.6 Maximally Entangled Pure State

A pure state $\psi_{AB} = |\psi\rangle\langle\psi|_{AB}$, for two systems A and B of the same dimension d, is called *maximally entangled* if the Schmidt coefficients of $|\psi\rangle_{AB}$ are all equal to $\frac{1}{\sqrt{d}}$, with d being the Schmidt rank of $|\psi\rangle_{AB}$.

In other words, ψ_{AB} is called maximally entangled if $|\psi\rangle_{AB}$ has the Schmidt decomposition

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |e_k\rangle_A \otimes |f_k\rangle_B \tag{3.2.39}$$

for some orthonormal sets $\{|e_k\rangle_A : 1 \leq k \leq d\}$ and $\{|f_k\rangle_B : 1 \leq k \leq d\}$. Observe then that

$$\mathrm{Tr}_A[|\psi\rangle\langle\psi|_{AB}] = \frac{\mathbb{1}_d}{d} = \mathrm{Tr}_B[|\psi\rangle\langle\psi|_{AB}]. \tag{3.2.40}$$

In other words, like the state Φ_{AB}, the marginal states of every maximally entangled state are maximally mixed for A and B.

Maximally entangled states provide a good example of why entangled quantum systems are, in a sense, greater than the sum of their parts. Since maximally entangled states have maximally mixed marginal states, the individual quantum systems can be viewed as being in a completely random state; indeed, as we have seen, maximally mixed states can be written as the expected state of every ensemble of orthonormal pure states with uniform probability distribution. However, intriguingly, the overall composite system is in a pure, definite state.

Exercise 3.7

Prove that every state vector of the form $(\mathbb{1}_d \otimes U)|\Phi_d\rangle = \frac{1}{\sqrt{d}}\mathrm{vec}(U)$ and $(U \otimes \mathbb{1}_d)|\Phi_d\rangle$, with $d \geq 2$ and U a unitary operator, corresponds to a maximally entangled pure state. Conversely, given a maximally entangled pure state $|\psi\rangle\langle\psi|_{AB}$, prove that there exists a unitary U_A such that $|\psi\rangle_{AB} = (U_A \otimes \mathbb{1}_B)|\Phi\rangle_{AB}$.

3.2.4 Bell States

In Exercise 3.7, we learned that every state vector of the form $(\mathbb{1} \otimes U)|\Phi\rangle$ and $(U \otimes \mathbb{1})|\Phi\rangle$, with U a unitary operator, is a maximally entangled state. We now provide an important example of a class of maximally entangled states, known as Bell states, for every dimension $d \geq 2$. These states are defined by particular choices for the unitary U. The Bell states are an important element of many quantum information processing tasks, most prominently quantum teleportation and super-dense coding, which we discuss in Chapter 5.

We start with dimension $d = 2$. Recall the Pauli operators X and Z from (3.2.6):

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.2.41}$$

Observe that, in addition to being Hermitian, these operators are unitary, which is due to the fact that $X^2 = Z^2 = \mathbb{1}$. The operator Y defined in (3.2.6) is also unitary, since $Y^2 = \mathbb{1}$, from which it follows that the operator $ZX = \mathrm{i}Y$ is also unitary. Using the operators X, Z, and ZX, we define the following set of four entangled, two-qubit state vectors:

$$|\Phi_{z,x}\rangle := (Z^z X^x \otimes \mathbb{1})|\Phi\rangle, \tag{3.2.42}$$

for $z, x \in \{0, 1\}$, where we recall from (3.2.38) that $|\Phi\rangle := \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$. The corresponding density operators $\Phi_{z,x}$ are known as the *two-qubit Bell states*. The following notation is commonly used:

$$|\Phi^+\rangle \equiv |\Phi_{0,0}\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle), \tag{3.2.43}$$

$$|\Phi^-\rangle \equiv |\Phi_{1,0}\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle - |1,1\rangle), \tag{3.2.44}$$

$$|\Psi^+\rangle \equiv |\Phi_{0,1}\rangle = \frac{1}{\sqrt{2}}(|0,1\rangle + |1,0\rangle), \tag{3.2.45}$$

$$|\Psi^-\rangle \equiv |\Phi_{1,1}\rangle = \frac{1}{\sqrt{2}}(|0,1\rangle - |1,0\rangle). \tag{3.2.46}$$

Exercise 3.8

1. Prove that the two-qubit Bell state vectors defined in (3.2.42) form an orthonormal basis for $\mathbb{C}^2 \otimes \mathbb{C}^2$.

2. Prove that the state vectors $|\Phi^+\rangle$, $|\Phi^-\rangle$, and $|\Psi^+\rangle$ form an orthonormal basis for $\mathrm{Sym}_2(\mathbb{C}^2)$. (*Hint*: See (2.5.13) and Exercise 2.35.) For this reason, the subspace $\mathrm{Sym}_2(\mathbb{C}^2)$ is called the triplet subspace.

3. Prove that $\mathrm{ASym}_2(\mathbb{C}^2) = \mathrm{span}\{|\Psi^-\rangle\}$. For this reason, the subspace $\mathrm{ASym}_2(\mathbb{C}^2)$ is called the singlet subspace and the state $|\Psi^-\rangle$ is called the singlet state vector.

We can generalize the Bell state vectors in (3.2.42) to systems with dimension $d > 2$. Doing so requires a generalization of the qubit Pauli operators X and Z to unitary operators for qudits[2].

[2] The qudit operators defined in (2.2.46)–(2.2.49) represent one generalization of the qubit Pauli operators. Although they are Hermitian, they are not generally unitary. What we require here is a generalization to qudit operators that are unitary.

Definition 3.7 Heisenberg–Weyl Operators

Let $d \geq 2$. The *Heisenberg–Weyl operators* make up a set $\{W_{z,x} : 0 \leq z, x \leq d-1\}$ of d^2 unitary operators acting on $\mathbb{C}^d$. They are defined as follows:

$$W_{z,x} = Z(z)X(x), \tag{3.2.47}$$

$$Z(z) := \sum_{k=0}^{d-1} e^{\frac{2\pi i k z}{d}} |k\rangle\langle k|, \tag{3.2.48}$$

$$X(x) := \sum_{k=0}^{d-1} |k+x\rangle\langle k|, \tag{3.2.49}$$

where the addition operation in the definition of $X(x)$ is performed modulo d.

Exercise 3.9

1. Verify that when $d = 2$, the Heisenberg–Weyl operators reduce to the qubit Pauli operators Z, X, and ZX.

2. Prove that the operators $Z(z)$ and $X(x)$ defined in (3.2.48) and (3.2.49) satisfy the commutation relation

$$Z(z)X(x) = e^{\frac{2\pi i x z}{d}} X(x)Z(z), \tag{3.2.50}$$

for all $z, x \in \{0, 1, \dots, d-1\}$.

The Heisenberg–Weyl operators are unitary, just like the Pauli operators; however, unlike the Pauli operators, they are *not* Hermitian. In particular,

$$W_{z,x}^\dagger = e^{-\frac{2\pi i x z}{d}} W_{-z,-x}. \tag{3.2.51}$$

It is also straightforward to show that

$$W_{z_1,x_1} W_{z_2,x_2} = e^{-\frac{2\pi i x_1 z_2}{d}} W_{z_1+z_2,x_1+x_2}. \tag{3.2.52}$$

Furthermore, the Heisenberg–Weyl operators are orthogonal with respect to the Hilbert–Schmidt inner product, meaning that

$$\langle W_{z_1,x_1}, W_{z_2,x_2} \rangle = \mathrm{Tr}[W_{z_1,x_1}^\dagger W_{z_2,x_2}] = d\delta_{z_1,z_2}\delta_{x_1,x_2} \tag{3.2.53}$$

for all $0 \leq z_1, z_2, x_1, x_2 \leq d-1$. This implies that the scaled Heisenberg–Weyl operators $\left\{\frac{1}{\sqrt{d}}W_{z,x} : 0 \leq z, x \leq d-1\right\}$ form an orthonormal basis for $\mathrm{L}(\mathbb{C}^d)$ for all $d \geq 2$.

Exercise 3.10

Prove (3.2.51), (3.2.52), and (3.2.53).

Exercise 3.11

Let $d \geq 2$, and consider the operator Q_d defined as

$$Q_d := \frac{1}{\sqrt{d}} \sum_{k,\ell=0}^{d-1} \mathrm{e}^{\frac{2\pi \mathrm{i} k\ell}{d}} |k\rangle\langle\ell|. \tag{3.2.54}$$

1. Show that Q_d is a unitary operator.
2. Prove that

$$Q_d X(x) Q_d^\dagger = Z(x), \tag{3.2.55}$$

$$Q_d Z(z) Q_d^\dagger = X(z)^\dagger, \tag{3.2.56}$$

   for all $z, x \in \{0, 1, \ldots, d-1\}$.

The unitary operator Q_d is known as the *(discrete) quantum Fourier transform* operator.

Using the Heisenberg–Weyl operators, we now define the set of qudit Bell states in a manner analogous to (3.2.42).

Definition 3.8 Qudit Bell States

Let $d \geq 2$. The *qudit Bell states* are d^2 pure quantum states $\Phi_{z,x} := |\Phi_{z,x}\rangle\langle\Phi_{z,x}|$ in $\mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$, where

$$|\Phi_{z,x}\rangle := (W_{z,x} \otimes \mathbb{1}_d)|\Phi_d\rangle \tag{3.2.57}$$

for all $z, x \in \{0, 1, \ldots, d-1\}$, and $|\Phi_d\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j, j\rangle$.

Exercise 3.12

Prove that the two-qudit Bell state vectors defined in (3.2.57) form an orthonormal basis for $\mathbb{C}^d \otimes \mathbb{C}^d$ for all $d \geq 2$.

The fact that the two-qubit Bell state vectors form an orthonormal basis for $\mathbb{C}^d \otimes \mathbb{C}^d$ implies, from Exercise 2.2, that

$$\sum_{z,x=0}^{d-1} |\Phi_{z,x}\rangle\langle\Phi_{z,x}| = \mathbb{1}_d \otimes \mathbb{1}_d. \tag{3.2.58}$$

A two-qudit state ρ_{AB} is known as a *Bell-diagonal state* if it is diagonal in the two-qudit Bell basis, so that it is of the form

$$\sum_{z,x=0}^{d-1} p(z,x)|\Phi_{z,x}\rangle\langle\Phi_{z,x}| \tag{3.2.59}$$

for some probability distribution $p : \{0,1,\ldots,d-1\}^2 \to [0,1]$, meaning that $0 \leq p(z,x) \leq 1$ for all $z,x \in \{0,1,\ldots,d-1\}$ and $\sum_{z,x=0}^{d-1} p(z,x) = 1$.

3.2.5 Purifications and Extensions

One of the most useful concepts in quantum information is the notion of purification. There is no strong classical analogue of this concept, and thus this notion represents another distinction between the classical and quantum theories of information.

Definition 3.9 Purification

Let ρ_A be a state of a system A. A *purification* of ρ_A is a pure state $|\psi\rangle\langle\psi|_{RA}$ for a bipartite system RA such that

$$\mathrm{Tr}_R[|\psi\rangle\langle\psi|_{RA}] = \rho_A. \tag{3.2.60}$$

We often call the reference system R the “purifying system.”

The following simple theorem establishes that every state ρ_A has a purification.

Theorem 3.10 State Purification

For every state ρ_A and every reference system R with $d_R \geq \text{rank}(\rho_A)$, there exists a purification $|\psi\rangle\langle\psi|_{RA}$ of ρ_A on RA.

PROOF: Consider a spectral decomposition of ρ_A

$$\rho_A = \sum_{k=1}^{r} \lambda_k |\phi_k\rangle\langle\phi_k|, \tag{3.2.61}$$

where $r = \text{rank}(\rho_A)$. Consider a reference system R with $d_R \geq r$ and an arbitrary set $\{|\varphi_k\rangle_R : 1 \leq k \leq r\}$ of orthonormal states. The unit vector

$$|\psi\rangle_{RA} \coloneqq \sum_{k=1}^{r} \sqrt{\lambda_k} |\varphi_k\rangle_R \otimes |\phi_k\rangle_A \tag{3.2.62}$$

then satisfies

$$\text{Tr}_R[|\psi\rangle\langle\psi|_{RA}] = \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} \underbrace{\text{Tr}[|\varphi_k\rangle\langle\varphi_{k'}|_R]}_{\delta_{k,k'}} |\phi_k\rangle\langle\phi_{k'}|_A \tag{3.2.63}$$

$$= \sum_{k=1}^{r} \lambda_k |\phi_k\rangle\langle\phi_k|_A \tag{3.2.64}$$

$$= \rho_A, \tag{3.2.65}$$

so that $|\psi\rangle\langle\psi|_{RA}$ is a purification of ρ_A, as required. ■

REMARK: The theorem above states that the condition $d_R \geq \text{rank}(\rho_A)$ on the dimension of the purifying system R is sufficient to guarantee the existence of a purification. This condition is also necessary, meaning that it is not possible to have a purifying system whose dimension is less than the rank of ρ_A.

The proof of the theorem above not only tells us that every state has a purification, but it also tells us explicitly how to construct one such purification. We can also construct a purification of every state ρ_A as follows:

$$|\psi\rangle_{RA} = (\mathbb{1}_R \otimes \sqrt{\rho_A})|\Gamma\rangle_{RA} = \text{vec}(\sqrt{\rho_A}), \tag{3.2.66}$$

where $|\Gamma\rangle_{RA} = \sum_{i=0}^{d_A-1} |i,i\rangle_{RA}$ and where the operation vec is defined in (2.2.33). We often call the state $|\psi\rangle\langle\psi|_{RA}$ the *canonical purification* of ρ_A. Note that the canonical purification is very closely related to the purification used in the proof of Theorem 3.10. Indeed, if

$$\rho_A = \sum_{k=1}^{r} \lambda_k |\phi_k\rangle\langle\phi_k| \tag{3.2.67}$$

is a spectral decomposition of ρ_A, with $r = \text{rank}(\rho_A)$, then

$$\text{vec}(\sqrt{\rho_A}) = \sum_{k=1}^{r} \sqrt{\lambda_k}\, \overline{|\phi_k\rangle}_R \otimes |\phi_k\rangle_A, \tag{3.2.68}$$

where we have made use of (2.2.37).

Physically, the fact that every state ρ_A has a purification means that every quantum system A in a mixed state can be viewed as being entangled with *some* system R to which we do not have access, such that the global state is a pure state $|\psi\rangle\langle\psi|_{RA}$. Since we do not have access to R, our description of the state of system A has to be as the partial trace of $|\psi\rangle\langle\psi|_{RA}$ over R, i.e., by ρ_A.

Observe that if the state ρ_A is pure, i.e., if $\rho_A = |\phi\rangle\langle\phi|_A$, then the only possible purification of it is of the form

$$|\psi\rangle\langle\psi|_{RA} = |\varphi\rangle\langle\varphi|_R \otimes |\phi\rangle\langle\phi|_A, \tag{3.2.69}$$

with $|\varphi\rangle\langle\varphi|_R$ a pure state of the system R. In other words, purifications of pure states can only be *pure product states*. Somewhat technically, according to Theorem 3.10, the dimension of system R need only satisfy $d_R \geq \text{rank}(\rho_A)$. In the case of a pure state, the rank is equal to one, so that the reference system can be a trivial system of dimension one. Thus, in this technical sense, pure states already purify themselves.

Purifications of states are not unique. In fact, given a state ρ_A and a purification $|\psi\rangle\langle\psi|_{RA}$ of ρ_A as in (3.2.62), let $V_{R\to R'}$ be an isometry (i.e., a linear operator satisfying $V^\dagger V = \mathbb{1}_R$) acting on the R system. Defining

$$|\psi'\rangle_{R'A} = (V_{R\to R'} \otimes \mathbb{1}_A)|\psi\rangle_{RA}, \tag{3.2.70}$$

we find that

$$\text{Tr}_{R'}[|\psi'\rangle\langle\psi'|_{R'A}] = \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} \text{Tr}[V|\varphi_k\rangle\langle\varphi_{k'}|_R V^\dagger]|\phi_k\rangle\langle\phi_k|_A \tag{3.2.71}$$

$$= \sum_{k=1}^{r} \lambda_k |\phi_k\rangle\langle\phi_k|_A \tag{3.2.72}$$

$$= \rho_A, \tag{3.2.73}$$

where we conclude that $\text{Tr}[V|\varphi_k\rangle\langle\varphi_{k'}|V^\dagger] = \delta_{k,k'}$ from cyclicity of the trace and $V^\dagger V = \mathbb{1}_R$. So $|\psi'\rangle\langle\psi'|_{R'A}$ is also a purification of ρ_A.

A converse statement holds as well by employing the Schmidt decomposition (Theorem 2.2): if $|\psi\rangle\langle\psi|_{RA}$ and $|\psi'\rangle\langle\psi'|_{R'A}$ are two purifications of the state ρ_A, then they are related by an isometry taking one reference system to the other. By combining this statement and the previous one, we can thus say that purifications are unique "up to isometries acting on the reference system."

A purification is an example of an "extension" of a quantum state.

Definition 3.11 Extension

An extension of a quantum state ρ_A is a state ω_{RA} satisfying $\text{Tr}_R[\omega_{RA}] = \rho_A$, where R is a reference system.

Remark: In order for a purification to exist, it is necessary that $d_R \geq \text{rank}(\rho_A)$. However, there is no such requirement for an extension.

Note that if the state ρ_A is pure, i.e., if $\rho_A = |\phi\rangle\langle\phi|_A$, then every extension ω_{RA} of ρ_A must be a product state, i.e., we must have $\omega_{RA} = \sigma_R \otimes |\phi\rangle\langle\phi|_A$ for some state σ_R.

3.2.6 Multipartite States and Permutations

A multipartite quantum state is a quantum state of more than two quantum systems. Let $A_1, \ldots, A_n$ denote $n \geq 2$ quantum systems. Then, every quantum state $\rho_{A_1 \cdots A_n}$ can be represented as

$$\rho_{A_1\cdots A_n} = \sum_{i_1,i_1'=0}^{d_{A_1}-1} \cdots \sum_{i_n,i_n'=0}^{d_{A_n}-1} \beta_{i_1,\ldots,i_n;i_1',\ldots,i_n'} |i_1,\ldots,i_n\rangle\langle i_1',\ldots,i_n'|_{A_1\cdots A_n}, \tag{3.2.74}$$

where $\beta_{i_1,\ldots,i_n;i_1',\ldots,i_n'} = \langle i_1,\ldots,i_n|\rho_{A_1\cdots A_n}|i_1',\ldots,i_n'\rangle$. This representation is simply the generalization of the representation in (3.2.12) to $n \geq 2$ quantum systems.

Similarly, the generalization of the representation in (3.2.13) to $n \geq 2$ quantum systems is

$$\rho_{A_1 \cdots A_n} = \frac{1}{d_{A_1} \cdots d_{A_n}} \sum_{k_1=0}^{d_{A_1}^2 - 1} \cdots \sum_{k_n=0}^{d_{A_n}^2 - 1} r_{k_1,\ldots,k_n} S_{A_1}^{k_1} \otimes \cdots \otimes S_{A_n}^{k_n}. \tag{3.2.75}$$

If the quantum systems $A_1, A_2, \ldots, A_n$ are identical, meaning that the Hilbert spaces $\mathcal{H}_{A_1}, \ldots, \mathcal{H}_{A_n}$ are isomorphic to each other, so that $d_{A_1} = d_{A_2} = \cdots = d_{A_n}$, then we can identify them with a single quantum system A of dimension d_A. In this case, for brevity, we often write $\rho_{A^n} \equiv \rho_{A_1 \cdots A_n}$ to denote a quantum state for the n identical systems. For the remainder of this section, we assume that the systems $A_1, A_2, \ldots, A_n$ are identical.

Unlike for bipartite systems, the entanglement of multipartite systems is less straightforward to define, because there are different notions of separability that one can define. A discussion of these different notions of separability for multipartite quantum systems is beyond of the scope of this book. Please see the Bibliographic Notes (Section 3.4) for references.

An important consideration for a collection of identical quantum systems is permutations. For example, many quantum systems, such as bosons and fermions, have quantum states that are symmetric and anti-symmetric, respectively, under permutation of the individual systems. This is due to the fact that bosons and fermions are identical particles, and as such they are fundamentally indistinguishable. (See the Bibliographic Notes in Section 3.4 for more information about the quantum theory of bosons and fermions.) For our purposes, in quantum information, permutations are a useful tool for establishing certain information quantities as upper bounds on the rates of some quantum communication tasks.

Recall that we discussed the notion of permutations in Section 2.5. Specifically, in (2.5.1), we defined a unitary operator $W_{A^n}^{\pi}$ acting on $\mathcal{H}_A^{\otimes n}$, for every permutation $\pi \in \mathcal{S}_n$, as follows:

$$W_{A^n}^{\pi} |i_1, i_2, \ldots, i_n\rangle_{A^n} = |i_{\pi^{-1}(1)}, i_{\pi^{-1}(2)}, \ldots, i_{\pi^{-1}(n)}\rangle_{A^n}, \tag{3.2.76}$$

for all $0 \leq i_1, i_2, \ldots, i_n \leq d - 1$. Physically, the operators $W_{A^n}^{\pi}$ correspond to permuting the states of the (identical) systems $A_1, A_2, \ldots, A_n$. As an example, consider n quantum states $\rho^1, \rho^2, \ldots, \rho^n \in \mathrm{D}(\mathcal{H}_A)$. Then, for every permutation $\pi \in \mathcal{S}_n$,

$$W_{A^n}^{\pi} \left(\rho_{A_1}^1 \otimes \rho_{A_2}^2 \cdots \otimes \rho_{A_n}^n \right) W_{A^n}^{\pi\dagger} = \rho_{A_1}^{\pi^{-1}(1)} \otimes \rho_{A_2}^{\pi^{-1}(2)} \cdots \otimes \rho_{A_n}^{\pi^{-1}(n)}, \tag{3.2.77}$$

which follows straightforwardly from the definition in (3.2.76).

Exercise 3.13

1. Verify (3.2.77).

2. Let $n \geq 2$, and consider the cyclic permutation $\pi = (1\,2\,\ldots\,n)$, which satisfies $\pi(i) = i + 1$ for all $i \in \{1, 2, \ldots, n-1\}$ and $\pi(n) = 1$. Prove that for all quantum states $\rho^1_{A_1}, \rho^2_{A_2}, \ldots, \rho^n_{A_n}$, with $A_1, A_2, \ldots, A_n$ being identical quantum systems,

$$\mathrm{Tr}[W^{\pi}_{A^n}(\rho^1_{A_1} \otimes \rho^2_{A_2} \otimes \cdots \otimes \rho^n_{A_n})] = \mathrm{Tr}[\rho^1_{A_1}\rho^2_{A_2} \cdots \rho^n_{A_n}]. \tag{3.2.78}$$

If, in (3.2.77), we have that $\rho^1 = \rho^2 = \cdots = \rho^n = \rho$, then the state $\rho^{\otimes n}$ is invariant under every permutation, i.e.,

$$W^{\pi}_{A^n}\rho^{\otimes n}_A W^{\pi\dagger}_{A^n} = \rho^{\otimes n}_A \tag{3.2.79}$$

for all $\pi \in \mathcal{S}_n$.

Definition 3.12 Permutation-Invariant State

A state $\rho \in \mathrm{D}(\mathcal{H}^{\otimes n})$ is called *permutation invariant* if

$$\rho = W^{\pi}\rho W^{\pi\dagger} \tag{3.2.80}$$

for every permutation $\pi \in \mathcal{S}_n$, where the unitary permutation operator W^{π} is defined in (3.2.76).

Remark: Note that the permutation-invariance condition in (3.2.80) does *not* imply that the state ρ is supported on the symmetric subspace $\mathrm{Sym}_n(\mathcal{H})$ of $\mathcal{H}^{\otimes n}$. In other words, the condition in (3.2.80) does not imply that

$$\Pi_{\mathrm{Sym}_n(\mathcal{H})}\rho\Pi_{\mathrm{Sym}_n(\mathcal{H})} = \rho. \tag{3.2.81}$$

As a simple example, suppose that $\mathcal{H} = \mathbb{C}^2$, and let $\rho = |\Psi^-\rangle\langle\Psi^-|$, where $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0,1\rangle - |1,0\rangle)$ is the two-qubit Bell state defined in (3.2.46). Then, it is easy to see that $W^{\pi}\rho W^{\pi\dagger} = \rho$ for all $\pi \in \mathcal{S}_2$, while $\Pi_{\mathrm{Sym}_2(\mathcal{H})}\rho\Pi_{\mathrm{Sym}_2(\mathcal{H})} = 0$. The latter is true because $|\Psi^-\rangle$ is an *anti-symmetric state*, i.e., $|\Psi^-\rangle \in \mathrm{ASym}_2(\mathcal{H})$. The state ρ is thus supported on the anti-symmetric subspace, even though it is permutation invariant.

Exercise 3.14

Let $\rho_{AB} = \sum_{x\in\mathcal{X}}^{n} p(x)\sigma_A^x \otimes \tau_B^x$, where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\sigma_A^x\}_{x\in\mathcal{X}}$, $\{\tau_B^x\}_{x\in\mathcal{X}}$ are sets of quantum states.

1. Prove that $\widetilde{\rho}_{ABB'} := \sum_{x\in\mathcal{X}} p(x)\sigma_A^x \otimes \tau_B^x \otimes \tau_{B'}^x$ is an extension of ρ_{AB}, with B' being the reference system, in accordance with Definition 3.11, such that $d_{B'} = d_B$. Prove also that $\widetilde{\rho}_{AB'} := \mathrm{Tr}_B[\widetilde{\rho}_{ABB'}] = \rho_{AB}$.

2. Now, let $\widetilde{\rho}_{AB_1B_2\cdots B_k} := \sum_{x\in\mathcal{X}} p(x)\sigma_A^x \otimes \tau_{B_1}^x \otimes \tau_{B_2}^x \otimes \cdots \otimes \tau_{B_k}^x$, where $k \in \mathbb{N}$. Prove that $\widetilde{\rho}_{AB_\ell} := \mathrm{Tr}_{B_j:j\neq\ell}[\widetilde{\rho}_{AB_1B_2\cdots B_k}] = \rho_{AB}$ for all $\ell \in \{1, 2, \ldots, k\}$, and that $W^\pi_{B_1\cdots B_k}\widetilde{\rho}_{AB_1\cdots B_k}W^{\pi\dagger}_{B_1\cdots B_k} = \widetilde{\rho}_{AB_1\cdots B_k}$ for all $\pi \in \mathcal{S}_k$. The notation $\mathrm{Tr}_{B_j:j\neq\ell}$ indicates to take the partial trace over all of the B systems except for B_ℓ.

In the case $n = 2$, meaning that there are only two quantum systems under consideration, there is only one non-trivial permutation, $\pi = (1\ 2)$, which swaps the two elements of the set $\{1, 2\}$. Recall from Exercise 2.34 that

$$W^{(1\ 2)} = F := \sum_{k,k'=0}^{d-1} |k, k'\rangle\langle k', k|. \tag{3.2.82}$$

We call F the *swap operator*, because $F(\rho \otimes \sigma)F^\dagger = \sigma \otimes \rho$ for all quantum states ρ and σ, which is a simple consequence of (3.2.77). In other words, the two states ρ and σ become "swapped" with respect to the quantum systems after the action of the operator F. The swap operator is Hermitian and satisfies $F^2 = \mathbb{1}$, meaning that it is also unitary and self-inverse. Also, as a consequence of (3.2.78), we have

$$\mathrm{Tr}[F(\rho \otimes \sigma)] = \mathrm{Tr}[\rho\sigma] \tag{3.2.83}$$

for all quantum states ρ and σ.

Exercise 3.15

1. Verify that $F|\Phi_d\rangle = |\Phi_d\rangle$ for all $d \geq 2$.

2. For the two-qubit Bell state vectors defined in (3.2.42), prove that $F|\Phi_{z,x}\rangle = (-1)^{zx}|\Phi_{z,x}\rangle$ for all $z, x \in \{0, 1\}$.

3. More generally, for the two-qudit Bell state vectors defined in (3.2.57),

prove that

$$F|\Phi_{z,x}\rangle = e^{\frac{2\pi i z x}{d}}|\Phi_{z,-x}\rangle \tag{3.2.84}$$

for all $z, x \in \{0, 1, \dots, d-1\}$.

Exercise 3.16

Let $\rho_{A^n} \in \mathrm{D}(\mathcal{H}^{\otimes n})$ be an arbitrary quantum state, and consider the state

$$\sigma_{A^n} := \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} W^{\pi}_{A^n}\rho_{A^n}W^{\pi\dagger}_{A^n}. \tag{3.2.85}$$

1. Prove that σ_{A^n} is a permutation-invariant state.
2. Let $|\phi\rangle_{RA^n}$ be a purification of ρ_{A^n}. Verify that

$$|\psi\rangle_{XRA^n} := \frac{1}{\sqrt{n!}}\sum_{\pi\in\mathcal{S}_n}|\pi\rangle_X \otimes (\mathbb{1}_R \otimes W^{\pi}_{A^n})|\phi\rangle_{RA} \tag{3.2.86}$$

is a purification of σ_{A^n}, where $\{|\pi\rangle\}_{\pi\in\mathcal{S}_n}$ is an orthonormal basis indexed by the elements of $\mathcal{S}_n$, such that $\langle\pi|\pi'\rangle = \delta_{\pi,\pi'}$ for all $\pi, \pi' \in \mathcal{S}_n$.

It turns out that for permutation-invariant states, we can construct a purification that is itself permutation invariant.

Lemma 3.13 Purification of Permutation-Invariant States

Let $\rho_{A^n} \in \mathrm{D}(\mathcal{H}_A^{\otimes n})$ be a permutation-invariant state, i.e.,

$$\rho_{A^n} = W^{\pi}_{A^n}\rho_{A^n}W^{\pi\dagger}_{A^n} \tag{3.2.87}$$

for all $\pi \in \mathcal{S}_n$, where the unitary operators in the set $\{W^{\pi}_{A^n}\}_{\pi\in\mathcal{S}_n}$ are defined in (2.5.1). Then, there exists a permutation-invariant purification $|\psi^\rho\rangle\langle\psi^\rho|_{\hat{A}^nA^n}$ of ρ_{A^n}, such that $|\psi^\rho\rangle_{\hat{A}^nA^n} \in \mathrm{Sym}_n(\mathcal{H}_{\hat{A}A})$. This means that

$$|\psi^\rho\rangle_{\hat{A}^nA^n} = W^{\pi}_{\hat{A}^n} \otimes W^{\pi}_{A^n}|\psi^\rho\rangle_{\hat{A}^nA^n} \tag{3.2.88}$$

for all $\pi \in \mathcal{S}_n$, where the dimension of $\hat{A}$ is equal to the dimension of A.

Proof: Consider the canonical purification of ρ_{A^n} as defined in (3.2.66); i.e., let

$$|\psi^\rho\rangle_{\hat{A}^n A^n} := \left(\mathbb{1}_{\hat{A}^n} \otimes \sqrt{\rho_{A^n}}\right)|\Gamma\rangle_{\hat{A}^n A^n}. \tag{3.2.89}$$

Then, because the operators $W^\pi_{A^n}$ are real in the standard basis, meaning that $\overline{W^\pi_{\hat{A}^n}} = W^\pi_{\hat{A}^n}$ for all $\pi \in \mathcal{S}_n$, and using the transpose trick in (2.2.44), we obtain

$$W^\pi_{\hat{A}^n} \otimes W^\pi_{A^n}|\psi^\rho\rangle_{\hat{A}^n A^n} = \left(\overline{W^\pi_{\hat{A}^n}} \otimes W^\pi_{A^n}\right)\left(\mathbb{1}_{\hat{A}^n} \otimes \sqrt{\rho_{A^n}}\right)|\Gamma\rangle_{\hat{A}^n A^n} \tag{3.2.90}$$

$$= \left(\mathbb{1}_{\hat{A}^n} \otimes W^\pi_{A^n}\sqrt{\rho_{A^n}}W^{\pi\dagger}_{A^n}\right)|\Gamma\rangle_{\hat{A}^n A^n} \tag{3.2.91}$$

$$= \left(\mathbb{1}_{\hat{A}^n} \otimes \sqrt{W^\pi_{A^n}\rho_{A^n}W^{\pi\dagger}_{A^n}}\right)|\Gamma\rangle_{\hat{A}^n A^n} \tag{3.2.92}$$

$$= \left(\mathbb{1}_{\hat{A}^n} \otimes \sqrt{\rho_{A^n}}\right)|\Gamma\rangle_{\hat{A}^n A^n} \tag{3.2.93}$$

$$= |\psi^\rho\rangle_{\hat{A}^n A^n} \tag{3.2.94}$$

for all $\pi \in \mathcal{S}_n$, where the third equality follows from (2.2.72) and the fourth equality follows from the permutation invariance of ρ_{A^n}. ■

3.2.7 Group-Invariant States

So far, we have seen two special types of unitary operators: the Heisenberg–Weyl operators $\{W_{z,x}\}_{z,x=0}^{d-1}$, introduced in Definition 3.7, and the permutation operators $\{W^\pi\}_{\pi\in\mathcal{S}_n}$ defined in (3.2.76). Both sets of operators are examples of projective unitary group representations. Specifically, the Heisenberg–Weyl operators form a projective unitary representation of the group $\mathbb{Z}_d \times \mathbb{Z}_d$, and the permutation operators form a unitary representation of the symmetric group $\mathcal{S}_n$.

Let us now formally define the concepts of a group and group representations. A *group* G is a tuple $(\mathcal{G}, *)$ consisting of a set $\mathcal{G}$ of objects and an associative operation $*$ used to combine them. We write $g \in G$ to mean that the object g belongs to the set $\mathcal{G}$. Then, the operation $*$ is such that $g * g' \in G$ for all $g, g' \in G$. Furthermore, there is an identity element id such that $g * \mathrm{id} = g = \mathrm{id} * g$ for all $g \in G$, and corresponding to every element $g \in G$ is an inverse $g^{-1} \in G$ such that $g * g^{-1} = g^{-1} * g = \mathrm{id}$. We mostly consider finite groups in this book, and we use $|G|$ to denote the number of elements in the associated set $\mathcal{G}$. Please see the Bibliographic Notes (Section 3.4) for more information about groups.

A *unitary representation* of a group G is a set $\{U^g\}_{g\in G}$ of unitary operators, with one unitary operator associated to each element of G. The unitary operators respect the group operation $*$, in the sense that $U^g U^{g'} = U^{g*g'}$ for all $g \in G$. In particular, this implies that $U^{\text{id}} = \mathbb{1}$ and $U^{g^{-1}} = (U^g)^\dagger$ for all $g \in G$. A unitary representation of a group G is called *projective* if the unitaries respect the group operation up to a phase factor, i.e., if $U^g U^{g'} = \omega(g, g') U^{g*g'}$ for all $g, g' \in G$, where $\omega(g, g') \in \mathbb{C}$ satisfies $|\omega(g, g')| = 1$ for all $g, g' \in G$. Please see the Bibliographic Notes (Section 3.4) for more information about group representations.

The *action* of the group representation on a quantum system is defined by the channels $\rho \mapsto U^g \rho U^{g\dagger}$ for all $g \in G$, where ρ is a state of the quantum system and $\{U^g\}_{g\in G}$ is a (projective) unitary representation of the group G. Physically, groups are used to model certain types of operations on a system (such as permutations, translations, rotations, etc.). Mathematically, systems that have symmetries are such that the states of the system are invariant under the action of the corresponding group.

Definition 3.14 Group-Invariant State

Let G be a (finite) group and $\{U^g\}_{g\in G}$ a d-dimensional unitary representation of G, with $d \geq 2$. A quantum state $\rho \in \mathrm{D}(\mathbb{C}^d)$ is called *group invariant*, or *G-invariant*, if $\rho = U^g \rho U^{g\dagger}$ for all $g \in G$.

Exercise 3.17

Let ρ_A be a quantum state for a d-dimensional quantum system A, let G be a group, and let $\{U_A^g\}_{g\in G}$ be a d-dimensional unitary representation of G.

1. Prove that the state

$$\mathcal{T}^G(\rho_A) := \frac{1}{|G|} \sum_{g\in G} U_A^g \rho_A U_A^{g\dagger} \tag{3.2.95}$$

is group invariant. The quantum channel $\mathcal{T}^G$ is known as the *twirl channel* with respect to the unitary representation $\{U^g\}_{g\in G}$ of the group G.

2. Let $|\phi\rangle_{RA}$ be a purification of ρ_A. Verify that

$$\frac{1}{\sqrt{|G|}} \sum_{g\in G} |g\rangle_X \otimes (\mathbb{1}_R \otimes U_A^g)|\phi\rangle_{RA} \tag{3.2.96}$$

is a purification of $\mathcal{T}^G(\rho_A)$.

The twirl map in (3.2.95) corresponding to the Heisenberg–Weyl unitaries has the following special form.

Lemma 3.15 Heisenberg–Weyl Twirl

For every linear operator $M \in \mathrm{L}(\mathbb{C}^d)$,

$$\frac{1}{d^2}\sum_{z,x=0}^{d-1} W_{z,x} M W_{z,x}^\dagger = \mathrm{Tr}[M]\frac{\mathbb{1}_d}{d}. \tag{3.2.97}$$

PROOF: The Heisenberg–Weyl operators form an irreducible projective unitary representation of the group $\mathbb{Z}_d \times \mathbb{Z}_d$. This fact can be used to prove (3.2.97) via Schur's lemma (see Bibliographic Notes in Section 3.4). Alternatively, the result is immediate using orthonormality of the set $\{\frac{1}{\sqrt{d}} W_{z,x}\}_{z,x=0}^{d-1}$ (see (3.2.53)) along with Problem 6 in Section 2.7. For an alternative approach, see Exercise 3.18. ■

Exercise 3.18

Provide a direct proof of Lemma 3.15. Do this by first showing that $\frac{1}{d^2}\sum_{z,x=0}^{d-1} W_{z,x} M W_{z,x}^\dagger = (\mathcal{D}_X \circ \mathcal{D}_Z)(M)$, where

$$\mathcal{D}_X(M) := \frac{1}{d}\sum_{x=0}^{d-1} X(x) M X(x)^\dagger, \tag{3.2.98}$$

$$\mathcal{D}_Z(M) := \frac{1}{d}\sum_{z=0}^{d-1} Z(z) M Z(z)^\dagger. \tag{3.2.99}$$

Recall that the Heisenberg–Weyl operators reduce to the Pauli operators for $d = 2$. In this case, we obtain

$$\frac{1}{4}M + \frac{1}{4}XMX + \frac{1}{4}YMY + \frac{1}{4}ZMZ = \mathrm{Tr}[M]\frac{\mathbb{1}_2}{2}, \tag{3.2.100}$$

for every $M \in \mathrm{L}(\mathbb{C}^2)$.

Exercise 3.19

Let G be a group, and let $\{U^g\}_{g\in G}$ be a d-dimensional unitary representation of G, with $d \geq 2$. If $\rho \in \mathrm{D}(\mathbb{C}^d)$ is a group-invariant state, then prove that there exists a purification $|\psi^\rho\rangle$ of ρ such that $\overline{U^g} \otimes U^g|\psi^\rho\rangle = |\psi^\rho\rangle$ for all $g \in G$.

3.2.8 Ensembles and Classical–Quantum States

For a finite alphabet $\mathcal{X}$, an *ensemble* is a collection $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ consisting of a probability distribution $p : \mathcal{X} \to [0, 1]$ such that each probability $p(x)$ is paired with a quantum state ρ^x. Ensembles are used to describe quantum systems that are known to be in one of a given set of states with some probability.

Suppose that Alice is in possession of a quantum system and that she prepares the system in the state ρ^x with probability $p(x)$. The state of the system can thus be described by the ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$. If she sends the system to Bob without telling him in which of the states the system has been prepared, but Bob knows the ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ describing the system, then from Bob's perspective the state of the system is given by the expected state ρ of the ensemble, which is specified as

$$\rho = \sum_{x\in\mathcal{X}} p(x)\rho^x. \tag{3.2.101}$$

On the other hand, if Alice sends Bob classical information about which state she has prepared, then from Bob's perspective the state of the system can be described by the following *classical–quantum state*:

$$\rho_{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_B^x, \tag{3.2.102}$$

where X is the $|\mathcal{X}|$-dimensional quantum system corresponding to the register holding the information about which state was sent and $\{|x\rangle\}_{x\in\mathcal{X}}$ is an orthonormal basis for $\mathcal{H}_X$. Furthermore, the reduced state of Bob's system B, after discarding X, is

$$\mathrm{Tr}_X[\rho_{XB}] = \sum_{x\in\mathcal{X}} p(x)\rho^x = \rho, \tag{3.2.103}$$

consistent with (3.2.101).

Classical–quantum states have a block-diagonal structure, in the sense that they can equivalently be written as the following block-diagonal matrix:

$$\rho_{XB} = \begin{pmatrix} p(x_1)\rho^{x_1} & & & \\ & p(x_2)\rho^{x_2} & & \\ & & \ddots & \\ & & & p(x_{|\mathcal{X}|})\rho^{x_{|\mathcal{X}|}} \end{pmatrix}. \tag{3.2.104}$$

We can represent (3.2.104) more compactly as

$$\rho_{XB} = \bigoplus_{x\in\mathcal{X}} p(x)\rho^x. \tag{3.2.105}$$

Exercise 3.20

Construct a purification of the classical–quantum state ρ_{XB} in (3.2.102). (*Hint*: Consider a purification analogous to the one in (3.2.86).)

3.2.9 Partial Transpose and PPT States

In Section 3.2.2, we defined the partial trace superoperator as a generalization of the usual trace to the case that it acts only on one part of a composite quantum system. In a similar manner, in this section, we define the partial transpose superoperator. The partial transpose plays an important role in quantum information theory due to its connection with entanglement. In fact, as we show in this section, it leads to a sufficient condition for a bipartite state to be entangled.

Recall from (2.2.27) that the action of the transpose map T on a linear operator $X_{A\to A'}$ can be written as

$$\mathrm{T}(X) = \sum_{i=0}^{d_A-1}\sum_{j=0}^{d_{A'}-1} |i\rangle_A\langle j|_{A'} X |i\rangle_A\langle j|_{A'}, \tag{3.2.106}$$

where we have defined the transpose with respect to the orthonormal bases $\{|i\rangle_A : 0 \le i \le d_A - 1\}$ and $\{|j\rangle_{A'} : 0 \le j \le d_{A'} - 1\}$. This is consistent with the familiar definition of the transpose of a matrix X as being the matrix X^{T} with its rows and columns flipped relative to X. Indeed, if X has the matrix representation

$X = \sum_{j=0}^{d_{A'}-1} \sum_{i=0}^{d_A-1} X_{j,i}|j\rangle\langle_{|}A'i_A$, then it follows that the transpose of X is

$$X^{\mathsf{T}} = \sum_{j=0}^{d_{A'}-1} \sum_{i=0}^{d_A-1} X_{j,i}|i\rangle_A\langle j|_{A'} = \mathrm{T}(X). \tag{3.2.107}$$

Note that, unlike the trace or the conjugate transpose, the transpose depends on the choice of orthonormal bases used to evaluate it. Throughout the rest of this book, we use both $\mathrm{T}(X)$ and X^{T} to refer to the transpose of X with respect to the standard orthonormal basis.

Exercise 3.21

For all $d \geq 2$, prove that the transpose map can be realized as follows, in terms of the Heisenberg–Weyl operators from Definition 3.7:

$$\mathrm{T}(X) = \frac{1}{d} \sum_{z,x=0}^{d-1} \mathrm{e}^{\frac{2\pi \mathrm{i} z x}{d}} W_{z,x}^{\dagger} X W_{z,-x}, \tag{3.2.108}$$

where the equality holds for every linear operator $X \in \mathrm{L}(\mathbb{C}^d)$.

The transpose map is known as the *partial transpose* when it acts on one subsystem of a bipartite linear operator X_{AB}.

Definition 3.16 Partial Transpose

Given quantum systems A and B, the *partial transpose on B* is denoted by $\mathrm{T}_B \equiv \mathrm{id}_A \otimes \mathrm{T}_B$, and it is defined as

$$\mathrm{T}_B(X_{AB}) := \sum_{j,j'=0}^{d_B-1} (\mathbb{1}_A \otimes |j\rangle\langle j'|_B) X_{AB} (\mathbb{1}_A \otimes |j\rangle\langle j'|_B) \tag{3.2.109}$$

for every linear operator $X_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Similarly, the *partial transpose on A* is denoted by $\mathrm{T}_A \equiv \mathrm{T}_A \otimes \mathrm{id}_B$, and it is defined as

$$\mathrm{T}_A(X_{AB}) := \sum_{i,i'=0}^{d_A-1} (|i\rangle\langle i'|_A \otimes \mathbb{1}_B) X_{AB} (|i\rangle\langle i'|_A \otimes \mathbb{1}_B) \tag{3.2.110}$$

for all $X_{AB} \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

Given an expansion of X_{AB} as

$$X_{AB} = \sum_{i,j=0}^{d_B-1} X_A^{i,j} \otimes |i\rangle\langle j|_B, \tag{3.2.111}$$

where each $X_A^{i,j} := \langle i|_B X_{AB} |j\rangle_B$ is a linear operator acting on system A, the partial transpose map T_B has the action

$$\mathrm{T}_B(X_{AB}) = \sum_{i,j=0}^{d_B-1} X_A^{i,j} \otimes |j\rangle\langle i|_B = \sum_{i,j=0}^{d_B-1} X_A^{j,i} \otimes |i\rangle\langle j|_B. \tag{3.2.112}$$

The partial transpose map is self-inverse, i.e.,

$$(\mathrm{id}_A \otimes \mathrm{T}_B) \circ (\mathrm{id}_A \otimes \mathrm{T}_B) = \mathrm{id}_{AB}, \tag{3.2.113}$$

and it is self-adjoint with respect to the Hilbert–Schmidt inner product, in the sense that

$$\langle X_{AB}, \mathrm{T}_B(Y_{AB})\rangle = \langle \mathrm{T}_B(X_{AB}), Y_{AB}\rangle, \tag{3.2.114}$$

for all operators X_{AB} and Y_{AB}.

We also have the following generalization of the transpose trick from (2.2.42):

$$(X_{RA} \otimes \mathbb{1}_B)(\mathbb{1}_R \otimes |\Gamma\rangle_{AB}) = (\mathbb{1}_A \otimes \mathrm{T}_B(X_{RB}))(\mathbb{1}_R \otimes |\Gamma\rangle_{AB}), \tag{3.2.115}$$

where the Hilbert spaces corresponding to the systems A and B are isomorphic and $X_{RA} \in \mathrm{L}(\mathcal{H}_R \otimes \mathcal{H}_A)$.

Exercise 3.22

Verify (3.2.113), (3.2.114), and (3.2.115).

Definition 3.17 PPT State

A bipartite state ρ_{AB} is called a *positive partial transpose (PPT) state* if the partial transpose $\mathrm{T}_B(\rho_{AB})$ is positive semi-definite.

The set of PPT states is denoted by $\mathrm{PPT}(A\!:\!B)$, so that

$$\mathrm{PPT}(A\!:\!B) := \{\sigma_{AB} : \sigma_{AB} \geq 0,\ \mathrm{T}_B(\sigma_{AB}) \geq 0, \mathrm{Tr}[\sigma_{AB}] = 1\}. \tag{3.2.116}$$

Lemma 3.18

Given quantum systems A and B, the set $\mathrm{PPT}(A:B)$ does not depend on which system the transpose is taken, nor does it depend on which orthonormal basis is used to define the transpose map.

Proof: To see the first statement, suppose that $\rho_{AB} \in \mathrm{PPT}(A:B)$. This means that $\mathrm{T}_B(\rho_{AB}) \geq 0$. But since the eigenvalues are invariant under a full transpose $\mathrm{T}_A \otimes \mathrm{T}_B$, this means that $(\mathrm{T}_A \otimes \mathrm{T}_B)(\mathrm{T}_B(\rho_{AB})) \geq 0$, the latter being the same as $\mathrm{T}_A(\rho_{AB}) \geq 0$ due to the self-inverse property of the partial transpose. So $\mathrm{T}_B(\rho_{AB}) \geq 0$ implies $\mathrm{T}_A(\rho_{AB}) \geq 0$, and vice versa.

To see the second statement, let $\mathrm{T}_B(\rho_{AB}) \geq 0$, and let $\{|\phi_\ell\rangle_B\}_{\ell=0}^{d_B-1}$ be some other orthonormal basis for B. The partial transpose with respect to this basis is given by

$$\sum_{\ell,\ell'=0}^{d_B-1} (\mathbb{1}_A \otimes |\phi_\ell\rangle\langle\phi_{\ell'}|_B)\rho_{AB}(\mathbb{1}_A \otimes |\phi_\ell\rangle\langle\phi_{\ell'}|_B). \tag{3.2.117}$$

Now, consider that $\sum_{\ell=0}^{d_B-1} |\phi_\ell\rangle\langle\phi_\ell| = \mathbb{1}_B$, so that

$$\begin{aligned}
&\mathrm{T}_B(\rho_{AB}) \\
&= \sum_{j,j'=0}^{d_B-1} (\mathbb{1}_A \otimes |j\rangle\langle j'|_B)\rho_{AB}(\mathbb{1}_A \otimes |j\rangle\langle j'|_B) \\
&= \sum_{j,j',\ell,\ell'=0}^{d_B-1} (\mathbb{1}_A \otimes |j\rangle\langle j'|\phi_\ell\rangle\langle\phi_\ell|_B)\rho_{AB}(\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\langle\phi_{\ell'}|j\rangle\langle j'|_B) \\
&= \sum_{j,j',\ell,\ell'=0}^{d_B-1} (\mathbb{1}_A \otimes \langle\phi_{\ell'}|j\rangle|j\rangle\langle\phi_\ell|_B)\rho_{AB}(\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\langle j'|_B\langle j'|\phi_\ell\rangle) \\
&= \sum_{\ell,\ell'=0}^{d_B-1} \left(\mathbb{1}_A \otimes \left(\sum_{j=0}^{d_B-1} \langle\phi_{\ell'}|j\rangle|j\rangle\right)\langle\phi_\ell|_B\right)\rho_{AB}\left(\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\left(\sum_{j'=0}^{d_B-1} \langle j'|_B\langle j'|\phi_\ell\rangle\right)\right) \\
&= \sum_{\ell,\ell'=0}^{d_B-1} \left(\mathbb{1}_A \otimes |\overline{\phi_{\ell'}}\rangle\langle\phi_\ell|_B\right)\rho_{AB}\left(\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\langle\overline{\phi_\ell}|\right),
\end{aligned} \tag{3.2.118}$$

where in the last line we defined $|\overline{\phi_\ell}\rangle \coloneqq \sum_{j=0}^{d_B-1}\langle\phi_\ell|j\rangle|j\rangle$ for $0 \leq \ell \leq d_B - 1$. Note that the set $\{|\overline{\phi_\ell}\rangle\}_{\ell=0}^{d_B-1}$ is orthonormal, so that $U_B \coloneqq \sum_{\ell=0}^{d_B-1} |\phi_\ell\rangle\langle\overline{\phi_\ell}|_B$ is a unitary operator. Then we find that

$$\sum_{\ell,\ell'=0}^{d_B-1} (\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\langle\phi_\ell|_B)\, \rho_{AB}\, (\mathbb{1}_A \otimes |\phi_{\ell'}\rangle\langle\phi_\ell|_B) = U_B \mathrm{T}_B(\rho_{AB}) U_B^\dagger \geq 0, \quad (3.2.119)$$

where the last inequality follows from the condition $\mathrm{T}_B(\rho_{AB}) \geq 0$ and property 1 of Lemma 2.14. We thus conclude that the PPT property does not depend on which orthonormal basis is used to define the transpose map. ■

We mentioned at the beginning of this section that the partial transpose is useful in quantum information because it leads to a sufficient condition for a bipartite state to be entangled. We now derive the sufficient condition.

Suppose that a state σ_{AB} is a separable, unentangled state of the following form:

$$\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x, \quad (3.2.120)$$

where $p : \mathcal{X} \to [0, 1]$ is a probability distribution on a finite alphabet $\mathcal{X}$ and $\{\omega_A^x\}_{x\in\mathcal{X}}$ and $\{\tau_B^x\}_{x\in\mathcal{X}}$ are sets of quantum states. Then, the action of the partial transpose map T_B on σ_{AB} is as follows:

$$\mathrm{T}_B(\sigma_{AB}) = \sum_{x} p(x)\omega_A^x \otimes \mathrm{T}(\tau_B^x), \quad (3.2.121)$$

which is a separable quantum state, being the expected state of the ensemble $\{(p(x), \omega_A^x \otimes \mathrm{T}(\tau_B^x))\}_{x\in\mathcal{X}}$. Each element of the ensemble is indeed a quantum state because the transpose is a positive map, i.e., $\mathrm{T}(\tau_B^x) \geq 0$ if $\tau_B^x \geq 0$. Due to this fact, we conclude that $\mathrm{T}_B(\sigma_{AB}) \geq 0$, so that σ_{AB} is a PPT state. Thus, we conclude the following:

If a state is separable, then it is PPT.

This is called the *PPT criterion*. Equivalently, by taking the contrapositive of this statement, we obtain the following:

If a state is not PPT, then it is entangled.

So the condition of a state being NPT (non-positive partial transpose) is sufficient for detecting entanglement.

As an example, let us consider applying the partial transpose map T_B to the maximally entangled state Φ_{AB}, as defined in (3.2.38). In the case that the bases of the partial transpose and the maximally entangled state are the same, we find that

$$\mathrm{T}_B(\Phi_{AB}) = \frac{1}{d}\mathrm{T}_B\left(\sum_{i,i'=0}^{d-1} |i\rangle\langle i'|_A \otimes |i\rangle\langle i'|_B\right) \tag{3.2.122}$$

$$= \sum_{i,i'=0}^{d-1} |i\rangle\langle i'|_A \otimes |i'\rangle\langle i|_B \tag{3.2.123}$$

$$= \frac{1}{d}F_{AB}, \tag{3.2.124}$$

where F_{AB} is the swap operator defined in (3.2.82). If the bases are not the same, then we find, by applying the same development from (3.2.117)–(3.2.119), that there exists a unitary U_B such that

$$\mathrm{T}_B(\Phi_{AB}) = \frac{1}{d}U_B F_{AB} U_B^\dagger. \tag{3.2.125}$$

From (2.5.10) and (2.5.11), the swap operator has the following spectral decomposition:

$$F_{AB} = \Pi_{AB}^{\mathrm{Sym}} - \Pi_{AB}^{\mathrm{ASym}}, \tag{3.2.126}$$

where $\Pi_{AB}^{\mathrm{Sym}} \equiv \Pi_{\mathrm{Sym}_2(\mathbb{C}^d)}$ is the projection onto the symmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$ and $\Pi_{AB}^{\mathrm{ASym}} \equiv \Pi_{\mathrm{ASym}_2(\mathbb{C}^d)}$ is the projection onto the anti-symmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$. Indeed, we have that $\Pi_{AB}^{\mathrm{Sym}} + \Pi_{AB}^{\mathrm{ASym}} = \mathbb{1}_{AB}$ and $\Pi_{AB}^{\mathrm{Sym}}\Pi_{AB}^{\mathrm{ASym}} = 0$. Thus, the swap operator has negative eigenvalues, which by the PPT criterion means that Φ_{AB} is an entangled state, as expected.

Although the PPT criterion is generally only a necessary condition for separability of a bipartite state, it is known to be also sufficient for every quantum state ρ_{AB} for which both A and B are qubits or A is a qubit and B is a qutrit; please consult the Bibliographic Notes in Section 3.4. In higher dimensions, however, there are PPT states that are entangled. These PPT entangled states turn out to be useless for the task of entanglement distillation (see Chapter 13), and thus they are called *bound entangled* (although they are entangled, they cannot be used to extract pure maximally entangled states).

Exercise 3.23

Prove that the swap operator F_{AB} possesses the following symmetry:

$$(U_A \otimes V_B)F_{AB} = F_{AB}(V_A \otimes U_B) \tag{3.2.127}$$

for all unitaries U and V.

3.2.10 Isotropic and Werner States

In Section 3.2.6, we defined permutation-invariant states, which are states that are invariant under the action of the unitary operator W^π for every permutation $\pi \in \mathcal{S}_n$. Another important class of quantum states in quantum information theory consists of bipartite states that are invariant under certain kinds of unitaries. There are two distinct such classes of states that we define in this section.

Definition 3.19 Isotropic States

Consider two quantum systems A and B, with $d_A = d_B = d \geq 2$. A quantum state ρ_{AB} is called an *isotropic state* if it is invariant under the action of $U \otimes \overline{U}$ for every unitary U, i.e., if

$$\rho_{AB} = (U \otimes \overline{U})\rho_{AB}(U \otimes \overline{U})^\dagger \tag{3.2.128}$$

for every unitary U. For every isotropic state ρ_{AB}, there exists $p \in [0, 1]$ such that $\rho_{AB} = \rho_{AB}^{\text{iso};p}$, where

$$\rho_{AB}^{\text{iso};p} := p|\Phi\rangle\langle\Phi|_{AB} + \frac{1-p}{d^2-1}\left(\mathbb{1}_{AB} - |\Phi\rangle\langle\Phi|_{AB}\right), \tag{3.2.129}$$

where $|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}|j,j\rangle_{AB}$.

Using (3.2.58), we can write every isotropic state as follows:

$$\rho_{AB}^{\text{iso};p} = p|\Phi\rangle\langle\Phi|_{AB} + \frac{1-p}{d^2-1}\sum_{\substack{0\leq z,x\leq d-1\\(z,x)\neq(0,0)}}|\Phi_{z,x}\rangle\langle\Phi_{z,x}|. \tag{3.2.130}$$

In other words, the isotropic state can be viewed as a probabilistic mixture of the qudit Bell states defined in (3.2.57), such that the state $|\Phi\rangle\langle\Phi|$ is prepared with probability p, and the states $|\Phi_{z,x}\rangle\langle\Phi_{z,x}|$, with $(z,x) \neq (0,0)$, are prepared with probability $\frac{1-p}{d^2-1}$. This implies that every isotropic state is a Bell-diagonal state (see (3.2.59)), that it has full rank for $p \in (0,1)$, and that its eigenvalues are p and $\frac{1-p}{d^2-1}$ (the latter with multiplicity d^2-1).

Exercise 3.24

1. Verify that the isotropic state in (3.2.129) is invariant under $U \otimes \overline{U}$ for every unitary U, i.e., verify that

$$(U_A \otimes \overline{U}_B)\rho_{AB}^{\text{iso};p}(U_A \otimes \overline{U}_B)^\dagger = \rho_{AB}^{\text{iso};p} \tag{3.2.131}$$

   for every unitary U and all $p \in [0,1]$.

2. Show that, for all $p \in [0,1]$, the isotropic state $\rho_{AB}^{\text{iso};p}$ can be represented as $a|\Phi\rangle\langle\Phi|_{AB} + (1-a)\frac{\mathbb{1}_{AB}}{d^2}$, where $a = \frac{pd^2-1}{d^2-1}$. Conclude that $a \in \left[\frac{-1}{d^2-1}, 1\right]$.

Just as every multipartite quantum state can be made permutation invariant via the construction in (3.2.85), every bipartite quantum state ρ_{AB}, with $d_A = d_B$, can be made invariant under $U \otimes \overline{U}$, i.e., isotropic, via the following construction:

$$\int_U (U \otimes \overline{U})\rho_{AB}(U \otimes \overline{U})^\dagger \, \mathrm{d}U. \tag{3.2.132}$$

This construction can be thought of as a uniform average over unitaries, analogous to the uniform average over permutations in (3.2.85). The object "dU" is known as the *Haar measure*. Intuitively, an integral is used due to the fact that the set of unitaries is continuous; however, the integral can be evaluated using a uniform average over a discrete set of unitaries known as a *unitary two-design*. In particular, for every state ρ_{AB},

$$\int_U (U \otimes \overline{U})\rho_{AB}(U \otimes \overline{U})^\dagger \, \mathrm{d}U = \rho_{AB}^{\text{iso};p}, \quad p = \langle\Phi|\rho_{AB}|\Phi\rangle. \tag{3.2.133}$$

Please see the Bibliographic Notes (Section 3.4) for more information about this result, as well as about integrals of functions of unitaries with respect to the Haar measure.

The isotropic states constitute one class of bipartite quantum states in which every state is invariant under the action of a unitary acting on the individual subsystems. We now define a second class of such states.

Definition 3.20 Werner States

Consider two quantum systems A and B, with $d_A = d_B = d \geq 2$. A quantum state ρ_{AB} is called a *Werner state* if it is invariant under the action of $U \otimes U$ for every unitary U, i.e., if

$$\rho_{AB} = (U \otimes U)\rho_{AB}(U \otimes U)^\dagger \tag{3.2.134}$$

for every unitary U. For every Werner state ρ_{AB}, there exists $p \in [0,1]$ such that $\rho_{AB} = \rho_{AB}^{\mathrm{W};p}$, where

$$\rho_{AB}^{\mathrm{W};p} := p\zeta_{AB} + (1-p)\zeta_{AB}^{\perp}. \tag{3.2.135}$$

Here, ζ_{AB} and $\zeta_{AB}^{\perp}$ are quantum states defined as

$$\zeta_{AB} := \frac{1}{d(d-1)}\left(\mathbb{1}_{AB} - F_{AB}\right), \tag{3.2.136}$$

$$\zeta_{AB}^{\perp} := \frac{1}{d(d+1)}\left(\mathbb{1}_{AB} + F_{AB}\right), \tag{3.2.137}$$

and $F_{AB} = \sum_{i,j=0}^{d-1} |i,j\rangle\langle j,i|_{AB}$ is the swap operator.

Observe that the states ζ_{AB} and $\zeta_{AB}^{\perp}$ in Definition 3.20 are proportional to the projections $\Pi_{AB}^{\mathrm{ASym}} \equiv \Pi_{\mathrm{ASym}_2(\mathbb{C}^d)}$ and $\Pi_{AB}^{\mathrm{Sym}} \equiv \Pi_{\mathrm{Sym}_2(\mathbb{C}^d)}$ onto the anti-symmetric and symmetric subspaces, respectively, of $\mathbb{C}^d \otimes \mathbb{C}^d$ (recall (2.5.10) and (2.5.11)). In particular,

$$\zeta_{AB} = \frac{2}{d(d-1)}\Pi_{AB}^{\mathrm{ASym}}, \quad \zeta_{AB}^{\perp} = \frac{2}{d(d+1)}\Pi_{AB}^{\mathrm{Sym}}. \tag{3.2.138}$$

Exercise 3.25

1. Verify that the Werner state in (3.2.135) is invariant under $U \otimes U$ for every

unitary U, i.e., verify that

$$(U_A \otimes U_B)\rho_{AB}^{\mathrm{W};p}(U_A \otimes U_B)^\dagger = \rho_{AB}^{\mathrm{W};p} \tag{3.2.139}$$

for every unitary U and all $p \in [0,1]$.

2. Show that, for all $p \in [0,1]$, the Werner state $\rho_{AB}^{\mathrm{W};p}$ can be represented as $\frac{1}{d^2-da}(\mathbb{1}_{AB} - aF_{AB})$, where $a = \frac{d(2p-1)+1}{2p-1+d}$. Conclude that $a \in [-1,1]$.
3. Prove that for $d=2$, $\zeta_{AB} = \Pi_{AB}^{\mathrm{ASym}} = |\Psi^-\rangle\langle\Psi^-|$, where $|\Psi^-\rangle$ is defined in (3.2.46).

As with the isotropic states, every bipartite quantum state ρ_{AB}, with $d_A = d_B$, can be made into a Werner state via the following construction:

$$\int_U (U \otimes U)\rho_{AB}(U \otimes U)^\dagger \, \mathrm{d}U. \tag{3.2.140}$$

As before, the integral represents the uniform average over all unitaries, which can be evaluated using a unitary two-design. In particular, for every state ρ_{AB},

$$\int_U (U \otimes U)\rho_{AB}(U \otimes U)^\dagger \, \mathrm{d}U = \rho_{AB}^{\mathrm{W};p}, \quad p = \mathrm{Tr}\left[\Pi_{AB}^{\mathrm{ASym}}\rho_{AB}\right]. \tag{3.2.141}$$

Please see the Bibliographic Notes (Section 3.4) for more information.

3.3 Measurements

Measurements in quantum mechanics are described by *positive operator-valued measures (POVMs)*.

Definition 3.21 Positive Operator-Valued Measure (POVM)

A POVM is a set $\{M_x\}_{x\in\mathcal{X}}$ of operators satisfying

$$M_x \geq 0 \quad \forall\, x \in \mathcal{X}, \qquad \sum_{x\in\mathcal{X}} M_x = \mathbb{1}. \tag{3.3.1}$$

For our purposes, it suffices to consider finite sets of such operators. The elements of the finite alphabet $\mathcal{X}$ are used to label the outcomes of the measurement.

The measurement of a quantum system in the state ρ according to the POVM $\{M_x\}_{x\in\mathcal{X}}$ induces a probability distribution $p_X : \mathcal{X} \to [0,1]$. This distribution corresponds to a random variable X that takes values in the alphabet $\mathcal{X}$ and is defined by the *Born rule*:

$$p_X(x) = \Pr[X = x] = \mathrm{Tr}[M_x\rho]. \tag{3.3.2}$$

If every element M_x of the POVM is a projection, then the corresponding measurement is called a *projective measurement*. If the POVM of a projective measurement consists solely of rank-one projections, then the measurement is sometimes called a *von Neumann measurement*. Observe using (2.2.3) that every orthonormal basis $\{|e_k\rangle\}_{k=1}^d$ for a d-dimensional Hilbert space $\mathcal{H}$ defines a von Neumann measurement via the POVM $\{|e_k\rangle\langle e_k|\}_{k=1}^d$. Furthermore, as stated at the beginning of this chapter, every Hermitian operator defines a projective measurement, with the corresponding POVM given by its spectral projections.

Projective measurements, in particular von Neumann measurements, are viewed as the simplest type of measurement that can be performed on a quantum system. However, as it turns out, we can implement every measurement as a von Neumann measurement if we have access to an auxiliary quantum system, and this is the content of *Naimark's theorem*.

Theorem 3.22 Naimark's Theorem

For every POVM $\{M_x\}_{x\in\mathcal{X}}$, there exists an isometry V such that

$$M_x = V^\dagger(\mathbb{1} \otimes |x\rangle\langle x|)V \quad \forall\, x \in \mathcal{X}, \tag{3.3.3}$$

where $\{|x\rangle\}_{x\in\mathcal{X}}$ is an orthonormal set.

Proof: This follows immediately by defining the isometry V as

$$V = \sum_{x\in\mathcal{X}} \sqrt{M_x} \otimes |x\rangle. \tag{3.3.4}$$

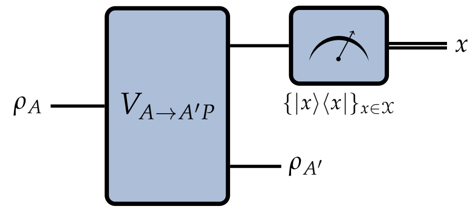


FIGURE 3.3: The system-probe model of measurement. In order to measure the system A of interest, we allow it to first interact with a probe system P via an isometry $V_{A\to A'P}$. We then measure the probe with a projective measurement consisting of rank-one elements.

This is indeed an isometry because

$$V^\dagger V = \left(\sum_{x\in\mathcal{X}} \sqrt{M_x} \otimes \langle x|\right)\left(\sum_{x'\in\mathcal{X}} \sqrt{M_{x'}} \otimes |x'\rangle\right) \tag{3.3.5}$$

$$= \sum_{x,x'\in\mathcal{X}} \sqrt{M_x}\sqrt{M_{x'}}\langle x|x'\rangle \tag{3.3.6}$$

$$= \sum_{x\in\mathcal{X}} M_x \tag{3.3.7}$$

$$= \mathbb{1}, \tag{3.3.8}$$

where the last equality holds because $\{M_x\}_{x\in\mathcal{X}}$ is a POVM. It is straightforward to check that (3.3.3) is satisfied for the choice of V in (3.3.4). ■

The physical relevance of Naimark's theorem is illustrated in Figure 3.3. We model the measurement of the system A of interest as an interaction of the system with a *probe* system P followed by a projective measurement of the probe system described by the POVM $\{|x\rangle\langle x| : x \in \mathcal{X}\}$. The interaction is described by an isometry $V_{A\to A'P}$, and if $\rho_{AP} = V\rho_A V^\dagger$ is the joint state of the system and the probe after the interaction, then the measurement outcome probabilities are

$$p_X(x) = \mathrm{Tr}[(\mathbb{1}_{A'} \otimes |x\rangle\langle x|)\rho_{A'P}] \tag{3.3.9}$$

$$= \mathrm{Tr}[(\mathbb{1}_{A'} \otimes |x\rangle\langle x|)V\rho_A V^\dagger] \tag{3.3.10}$$

$$= \mathrm{Tr}[V^\dagger(\mathbb{1}_{A'} \otimes |x\rangle\langle x|)V\rho_A] \tag{3.3.11}$$

$$= \mathrm{Tr}[M_x\rho_A], \tag{3.3.12}$$

where we let $M_x := V^\dagger(\mathbb{1}_{A'} \otimes |x\rangle\langle x|)V$. This shows us that the system-probe model of measurement, in which the probe is measured after it interacts with the system of interest, can effectively be described by a POVM. Naimark's theorem, on the other hand, tells us the converse: for every measurement described by a POVM, there exists an isometry such that the measurement can be described in the system-probe model, as depicted in Figure 3.3.

Exercise 3.26

Consider the qubit state vectors

$$|\psi_k\rangle := \cos\left(\frac{2\pi k}{5}\right)|0\rangle + \sin\left(\frac{2\pi k}{5}\right)|1\rangle, \quad k \in \{0,1,2,3,4\}. \tag{3.3.13}$$

Verify that the set $\left\{\frac{2}{5}|\psi_k\rangle\langle\psi_k|\right\}_{k=0}^{4}$ is a POVM. Note that this POVM gives us an example of a non-projective measurement.

When performing measurements in quantum mechanics, we are typically interested not only in the measurement outcomes and their probabilities but also with the so-called *post-measurement* states of the system being measured. That is, we are interested in knowing the state of the system after we have measured it and observed the outcome.

Since every POVM element M_x is positive semi-definite, there exists an operator K_x such that $M_x = K_x^\dagger K_x$ for all $x \in \mathcal{X}$. For example, we could let K_x be the square root of M_x, so that $K_x = \sqrt{M_x}$. Then, the Born rule in (3.3.2) for the probability of the measurement outcome $x \in \mathcal{X}$ can be written as $p_X(x) = \mathrm{Tr}[K_x \rho K_x^\dagger]$. In this case, the post-measurement state corresponding to the outcome $x \in \mathcal{X}$ is as follows:

$$\rho^x := \frac{K_x \rho K_x^\dagger}{\mathrm{Tr}[K_x \rho K_x^\dagger]}. \tag{3.3.14}$$

The state ρ^x can be understood to capture the experimenter's description of the state of the system given that the measurement outcome was observed to be x.

The post-measurement states ρ^x give rise to the ensemble $\{(p_X(x), \rho^x)\}_{x\in\mathcal{X}}$. The expected density operator of the ensemble is

$$\rho_{\mathcal{M}} := \sum_{x\in\mathcal{X}} p_X(x)\rho^x = \sum_{x\in\mathcal{X}} K_x \rho K_x^\dagger. \tag{3.3.15}$$

This expected density operator is the state of the system after measurement if the measurement outcome is not available. It can be interpreted as the state of the system after measurement if the experimenter does not have access to the measurement outcome.

Due to the unitary freedom in the decomposition $M_x = K_x^\dagger K_x$, other choices of K_x are given by $K_x = U_x\sqrt{M_x}$ for some unitary U_x, so that there is not a unique way to determine the post-measurement state when starting from a POVM.

Suppose now that we perform a measurement on a subsystem of a composite system. Specifically, consider measuring a system A that is in a joint state ρ_{RA} with a reference system R, and let the measurement be described by the POVM $\{M_A^x\}_{x\in\mathcal{X}}$ for some finite alphabet $\mathcal{X}$. If we let $M_A^x = K_A^{x\dagger}K_A^x$, then according to (3.3.14), the measurement probabilities are given by $p_X(x) = \mathrm{Tr}[(\mathbb{1}_R \otimes K_A^x)\rho_{RA}(\mathbb{1}_R \otimes K_A^{x\dagger})]$ and the post-measurement states are as follows:

$$\rho_{RA}^x = \frac{(\mathbb{1}_R \otimes K_A^x)\rho_{RA}(\mathbb{1}_R \otimes K_A^{x\dagger})}{\mathrm{Tr}[(\mathbb{1}_R \otimes K_A^x)\rho_{RA}(\mathbb{1}_R \otimes K_A^{x\dagger})]} \tag{3.3.16}$$

$$= \frac{1}{p_X(x)}(\mathbb{1}_R \otimes K_A^x)\rho_{RA}(\mathbb{1}_R \otimes K_A^{x\dagger}) \tag{3.3.17}$$

for all $x \in \mathcal{X}$. The state of the system R conditioned on the measurement outcome x is then

$$\rho_R^x := \mathrm{Tr}_A[\rho_{RA}^x] \tag{3.3.18}$$

$$= \frac{1}{p_X(x)}\mathrm{Tr}_A[(\mathbb{1}_R \otimes K_A^x)\rho_{RA}(\mathbb{1}_R \otimes K_A^{x\dagger})] \tag{3.3.19}$$

$$= \frac{1}{p_X(x)}\mathrm{Tr}_A[(\mathbb{1}_R \otimes K_A^{x\dagger}K_A)\rho_{RA}] \tag{3.3.20}$$

$$= \frac{1}{p_X(x)}\mathrm{Tr}_A[(\mathbb{1}_R \otimes M_A^x)\rho_{RA}]. \tag{3.3.21}$$

We thus see that, although the post-measurement state on the system A being measured is not uniquely defined due to the unitary freedom in the decomposition $M_A^x = K_A^{x\dagger}K_A^x$, as described earlier, the post-measurement state on the reference system R not being measured is uniquely defined because it depends directly on each POVM element M_A^x. If the system A undergoes a measurement for which $M_A^x = |\psi^x\rangle\langle\psi^x|_A$, then (3.3.21) can be written as

$$\rho_R^x = \frac{1}{p_X(x)}\langle\psi^x|_A\rho_{RA}|\psi^x\rangle_A. \tag{3.3.22}$$

Exercise 3.27

Consider a finite set $\{\rho^x\}_{x\in\mathcal{X}}$ of quantum states, and let $R := \sum_{x\in\mathcal{X}} \rho^x$.

1. If R is invertible, let

$$M_x := R^{-\frac{1}{2}} \rho^x R^{-\frac{1}{2}} \quad \forall\, x \in \mathcal{X}. \tag{3.3.23}$$

   Prove that $\{M_x\}_{x\in\mathcal{X}}$ is a POVM.

2. If R is not invertible, we can still define $M_x := R^{-\frac{1}{2}} \rho^x R^{-\frac{1}{2}}$ by taking the inverse of R on its support. However, in this way the operators M_x no longer add up to the identity. To complete them, we can define

$$M_\perp := \mathbb{1} - \Pi_R, \tag{3.3.24}$$

   where Π_R is the projection onto the support of R. Prove that $\{M_x\}_{x\in\mathcal{X}\cup\{\perp\}}$ is a POVM. (*Hint*: Recall the defnitions from Section 2.2.8.1.)

3.4 Bibliographic Notes

More details on many of the concepts that have been presented in this chapter can be found in the books of Nielsen and Chuang (2000), Holevo (2013), Hayashi (2017), Wilde (2017a), and Watrous (2018). For a treatment of these concepts in infinite-dimensional Hilbert spaces, see the books by Holevo (2013) and Heinosaari and Ziman (2012).

The mathematical theory of quantum mechanics was developed by von Neumann (1927a, 1932) and Landau (1927). The book by Helstrom (1976) also contains early developments in the theory of quantum measurements.

We have used quantum-optical modes as a concrete example of qubit encodings throughout this chapter. For a detailed reference on this topic, see the review by Kok et al. (2007).

Recall the vector $\vec{r}_\rho$ in (3.2.4) of coefficients of a quantum state in terms of the orthogonal basis defined via the operators in (2.2.46)–(2.2.49). This vector is known as both the "coherence vector" and the "Bloch vector" of ρ. Perhaps the earliest use of the term "coherence vector" is in the work of Hioe and Eberly (1981), in which the orthogonality convention for the operators in (2.2.46)–(2.2.49) differs from the convention used in this book. Positivity of linear operators in terms of the coherence vector was presented by Byrd and Khaneja (2003); Kimura (2003). The latter work by Kimura (2003) uses the term "Bloch vector", which has also been used by Bertlmann and Krammer (2008).

Lemma 3.13, regarding purifications of permutation-invariant states, is due to Renner (2005). Separable and entangled states were defined by Werner (1989b). For an in-depth review of the properties of entanglement and its various applications in quantum information theory, including a discussion of multipartite entanglement, we refer to the article by Horodecki et al. (2009b). For an in-depth discussion of multipartite entanglement, we refer to (Walter et al., 2016). The relevance of the partial transpose for characterizing entanglement in quantum information was pointed out by Peres (1996); Horodecki et al. (1996), and the set of positive-partial-transpose states was discussed by Horodecki (1997); Horodecki et al. (1998). In particular, the existence of PPT entangled states was found by Horodecki (1997).

Isotropic states were defined by Horodecki and Horodecki (1999), and Werner states were defined by Werner (1989b). Proofs of formulas for the integration of unitary operators with respect to the Haar measure, including proofs of (3.2.133) and (3.2.141), can be found in (Collins, 2003; Collins and Śniady, 2006; Roy and Scott, 2009).

Appendix 3.A Proof of Lemma 3.3

PROOF: First suppose that X_{AB} is rank one, so that $X_{AB} = |\Psi\rangle\langle\Psi|_{AB}$ for some vector $|\Psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. Due to the Schmidt decomposition theorem (Theorem 2.2), we have that

$$|\Psi\rangle_{AB} = \sum_{z\in\mathcal{Z}} \gamma_z |\theta_z\rangle_A \otimes |\xi_z\rangle_B, \tag{3.A.1}$$

where $|\mathcal{Z}| \leq \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}$, the set $\{\gamma_z\}_z$ is a set of strictly positive numbers, and $\{|\theta_z\rangle_A\}_z$ and $\{|\xi_z\rangle_B\}_z$ are orthonormal bases. Then

$$\begin{aligned} \mathrm{supp}(X_{AB}) &= \mathrm{span}\{|\Psi\rangle_{AB}\} && (3.A.2)\\ &\subseteq \mathrm{span}\{|\theta_z\rangle_A : z \in \mathcal{Z}\} \otimes \mathrm{span}\{|\xi_z\rangle_B : z \in \mathcal{Z}\}. && (3.A.3) \end{aligned}$$

The statement then follows for this case because $\mathrm{supp}(X_A) = \mathrm{span}\{|\theta_z\rangle_A : z \in \mathcal{Z}\}$ and $\mathrm{supp}(X_B) = \mathrm{span}\{|\xi_z\rangle_B : z \in \mathcal{Z}\}$.

Now suppose that X_{AB} is not rank one. It admits a decomposition into rank-one vectors of the following form:

$$X_{AB} = \sum_{x \in \mathcal{X}} |\Psi^x\rangle\!\langle\Psi^x|_{AB}, \tag{3.A.4}$$

where $|\Psi^x\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ for all $x \in \mathcal{X}$. Set $\Psi^x_{AB} = |\Psi^x\rangle\!\langle\Psi^x|_{AB}$, and let $\Psi^x_A \coloneqq \mathrm{Tr}_B[\Psi^x_{AB}]$ and $\Psi^x_B \coloneqq \mathrm{Tr}_A[\Psi^x_{AB}]$. Then

$$\begin{aligned}
\mathrm{supp}(X_{AB}) &= \mathrm{span}\{|\Psi^x\rangle_{AB} : x \in \mathcal{X}\} && (3.A.5)\\
&\subseteq \mathrm{span}\left[\bigcup_{x \in \mathcal{X}} \left[\mathrm{supp}(\Psi^x_A) \otimes \mathrm{supp}(\Psi^x_B)\right]\right] && (3.A.6)\\
&\subseteq \mathrm{span}\left[\bigcup_{x \in \mathcal{X}} \mathrm{supp}(\Psi^x_A)\right] \otimes \mathrm{span}\left[\bigcup_{x \in \mathcal{X}} \mathrm{supp}(\Psi^x_B)\right] && (3.A.7)\\
&= \mathrm{supp}(X_A) \otimes \mathrm{supp}(X_B), && (3.A.8)
\end{aligned}$$

concluding the proof. ■

Appendix 3.B Proof of Lemma 3.4

Proof: First suppose that X_{AB} is rank one, as in the first part of the proof of the previous lemma, and let us use the same notation as given there. Applying the same lemma gives that

$$\mathrm{supp}(X_{AB}) \subseteq \mathrm{supp}(Y_{AB}) \subseteq \mathrm{supp}(Y_A) \otimes \mathrm{supp}(Y_B), \tag{3.B.1}$$

which in turn implies that $\mathrm{supp}(X_{AB}) = \mathrm{span}\{|\Psi\rangle_{AB}\} \subseteq \mathrm{supp}(Y_A) \otimes \mathrm{supp}(Y_B)$. This implies that $|\theta_z\rangle_A \in \mathrm{supp}(Y_A)$ for all $z \in \mathcal{Z}$, and thus that $\mathrm{span}\{|\theta_z\rangle_A\} \subseteq \mathrm{supp}(Y_A)$. We can then conclude the statement in this case because $\mathrm{span}\{|\theta_z\rangle_A\} = \mathrm{supp}(X_A)$.

Now suppose that X_{AB} is not rank one. Then it admits a decomposition as given in the proof of the previous lemma. Using the same notation, we have that $\mathrm{supp}(\Psi^x_{AB}) \subseteq \mathrm{supp}(Y_{AB})$ holds for all $x \in \mathcal{X}$. Since we have proven the lemma for rank-one operators, we can conclude that $\mathrm{supp}(\Psi^x_A) \subseteq \mathrm{supp}(Y_A)$ holds for all $x \in \mathcal{X}$. As a consequence, we find that

$$\mathrm{supp}(X_A) = \mathrm{span}\left[\bigcup_{x \in \mathcal{X}} \mathrm{supp}(\Psi^x_A)\right] \subseteq \mathrm{supp}(Y_A), \tag{3.B.2}$$

concluding the proof. ■

Chapter 4

Quantum Channels

In the previous chapter, we studied quantum states and measurements, which are regulated by the first three axioms of quantum mechanics, as presented in Section 3.1. The fourth and final axiom is about the evolution of quantum systems, which is the subject of this chapter. Mathematically, the evolution is described by a *quantum channel*. As quantum communication necessarily involves the evolution of quantum systems (such as the evolution of photons when travelling through an optical fiber), quantum channels are the primary objects of study in this book. This chapter is devoted to a detailed study of quantum channels, including their properties, representations, and various examples that are relevant for quantum communication and quantum information more broadly.

The fourth axiom in Section 3.1 states that a quantum channel is a "linear, completely positive, and trace-preserving map acting on the state of the system." At first glance, this appears to be a purely mathematical statement (which we elaborate upon in Section 4.1), with seemingly little connection to physics. However, we can connect this statement to the axiom of evolution of quantum systems as taught in a basic quantum physics course. There, one learns that the evolution of a (non-relativistic) quantum system is governed by the *Schrödinger equation*:

$$i\hbar \frac{\partial}{\partial t}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \tag{4.0.1}$$

where $|\psi(t)\rangle$ is the state vector of the system at time $t \geq 0$ and $H(t)$ is the Hamiltonian operator of the system at time t. The Hamiltonian operator is a Hermitian operator that describes the energy of the system. Now, we know from Chapter 3 that the state of a quantum system is described more generally by a density operator. The analogue of (4.0.1) for density operators is known as the *von Neumann equation*:

$$i\hbar \frac{\partial \rho(t)}{\partial t} = [H(t), \rho(t)], \tag{4.0.2}$$

where $\rho(t)$ is the density operator describing the state of the system at time $t \geq 0$, and $[H(t), \rho(t)] = H(t)\rho(t) - \rho(t)H(t)$ is the commutator of the Hamiltonian $H(t)$ and the state $\rho(t)$.

Both (4.0.1) and (4.0.2) describe the evolution of so-called *closed* quantum systems, and this evolution is given by unitary maps. In other words, the solution to (4.0.1) is $|\psi(t)\rangle = U(t)|\psi_0\rangle$ for all $t \geq 0$, where $|\psi_0\rangle$ is an initial state vector of the system (at time $t = 0$) and $U(t)$ is a unitary operator. Similarly, the solution to (4.0.2) is $\rho(t) = U(t)\rho_0 U(t)^\dagger$ for all $t \geq 0$, where ρ_0 is an initial quantum state of the system (at time $t = 0$) and $U(t)$ is a unitary operator. We refer to the Bibliographic Notes in Section 4.8 for references on explicit forms for the unitary $U(t)$. We show in this chapter that unitary maps are quantum channels. This fact provides a connection between the mathematical statement of the evolution axiom in Section 3.1 and the statement of the evolution axiom typically taught in quantum physics courses.

More generally, we are interested in the evolution of *open* quantum systems, i.e., quantum systems that interact with an external environment that is out of our control. For such systems, the same connection as before holds. In fact, the evolution is given by a joint unitary evolution of the system and environment followed by discarding the state of the environment, and as we show in Section 4.3, every completely positive trace-preserving map (i.e., every quantum channel) can be viewed in terms of a joint unitary evolution with an environment followed by discarding the state of the environment. (Please see the Bibliographic Notes in Section 4.8 for references on open quantum systems.) Thus, from an abstract, information-theoretic perspective, the evolution of a quantum system is given simply by a quantum channel, and the details of the actual physical system of interest (which would be given by the Hamiltonian operator) are unimportant. This viewpoint is powerful: with it, we realize that virtually every operation on quantum states, including measurements, is a quantum channel.

4.1 Definition

We can motivate the definition of a quantum channel by using the following basic mathematical facts that should be satisfied by a map $\mathcal{N} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ that represents the evolution of a quantum system:

1. If $\mathcal{N}$ acts on a mixture of quantum states, then the output state should be equal to the mixture of the individual outputs. That is,

$$\mathcal{N}(\lambda\rho + (1-\lambda)\sigma) = \lambda\mathcal{N}(\rho) + (1-\lambda)\mathcal{N}(\sigma) \tag{4.1.1}$$

   for all states ρ and σ and $\lambda \in [0,1]$. This requirement is called convex linearity on density operators. Importantly, for each convex linear map acting on the convex set of density operators it is possible to define a unique linear map acting on the space of all linear operators that is an extension of the original map, i.e., that reproduces its action on density operators. The latter is the mathematical object that we employ, and so we require that $\mathcal{N}$ be a linear map, i.e., a superoperator. (Recall the definition of a superoperator from Section 2.2.11.)

2. The map $\mathcal{N}$ should accept a quantum state (or a mixture of quantum states) as input and output a legitimate quantum state. This means that $\mathcal{N}$ should be *trace preserving* and *positive*. However, it is furthermore reasonable to demand that if the channel acts on one share A of a bipartite quantum state ρ_{RA}, then the output should be a legitimate bipartite quantum state. So we demand additionally that a quantum channel should be not just positive, but additionally *completely positive*. Let us now define these terms.

   (a) $\mathcal{N}$ is called *trace preserving* if $\mathrm{Tr}[\mathcal{N}(X)] = \mathrm{Tr}[X]$ for every linear operator X. More generally, $\mathcal{N}$ is called *trace non-increasing* if $\mathrm{Tr}[\mathcal{N}(X)] \leq \mathrm{Tr}[X]$ for every positive semi-definite operator X.

   (b) $\mathcal{N}$ is called *positive* if it maps positive semi-definite operators to positive semi-definite operators, i.e., $\mathcal{N}(X) \geq 0$ for all $X \geq 0$. It is called *k-positive*, with $k \geq 1$, if the map $\mathrm{id}_k \otimes \mathcal{N}$ is positive. Note that if $\mathcal{N}$ is a map acting on linear operators in $\mathrm{L}(\mathbb{C}^d)$, then the map $\mathrm{id}_k \otimes \mathcal{N}$ acts on linear operators in $\mathrm{L}(\mathbb{C}^{kd})$. In other words, for every linear operator X acting on a kd-dimensional Hilbert space, which we can decompose as the block matrix

$$X = \begin{pmatrix} X_{0,0} & \cdots & X_{0,k-1} \\ \vdots & \ddots & \vdots \\ X_{k-1,0} & \cdots & X_{k-1,k-1} \end{pmatrix}, \tag{4.1.2}$$

such that $X_{i,j}$ is a $d \times d$ matrix for all $0 \le i, j \le k-1$, the action of the map $\mathrm{id}_k \otimes \mathcal{N}$ is defined as

$$(\mathrm{id}_k \otimes \mathcal{N})(X) = \begin{pmatrix} \mathcal{N}(X_{0,0}) & \cdots & \mathcal{N}(X_{0,k-1}) \\ \vdots & \ddots & \vdots \\ \mathcal{N}(X_{k-1,0}) & \dots & \mathcal{N}(X_{k-1,k-1}) \end{pmatrix}. \tag{4.1.3}$$

We can write this more compactly as follows. Noting that $\mathrm{L}(\mathbb{C}^{kd})$ is isomorphic to $\mathrm{L}(\mathbb{C}^k) \otimes \mathrm{L}(\mathbb{C}^d)$, we can write $X \in \mathrm{L}(\mathbb{C}^{kd})$ as

$$X = \sum_{i,j=0}^{k-1} |i\rangle\langle j| \otimes X_{i,j}. \tag{4.1.4}$$

Then the action of $\mathrm{id}_k \otimes \mathcal{N}$ is defined as

$$(\mathrm{id}_k \otimes \mathcal{N})(X) = \sum_{i,j=0}^{k-1} |i\rangle\langle j| \otimes \mathcal{N}(X_{i,j}). \tag{4.1.5}$$

The superoperator $\mathcal{N}$ is called *completely positive* if $\mathrm{id}_k \otimes \mathcal{N}$ is positive for every integer $k \ge 1$.

Physically, the complete positivity of $\mathcal{N}$ takes into account the fact that the system of interest might be entangled with another system that is outside of our control, so that simply letting $\mathcal{N}$ be positive is not sufficient to ensure that all positive semi-definite operators get mapped to positive semi-definite operators. Letting $\mathcal{N}$ be completely positive means that positive semi-definite operators are mapped to positive semi-definite operators even in this more general setting.

The defining properties of linearity, complete positivity, and trace preservation for quantum channels together ensure that quantum states for the systems of interest get mapped to quantum states, even if they happen to be entangled with other external systems outside of our control. These properties are consistent with what is observed in real physical systems.

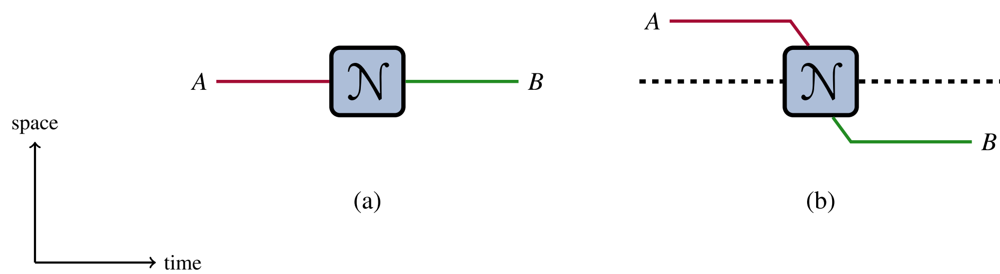


FIGURE 4.1: Our convention for drawing quantum channels throughout this book, with time increasing horizontally towards the right, and spatial separations indicated vertically. In (a), the input and output systems A and B, respectively, of the quantum channel $\mathcal{N}$ are temporally separated but not spatially separated. In (b), A and B are both spatially and temporally separated. We often draw a dashed line to indicate the spatial separation explicitly.

Throughout this book, we write $\mathcal{N}_{A\to B}$ to denote a map $\mathcal{N} : \mathrm{L}(\mathcal{H}_A) \to \mathrm{L}(\mathcal{H}_B)$ taking a quantum system A to a quantum system B. We sometimes write $\mathcal{N}_A$ if the input and output systems of the channel have the same dimension. We drop the subscript indicating the input and output systems if they are not important in the context being considered. Physically, a quantum channel, being a description of the time evolution of a quantum system, describes the transition of a quantum system between two distinct points in time. The systems A and B at the input and output of the channel, respectively, thus represent quantum systems at two distinct points in time, as shown in Figure 4.1(a). However, in the context of communication, we regard the systems A and B as being separated both in time as well as in space, with A belonging to an individual "Alice" and B belonging to "Bob." We show this physical separation explicitly throughout this book according to the convention shown in Figure 4.1(b).

Exercise 4.1

Let $\mathcal{N}$ be a k-positive superoperator, for an integer $k \geq 1$. Prove that $\mathcal{N}$ is Hermiticity preserving (recall Definition 2.17). (*Hint*: See Exercise 2.17 and use the Jordan–Hahn decomposition.)

Exercise 4.2

Prove that a superoperator $\mathcal{N}_{A\to B}$ is trace-non-increasing if and only if its adjoint is *subunital*, meaning that $\mathcal{N}^\dagger(\mathbb{1}_B) \leq \mathbb{1}_A$. Prove that the inequality is saturated, i.e., that $\mathcal{N}^\dagger(\mathbb{1}_B) = \mathbb{1}_A$ ($\mathcal{N}^\dagger$ is unital), if and only if $\mathcal{N}$ is trace preserving.

Exercise 4.3

1. Let $\mathcal{N}_{A\to B}$ be a positive superoperator. Starting with (2.2.171), prove that

$$\|\mathcal{N}\|_1 = \left\|\mathcal{N}^\dagger(\mathbb{1}_B)\right\|_\infty . \tag{4.1.6}$$

2. Using 1., conclude the following:
   (a) If $\mathcal{N}_{A\to B}$ is a k-positive, trace-non-increasing superoperator, then $\|\mathrm{id}_k \otimes \mathcal{N}\|_1 \leq 1$;
   (b) If $\mathcal{N}_{A\to B}$ is a k-positive, trace-preserving superoperator, then $\|\mathrm{id}_k \otimes \mathcal{N}\|_1 = 1$.
3. Using 1. and 2., conclude the following:
   (a) If $\mathcal{N}_{A\to B}$ is a completely positive, trace-non-increasing superoperator, then $\|\mathcal{N}\|_\diamond \leq 1$ (the diamond norm $\|\cdot\|_\diamond$ is introduced in Definition 2.20);
   (b) If $\mathcal{N}_{A\to B}$ is a quantum channel, then $\|\mathcal{N}\|_\diamond = 1$.

Combining the result of Exercise 4.3 and (2.2.170), we conclude that for every positive, trace-non-increasing superoperator $\mathcal{N}_{A\to B}$, and every linear operator $X \in \mathrm{L}(\mathcal{H}_A)$,

$$\|\mathcal{N}(X)\|_1 \leq \|X\|_1 . \tag{4.1.7}$$

An inequality of this type is called a *data-processing inequality*, for which we provide an interpretation later on in Section 6.1. We encounter numerous such inequalities with respect to various different quantities throughout the rest of this book, and they turn out to be of central importance in the analysis of quantum communication protocols, and in quantum information theory more generally.

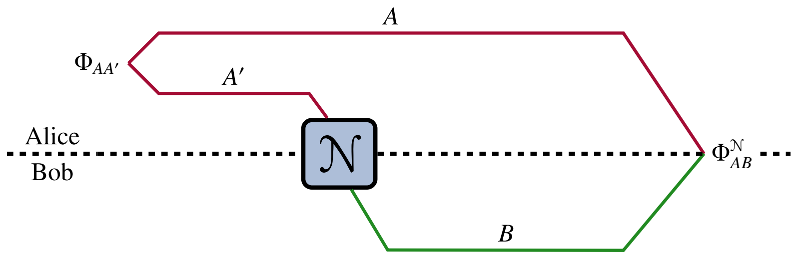


FIGURE 4.2: The normalized Choi representation $\Phi^{\mathcal{N}}_{AB}$ of a superoperator $\mathcal{N}_{A\to B}$ is the bipartite operator resulting from sending one share of the maximally entangled state $\Phi_{AA'}$, defined in (3.2.38), through $\mathcal{N}$.

4.2 Choi Representation

The Choi representation gives a way to represent a quantum channel as a bipartite operator and is an essential concept in quantum information theory.

Definition 4.1 Choi Representation

For every superoperator $\mathcal{N}_{A\to B}$, its *Choi representation*, or *Choi operator*, is defined as

$$\Gamma^{\mathcal{N}}_{AB} := \mathcal{N}_{A'\to B}(|\Gamma\rangle\langle\Gamma|_{AA'}) = \sum_{i,j=0}^{d_A-1} |i\rangle\langle j|_A \otimes \mathcal{N}(|i\rangle\langle j|_{A'}), \tag{4.2.1}$$

where $\mathcal{H}_{A'}$ is isomorphic to the Hilbert space $\mathcal{H}_A$ corresponding to the channel input system A. We also define the operator

$$\Phi^{\mathcal{N}}_{AB} := \frac{1}{d_A}\Gamma^{\mathcal{N}}_{AB} = \mathcal{N}_{A'\to B}(\Phi_{AA'}), \tag{4.2.2}$$

where $\Phi_{AA'} = |\Phi\rangle\langle\Phi|_{AA'}$ is the maximally entangled state defined in (3.2.38).

The Choi representation $\Gamma^{\mathcal{N}}_{AB}$ of the superoperator $\mathcal{N}_{A\to B}$ is an operator acting on $\mathcal{H}_{AB}$ and uniquely characterizes the map because it specifies the action of $\mathcal{N}$ on the basis $\{|i\rangle\langle j|_A : 0 \le i, j \le d_A - 1\}$ of linear operators acting on $\mathcal{H}_A$. As shown in Figure 4.2, the operator $\Phi^{\mathcal{N}}_{AB}$ is simply the Choi representation normalized by the dimension d_A of the input system A of the superoperator $\mathcal{N}$, and it is the linear operator resulting from sending one share of a maximally entangled state through $\mathcal{N}$. When $\mathcal{N}$ is a quantum channel, we refer to $\Phi^{\mathcal{N}}_{AB}$ as the *Choi state* of $\mathcal{N}$, because $\Phi^{\mathcal{N}}_{AB}$ is positive semi-definite and has unit trace; see Theorem 4.3 below.

Exercise 4.4

Let $\mathcal{N}_{A\to B}$ be a superoperator.

1. Prove that $\mathrm{Tr}_A[\Gamma^{\mathcal{N}}_{AB}] = \mathcal{N}_{A\to B}(\mathbb{1}_A)$.

2. Prove that $\mathcal{N}_{A\to B}$ is trace preserving if and only if $\mathrm{Tr}_B[\Gamma^{\mathcal{N}}_{AB}] = \mathbb{1}_A$. Conclude that $\mathrm{Tr}[\Phi^{\mathcal{N}}_{AB}] = 1$.

3. Prove that $\left\langle \overline{X_A} \otimes Y_B, \Gamma^{\mathcal{N}}_{AB} \right\rangle = \langle Y_B, \mathcal{N}_{A\to B}(X_A)\rangle$ for all $X_A \in \mathrm{L}(\mathcal{H}_A)$ and $Y_B \in \mathrm{L}(\mathcal{H}_B)$.

4. Using 3., prove that the Choi representation of $\mathcal{N}$ can be expressed using the adjoint $\mathcal{N}^\dagger$ as follows:

$$\Gamma^{\mathcal{N}}_{AB} = \sum_{k,\ell=0}^{d_B-1} \overline{\mathcal{N}^\dagger(|k\rangle\langle\ell|_B)} \otimes |k\rangle\langle\ell|_B. \tag{4.2.3}$$

   Conclude that $\mathrm{Tr}_B[\Gamma^{\mathcal{N}}_{AB}] = \overline{\mathcal{N}^\dagger(\mathbb{1}_B)}$.

5. Prove that, for every unitary operator U_A,

$$\Gamma^{\mathcal{N}}_{AB} = \big(\mathcal{U}_A \otimes \big(\mathcal{N}_{A'\to B} \circ \overline{\mathcal{U}}_{A'}\big)\big)(\Gamma_{AA'}), \tag{4.2.4}$$

   where $\mathcal{U}_A(\cdot) \coloneqq U_A(\cdot)U_A^\dagger$ and $\overline{\mathcal{U}}_A(\cdot) \coloneqq \overline{U}_A(\cdot)\overline{U}_A^\dagger$.

Proposition 4.2 Quantum Dynamics from Choi Operator

Let $\mathcal{N}_{A\to B}$ be a superoperator, and let X_{RA} be a bipartite operator, with R an arbitrary reference system. Then the action of the superoperator $\mathrm{id}_R \otimes \mathcal{N}_{A\to B}$ on X_{RA} can be expressed in terms of the Choi operator $\Gamma^{\mathcal{N}}_{AB}$ as follows:

$$(\mathrm{id}_R \otimes \mathcal{N}_{A\to B})(X_{RA}) = \mathrm{Tr}_A[(\mathrm{T}_A(X_{RA}) \otimes \mathbb{1}_B)(\mathbb{1}_R \otimes \Gamma^{\mathcal{N}}_{AB})], \tag{4.2.5}$$

$$= \langle\Gamma|_{A'A}(X_{RA'} \otimes \Gamma^{\mathcal{N}}_{AB})|\Gamma\rangle_{A'A}, \quad (4.2.6)$$

where T_A denotes the partial transpose from Definition 3.16.

PROOF: Observe that (4.2.1) implies that

$$\mathcal{N}_{A\to B}(|i\rangle\langle j|_A) = (\langle i|_A \otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{AB}(|j\rangle_A \otimes \mathbb{1}_B) \quad (4.2.7)$$

for all $0 \leq i, j \leq d_A - 1$. We extend this by linearity to apply to every input operator X_A by the following reasoning. Expanding X_A as $X_A = \sum_{i,j=0}^{d_A-1} X_{i,j}|i\rangle\langle j|_A$, we find that

$$\mathcal{N}_{A\to B}(X_A) = \sum_{i,j=0}^{d_A-1} X_{i,j}\mathcal{N}_{A\to B}(|i\rangle\langle j|_A) \quad (4.2.8)$$

$$= \sum_{i,j=0}^{d_A-1} X_{i,j}(\langle i|_A \otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{AB}(|j\rangle_A \otimes \mathbb{1}_B) \quad (4.2.9)$$

$$= \sum_{i,j=0}^{d_A-1} X_{i,j}\mathrm{Tr}_A[(|j\rangle\langle i|_A \otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{AB}] \quad (4.2.10)$$

$$= \mathrm{Tr}_A[(X_A^{\mathsf{T}} \otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{AB}]. \quad (4.2.11)$$

So we conclude that the action of $\mathcal{N}_{A\to B}$ on every linear operator X_A can be expressed using the Choi representation as

$$\mathcal{N}_{A\to B}(X_A) = \mathrm{Tr}_A[(X_A^{\mathsf{T}} \otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{AB}]. \quad (4.2.12)$$

Now employing (2.2.43) and (2.2.42), we conclude that the action of $\mathcal{N}_{A\to B}$ on every linear operator X_A can be expressed alternatively as

$$\mathcal{N}_{A\to B}(X_A) = \langle\Gamma|_{A'A}(X_{A'} \otimes \Gamma^{\mathcal{N}}_{AB})|\Gamma\rangle_{A'A}. \quad (4.2.13)$$

The identities in (4.2.5) and (4.2.13) extend more generally to the case of the superoperator $\mathrm{id}_R \otimes \mathcal{N}_{A\to B}$ acting on a bipartite operator X_{RA} by expanding X_{RA} as $X_{RA} = \sum_{i,j=0}^{d_R-1} |i\rangle\langle j|_R \otimes X_A^{i,j}$ and using linearity. We thus conclude (4.2.5) and (4.2.6). ■

Exercise 4.5

Prove that a superoperator $\mathcal{N}_{A\to B}$ is Hermiticity preserving (recall the definition in Section 2.2.11) if and only if its Choi representation $\Gamma^{\mathcal{N}}_{AB}$ is Hermitian.

Using the definition of the Choi state and the maximally entangled state $\Phi_{A'A}$, we can write (4.2.6) as

$$\langle\Phi|_{A'A}(X_{RA'}\otimes\Phi^{\mathcal{N}}_{AB})|\Phi\rangle_{A'A}=\frac{1}{d_A^2}\mathcal{N}_{A'\to B}(X_{RA'}). \tag{4.2.14}$$

Comparing this equation with (3.3.22), we see that it has the following physical interpretation: if we start with the systems R, A', A and B in the state $\rho_{RA'}\otimes\Phi^{\mathcal{N}}_{AB}$ and we measure A' and A according to the POVM $\{|\Phi\rangle\langle\Phi|_{A'A},\mathbb{1}_{A'A}-|\Phi\rangle\langle\Phi|_{A'A}\}$, then the outcome corresponding to $|\Phi\rangle\langle\Phi|_{A'A}$ occurs with probability $\frac{1}{d_A^2}$ and the post-measurement state on systems R and B is $\mathcal{N}_{A'\to B}(\rho_{RA'})$. The Choi state $\Phi^{\mathcal{N}}_{AB}$ can thus be viewed as a *resource state* for the probabilistic implementation of the channel $\mathcal{N}$. We return to this point in Section 5.1 when we discuss post-selected quantum teleportation.

The concept of Choi state allows us to associate to each quantum channel $\mathcal{N}_{A\to B}$ a bipartite quantum state. Conversely, given a bipartite state ρ_{AB}, we can associate a map given by

$$X_A\mapsto d_A\mathrm{Tr}_A[(X_A^{\mathsf{T}}\otimes\mathbb{1}_B)\rho_{AB}]. \tag{4.2.15}$$

It is straightforward to see that this map is completely positive; however, it is trace preserving if and only if $\mathrm{Tr}_B[\rho_{AB}]=\pi_A$. On the other hand, the map $\mathcal{N}^{\rho}_{A\to B}$ defined as

$$\mathcal{N}^{\rho}_{A\to B}(X_A)\coloneqq\mathrm{Tr}_A\left[(X_A^{\mathsf{T}}\otimes\mathbb{1}_B)\rho_A^{-\frac{1}{2}}\rho_{AB}\rho_A^{-\frac{1}{2}}\right] \tag{4.2.16}$$

is a quantum channel whenever ρ_A is positive definite, where $\rho_A=\mathrm{Tr}_B[\rho_{AB}]$. The operator $\rho_A^{-\frac{1}{2}}\rho_{AB}\rho_A^{-\frac{1}{2}}$ is sometimes called a "conditional state," motivated by the fact that it reduces to a conditional probability distribution when ρ_{AB} is a fully classical state, so that it can be written as $\rho_{AB}=\sum_{x,y}p(x,y)|x\rangle\langle x|_A\otimes|y\rangle\langle y|_B$ where $p(x,y)$ is a probability distribution. Note that if ρ_A is not invertible, then the inverse in (4.2.16) should be taken on the support of ρ_A, in which case the channel $\mathcal{N}^{\rho}_{A\to B}$ is defined as in (4.2.16) only on the support of ρ_A.

Exercise 4.6

1. Given two superoperators $(\mathcal{N}_1)_{A_1\to B_1}$ and $(\mathcal{N}_2)_{A_2\to B_2}$, prove that the Choi representation of the tensor-product superoperator $\mathcal{N}_1 \otimes \mathcal{N}_2$ is given by

$$\Gamma^{\mathcal{N}_1\otimes\mathcal{N}_2}_{A_1A_2B_1B_2} = \Gamma^{\mathcal{N}_1}_{A_1B_1} \otimes \Gamma^{\mathcal{N}_2}_{A_2B_2}. \tag{4.2.17}$$

2. Given two superoperators $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{B\to C}$, prove that the Choi representation of the composition $(\mathcal{M}\circ\mathcal{N})_{A\to C}$ is given by

$$\begin{aligned}\Gamma^{\mathcal{M}\circ\mathcal{N}}_{AC} &= \mathcal{M}_{B\to C}(\Gamma^{\mathcal{N}}_{AB}) &(4.2.18)\\ &= \mathrm{Tr}_B[\mathrm{T}_B(\Gamma^{\mathcal{N}}_{AB})\Gamma^{\mathcal{M}}_{BC}] &(4.2.19)\\ &= \langle\Gamma|_{BB'}\Gamma^{\mathcal{N}}_{AB}\otimes\Gamma^{\mathcal{M}}_{B'C}|\Gamma\rangle_{BB'}. &(4.2.20)\end{aligned}$$

Exercise 4.7

Let $\mathcal{N}_{A\to B}$ be a superoperator. Prove that

$$\frac{1}{d_A}\left\|\Gamma^{\mathcal{N}}_{AB}\right\|_1 \le \|\mathcal{N}\|_\diamond \le \left\|\Gamma^{\mathcal{N}}_{AB}\right\|_1. \tag{4.2.21}$$

(*Hint*: Start with Theorem 2.21. Then, for the right-most inequality, start with the discussion around (2.2.40) and then use (2.2.93).)

4.3 Characterizations of Channels: Choi, Kraus, Stinespring

The following theorem provides three useful ways to characterize quantum channels, and as such, it is one of the most important theorems in quantum information theory.

Theorem 4.3 Characterizations of Quantum Channels

Let $\mathcal{N}_{A\to B}$ be a linear map from $\mathrm{L}(\mathcal{H}_A)$ to $\mathrm{L}(\mathcal{H}_B)$. Then the following are equivalent:

1. $\mathcal{N}$ is a quantum channel.

2. *Choi*: The Choi representation $\Gamma^{\mathcal{N}}_{AB}$ is positive semi-definite and satisfies $\mathrm{Tr}_B[\Gamma^{\mathcal{N}}_{AB}] = \mathbb{1}_A$.

3. *Kraus*: There exists a set $\{K_i\}_{i=1}^r$ of finitely many operators, called *Kraus operators*, such that

$$\mathcal{N}(X_A) = \sum_{i=1}^{r} K_i X_A K_i^\dagger \tag{4.3.1}$$

for every linear operator X_A, where $K_i \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ for all $i \in \{1, \ldots, r\}$, and $\sum_{i=1}^r K_i^\dagger K_i = \mathbb{1}_A$.

4. *Stinespring*: There exists an isometry $V_{A\to BE}$, called an *isometric extension*, such that

$$\mathcal{N}(X_A) = \mathrm{Tr}_E[V X_A V^\dagger] \tag{4.3.2}$$

for every linear operator X_A.

Please consult the Bibliographic Notes in Section 3.4 for references containing a proof of this theorem.

Remark: The number r of Kraus operators in part 3 of Theorem 4.3 can be chosen to be equal to the rank of the Choi matrix, i.e., $r = \mathrm{rank}(\Gamma^{\mathcal{N}}_{AB})$. Note that this is the minimal number of Kraus operators needed in a Kraus decomposition of the channel. Similarly, the dimension d_E of the environment E appearing in the Stinespring extension in part 4 can be chosen to be equal to $\mathrm{rank}(\Gamma^{\mathcal{N}}_{AB})$, and this is the minimal dimension of an environment E needed in a Stinespring representation of the channel.

Remark: Theorem 4.3 holds for quantum channels, i.e., completely positive trace-preserving maps. More generally, if $\mathcal{N}_{A\to B}$ is completely positive and trace non-increasing, then the trace-preserving condition for the Choi, Kraus, and Stinespring representations of $\mathcal{N}_{A\to B}$ changes as follows.

- The trace-preserving condition $\mathrm{Tr}_B[\Gamma^{\mathcal{N}}_{AB}] = \mathbb{1}_A$ on the Choi representation of $\mathcal{N}_{A\to B}$ changes to $\mathrm{Tr}_B[\Gamma^{\mathcal{N}}_{AB}] \leq \mathbb{1}_A$ when $\mathcal{N}_{A\to B}$ is trace non-increasing.
- The trace-preserving condition $\sum_{i=1}^r K_i^\dagger K_i = \mathbb{1}_A$ on a set of Kraus operators changes to $\sum_{i=1}^r K_i^\dagger K_i \leq \mathbb{1}_A$ when $\mathcal{N}_{A\to B}$ is trace non-increasing.
- The isometric property of the operator V in (4.3.2), which corresponds to the trace-preserving property of every quantum channel, changes to $V^\dagger V \leq \mathbb{1}_A$ when $\mathcal{N}_{A\to B}$ is trace non-increasing.

Completely positive trace-non-increasing maps arise in the context of quantum instruments, which we discuss in Section 4.4.5.

The Kraus operators K_i in (4.3.1) have an interpretation in quantum error correction as "error operators" that characterize various errors that a quantum system undergoes. Kraus operators for a given quantum channel are, however, not unique in general. If $\{K_i\}_{i=1}^r$ is a set of Kraus operators for the channel $\mathcal{N}$, then, given an $s \times r$ isometric matrix V with elements $\{V_{i,j} : 1 \leq i \leq s,\ 1 \leq j \leq r\}$, the operators $\{K_i'\}_{i=1}^s$ defined as

$$K_i' = \sum_{j=1}^{r} V_{i,j} K_j \tag{4.3.3}$$

are also Kraus operators for $\mathcal{N}$. Indeed, for every linear operator X, the following equality holds

$$\begin{aligned}
\sum_{i=1}^{s} K_i' X (K_i')^\dagger &= \sum_{i=1}^{s} \sum_{j,j'=1}^{r} V_{i,j} \overline{V_{i,j'}} K_j X K_{j'}^\dagger, &(4.3.4)\\
&= \sum_{j,j'=1}^{r} \underbrace{\left(\sum_{i=1}^{s} (V^\dagger)_{j',i} V_{i,j} \right)}_{(V^\dagger V)_{j',j} = \delta_{j',j}} K_j X K_{j'}^\dagger, &(4.3.5)\\
&= \sum_{j=1}^{r} K_j X K_j^\dagger &(4.3.6)\\
&= \mathcal{N}(X), &(4.3.7)
\end{aligned}$$

where the second equality follows because $\overline{V_{i,j'}} = (V^\dagger)_{j',i}$.

We note also that a converse statement holds: if $\{K_i\}_{i=1}^r$ and $\{K_i'\}_{i=1}^s$ are two sets of Kraus operators that realize the same quantum channel, then they are related by an isometry as in (4.3.3). This is a dynamical version of the statement made earlier in Section 3.2.5, the statement there being that all purifications of a state are related by an isometry acting on the purifying system.

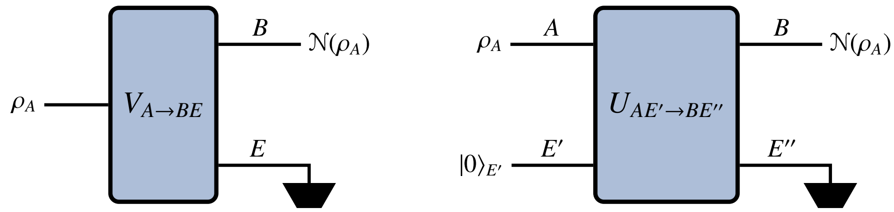


FIGURE 4.3: (Left) According to Stinespring's theorem, the evolution of every quantum system A via a quantum channel $\mathcal{N}_{A\to B}$ can be described as an interaction of the system A with its environment E via an isometry $V_{A\to BE}$, followed by discarding the environment. (Right) The isometry $V_{A\to BE}$ can be extended to a unitary $U_{AE'\to BE''}$ using, e.g., the construction in (4.3.27), such that A and E' are initially in a product state, with E' starting in a pure state. The two systems then interact accoring to U, and after discarding the environment E'', the resulting state is the output of the channel.

Exercise 4.8

Show that the Choi representation of a quantum channel $\mathcal{N}_{A\to B}$ can be expressed using a set $\{K_i\}_{i=1}^r$ of its Kraus operators, with $r \geq \text{rank}(\Gamma_{AB}^{\mathcal{N}})$, as

$$\Gamma_{AB}^{\mathcal{N}} = \sum_{i=1}^{r} \text{vec}(K_i)\text{vec}(K_i)^\dagger. \tag{4.3.8}$$

Physically, the isometric extension $V_{A\to BE}$ in (4.3.2) can be thought of as modelling the interaction of the quantum system of interest with its *environment*, i.e., anything external to the quantum system that is not under our control. Stinespring's theorem then tells us that the evolution of every quantum system can be represented as an interaction of the system with its environment, which is then later discarded; see Figure 4.3. Given a set $\{K_i\}_{i=1}^r$ of Kraus operators for $\mathcal{N}$, we can let the environment E correspond to a space of dimension r and define the isometry $V_{A\to BE}$ as

$$V_{A\to BE} = \sum_{j=1}^{r} K_j \otimes |j-1\rangle_E. \tag{4.3.9}$$

It is straightforward to show that this is indeed an isometric extension of $\mathcal{N}$ since $\text{Tr}_E[VXV^\dagger] = \sum_{j=1}^r K_j X K_j^\dagger = \mathcal{N}(X)$.

Exercise 4.9

Show that the Choi representation of a quantum channel $\mathcal{N}_{A\to B}$ can be expressed using an isometric extension $V_{A\to BE}$ of $\mathcal{N}$, with $d_E \geq \operatorname{rank}(\Gamma^{\mathcal{N}}_{AB})$, as

$$\Gamma^{\mathcal{N}}_{AB} = \operatorname{Tr}_E[\operatorname{vec}(V)\operatorname{vec}(V)^\dagger]. \tag{4.3.10}$$

Hence, conclude that $\frac{1}{\sqrt{d_A}}\operatorname{vec}(V) = (\mathbb{1}_A \otimes V_{A'\to BE})|\Phi\rangle_{AA'}$ is a purification of the Choi state $\Phi^{\mathcal{N}}_{AB}$ of $\mathcal{N}$.

Exercise 4.10

Let $\mathcal{N}_{A\to B}$ be a quantum channel with the following Kraus and Stinespring representations:

$$\mathcal{N}(X) = \sum_{i=1}^{r} K_i X K_i^\dagger = \operatorname{Tr}_E[VXV^\dagger]. \tag{4.3.11}$$

1. Verify using (2.2.167) that the adjoint map $\mathcal{N}^\dagger$ can be represented in the following two ways:

$$\mathcal{N}^\dagger(Y) = \sum_{i=1}^{r} K_i^\dagger Y K_i = V^\dagger(Y \otimes \mathbb{1}_E)V. \tag{4.3.12}$$

2. Using 1., verify the following facts:
   (a) The adjoint of a completely positive map is completely positive.
   (b) The adjoint of a trace preserving map is a unital map (recall Definition 2.19). More generally, the adjoint of a trace non-increasing map is subunital, meaning that $\mathcal{N}^\dagger(\mathbb{1}_B) \leq \mathbb{1}_A$.
   (c) The adjoint of a unital map is trace preserving, and the adjoint of a subunital map is trace non-increasing.

Exercise 4.11

1. Let $\mathcal{N}_{A\to B}$ be a positive trace-preserving map. Prove that the set $\{\mathcal{N}^\dagger(|i\rangle\langle i|)\}_{i=0}^{d_B-1}$ is a POVM. More generally, prove that the set

$\{\mathcal{N}^\dagger(E_j\rho E_j^\dagger)\}_{j=1}^{d_B^2}$ is a POVM for every orthonormal basis $\{E_j\}_{j=1}^{d_B^2}$ for $\mathrm{L}(\mathcal{H}_B)$ and every quantum state $\rho \in \mathrm{D}(\mathcal{H}_B)$.

2. Conversely, let $\{M_x\}_{x\in\mathcal{X}}$ be a POVM, where $\mathcal{X}$ is a finite set. Prove that there exists a quantum channel $\mathcal{N}$ such that $M_x = \mathcal{N}^\dagger(|x\rangle\langle x|)$ for all $x \in \mathcal{X}$, where $\{|x\rangle\}_{x\in\mathcal{X}}$ is an orthonormal set. (*Hint*: Recall Naimark's Theorem (Theorem 3.22).)

4.3.1 Relating Quantum State Extensions and Purifications

We can now establish a fundamental relationship between a purification ψ_{RA} of a state ρ_A and an extension $\omega_{R'A}$ of ρ_A.

Proposition 4.4

Let ρ_A be a quantum state with purification ψ_{RA}. For every extension $\omega_{R'A}$ of ρ_A, there exists a quantum channel $\mathcal{N}_{R\to R'}$ such that

$$\mathcal{N}_{R\to R'}(\psi_{RA}) = \omega_{R'A}. \tag{4.3.13}$$

PROOF: Consider a purification $\phi_{R''R'A}$ of $\omega_{R'A}$, with purifying system R'' satisfying $d_{R''} \geq \mathrm{rank}(\omega_{R'A})$. Since $\omega_{R'A}$ is an extension of ρ_A, we have that

$$\mathrm{Tr}_{R''R'}[\phi_{R''R'A}] = \rho_A, \tag{4.3.14}$$

which means that $\phi_{R''R'A}$ is a purification of ρ_A. The state ψ_{RA} is also a purification of ρ_A, which means that, by the isometric equivalence of purifications (see (3.2.70)–(3.2.73) and the paragraph thereafter), there exists an isometry $V_{R\to R''R'}$ such that

$$V_{R\to R''R'}|\psi\rangle_{RA} = |\phi\rangle_{R''R'A}. \tag{4.3.15}$$

Now, let us use this isometry to define the channel $\mathcal{N}_{R\to R'}$:

$$\mathcal{N}_{R\to R'}(\cdot) = \mathrm{Tr}_{R''}[V_{R\to R''R'}(\cdot)V_{R\to R''R'}^\dagger]. \tag{4.3.16}$$

It then follows that

$$\mathcal{N}_{R\to R'}(\psi_{RA}) = \mathrm{Tr}_{R''}[V_{R\to R''R'}\psi_{RA}V_{R\to R''R'}^\dagger] = \mathrm{Tr}_{R''}[\phi_{R''R'A}] = \omega_{R'A}, \tag{4.3.17}$$

as required. ■

Proposition 4.4 tells us that every extension of a quantum state can be "reached" via a quantum channel acting on a purification of the state. In this sense, a purification can be viewed as the "strongest" extension of a state.

4.3.2 Complementary Channels

As stated earlier, the Stinespring representation $\mathcal{N}_{A\to B}(X_A) = \mathrm{Tr}_E[VXV^\dagger]$ of a quantum channel $\mathcal{N}_{A\to B}$, where $V_{A\to BE}$ is an isometry, can be interpreted as an interaction of the quantum system A of interest with its environment E followed by discarding E. If instead we discard the system B of the isometric channel $\mathcal{V}_{A\to BE}(X_A) = VXV^\dagger$, then we obtain the state of the environment after the interaction with A. This defines a channel complementary to $\mathcal{N}_{A\to B}$.

Definition 4.5 Complementary Channel

Let $V_{A\to BE}$ be an isometry and $\mathcal{N}_{A\to B}$ a quantum channel defined as

$$\mathcal{N}_{A\to B}(X_A) = \mathrm{Tr}_E[VXV^\dagger]. \tag{4.3.18}$$

The *complementary channel* for $\mathcal{N}_{A\to B}$ associated with the isometric extension $V_{A\to BE}$ is denoted by $\mathcal{N}^c_{A\to E}$ and is defined as

$$\mathcal{N}^c_{A\to E}(X_A) \coloneqq \mathrm{Tr}_B[VXV^\dagger]. \tag{4.3.19}$$

Related to the above, the channel $\mathcal{M}^c_{A\to E}$ is a complementary channel for the channel $\mathcal{M}_{A\to B}$ if there exists an isometric channel $\mathcal{W}_{A\to BE}$ such that

$$\mathcal{M}_{A\to B} = \mathrm{Tr}_E \circ \mathcal{W}_{A\to BE} \tag{4.3.20}$$

$$\mathcal{M}^c_{A\to E} = \mathrm{Tr}_B \circ \mathcal{W}_{A\to BE}. \tag{4.3.21}$$

A quantum channel does not have, in general, a uniquely defined complementary channel, just as it does not have a unique Kraus representation; analogously, a quantum state does not have a unique purification. Similar to the latter scenarios, however, it is possible to show that all complementary channels for $\mathcal{N}_{A\to B}$ are related by an isometric channel acting on their output. That is, let us suppose that $(\mathcal{N}_1)^c_{A\to E}$ and $(\mathcal{N}_2)^c_{A\to E'}$ are complementary channels for $\mathcal{N}_{A\to B}$. Then there exists an isometric channel $\mathcal{S}_{E\to E'}$ such that

$$(\mathcal{N}_2)^c_{A\to E'} = \mathcal{S}_{E\to E'} \circ (\mathcal{N}_1)^c_{A\to E}. \tag{4.3.22}$$

Exercise 4.12

Let $\mathcal{N}_{A\to B}$ be a quantum channel with a set $\{K_i\}_{i=1}^r$ of Kraus operators, where $r \geq \operatorname{rank}(\Gamma^{\mathcal{N}}_{AB})$.

1. Using (4.3.9), show that a channel complementary to $\mathcal{N}$ is given by

$$\mathcal{N}^c(X) = \sum_{i,i'=1}^{r} \operatorname{Tr}[K_i X K_{i'}^\dagger] |i-1\rangle\langle i'-1|_E \tag{4.3.23}$$

   for all $X \in \mathrm{L}(\mathcal{H}_A)$.

2. Let $W := \sum_{i=1}^r K_i^\dagger \otimes |i-1\rangle$. Show that the Choi representation of the complementary channel in (4.3.23) is

$$\Gamma^{\mathcal{N}^c}_{AE} = (WW^\dagger)^{\mathsf{T}}. \tag{4.3.24}$$

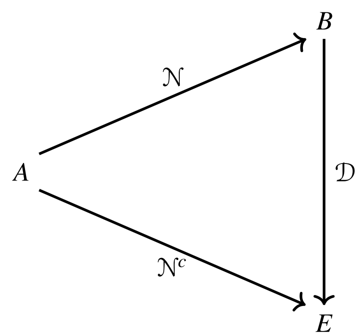
The notion of complementary channel arises in the scenario in which two parties, Alice and Bob, wish to communicate to each other using a channel $\mathcal{N}_{A\to B}$ in the presence of an eavesdropper Eve. In this scenario, we can naturally identify Alice and Bob with the quantum systems A and B and Eve with the system E, where E is as given in Definition 4.5. Any signals sent through the quantum channel by Alice are received by Bob via the action of $\mathcal{N}_{A\to B}$, while Eve receives a signal via the action of $\mathcal{N}^c_{A\to E}$. These concepts are important for private communication over quantum channels, which is the topic of Chapter 16. Two important classes of channels are relevant in this context.

Definition 4.6 Degradable and Anti-Degradable Channels

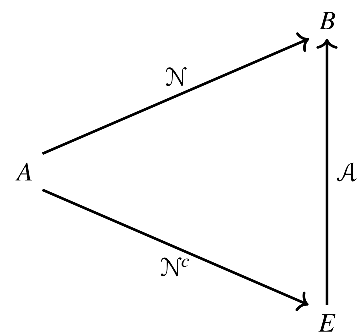
A channel $\mathcal{N}_{A\to B}$ is called *degradable* if there exists a channel $\mathcal{D}_{B\to E}$, called a *degrading channel*, such that

$$\mathcal{D}_{B\to E} \circ \mathcal{N}_{A\to B} = \mathcal{N}^c_{A\to E}, \tag{4.3.25}$$

where $\mathcal{N}^c_{A\to E}$ is a complementary channel of $\mathcal{N}_{A\to B}$. The channel $\mathcal{N}_{A\to B}$ is called *anti-degradable* if a complementary channel $\mathcal{N}^c_{A\to E}$ of it is degradable,



(a) $\mathcal{N}$ degradable.



(b) $\mathcal{N}$ anti-degradable.

FIGURE 4.4: Degradable and anti-degradable channels. In (a), the channel $\mathcal{N}$ is degradable, meaning that there exists a channel $\mathcal{D}$ that Bob can apply to the output he receives via $\mathcal{N}$ that can be used to simulate what Eve receives via $\mathcal{N}^c$. In (b) on the other hand, $\mathcal{N}$ is anti-degradable since Eve can simulate, using the channel $\mathcal{E}$, what Bob receives.

i.e., if there exists a channel $\mathcal{A}_{E\to B}$, called an *anti-degrading channel*, such that

$$\mathcal{A}_{E\to B} \circ \mathcal{N}^c_{A\to E} = \mathcal{N}_{A\to B}. \tag{4.3.26}$$

See Figure 4.4 for a schematic depiction of degradable and anti-degradable channels. A degradable channel is one whose complement can be simulated (via $\mathcal{D}$) using the output of $\mathcal{N}$. This means that Bob can simulate Eve's received signal. On the other hand, an anti-degradable channel is such that $\mathcal{N}$ can be simulated (via $\mathcal{A}$) using the output of $\mathcal{N}^c$, which means that Eve can simulate Bob's received signal. In the above precise sense, therefore, the output of a degradable channel is "less noisy" than the environment with which it has interacted; that is, it contains more quantum information. For an anti-degradable channel, instead, the opposite is true.

Exercise 4.13

Prove that a quantum channel $\mathcal{N}_{A\to B}$ is anti-degradable if and only if its Choi state $\Phi^{\mathcal{N}}_{AB}$ is *two-extendible*, meaning that there exists a state $\sigma_{ABB'}$, with $\mathcal{H}_B$ and $\mathcal{H}_{B'}$ isomorphic (and thus $d_{B'} = d_B$), such that $\mathrm{Tr}_{B'}[\sigma_{ABB'}] = \mathrm{Tr}_B[\sigma_{ABB'}] = \Phi^{\mathcal{N}}_{AB}$. (*Hint*: Use Proposition 4.4.)

4.3.3 Unitary Extensions of Quantum Channels from Isometric Extensions

We can always extend every isometric extension $V_{A\to BE}$ of a channel $\mathcal{N}_{A\to B}$ to a unitary $U_{AE'\to BE''}$ in a way similar to what we used in (2.2.152) in the proof of the operator Jensen inequality (Theorem 2.16). Let the unitary $U_{AE'\to BE''}$ be defined as the following block matrix:

$$U = \begin{pmatrix} V & \mathbb{1} - VV^\dagger & 0_{d_B d_E \times d'} \\ 0_{d_A\times d_A} & V^\dagger & 0_{d_A\times d'} \\ 0_{d'\times d_A} & 0_{d'\times d_B d_E} & \mathbb{1}_{d'} \end{pmatrix}, \tag{4.3.27}$$

where we set $d' := (d_B - 1)d_A$. Without the various dimensions indicated, U is more simply expressed as

$$U = \begin{pmatrix} V & \mathbb{1} - VV^\dagger & 0 \\ 0 & V^\dagger & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix}. \tag{4.3.28}$$

Note that V is a $d_B d_E \times d_A$ matrix and $VV^\dagger$ is a $d_B d_E \times d_B d_E$ matrix. Furthermore, let us suppose that $d_E = d_A d_B$. Note that, by the Stinespring theorem, it is always possible to pick this as the dimension of $\mathcal{H}_E$ since $0 < \text{rank}(\Gamma^{\mathcal{N}}_{AB}) \leq d_A d_B$. Then

$$d_B d_E + d_A + d' = d_A d_B(d_B + 1), \tag{4.3.29}$$

and we conclude that U is a $(d_A d_B(d_B + 1)) \times (d_A d_B(d_B + 1))$ matrix. It is also indeed a unitary because

$$UU^\dagger = \begin{pmatrix} V & \mathbb{1} - VV^\dagger & 0 \\ 0 & V^\dagger & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix}\begin{pmatrix} V^\dagger & 0 & 0 \\ \mathbb{1} - VV^\dagger & V & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix} \tag{4.3.30}$$

$$= \begin{pmatrix} VV^\dagger + (\mathbb{1} - VV^\dagger)(\mathbb{1} - VV^\dagger) & (\mathbb{1} - VV^\dagger)V & 0 \\ V^\dagger(\mathbb{1} - VV^\dagger) & V^\dagger V & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix} \tag{4.3.31}$$

$$= \begin{pmatrix} \mathbb{1} & 0 & 0 \\ 0 & \mathbb{1} & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix}. \tag{4.3.32}$$

Since U is a square matrix, this already implies that $U^{-1} = U^\dagger$; i.e., U is unitary. By defining the system E' with dimension $d_{E'} = d_B(d_B + 1)$, we can think of U as acting on the input tensor-product space $\mathcal{H}_{E'} \otimes \mathcal{H}_A$. Then, we can embed the state ρ_A into this larger space as

$$|0\rangle\langle 0|_{E'} \otimes \rho_A = \begin{pmatrix} \rho_A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.3.33}$$

so that

$$U \begin{pmatrix} \rho_A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} U^\dagger = \begin{pmatrix} V\rho V^\dagger & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.3.34}$$

By defining the system E'' with dimension $d_{E''} = d_A(d_B + 1)$, we can think of the output space of U as the tensor-product space $\mathcal{H}_B \otimes \mathcal{H}_{E''}$, so that

$$\mathcal{N}(\rho_A) = \mathrm{Tr}_E[V\rho V] = \mathrm{Tr}_{E''}[U(|0\rangle\langle 0|_{E'} \otimes \rho_A)U^\dagger]. \tag{4.3.35}$$

The above construction is not necessarily an efficient construction (using as few extra degrees of freedom as possible), but it illustrates the principle that every quantum channel can be thought of as arising from

1. adjoining an environment state $|0\rangle\langle 0|_{E'}$ to the input system A,

2. performing a unitary interaction U, and then

3. tracing over an output environment system E'',

thus assigning a strong physical meaning to the notion of isometric extension of a quantum channel; see Figure 4.3.

Another construction of a unitary operator that is less explicit but more efficient is as follows. Let $V_{A\to BE}$ be the isometric extension from (4.3.9), and let $r = \mathrm{rank}(\Gamma^{\mathcal{N}}_{AB})$. Since $r \leq d_A d_B$, without loss of generality, we can take the output environment system E to have dimension $d_E = d_A d_B$. Let $\{|k\rangle_A \otimes |\ell\rangle_{E'}\}_{k,\ell}$ be an orthonormal basis for the input system A and an input environment system E', with $d_{E'} = d_B^2$, where $0 \leq k \leq d_A - 1$ and $0 \leq \ell \leq d_{E'} - 1$. Then define the orthonormal vectors $|\phi_{k,0}\rangle_{BE}$ for $0 \leq k \leq d_A - 1$ as follows:

$$\begin{aligned} |\phi_{k,0}\rangle_{BE} &:= \left(\sum_{j=1}^{d_E} K_j \otimes |j-1\rangle_E \langle 0|_{E'} \right) |k\rangle_A \otimes |0\rangle_{E'} && (4.3.36) \\ &= V_{A\to BE}|k\rangle_A. && (4.3.37) \end{aligned}$$

The fact that these vectors form an orthonormal set is a consequence of the facts that $V^\dagger V = \mathbb{1}_A$ and $\{|k\rangle_A\}_{k=0}^{d_A-1}$ is an orthonormal set. We then define the action of the unitary $U_{AE'\to BE}$ on these vectors $|k\rangle_A \otimes |0\rangle_{E'}$ as follows:

$$U_{AE'\to BE}|k\rangle_A \otimes |0\rangle_{E'} = |\phi_{k,0}\rangle_{BE}, \tag{4.3.38}$$

for $0 \le k \le d_A - 1$. The set of orthonormal vectors $\{|\phi_{k,0}\rangle_{BE}\}_{k=0}^{d_A-1}$ can be completed to an orthonormal basis of the output space by adding suitably chosen $d_B d_E - d_A = (d_B^2 - 1)d_A$ vectors. Let us denote the basis of the Hilbert space of BE obtained in this way as $\{|\phi_{k,\ell}\rangle_{BE}\}_{k,\ell}$, where $0 \le k \le d_A - 1$ and $0 \le \ell \le d_B^2 - 1$. By employing the vectors in the set $\{|\phi_{k,\ell}\rangle_{BE}\}_{k,\ell}$, we can now extend the action of the unitary $U_{AE'\to BE}$ on the remaining input space vectors as follows:

$$U_{AE'\to BE}|k\rangle_A \otimes |\ell\rangle_{E'} = |\phi_{k,\ell}\rangle_{BE}, \tag{4.3.39}$$

for $0 \le k \le d_A - 1$ and $1 \le \ell \le d_B^2 - 1$. Thus, the full unitary $U_{AE'\to BE}$ is specified as

$$U_{AE'\to BE} := \sum_{k=0}^{d_A-1} \sum_{\ell=0}^{d_B^2-1} |\phi_{k,\ell}\rangle_{BE}\langle k|_A \otimes \langle \ell|_{E'}. \tag{4.3.40}$$

By construction, the following identity holds for every input state ρ_A:

$$U_{AE'\to BE}\,(\rho_A \otimes |0\rangle\langle 0|_{E'})\,(U_{AE'\to BE})^\dagger = V\rho_A V^\dagger, \tag{4.3.41}$$

so that we can realize the isometric channel $\rho_A \mapsto V\rho_A V^\dagger$ by tensoring in an environment state $|0\rangle\langle 0|_{E'}$ and applying the unitary $U_{AE'\to BE}$. We then realize the original channel $\mathcal{N}_{A\to B}$ by applying a final partial trace over the environment system E:

$$\mathcal{N}_{A\to B}(\rho_A) = \text{Tr}_E[U_{AE'\to BE}\,(\rho_A \otimes |0\rangle\langle 0|_{E'})\,(U_{AE'\to BE})^\dagger]. \tag{4.3.42}$$

4.4 General Types of Channels

4.4.1 Preparation, Appending, and Replacement Channels

The preparation of a quantum system in a given (fixed) state, as well as taking the tensor product of a state with a given (fixed) state, can both be viewed as quantum channels.

Definition 4.7 Preparation and Appending Channels

For a quantum system A and a state ρ_A, the *preparation channel* $\mathcal{P}_{\rho_A}$ is defined for $\alpha \in \mathbb{C}$ as

$$\mathcal{P}_{\rho_A}(\alpha) \coloneqq \alpha\rho_A. \tag{4.4.1}$$

When acting in parallel with the identity channel on a linear operator Y_B of another quantum system B, the preparation channel $\mathcal{P}_{\rho_A}$ is called the *appending channel* $\mathcal{P}_{\rho_A}$ and is defined as

$$\mathcal{P}_{\rho_A}(Y_B) \equiv (\mathcal{P}_{\rho_A} \otimes \mathrm{id}_B)(Y_B) = \rho_A \otimes Y_B. \tag{4.4.2}$$

In other words, the appending channel $\mathcal{P}_{\rho_A} \otimes \mathrm{id}_B$ takes the tensor product of its argument with ρ_A.

One way to determine whether a map $\mathcal{N}$ on quantum states is completely positive is to find a Kraus representation for it, i.e., a set $\{K_i\}_i$ of operators such that $\mathcal{N}(X) = \sum_i K_i X K_i^\dagger$ for every operator X; if that is possible, then $\mathcal{N}$ is completely positive by Theorem 4.3. If in addition the Kraus operators satisfy $\sum_i K_i^\dagger K_i = \mathbb{1}$, then $\mathcal{N}$ is also trace preserving and therefore a quantum channel.

Supposing that ρ_A has a spectral decomposition of the form

$$\rho_A = \sum_k \lambda_k |\phi_k\rangle\langle\phi_k|_A, \tag{4.4.3}$$

one set of Kraus operators for the preparation channel $\mathcal{P}_{\rho_A}$ is $\{\sqrt{\lambda_k}|\phi_k\rangle_A\}_k$. A set of Kraus operators for the appending channel $\mathcal{P}_{\rho_A} \otimes \mathrm{id}_B$ is then $\{\sqrt{\lambda_k}|\phi_k\rangle \otimes \mathbb{1}_B\}_k$.

Exercise 4.14

Determine the Choi representation and a Stinespring representation of the preparation channel $\mathcal{P}_{\rho_A}$ corresponding to the quantum state ρ_A.

Definition 4.8 Replacement Channel

For a state σ_B, the *replacement channel* $\mathcal{R}^{\sigma_B}_{A\to B}$ is defined as the channel that traces out its input and replaces it with the state σ_B; i.e.,

$$\mathcal{R}^{\sigma_B}_{A\to B}(X_A) = \mathrm{Tr}[X_A]\sigma_B \tag{4.4.4}$$

for every linear operator X_A. When acting on one share of a bipartite state ρ_{RA}, the replacement channel $\mathcal{R}^{\sigma_B}_{A\to B}$ has the following action:

$$\mathcal{R}^{\sigma_B}_{A\to B}(\rho_{RA}) = \mathrm{Tr}_A[\rho_{RA}] \otimes \sigma_B = \rho_R \otimes \sigma_B. \tag{4.4.5}$$

Observe that we can write the replacement channel $\mathcal{R}^{\sigma_B}_{A\to B}$ as the composition of the partial trace over A followed by the preparation/appending channel $\mathcal{P}_{\sigma_B}$:

$$\mathcal{R}^{\sigma_B}_{A\to B} = \mathcal{P}_{\sigma_B} \circ \mathrm{Tr}_A. \tag{4.4.6}$$

We often omit the superscript on $\mathcal{R}^{\sigma_B}_{A\to B}$ when it is clear from the context that the replacement state is σ_B.

Exercise 4.15

Given a quantum state σ_B, determine the Choi representation, as well as Kraus and Stinespring representations, of the replacement channel $\mathcal{R}^{\sigma_B}_{A\to B}$.

4.4.2 Trace and Partial-Trace Channels

Recall the trace and partial trace of a linear operator from Definition 3.2. As a map acting on linear operators, one can ask whether the partial trace is a channel. The answer, perhaps not surprisingly, is "yes." In fact, observe that the definition in (3.2.17) of the partial trace Tr_B over B is already in Kraus form, with Kraus operators $K_j = \mathbb{1}_A \otimes \langle j|_B$. This means that Tr_B is completely positive. It is also trace preserving because

$$\sum_{j=0}^{d_B-1} K_j^\dagger K_j = \sum_{j=0}^{d_B-1} (\mathbb{1}_A \otimes |j\rangle_B)(\mathbb{1}_A \otimes \langle j|_B) = \mathbb{1}_A \otimes \sum_{j=0}^{d_B-1} |j\rangle\langle j|_B = \mathbb{1}_{AB}, \tag{4.4.7}$$

where we used the fact that $\sum_{j=0}^{d_B-1} |j\rangle\langle j|_B = \mathbb{1}_B$.

Exercise 4.16

1. Determine the Choi representation, as well as a Stinespring representation, of the partial trace channel Tr_B.

2. Prove that the adjoint of the partial trace channel Tr_B is

$$\mathrm{Tr}_B^\dagger(X_A) = X_A \otimes \mathbb{1}_B. \tag{4.4.8}$$

Unlike the trace and partial trace, the transpose and partial transpose are trace-preserving maps but *not* completely positive. Indeed, for the latter, recall from (3.2.82) that $\mathrm{T}_B(\Phi_{AB}) = \frac{1}{d}F_{AB}$, so that its Choi representation is $\Gamma_{AB}^{\mathrm{T}_B} = F_{AB}$, which we know has negative eigenvalues, as shown in (3.2.126). So, by Theorem 4.3, the transpose map T_B is not completely positive.

4.4.3 Isometric and Unitary Channels

Two more simple examples of quantum channels are *isometric* and *unitary channels*. An isometric channel conjugates the channel input by an isometry, and a unitary channel conjugates the channel input by a unitary. Specifically, the isometric channel $\mathcal{V}$ corresponding to an isometry V is

$$\mathcal{V}(X) \coloneqq VXV^\dagger. \tag{4.4.9}$$

Similarly, the unitary channel $\mathcal{U}$ corresponding to a unitary U is

$$\mathcal{U}(X) \coloneqq UXU^\dagger. \tag{4.4.10}$$

Since every unitary is also an isometry, it follows that every unitary channel is an isometric channel. Isometric channels are completely positive because they can be described using only one Kraus operator, the isometry V. In fact, a quantum channel is isometric if and only if it has a single Kraus operator.

Observe that by the unitarity of U, the map

$$\mathcal{U}^\dagger(Y) \coloneqq U^\dagger YU, \tag{4.4.11}$$

i.e., the conjugation by $U^\dagger$, is also a channel. In particular,

$$\mathcal{U}^\dagger \circ \mathcal{U} = \mathcal{U} \circ \mathcal{U}^\dagger = \mathrm{id}, \tag{4.4.12}$$

so that $\mathcal{U}^\dagger$ is the inverse channel of $\mathcal{U}$.

On the other hand, for an isometry V, conjugation by $V^\dagger$ is *not* necessarily a channel: although the map $\mathcal{V}^\dagger(Y) := V^\dagger Y V$ is completely positive, it is not necessarily trace preserving because $VV^\dagger \neq \mathbb{1}$ in general. However, by defining the reversal channel $\mathcal{R}_V$ as

$$\mathcal{R}_V(Y) := \mathcal{V}^\dagger(Y) + \mathrm{Tr}[(\mathbb{1} - VV^\dagger)Y]\sigma, \tag{4.4.13}$$

where σ is an arbitrary (but fixed) state, we find that $\mathcal{R}_V$ is trace preserving:

$$\begin{aligned} \mathrm{Tr}[\mathcal{R}_V(Y)] &= \mathrm{Tr}\left[\mathcal{V}^\dagger(Y) + \mathrm{Tr}[(\mathbb{1} - VV^\dagger)Y]\sigma\right] && (4.4.14)\\ &= \mathrm{Tr}[V^\dagger Y V] + \mathrm{Tr}[Y] - \mathrm{Tr}[VV^\dagger Y] && (4.4.15)\\ &= \mathrm{Tr}[Y]. && (4.4.16) \end{aligned}$$

Since it is also completely positive, being the sum of completely positive maps, the map $\mathcal{R}_V$ is indeed a quantum channel. Like $\mathcal{U}^\dagger$, the reversal channel $\mathcal{R}_V$ reverses the action of $\mathcal{V}$:

$$\begin{aligned} (\mathcal{R}_V \circ \mathcal{V})(X) &= \mathcal{V}^\dagger(\mathcal{V}(X)) + \mathrm{Tr}[(\mathbb{1} - VV^\dagger)\mathcal{V}(X)]\sigma && (4.4.17)\\ &= V^\dagger V X V^\dagger V + \mathrm{Tr}[(\mathbb{1} - VV^\dagger)VXV^\dagger]\sigma && (4.4.18)\\ &= X + (\mathrm{Tr}[VXV^\dagger] - \mathrm{Tr}[VV^\dagger VXV^\dagger])\sigma && (4.4.19)\\ &= X. && (4.4.20) \end{aligned}$$

In the above sense, $\mathcal{R}_V$ is a left inverse of $\mathcal{V}$. Unlike $\mathcal{U}^\dagger$, however, $\mathcal{R}_V$ is not the right inverse of $\mathcal{V}$ because the equality $(\mathcal{V} \circ \mathcal{R}_V)(Y) = Y$ need not hold.

Exercise 4.17

Determine the Choi representation, as well as a Kraus and Stinespring representation, of the reversal channel $\mathcal{R}_V$ corresponding to an isometry V.

4.4.4 Classical–Quantum and Quantum–Classical Channels

Any classical probability distribution $p : \mathcal{X} \to [0, 1]$ over a finite alphabet $\mathcal{X}$ can be represented as a quantum state of a $|\mathcal{X}|$-dimensional system X that is diagonal in an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of X. Specifically, the probability distribution can be represented as the state

$$\rho_X = \sum_{x \in \mathcal{X}} p(x)|x\rangle\langle x|_X. \tag{4.4.21}$$

States that are diagonal in a preferred basis, as in the equation above, are typically called *classical states*. In addition to being in one-to-one correspondence with classical probability distributions, classical states do not exhibit the quantum properties of coherence and entanglement.

In Chapter 12, however, we are interested in classical communication over quantum channels. We are then interested in so-called *classical–quantum channels*, which we discuss in this section, that take a classical state as input and output a quantum state.

Definition 4.9 Classical–Quantum Channel

A *classical–quantum channel* is a map from a system X with alphabet $\mathcal{X}$ and orthonormal basis $\{|x\rangle : x \in \mathcal{X}\}$ to a quantum system A with a specified set $\{\sigma_A^x : x \in \mathcal{X}\}$ of states such that

$$|x\rangle\langle x'| \mapsto \delta_{x,x'}\sigma_A^x \quad \forall\, x, x' \in \mathcal{X}. \tag{4.4.22}$$

If $\mathcal{N}^{\text{cq}}$ is a classical–quantum channel, then for every classical state $\rho_X = \sum_{x \in \mathcal{X}} p(x)|x\rangle\langle x|$, we have that

$$\mathcal{N}^{\text{cq}}(\rho_X) = \sum_{x \in \mathcal{X}} p(x)\sigma_A^x. \tag{4.4.23}$$

More generally, for every $|\mathcal{X}|$-dimensional system A and every state ρ_A that is not necessarily classical and is expressed in the computational basis as $\rho_A = \sum_{x,x' \in \mathcal{X}} \langle x|\rho_A|x'\rangle |x\rangle\langle x'|$, we find that

$$\mathcal{N}^{\text{cq}}(\rho_A) = \sum_{x,x' \in \mathcal{X}} \langle x|\rho_A|x'\rangle \delta_{x,x'}\sigma_A^x = \sum_{x \in \mathcal{X}} \langle x|\rho_A|x\rangle \sigma_A^x. \tag{4.4.24}$$

Therefore, for every quantum state, the classical–quantum channel $\mathcal{N}^{\text{cq}}$ takes the input state, measures it in the computational basis $\{|x\rangle\}_{x \in \mathcal{X}}$, and outputs the state σ_A^x corresponding to the outcome $x \in \mathcal{X}$.

Exercise 4.18

Show that the Choi state of a classical–quantum channel $\mathcal{N}^{\mathrm{cq}}$ is

$$\Phi_{XA}^{\mathcal{N}^{\mathrm{cq}}} = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \sigma_A^x. \tag{4.4.25}$$

In other words, the Choi state of a classical–quantum channel is a classical–quantum state.

If the states σ_A^x have spectral decomposition

$$\sigma_A^x = \sum_{j=1}^{r_x} \lambda_j^x |\varphi_j^x\rangle\langle\varphi_j^x|, \tag{4.4.26}$$

where $r_x = \mathrm{rank}(\sigma_A^x)$, then $\{K_j^x : x \in \mathcal{X},\ 1 \le j \le r_x\}$ is a set of Kraus operators for $\mathcal{N}^{\mathrm{cq}}$, where $K_j^x = \sqrt{\lambda_j^x}|\varphi_j^x\rangle_A\langle x|_X$. Indeed, for all $x, x' \in \mathcal{X}$,

$$\begin{aligned}
&\sum_{x''\in\mathcal{X}}\sum_{j=1}^{r_{x''}} K_j^{x''}|x\rangle\langle x'|(K_j^{x''})^\dagger \\
&= \sum_{x''\in\mathcal{X}}\sum_{j=1}^{r_{x''}} \sqrt{\lambda_j^{x''}}|\varphi_j^{x''}\rangle\langle x''|x\rangle\langle x'|x''\rangle\langle\varphi_j^{x''}|\sqrt{\lambda_j^{x''}} && (4.4.27)\\
&= \delta_{x,x'}\sum_{j=1}^{r_x} \lambda_k^x|\varphi_j^x\rangle\langle\varphi_j^x| && (4.4.28)\\
&= \delta_{x,x'}\sigma_A^x && (4.4.29)\\
&= \mathcal{N}^{\mathrm{cq}}(|x\rangle\langle x'|), && (4.4.30)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x}(K_j^x)^\dagger K_j^x &= \sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x} \lambda_j^x|x\rangle\underbrace{\langle\varphi_j^x|\varphi_j^x\rangle}_{=1}\langle x| && (4.4.31)\\
&= \sum_{x\in\mathcal{X}}\underbrace{\sum_{j=1}^{r_x} \lambda_j^x}_{=1\ \forall\ x}|x\rangle\langle x| && (4.4.32)
\end{aligned}$$

$$= \mathbb{1}_X. \tag{4.4.33}$$

Also, observe from the construction above that every classical–quantum channel has a Kraus representation with unit-rank Kraus operators.

Having described classical–quantum channels, let us now describe channels for which the situation is opposite, such that they accept quantum inputs and provide classical outputs.

Definition 4.10 Quantum–Classical Channel

Given a quantum system A and a measurement on A with corresponding POVM $\{M_x\}_{x\in\mathcal{X}}$ indexed by elements of a finite alphabet $\mathcal{X}$, a *quantum–classical channel*, or *measurement channel*, is a map $\mathcal{M}_{A\to X}$ from the quantum system A to a classical system X with alphabet $\mathcal{X}$ such that

$$\mathcal{M}_{A\to X}(\rho_A) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho_A]|x\rangle\langle x|_X \tag{4.4.34}$$

for every state ρ_A on A, where $\{|x\rangle : x \in \mathcal{X}\}$ is an orthonormal basis for X.

A measurement channel thus takes the measurement outcome probabilities $\mathrm{Tr}[M_x\rho_A]$ and arranges them into a classical state.

Exercise 4.19

Prove that the Choi state of a measurement channel $\mathcal{M}_{A\to X}$, as defined in (4.4.34), is

$$\Phi^{\mathcal{M}}_{AX} = \frac{1}{d_A}\sum_{x\in\mathcal{X}}(M_x^{\mathsf{T}})_A \otimes |x\rangle\langle x|_X. \tag{4.4.35}$$

By writing a spectral decomposition of M_x as

$$M_x = \sum_{j=1}^{r_x} \mu_j^x |\phi_j^x\rangle\langle\phi_j^x|, \tag{4.4.36}$$

where $r_x = \mathrm{rank}(M_x)$, we can write

$$\mathrm{Tr}[M_x\rho] = \sum_{j=1}^{r_x} \mu_j^x \langle\phi_j^x|\rho|\phi_j^x\rangle. \tag{4.4.37}$$

Therefore, the action of a quantum–classical channel $\mathcal{N}^{\text{qc}}$ can be written as

$$\mathcal{N}^{\text{qc}}(\rho) = \sum_{x\in\mathcal{X}} \text{Tr}[M_x\rho]|x\rangle\langle x| \tag{4.4.38}$$

$$= \sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x} \mu_j^x\langle\phi_j^x|\rho|\phi_j^x\rangle|x\rangle\langle x| \tag{4.4.39}$$

$$= \sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x} \sqrt{\mu_j^x}|x\rangle\langle\phi_j^x|\rho|\phi_j^x\rangle\langle x|\sqrt{\mu_j^x} \tag{4.4.40}$$

$$= \sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x} K_j^x\rho(K_j^x)^\dagger, \tag{4.4.41}$$

where

$$K_j^x \coloneqq \sqrt{\mu_j^x}|x\rangle\langle\phi_j^x| \quad \forall\, x\in\mathcal{X},\ 1\le j\le r_x. \tag{4.4.42}$$

Since

$$\sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x}(K_j^x)^\dagger K_j^x = \sum_{x\in\mathcal{X}}\sum_{j=1}^{r_x} \mu_j^x|\phi_j^x\rangle\langle x|x\rangle\langle\phi_j^x| \tag{4.4.43}$$

$$= \sum_{x\in\mathcal{X}}\underbrace{\sum_{j=1}^{r_x} \mu_j^x|\phi_j^x\rangle\langle\phi_j^x|}_{M_x} \tag{4.4.44}$$

$$= \sum_{x\in\mathcal{X}} M_x \tag{4.4.45}$$

$$= \mathbb{1}_A, \tag{4.4.46}$$

the operators K_j^x, where $x\in\mathcal{X}$ and $1\le j\le r_x$, form a set of Kraus operators for $\mathcal{N}^{\text{qc}}$. This means that all quantum–classical channels have a Kraus representation with unit-rank Kraus operators.

4.4.5 Quantum Instruments

Recall from Section 3.3 that, for an arbitrary POVM $\{M_x\}_{x\in\mathcal{X}}$, a set of post-measurement states corresponding to an initial state ρ can be given by (3.3.14),

$$\rho^x = \frac{K_x\rho K_x^\dagger}{\text{Tr}[K_x\rho K_x^\dagger]} \quad \forall\, x\in\mathcal{X}, \tag{4.4.47}$$

where $\{K_x\}_{x\in\mathcal{X}}$ is a set of operators such that $M_x = K_x^\dagger K_x$ for all $x \in \mathcal{X}$. Also recall that the expected density operator of the measurement is given by (3.3.15),

$$\sum_{x\in\mathcal{X}} K_x \rho K_x^\dagger. \tag{4.4.48}$$

This expected state can be seen as arising from a quantum channel with Kraus operators $\{K_x\}_{x\in\mathcal{X}}$. Note that this map is indeed a channel since $\sum_{x\in\mathcal{X}} K_x^\dagger K_x = \sum_{x\in\mathcal{X}} M_x = \mathbb{1}$. We can view the channel as being the sum of the completely positive and trace-non-increasing maps $\mathcal{M}_x$ defined as

$$\mathcal{M}_x(\rho) = K_x \rho K_x^\dagger. \tag{4.4.49}$$

A quantum instrument generalizes this notion to maps $\mathcal{M}_x$ that are completely positive and trace non-increasing with an arbitrary number of Kraus operators, not just one.

Definition 4.11 Quantum Instrument

A *quantum instrument* is a collection $\{\mathcal{M}_x\}_{x\in\mathcal{X}}$ of completely positive maps indexed by elements of a finite alphabet $\mathcal{X}$, such that the sum map $\sum_{x\in\mathcal{X}} \mathcal{M}_x$ is a quantum channel.

REMARK: In order for the sum map $\sum_{x\in\mathcal{X}} \mathcal{M}_x$ to be a quantum channel, and hence trace preserving, each $\mathcal{M}_x$ needs to be a trace-non-increasing map.

A quantum instrument corresponds to a more general notion of a measurement in which, as before, the elements of the alphabet $\mathcal{X}$ label the measurement outcomes. If the initial state of the system is ρ, then the corresponding measurement outcome probabilities $p(x)$ are given by

$$p(x) = \mathrm{Tr}[\mathcal{M}_x(\rho)], \tag{4.4.50}$$

and the corresponding post-measurement state is

$$\rho^x = \frac{\mathcal{M}_x(\rho)}{\mathrm{Tr}[\mathcal{M}_x(\rho)]}. \tag{4.4.51}$$

The expected state of the ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ is then

$$\sum_{x\in\mathcal{X}} \mathcal{M}_x(\rho). \tag{4.4.52}$$

It is customary to define the output of a quantum instrument $\{\mathcal{M}_x\}_{x\in\mathcal{X}}$ as the channel $\mathcal{M}$ defined by

$$\mathcal{M}(\rho) = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \mathcal{M}_x(\rho). \tag{4.4.53}$$

That is, the output of a quantum instrument is a classical–quantum state in which the classical register X stores the outcome of the measurement. This is unlike the expected state in (4.4.52), which represents a lack of knowledge of which measurement outcome occurred.

Note that the channel in (4.4.53) corresponding to a quantum instrument reduces to the measurement channel defined in (4.4.34) if we consider a measurement with POVM $\{M_x\}_{x\in\mathcal{X}}$ and we define the maps $\mathcal{M}_x$ as $\mathcal{M}_x(\rho) = \mathrm{Tr}[M_x\rho]$ for all $x \in \mathcal{X}$. In this case, the channel in (4.4.53) becomes

$$\mathcal{M}(\rho) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho]|x\rangle\langle x|_X, \tag{4.4.54}$$

which is precisely the measurement channel in (4.4.34).

4.4.6 Entanglement-Breaking Channels

An important class of channels in quantum communication consists of those that, when acting on one share of a bipartite state, eliminate any entanglement between the two systems, such that the resulting state is separable (recall Definition 3.5). Such channels are called *entanglement breaking*, and we define them as follows.

Definition 4.12 Entanglement-Breaking Channel

A channel $\mathcal{N}_{A\to B}$ is called *entanglement breaking* if $(\mathrm{id}_R \otimes \mathcal{N}_{A\to B})(\rho_{RA})$ is a separable state for every state ρ_{RA}, where R is an arbitrary reference system.

Although Definition 4.12 suggests that it is necessary to check an infinite number of input states to determine whether a channel is entanglement breaking, the following proposition states that it is only necessary to check the output of the channel on the maximally entangled state.

Proposition 4.13

A channel $\mathcal{N}$ is entanglement breaking if and only if its Choi state $\Phi^{\mathcal{N}}_{AB}$ is separable.

PROOF: Observe that if $\mathcal{N}_{A\to B}$ is entanglement breaking, then its Choi state $\Phi^{\mathcal{N}}_{AB}$ is separable. On the other hand, if the Choi state $\Phi^{\mathcal{N}}_{AB}$ of a given channel $\mathcal{N}$ is separable, then it is of the form

$$\Phi^{\mathcal{N}}_{AB} = \sum_{x\in\mathcal{X}} p(x)\sigma^x_A \otimes \tau^x_B \tag{4.4.55}$$

for some probability distribution $p : \mathcal{X} \to [0,1]$ on a finite alphabet $\mathcal{X}$ and sets $\{\sigma^x_A : x \in \mathcal{X}\}$ and $\{\tau^x_B : x \in \mathcal{X}\}$ of states. We note that the property $\text{Tr}_B[\Phi^{\mathcal{N}}_{AB}] = \pi_A$ of the Choi state translates to

$$\pi_A = \frac{\mathbb{1}_A}{d_A} = \sum_{x\in\mathcal{X}} p(x)\sigma^x_A. \tag{4.4.56}$$

Then, for every reference system R and state ξ_{RA} acting on $\mathcal{H}_{RA}$, we find, by using (4.2.5), that

$$\begin{aligned}(\text{id}_R \otimes \mathcal{N})(\xi_{RA}) &= d_A \text{Tr}_A[(\text{T}_A(\xi_{RA}) \otimes \mathbb{1}_B)(\mathbb{1}_R \otimes \Phi^{\mathcal{N}}_{AB})] &\quad (4.4.57)\\ &= \sum_{x\in\mathcal{X}} q(x)\omega^x_R \otimes \tau^x_B, &\quad (4.4.58)\end{aligned}$$

where

$$\begin{aligned}\omega^x_R &\coloneqq \frac{\text{Tr}_A[\text{T}_A(\xi_{RA})(\mathbb{1}_R \otimes d_A p(x)\sigma^x_A)]}{q(x)}, &\quad (4.4.59)\\ q(x) &\coloneqq p(x) d_A \text{Tr}[\xi^{\mathsf{T}}_A \sigma^x_A]. &\quad (4.4.60)\end{aligned}$$

Now, the map $x \mapsto q(x)$ is a probability distribution on $\mathcal{X}$ since $q(x) \geq 0$ for all $x \in \mathcal{X}$ and

$$\begin{aligned}\sum_{x\in\mathcal{X}} q(x) &= d_A \text{Tr}\left[\xi^{\mathsf{T}}_A \left(\sum_{x\in\mathcal{X}} p(x)\sigma^x_A\right)\right] &\quad (4.4.61)\\ &= d_A \text{Tr}[\xi^{\mathsf{T}}_A \pi_A] &\quad (4.4.62)\\ &= \text{Tr}[\xi_A] &\quad (4.4.63)\\ &= 1, &\quad (4.4.64)\end{aligned}$$

where we have made use of (4.4.56). So $(\text{id}_R \otimes \mathcal{N})(\xi_{RA})$ is a separable state. ■

Another useful characterization of entanglement breaking channels is through their Kraus representations, as shown in the following proposition:

Proposition 4.14

A channel $\mathcal{N}$ is entanglement breaking if and only if there exists a set of Kraus operators for $\mathcal{N}$, with each Kraus operator having unit rank.

PROOF: First suppose that the Kraus operators of $\mathcal{N}$ have unit rank. They are therefore of the form $|\phi_j\rangle_B\langle\psi_j|_A =: K_j$ for $1 \leq j \leq r$. Without loss of generality, we can let each vector in the set $\{|\phi_j\rangle\}_{j=1}^r$ be normalized. Then, since $\mathcal{N}$ is trace preserving, it holds that

$$\mathbb{1}_A = \sum_{j=1}^r K_j^\dagger K_j = \sum_{j=1}^r |\psi_j\rangle_A\langle\phi_j|\phi_j\rangle\langle\psi_j|_A = \sum_{j=1}^r |\psi_j\rangle\langle\psi_j|_A. \quad (4.4.65)$$

Now, for every reference system R of arbitrary dimension and every state ρ_{RA}, we find that

$$(\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA}) = \sum_{j=1}^r (\mathbb{1}_R \otimes K_j)\rho_{RA}(\mathbb{1}_R \otimes K_j^\dagger) \quad (4.4.66)$$

$$= \sum_{j=1}^r (\mathbb{1}_R \otimes |\phi_j\rangle_B\langle\psi_j|_A)\rho_{RA}(\mathbb{1}_R \otimes |\psi_j\rangle_A\langle\phi_j|_B) \quad (4.4.67)$$

$$= \sum_{j=1}^r (\mathbb{1}_R \otimes \langle\psi_j|_A)(\rho_{RA})(\mathbb{1}_R \otimes |\psi_j\rangle_A) \otimes |\phi_j\rangle\langle\phi_j|_B \quad (4.4.68)$$

$$= \sum_{j=1}^r p(j)\sigma_R^j \otimes |\phi_j\rangle\langle\phi_j|_B, \quad (4.4.69)$$

where

$$\sigma_R^j := \frac{(\mathbb{1}_R \otimes \langle\psi_j|_A)(\rho_{RA})(\mathbb{1}_R \otimes |\psi_j\rangle_A)}{p(j)}, \quad p(j) := \langle\psi_j|_A\rho_A|\psi_j\rangle_A. \quad (4.4.70)$$

Note that $j \mapsto p(j)$ is a probability distribution since $p(j) \geq 0$ for all j, and

$$\sum_{j=1}^r p(j) = \sum_{j=1}^r \langle\psi_j|_A\rho_A|\psi_j\rangle_A \quad (4.4.71)$$

$$= \sum_{j=1}^{r} \mathrm{Tr}[|\psi_j\rangle\langle\psi_j|\rho_A] \tag{4.4.72}$$

$$= \mathrm{Tr}\left[\left(\sum_{j=1}^{r} |\psi_j\rangle\langle\psi_j|_A\right)\rho_A\right] \tag{4.4.73}$$

$$= \mathrm{Tr}[\rho_A] \tag{4.4.74}$$

$$= 1, \tag{4.4.75}$$

where we have made use of (4.4.65). Therefore, $(\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA})$ is a separable state, so that $\mathcal{N}$ is entanglement breaking.

Now, suppose that $\mathcal{N}$ is entanglement breaking. This means that its Choi state $\Phi^{\mathcal{N}}_{AB}$ is a separable state, which means that it can be written as

$$\Phi^{\mathcal{N}}_{AB} = \sum_{x \in \mathcal{X}} p(x) |\psi_x\rangle\langle\psi_x|_A \otimes |\phi_x\rangle\langle\phi_x|_B \tag{4.4.76}$$

for some probability distribution $p : \mathcal{X} \to [0, 1]$ on a finite alphabet $\mathcal{X}$ and sets of pure states $\{|\psi_x\rangle_A : x \in \mathcal{X}\}$, $\{|\phi_x\rangle_B : x \in \mathcal{X}\}$. Define the unit-rank operators

$$K_x := \sqrt{d_A p(x)} |\phi_x\rangle_B \langle\overline{\psi_x}|_A, \quad x \in \mathcal{X}. \tag{4.4.77}$$

Then, for every orthonormal basis state $|i\rangle_A$ on A, we have

$$K_x |i\rangle_A = \sqrt{d_A p(x)} |\phi_x\rangle_B \langle\overline{\psi_x}|i\rangle \tag{4.4.78}$$

$$= \sqrt{d_A p(x)} \langle i|\psi_x\rangle |\phi_x\rangle_B \tag{4.4.79}$$

$$= \sqrt{d_A p(x)} (\langle i|_A \otimes \mathbb{1}_B)(|\psi_x\rangle_A \otimes |\phi_x\rangle_B). \tag{4.4.80}$$

Using this, we find that

$$\begin{aligned}
\mathcal{N}(|i\rangle\langle i'|_A) &= d_A(\langle i|_A \otimes \mathbb{1}_B)\Phi^{\mathcal{N}}_{AB}(|i'\rangle_A \otimes \mathbb{1}_B) \\
&= \sum_{x \in \mathcal{X}} \sqrt{d_A p(x)} (\langle i|_A \otimes \mathbb{1}_B)(|\psi_x\rangle_A \otimes |\phi_x\rangle_B) \\
&\qquad \times (\langle\psi_x|_A \otimes \langle\phi_x|_B)(|i'\rangle_A \otimes \mathbb{1}_B)\sqrt{d_A p(x)} \\
&= \sum_{x \in \mathcal{X}} K_x |i\rangle\langle i'|_A K_x^\dagger.
\end{aligned} \tag{4.4.81}$$

This holds for all $0 \leq i, i' \leq d_A - 1$, which means that $\{K_x\}_{x \in \mathcal{X}}$ is a set of Kraus operators for $\mathcal{N}$, each of which has unit rank. ■

From this proposition, we immediately see that both quantum–classical and classical–quantum channels are entanglement breaking, because each one has a Kraus representation with unit-rank Kraus operators. This implies that a composition of a quantum–classical channel followed by a classical-quantum channel is also entanglement-breaking, and every such map can be written as

$$\rho \mapsto \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho]\sigma^x, \tag{4.4.82}$$

where $\mathcal{X}$ is a finite alphabet, $\{\sigma^x\}_{x\in\mathcal{X}}$ is a set of quantum states, and $\{M_x\}_{x\in\mathcal{X}}$ is a POVM. Indeed, if each POVM element M_x has a spectral decomposition of the form

$$M_x = \sum_{k=1}^{r_x} \lambda_k^x |\psi_k^x\rangle\langle\psi_k^x|, \tag{4.4.83}$$

where $r_x = \mathrm{rank}(M_x)$, and each state σ^x has a spectral decomposition of the form

$$\sigma^x = \sum_{\ell=1}^{s_x} \alpha_\ell^x |\phi_\ell^x\rangle\langle\phi_\ell^x|, \tag{4.4.84}$$

where $s_x = \mathrm{rank}(\sigma^x)$, then (4.4.82) can be written as

$$\rho \mapsto \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_x} \sum_{\ell=1}^{s_x} \lambda_k^x \alpha_\ell^x |\phi_\ell^x\rangle\langle\phi_\ell^x|\langle\psi_k^x|\rho|\psi_k^x\rangle \tag{4.4.85}$$

$$= \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_x} \sum_{\ell=1}^{s_x} \sqrt{\lambda_k^\ell \alpha_\ell^x} |\phi_\ell^x\rangle\langle\psi_k^x|\rho|\psi_k^x\rangle\langle\phi_\ell^x| \sqrt{\lambda_k^x \alpha_\ell^x} \tag{4.4.86}$$

$$= \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_x} \sum_{\ell=1}^{s_x} K_{k,\ell}^x \rho (K_{k,\ell}^x)^\dagger, \tag{4.4.87}$$

where $K_{k,\ell}^x := \sqrt{\lambda_k^x \alpha_\ell^x} |\phi_\ell^x\rangle\langle\psi_k^x|$. Since $\{K_{k,\ell}^x : x \in \mathcal{X},\ 1 \le k \le r_x,\ 1 \le \ell \le s_x\}$ is a set of Kraus operators for the map, with each Kraus operator having unit rank, it holds by Proposition 4.14 that the map in (4.4.82) is entanglement breaking.

A channel of the form (4.4.82) is sometimes called a "measure-and-prepare channel" or an "intercept-resend channel" since the input to the channel is first measured then replaced by a new state conditioned on the measurement outcome. Remarkably, *any* entanglement breaking channel can be written in the form (4.4.82), as we now show.

Theorem 4.15

For every entanglement-breaking channel $\mathcal{N}$, there exists a finite alphabet $\mathcal{X}$, a set $\{\sigma^x\}_{x\in\mathcal{X}}$ of states, and a POVM $\{M_x\}_{x\in\mathcal{X}}$ such that the action of $\mathcal{N}$ can be written as

$$\mathcal{N}(\rho) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho]\sigma^x \tag{4.4.88}$$

for every state ρ.

Proof: By Proposition 4.14, the action of $\mathcal{N}$ can be written as

$$\mathcal{N}(\rho) = \sum_{j=1}^{r} K_j\rho K_j^\dagger, \tag{4.4.89}$$

where $r = \mathrm{rank}(\Gamma^{\mathcal{N}})$ and $\{K_j\}_{j=1}^r$ is a set of Kraus operators for $\mathcal{N}$, with each Kraus operator having unit rank. Since each Kraus operator has unit rank, it holds that $K_j = |\phi_j\rangle\langle\psi_j|$ for all $1 \leq j \leq r$, where $\{|\phi_j\rangle\}_{j=1}^r$ and $\{|\psi_j\rangle\}_{j=1}^r$ are sets of vectors (without loss of generality, we can take $\{|\phi_j\rangle\}_{j=1}^r$ to be a set of pure states). Since $\mathcal{N}$ is trace preserving, it holds that

$$\sum_{j=1}^{r} K_j^\dagger K_j = \sum_{j=1}^{r} |\psi_j\rangle\langle\phi_j|\phi_j\rangle\langle\psi_j| = \sum_{j=1}^{r} |\psi_j\rangle\langle\psi_j| = \mathbb{1}. \tag{4.4.90}$$

This implies that $\{|\psi_j\rangle\langle\psi_j|\}_{j=1}^r$ is a POVM. Therefore, defining the alphabet $\mathcal{X} = \{1, 2, \ldots, r\}$, the POVM elements $M_x := |\psi_x\rangle\langle\psi_x|$ and states $\sigma^x := |\phi_x\rangle\langle\phi_x|$, we have that

$$\mathcal{N}(\rho) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho]\sigma^x, \tag{4.4.91}$$

as required. ■

An extreme example of a measure-and-prepare channel, as in (4.4.88), is one for which the output is a fixed state σ for every outcome of the measurement described by the POVM $\{M_x\}_{x\in\mathcal{X}}$. In this case, the channel in (4.4.88) takes the form

$$\mathcal{N}(\rho) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[M_x\rho]\sigma = \mathrm{Tr}[\rho]\sigma = \mathcal{R}^\sigma(\rho), \tag{4.4.92}$$

where the second equality follows because $\sum_{x\in\mathcal{X}} M_x = \mathbb{1}$, and the last equality from the definition of the replacement channel for σ in Definition 4.8. This means that every replacement channel is a measure-and-prepare channel, and in particular an entanglement-breaking channel.

The development above Proposition 4.14 tells us that to every entanglement breaking channel there is associated a separable state, namely, the Choi state. The converse statement also holds; see Exercise 4.20.

Exercise 4.20

Given a separable state $\rho_{AB} = \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x$, show that the channel $\mathcal{N}_{A\to B}^{\rho}$ defined in (4.2.16) has the form

$$\mathcal{N}_{A\to B}^{\rho}(X_A) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[X_A M_A^x]\tau_B^x, \tag{4.4.93}$$

where

$$M_A^x = p(x)\left(\rho_A^{-\frac{1}{2}}\omega_A^x\rho_A^{-\frac{1}{2}}\right)^{\mathsf{T}} \tag{4.4.94}$$

for all $x \in \mathcal{X}$. In other words, by the discussion after (4.4.82), every separable state can be associated with an entanglement-breaking channel.

4.4.7 Hadamard Channels

It turns out that every entanglement-breaking channel can be regarded as the complement of what is called a *Hadamard channel*, which we now define.

Definition 4.16 Hadamard Channel

A channel $\mathcal{N}$ is called a *Hadamard channel* or a *Schur channel* if there exists a positive semi-definite operator N with unit diagonal elements (in the standard basis) and an isometry V such that

$$\mathcal{N}(X) = N * VXV^{\dagger} \tag{4.4.95}$$

for every linear operator X, where $N * VXV^{\dagger}$ is the *Hadamard product*, also called the *Schur product*, which is defined as the element-wise product of the operators N and $VXV^{\dagger}$ when represented as matrices with respect to the standard basis.

Given linear operators X and Y acting on a d-dimensional Hilbert space, with matrix representations in the standard basis as

$$X = \sum_{i,j=0}^{d-1} X_{i,j} |i\rangle\langle j|, \quad Y = \sum_{i,j=0}^{d-1} Y_{i,j} |i\rangle\langle j|, \tag{4.4.96}$$

the Hadamard product $X * Y$ is the operator with matrix representation in the standard basis given by the product of the matrix elements of X and Y:

$$X * Y = \sum_{i,j=0}^{d-1} X_{i,j} Y_{i,j} |i\rangle\langle j|. \tag{4.4.97}$$

Note that the Hadamard product can be defined as the element-wise product with respect to an arbitrary orthonormal basis; however, in this book, we only consider the Hadamard product with respect to the standard basis.

The positive semi-definiteness of N in the definition of a Hadamard channel is necessary and sufficient for the map $\mathcal{N}$ defined in (4.4.95) to be completely positive, while the fact that N has unit diagonal elements in the standard basis and that V is an isometry ensures that $\mathcal{N}$ is trace preserving.

Exercise 4.21

1. Show that a dephasing channel, as defined in (4.5.35), is a Hadamard channel.
2. Let $\{K_j\}_{j=1}^r$ be a set of Kraus operators. Prove that the channel, defined by the set $\{K_j \otimes |j\rangle\}_{j=1}^r$ of Kraus operators, is a Hadamard channel.

The following fact about complements of Hadamard channels provides an important characterization of Hadamard channels:

Proposition 4.17

Any complement of a Hadamard channel is entanglement breaking.

PROOF: Suppose $\mathcal{N}_{A\to B}$ is a Hadamard channel between systems A and B with associated positive semi-definite operator N having unit diagonal elements in the standard basis and with associated isometry V. Since N is positive semi-definite and it has unit diagonal elements, it can be expressed as the Gram matrix of some set $\{|\psi_i\rangle : 1 \leq i \leq d\}$ of normalized vectors, so that

$$N = \sum_{i,j=0}^{d-1} \langle\psi_i|\psi_j\rangle |i\rangle\langle j|. \tag{4.4.98}$$

Therefore, using (4.4.97) and Definition 4.16, the action of $\mathcal{N}$ can be written as

$$\mathcal{N}(X) = \sum_{i,j=0}^{d-1} \langle\psi_i|\psi_j\rangle\langle i|VXV^\dagger|j\rangle |i\rangle\langle j|_B. \tag{4.4.99}$$

Now, set

$$|\phi_i\rangle := V^\dagger|i\rangle \Rightarrow \langle\phi_i| = \langle i|V, \tag{4.4.100}$$

so that

$$\mathcal{N}(X) = \sum_{i,j=0}^{d-1} \langle\psi_i|\psi_j\rangle\langle\phi_i|X|\phi_j\rangle |i\rangle\langle j|_B. \tag{4.4.101}$$

Consider the operator $(U^{\mathcal{N}})_{A\to BE}$ defined as

$$U^{\mathcal{N}} := \sum_{i=0}^{d-1} |i\rangle_B\langle\phi_i|_A \otimes |\psi_i\rangle_E. \tag{4.4.102}$$

Since V is an isometry, and the vectors $\{|\psi_i\rangle : 1 \leq i \leq d\}$ are normalized, it follows that

$$(U^{\mathcal{N}})^\dagger U^{\mathcal{N}} = \sum_{i=0}^{d-1} |\phi_i\rangle\langle\phi_i| = \sum_{i=0}^{d-1} V^\dagger|i\rangle\langle i|V = V^\dagger V = \mathbb{1}. \tag{4.4.103}$$

This means that $U^{\mathcal{N}}$ is an isometry. Furthermore,

$$U^{\mathcal{N}} X (U^{\mathcal{N}})^\dagger = \sum_{i,j=0}^{d-1} \langle\phi_i|X|\phi_j\rangle |\psi_i\rangle\langle\psi_j|_E \otimes |i\rangle\langle j|_B, \tag{4.4.104}$$

so that $\text{Tr}_E[U^{\mathcal{N}} X (U^{\mathcal{N}})^\dagger] = \mathcal{N}(X)$. The operator $U^{\mathcal{N}}$ is therefore an isometric extension of $\mathcal{N}$. A complementary channel then results from tracing out B in (4.4.104), i.e.,

$$\mathcal{N}^c(X) = \text{Tr}_B[U^{\mathcal{N}} X (U^{\mathcal{N}})^\dagger] \tag{4.4.105}$$

$$= \sum_{i=0}^{d-1} \langle\phi_i|X|\phi_i\rangle |\psi_i\rangle\!\langle\psi_i|_E \tag{4.4.106}$$

$$= \sum_{i=0}^{d-1} |\psi_i\rangle\!\langle\phi_i|X|\phi_i\rangle\!\langle\psi_i| \tag{4.4.107}$$

$$= \sum_{i=0}^{d-1} K_i X K_i^\dagger, \tag{4.4.108}$$

where $K_i := |\psi_i\rangle\!\langle\phi_i|$. So $\mathcal{N}^c$ has a Kraus representation with unit-rank Kraus operators, which means, by Proposition 4.14, that $\mathcal{N}^c$ is entanglement breaking. Every complement of a channel is related to another complement by an isometric channel acting on the output of the complement, and this does not change the entanglement-breaking property. ■

By following the proof above backwards, we find that every entanglement-breaking channel is the complement of some Hadamard channel.

4.4.8 Covariant Channels

In Section 3.2.7, we defined states that are invariant under the action of a unitary representation of a group. We now define an analogous notion of invariance for quantum channels.

Definition 4.18 Group-Covariant Channel

Let G be a group. A channel $\mathcal{N}_{A\to B}$ is called *covariant with respect to* G, *group-covariant*, G*-covariant*, or simply *covariant*, if there exist unitary representations $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$ of G such that for every state ρ_A,

$$\mathcal{N}_{A\to B}(U_A^g \rho_A U_A^{g\dagger}) = V_B^g \mathcal{N}_{A\to B}(\rho_A) V_B^{g\dagger} \tag{4.4.109}$$

for all $g \in G$.

Exercise 4.22

Let $\mathcal{N}_{A\to B}$ be a group-covariant channel, as per Definition 4.18.

1. Show that the condition in (4.4.109) can be written more compactly as follows:
$$\mathcal{N}_{A\to B} = \mathcal{V}_B^{g\dagger} \circ \mathcal{N}_{A\to B} \circ \mathcal{U}_A^g \tag{4.4.110}$$
for all $g \in G$, where $\mathcal{V}_B^{g\dagger}(\cdot) := V_B^{g\dagger}(\cdot)V_B^g$ and $\mathcal{U}_A^g(\cdot) := U_A^g(\cdot)U_A^{g\dagger}$.

2. Show that the Choi representation of $\mathcal{N}_{A\to B}$ is invariant under the action of $U_A^{g\mathsf{T}} \otimes V_B^{g\dagger}$ for all $g \in G$; i.e., show that
$$\Gamma_{AB}^{\mathcal{N}} = (U_A^{g\mathsf{T}} \otimes V_B^{g\dagger})\Gamma_{AB}^{\mathcal{N}}(U_A^{g\mathsf{T}} \otimes V_B^{g\dagger})^\dagger \tag{4.4.111}$$
for all $g \in G$.

3. For every set $\{K_i\}_{i=1}^r$ of Kraus operators for $\mathcal{N}$, with $r \geq \text{rank}(\Gamma_{AB}^{\mathcal{N}})$, show that $\{K_i^g\}_{i=1}^r$, with $K_i^g := V_B^{g\dagger}K_iU_A^g$, is another set of Kraus operators for $\mathcal{N}$ for all $g \in G$.

4. For every isometric extension $W_{A\to BE}$ of $\mathcal{N}$, with $d_E \geq \text{rank}(\Gamma_{AB}^{\mathcal{N}})$, show that $W_{A\to BE}^g := V_B^{g\dagger}WU_A^g$ is another isometric extension of $\mathcal{N}$ for all $g \in G$.

Group covariant channels have group covariant isometric extensions, as the following lemma states.

Lemma 4.19 Isometric Extensions of Group Covariant Channels

Suppose that a channel $\mathcal{N}_{A\to B}$ is covariant with respect to a group G. For an isometric extension $U_{A\to BE}^{\mathcal{N}}$ of $\mathcal{N}_{A\to B}$, there is a set of unitaries $\{W_E^g\}_{g\in G}$ such that the following covariance holds for all $g \in G$:
$$U_{A\to BE}^{\mathcal{N}}U_A^g = \left(V_B^g \otimes W_E^g\right) U_{A\to BE}^{\mathcal{N}}. \tag{4.4.112}$$

Proof: Given is a group G and a quantum channel $\mathcal{N}_{A\to B}$ that is covariant in the following sense:
$$\mathcal{N}_{A\to B}(U_A^g\rho_AU_A^{g\dagger}) = V_B^g\mathcal{N}_{A\to B}(\rho_A)V_B^{g\dagger}, \tag{4.4.113}$$

for a set of unitaries $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$.

Let a Kraus representation of $\mathcal{M}_{A\to B}$ be given as

$$\mathcal{N}_{A\to B}(\rho_A) = \sum_j L^j \rho_A L^{j\dagger}. \tag{4.4.114}$$

We can rewrite (4.4.113) as

$$V_B^{g\dagger}\mathcal{N}_{A\to B}(U_A^g\rho_A U_A^{g\dagger})V_B^g = \mathcal{N}_{A\to B}(\rho_A), \tag{4.4.115}$$

which means that for all g, the following equality holds

$$\sum_j L^j \rho_A L^{j\dagger} = \sum_j V_B^{g\dagger} L^j U_A^g \rho_A \left(V_B^{g\dagger} L^j U_A^g\right)^\dagger. \tag{4.4.116}$$

Thus, the channel has two different Kraus representations $\{L^j\}_j$ and $\{V_B^{g\dagger} L^j U_A^g\}_j$, and these are necessarily related by a unitary with matrix elements w_{jk}^g (see the discussion after (4.3.7)):

$$V_B^{g\dagger} L^j U_A^g = \sum_k w_{jk}^g L^k. \tag{4.4.117}$$

A canonical isometric extension $U_{A\to BE}^{\mathcal{N}}$ of $\mathcal{N}_{A\to B}$ is given as

$$U_{A\to BE}^{\mathcal{N}} = \sum_j L^j \otimes |j\rangle_E, \tag{4.4.118}$$

where $\{|j\rangle_E\}_j$ is an orthonormal basis. Defining W_E^g as the following unitary

$$W_E^g|k\rangle_E = \sum_j w_{jk}^g |j\rangle_E, \tag{4.4.119}$$

where the states $|k\rangle_E$ are chosen from $\{|j\rangle_E\}_j$, consider that

$$U_{A\to BE}^{\mathcal{N}} U_A^g = \sum_j L^j U_A^g \otimes |j\rangle_E \tag{4.4.120}$$

$$= \sum_j V_B^g V_B^{g\dagger} L^j U_A^g \otimes |j\rangle_E \tag{4.4.121}$$

$$= \sum_j V_B^g \left[\sum_k w_{jk}^g L^k\right] \otimes |j\rangle_E \tag{4.4.122}$$

$$= V_B^g \sum_k L^k \otimes \sum_j w_{jk}^g |j\rangle_E \tag{4.4.123}$$

$$= V_B^g \sum_k L^k \otimes W_E^g |k\rangle_E \tag{4.4.124}$$

$$= \left(V_B^g \otimes W_E^g\right) U_{A\to BE}^{\mathcal{N}}. \tag{4.4.125}$$

This concludes the proof. ■

Recall the definition of the twirling map in Exercise 3.17:

$$\mathcal{T}^G(\rho) := \frac{1}{|G|}\sum_{g\in G} U^g \rho U^{g\dagger}. \tag{4.4.126}$$

This map, which is evidently a quantum channel, takes an arbitrary state ρ and makes it invariant under the action of the group G given by the unitary representation $\{U^g\}_{g\in G}$. Similarly, we can define the *twirl of a quantum channel*, which takes an arbitrary quantum channel $\mathcal{N}_{A\to B}$ and makes it group covariant, as per Definition 4.18:

$$\mathcal{N}_{A\to B}^G := \frac{1}{|G|}\sum_{g\in G} \mathcal{V}_B^g \circ \mathcal{N}_{A\to B} \circ \mathcal{U}_A^{g\dagger}. \tag{4.4.127}$$

Exercise 4.23

Given a quantum channel $\mathcal{N}_{A\to B}$, prove that the twirled channel $\mathcal{N}_{A\to B}^G$, as defined in (4.4.127), is group covariant.

Proposition 4.20

Let $\mathcal{N}_{A\to B}$ be a Hermiticity-preserving superoperator that is covariant with respect to a group G, as defined in Definition 4.18.

1. For every pure state $\psi_{A'A}$, with $d_{A'} = d_A$,

$$\|\mathcal{N}_{A\to B}(\psi_{A'A})\|_1 \le \|\mathcal{N}_{A\to B}(\phi_{A'A}^{\overline{\rho}})\|_1, \tag{4.4.128}$$

   where $\rho_A := \psi_A = \mathrm{Tr}_{A'}[\psi_{A'A}]$, $\overline{\rho}_A := \mathcal{T}_A^G(\rho_A) = \frac{1}{|G|}\sum_{g\in G} U_A^g \rho_A U_A^{g\dagger}$, and $\phi_{A'A}^{\overline{\rho}}$ is a purification of $\overline{\rho}$.

2. The diamond norm of $\mathcal{N}$ is given by

$$\|\mathcal{N}\|_\diamond = \sup_{\psi_{A'A}} \left\{ \|(\mathrm{id}_{A'} \otimes \mathcal{N}_{A\to B})(\psi_{A'A})\|_1 : \psi_A = \mathcal{T}_A^G(\psi_A) \right\}, \qquad (4.4.129)$$

where the optimization is with respect to pure states $\psi_{A'A}$, with $d_{A'} = d_A$, such that the reduced state ψ_A is invariant under the twirling channel $\mathcal{T}_A^G(\cdot) := \frac{1}{|G|}\sum_{g\in G} U_A^g(\cdot)U_A^{g\dagger}$ defined by the representation $\{U_A^g\}_{g\in G}$.

3. If the representation $\{U_A^g\}_{g\in G}$ is such that $\mathcal{T}_A^G(\cdot) = \mathrm{Tr}[\cdot]\frac{\mathbb{1}_A}{d_A}$, then

$$\|\mathcal{N}\|_\diamond = \left\|\Phi_{AB}^{\mathcal{N}}\right\|_1. \qquad (4.4.130)$$

Proof:

1. Let $\psi_{A'A}$ be an arbitrary pure state, $\rho_A = \psi_A$, $\overline{\rho}_A = \mathcal{T}_A^G(\rho_A)$, and let $\phi_{A'A}^{\overline{\rho}}$ be a purification of $\overline{\rho}_A$. Let us also consider the following purification of $\overline{\rho}_A$:

$$|\psi^{\overline{\rho}}\rangle_{RA'A} := \frac{1}{\sqrt{|G|}}\sum_{g\in G}|g\rangle_R \otimes U_A^g|\psi\rangle_{A'A}, \qquad (4.4.131)$$

where $\{|g\rangle\}_{g\in G}$ is an orthonormal set. (Recall Exercise 3.17.) Now, because all purifications of a state can be mapped to each other via isometries acting on the purifying system (see Section 3.2.5), there exists an isometry $W_{A'\to RA'}$ such that $|\psi^{\overline{\rho}}\rangle_{RA'A} = W_{A'\to RA'}|\phi^{\overline{\rho}}\rangle_{A'A}$. Therefore,

$$\left\|\mathcal{N}_{A\to B}(\psi_{RA'A}^{\overline{\rho}})\right\|_1 = \left\|W_{A'\to RA'}\mathcal{N}_{A\to B}(\phi_{A'A}^{\overline{\rho}})(W_{A'\to RA'})^\dagger\right\|_1 \qquad (4.4.132)$$

$$= \left\|\mathcal{N}_{A\to B}(\phi_{A'A}^{\overline{\rho}})\right\|_1, \qquad (4.4.133)$$

where the last line follows from isometric invariance of the trace norm (see (2.2.92)).

Now, let us apply the quantum channel $X \mapsto \sum_{g\in G}|g\rangle\langle g|X|g\rangle\langle g|$ to the system R. By the data-processing inequality in (4.1.7), we find that

$$\left\|\mathcal{N}_{A\to B}(\psi_{RA'A}^{\overline{\rho}})\right\|_1 \geq \left\|\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_R \otimes (\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{A'A})\right\|_1 \qquad (4.4.134)$$

$$= \left\| \frac{1}{|G|} \sum_{g \in G} |g\rangle\langle g|_R \otimes (\mathcal{V}_B^{g\dagger} \circ \mathcal{N}_{A \to B} \circ \mathcal{U}_A^g)(\psi_{A'A}) \right\|_1, \tag{4.4.135}$$

where the last line follows from applying the unitary channel defined by the unitary $\sum_{g \in G} |g\rangle\langle g|_R \otimes V_B^{g\dagger}$ and from unitary invariance of the trace norm. Finally, using the covariance of $\mathcal{N}$, in particular (4.4.110), along with (4.4.133), we obtain

$$\left\| \mathcal{N}_{A \to B}(\phi_{A'A}^{\overline{\rho}}) \right\|_1 = \left\| \mathcal{N}_{A \to B}(\psi_{RA'A}^{\overline{\rho}}) \right\|_1 \tag{4.4.136}$$

$$\geq \left\| \frac{1}{|G|} \sum_{g \in G} |g\rangle\langle g|_R \otimes \mathcal{N}_{A \to B}(\psi_{A'A}) \right\|_1 \tag{4.4.137}$$

$$= \|\mathcal{N}_{A \to B}(\psi_{A'A})\|_1, \tag{4.4.138}$$

where the last line follows from (2.2.95). The derived inequality is precisely (4.4.128),

2. Note that, by definition, every purification $\phi_{A'A}^{\overline{\rho}}$ of $\overline{\rho}_A$ is such that its reduced state on A is invariant under the channel $\mathcal{T}_A^G$. Therefore, using (4.4.128), for every pure state $\psi_{A'A}$, we obtain

$$\|\mathcal{N}_{A \to B}(\psi_{A'A})\|_1 \leq \left\| \mathcal{N}_{A \to B}(\phi_{A'A}^{\overline{\rho}}) \right\|_1 \tag{4.4.139}$$

$$\leq \sup_{\phi_{A'A}} \left\{ \|\mathcal{N}_{A \to B}(\phi_{A'A})\|_1 : \phi_A = \mathcal{T}_A^G(\phi_A) \right\}. \tag{4.4.140}$$

Since this inequality holds for every pure state $\psi_{A'A}$, and because $\mathcal{N}$ is Hermiticity preserving, we can use (2.2.175) to conclude that

$$\|\mathcal{N}\|_\diamond = \sup_{\psi_{A'A}} \|\mathcal{N}_{A \to B}(\psi_{A'A})\|_1 \tag{4.4.141}$$

$$\leq \sup_{\psi_{A'A}} \left\{ \|\mathcal{N}_{A \to B}(\psi_{A'A})\|_1 : \psi_A = \mathcal{T}_A^G(\psi_A) \right\}. \tag{4.4.142}$$

The opposite inequality is immediate, because the set $\{\psi_{A'A} : \psi_A = \mathcal{T}_A^G(\psi_A)\}$ is a subset of all pure states. This concludes the proof of (4.4.129).

3. If $\mathcal{T}_A^G(\cdot) = \mathrm{Tr}[\cdot] \frac{\mathbb{1}_A}{d_A}$, then the optimization in (4.4.128) is with respect to pure states $\psi_{A'A}$ whose reduced state on A is maximally mixed. All such pure

states are maximally entangled (see Definition 3.6 and the discussion below it), which means that there exists a unitary U such that $\psi_{A'A} = \mathcal{U}_{A'}(\Phi_{A'A})$, where $\Phi_{A'A} = |\Phi\rangle\langle\Phi|_{A'A}$ is the maximally entangled state defined by $|\Phi\rangle_{A'A} = \frac{1}{\sqrt{d_A}}\sum_{i=0}^{d_A-1}|i,i\rangle_{A'A}$ (see Exercise 3.7). By unitary invariance of the trace norm, we thus immediately obtain $\|\mathcal{N}\|_\diamond = \|\mathcal{N}_{A\to B}(\Phi_{A'A})\|_1 = \left\|\Phi^{\mathcal{N}}_{A'B}\right\|_1$, as required. ■

4.4.9 Bipartite and Multipartite Channels

4.5 Examples of Communication Channels

4.5.1 (Generalized) Amplitude Damping Channel

The *amplitude damping channel with decay parameter* $\gamma \in [0,1]$ is the channel $\mathcal{A}_\gamma$ given by $\mathcal{A}_\gamma(\rho) = A_1\rho A_1^\dagger + A_2\rho A_2^\dagger$, with the two Kraus operators A_1 and A_2 defined as

$$A_1 = \sqrt{\gamma}|0\rangle\langle 1|, \quad A_2 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|. \tag{4.5.1}$$

It is straightforward to verify that $A_1^\dagger A_1 + A_2^\dagger A_2 = \mathbb{1}$, so that $\mathcal{A}_\gamma$ is indeed trace preserving.

Let ρ be a state with a matrix representation in the standard basis $\{|0\rangle, |1\rangle\}$ as

$$\rho = \begin{pmatrix} 1-\lambda & \alpha \\ \overline{\alpha} & \lambda \end{pmatrix}. \tag{4.5.2}$$

In order for ρ to be a state (positive semi-definite with unit trace), the conditions $0 \le \lambda \le 1$ and $\lambda(1-\lambda) \ge |\alpha|^2$ should hold, where $\alpha \in \mathbb{C}$. The output state $\mathcal{A}_\gamma(\rho)$ has the matrix representation

$$\mathcal{A}_\gamma(\rho) = \begin{pmatrix} 1-(1-\gamma)\lambda & \sqrt{1-\gamma}\alpha \\ \sqrt{1-\gamma}\overline{\alpha} & (1-\gamma)\lambda \end{pmatrix}. \tag{4.5.3}$$

To obtain a physical interpretation of the amplitude damping channel, consider that it can written in the Stinespring form as

$$\mathcal{A}_\gamma(\rho) = \mathrm{Tr}_E[U^\eta(\rho \otimes |0\rangle\langle 0|_E)(U^\eta)^\dagger], \tag{4.5.4}$$

where E is a qubit environment system, $\eta \coloneqq 1-\gamma$, and

$$U^{\eta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{\eta} & \mathrm{i}\sqrt{1-\eta} & 0 \\ 0 & \mathrm{i}\sqrt{1-\eta} & \sqrt{\eta} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.5.5}$$

is unitary. Note that the action of $\mathcal{A}_\gamma$ on the pure states $|0\rangle$ and $|1\rangle$ is, respectively,

$$\begin{aligned} \mathcal{A}_{1-\eta}(|0\rangle\langle 0|_A) &= |0\rangle\langle 0|_B, \\ \mathcal{A}_{1-\eta}(|1\rangle\langle 1|_A) &= (1-\eta)|0\rangle\langle 0|_B + \eta|1\rangle\langle 1|_B. \end{aligned} \tag{4.5.6}$$

A complementary channel $\mathcal{A}_\gamma^c$ for $\mathcal{A}_\gamma$ is defined from

$$\mathcal{A}_\gamma^c(\rho) \coloneqq \mathrm{Tr}_B[U^{1-\gamma}(\rho \otimes |0\rangle\langle 0|_E)(U^{1-\gamma})^\dagger]. \tag{4.5.7}$$

Exercise 4.24

Prove that $\mathcal{A}_\gamma^c = \mathcal{A}_{1-\gamma}$ for all $\gamma \in [0,1]$.

Using the result of Exercise 4.24, we have that the action of the complementary channel $\mathcal{A}_{1-\eta}^c$ on these states is

$$\begin{aligned} \mathcal{A}_{1-\eta}^c(|0\rangle\langle 0|_A) &= |0\rangle\langle 0|_E, \\ \mathcal{A}_{1-\eta}^c(|1\rangle\langle 1|_A) &= \eta|0\rangle\langle 0|_E + (1-\eta)|1\rangle\langle 1|_E. \end{aligned} \tag{4.5.8}$$

We see that whenever the state $|0\rangle\langle 0|$ is input to the channel, the output systems B and E are both in the state $|0\rangle\langle 0|$. On the other hand, if the input state is $|1\rangle\langle 1|$, then B receives a mixed state: with probability $1-\eta$, the state is $|0\rangle\langle 0|$, and with probability η, the state is $|1\rangle\langle 1|$. The situation for E is reversed, receiving $|0\rangle\langle 0|$ with probability η and $|1\rangle\langle 1|$ with probability $1-\eta$. The unitary U^η can thus be viewed as a qubit analogue of a *beamsplitter*, and the amplitude damping channel $\mathcal{A}_{1-\eta}$ can be viewed as a qubit analogue of the *pure-loss bosonic channel*; see Figure 4.5.

A beamsplitter is an optical device that takes two beams of light as input and splits them into two separate output beams, with one of the output beams containing a fraction η of the intensity of the incoming beam and the other output beam

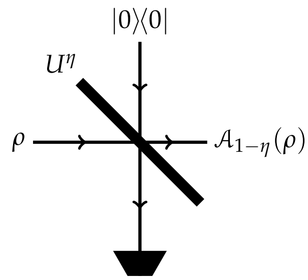


FIGURE 4.5: The amplitude damping channel $\mathcal{A}_{1-\eta}$ can be interpreted, using (4.5.4), as an interaction with a qubit analogue of a bosonic beamsplitter unitary U^η, followed by discarding the output state of the environment. The channel from the sender to the receiver is from the left to the right, while the input and output environment systems are at the top and the bottom, respectively. In the bosonic case, the state $|0\rangle\langle 0|$ of the environmental input arm of the beamsplitter corresponds to the vacuum state, which contains no photons, and the channel to the receiver's end is called the pure-loss bosonic channel.

containing the remaining fraction $1-\eta$ of the incoming intensity. When one of the input ports of the beamsplitter is empty, i.e., is in the vacuum state, the output to the receiver is by definition the pure-loss bosonic channel. In the case of a single incoming photon, the pure-loss channel either transmits the photon with probability η (allowing it to go to the receiver) or reflects it with probability $1-\eta$ (sending it to the environment).

To draw the correspondence between the qubit amplitude damping channel and the pure-loss bosonic channel described above, we can think of the qubit state $|0\rangle$ as the vacuum state and the qubit state $|1\rangle$ as the state of a single photon. It is possible to show that the output states of the amplitude damping channel in (4.5.6) and (4.5.8) for the receiver and environment, respectively, then exactly match the action of the bosonic pure-loss channel on the subspace spanned by $|0\rangle$ and $|1\rangle$ (please consult the Bibliographic Notes in Section 3.4 for further references on this connection). The amplitude damping channel can indeed, therefore, be viewed as the qubit analogue of the bosonic pure-loss channel.

By replacing the initial state $|0\rangle\langle 0|$ of the environment in (4.5.4) with the state

$$\theta_{N_B} \coloneqq (1-N_B)|0\rangle\langle 0| + N_B|1\rangle\langle 1|, \quad N_B \in [0,1], \tag{4.5.9}$$

we define the *generalized amplitude damping channel* $\mathcal{A}_{1-\eta,N_B}$ as

$$\mathcal{A}_{\gamma,N_B}(\rho) \equiv \mathcal{A}_{1-\eta,N_B}(\rho) \coloneqq \operatorname{Tr}_E[U^\eta(\rho \otimes \theta_{N_B})(U^\eta)^\dagger], \tag{4.5.10}$$

where we again use the relation $\gamma = 1 - \eta$. This channel has the following four Kraus operators:

$$A_1 = \sqrt{1 - N_B}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad A_2 = \sqrt{1 - N_B}\begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{4.5.11}$$

$$A_3 = \sqrt{N_B}\begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix}, \quad A_4 = \sqrt{N_B}\begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}. \tag{4.5.12}$$

Note that the amplitude damping channel is a special case of the generalized amplitude damping channel in which the thermal noise $N_B = 0$, so that $\mathcal{A}_{1-\eta} = \mathcal{A}_{1-\eta,0}$.

Exercise 4.25

1. Prove that $\mathcal{A}_{\gamma,N} = (1 - N)\mathcal{A}_{\gamma,0} + N\mathcal{A}_{\gamma,1}$ for all $\gamma, N \in [0, 1]$.
2. Prove that $\mathcal{A}_{\gamma,N} = \mathcal{A}_{\gamma_2,N_2} \circ \mathcal{A}_{\gamma_1,N_1}$, where $\gamma = \gamma_1 + \gamma_2 - \gamma_1\gamma_2$ and $N = \frac{\gamma_1(1-\gamma_2)N_1+\gamma_2 N_2}{\gamma_1+\gamma_2-\gamma_1\gamma_2}$.
3. Using 2., along with the result of Exercise 4.24, prove that the amplitude damping channel $\mathcal{A}_{\gamma,0}$ is degradable for all $\gamma \in \left[0, \frac{1}{2}\right)$, with degrading channel $\mathcal{A}_{\frac{1-2\gamma}{1-\gamma},0}$.

The state θ_{N_B} in (4.5.10) is a qubit thermal state and can be regarded as a qubit analogue of the bosonic thermal state. The latter is a state corresponding to a heat bath with an average number of photons equal to N_B. The generalized amplitude damping channel can then be seen as a qubit analogue of the bosonic thermal noise channel.

Exercise 4.26

Recall the Pauli operators X, Y, Z from (3.2.6), and consider the generalized amplitude-damping channel $\mathcal{A}_{\gamma,N}$, with $\gamma, N \in [0, 1]$. Show that

$$\mathcal{A}_{\gamma,N}(X) = \sqrt{1-\gamma}X, \tag{4.5.13}$$

$$\mathcal{A}_{\gamma,N}(Y) = \sqrt{1-\gamma}Y, \tag{4.5.14}$$

$$\mathcal{A}_{\gamma,N}(Z) = (1-\gamma)Z, \tag{4.5.15}$$

$$\mathcal{A}_{\gamma,N}(\mathbb{1}) = \mathbb{1} + \gamma(1-2N)Z. \tag{4.5.16}$$

From this, conclude that the generalized amplitude-damping channel is covariant with respect to the group defined by the operators $\{\mathbb{1}, Z\}$, so that for all $\gamma, N \in [0, 1]$,

$$\mathcal{A}_{\gamma,N}(Z\rho Z^\dagger) = Z\mathcal{A}_{\gamma,N}(\rho)Z^\dagger \tag{4.5.17}$$

for every state ρ.

4.5.2 Erasure Channel

The qudit *erasure channel* $\mathcal{E}_p$ with *erasure probability* $p \in [0, 1]$ is defined as follows for every $\rho \in \mathrm{L}(\mathbb{C}^d)$, with $d \geq 2$:

$$\mathcal{E}_p(\rho) = (1 - p)\rho + p\mathrm{Tr}[\rho]|e\rangle\langle e|, \tag{4.5.18}$$

where $|e\rangle$ is some unit vector that is orthogonal to all states in the input qudit Hilbert space and $|e\rangle\langle e|$ is called the *erasure state*. For example, if the input qudit Hilbert space is spanned by the standard basis $\{|0\rangle, |1\rangle, \ldots, |d - 1\rangle\}$, then we can set $|e\rangle = |d\rangle$. Thus, the erasure channel takes a linear operator acting on $\mathbb{C}^d$ and outputs a linear operator acting on $\mathbb{C}^{d+1}$.

Exercise 4.27

Verify that a set of Kraus operators for the erasure channel $\mathcal{E}_p$ is

$$\left\{\sqrt{1 - p}(|0\rangle\langle 0| + \cdots + |d - 1\rangle\langle d - 1|), \sqrt{p}|e\rangle\langle 0|, \ldots, \sqrt{p}|e\rangle\langle d - 1|\right\}. \tag{4.5.19}$$

Also, show that the Choi representation of $\mathcal{E}_p$ is

$$\Gamma^{\mathcal{E}_p} = (1 - p)\Gamma_d + p\mathbb{1}_d \otimes |e\rangle\langle e|. \tag{4.5.20}$$

We now discuss how the pure-loss bosonic channel restricted to acting on dual-rail single-photon inputs, an important model for transmission of single photons through an optical fiber, corresponds to a qubit erasure channel. The correspondence is illustrated in Figure 4.6, and recall our earlier discussion from Section 3.2.

Consider a dual-rail optical system, i.e., a quantum optical system with two distinct optical modes A_1 and A_2 representing the input to the channel. As described

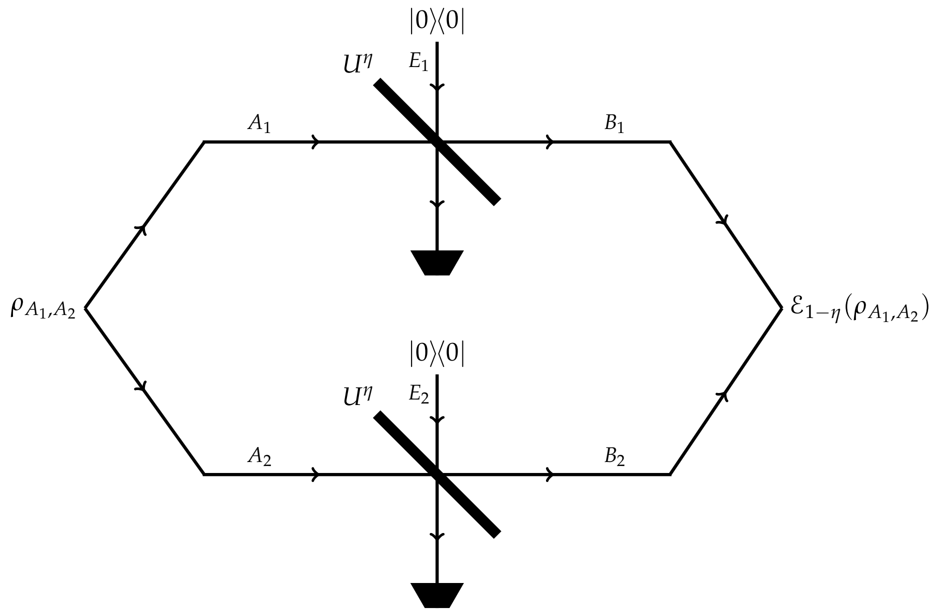


FIGURE 4.6: The qubit erasure channel $\mathcal{E}_{1-\eta}$, for $\eta \in [0,1]$, can be physically realized by using a photonic dual-rail qubit system and passing each of the two modes through a beamsplitter, modeled by the unitary U^η in (4.5.21), such that the input from the environment is the vacuum state.

at the beginning of Section 3.2, the two-dimensional subspace spanned by the states $\{|0,1\rangle_{A_1,A_2}, |1,0\rangle_{A_1,A_2}\}$, consisting of a total of one photon in either one of the two modes, constitutes a qubit system. We also let E_1 and E_2 be two distinct optical modes constituting a dual-rail qubit system representing the environment of the channel. Finally, let B_1 and B_2 be two distinct optical modes spanned by the states $\{|0,0\rangle_{B_1,B_2}, |0,1\rangle_{B_1,B_2}, |1,0\rangle_{B_1,B_2}\}$, so that together B_1 and B_2 constitute a qutrit system.

When the unitary corresponding to a quantum-optical beamsplitter acts on the space spanned by $\{|0,0\rangle, |0,1\rangle, |1,0\rangle\}$, it is equivalent to the upper left 3×3 matrix in (4.5.5). We can thus make use of this fact because the environment state for the bosonic pure-loss channel is prepared in the state $|0\rangle\langle 0|$. Let $U^\eta_{A_iE_i\to B_iE_i}$ then denote the beamsplitter unitary, for $i \in \{1,2\}$, with the following action on the basis $\{|0,0\rangle, |0,1\rangle, |1,0\rangle\}$ of the input and output modes (in that order):

$$U^\eta_{A_iE_i\to B_iE_i} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{\eta} & \mathrm{i}\sqrt{1-\eta} \\ 0 & \mathrm{i}\sqrt{1-\eta} & \sqrt{\eta} \end{pmatrix}. \tag{4.5.21}$$

Letting $\rho_{A_1A_2}$ be an arbitrary qubit state on the two modes A_1 and A_2 defined by

$$\begin{aligned}\rho_{A_1,A_2} = (1-\lambda)|0,1\rangle\langle 0,1|_{A_1,A_2} + \alpha|0,1\rangle\langle 1,0|_{A_1,A_2} + \overline{\alpha}|1,0\rangle\langle 0,1|_{A_1,A_2} \\ + \lambda|1,0\rangle\langle 1,0|_{A_1,A_2},\end{aligned} \tag{4.5.22}$$

the pure-loss bosonic channel on the dual-rail qubit system A_1A_2 is given by

$$\begin{aligned}\mathrm{Tr}_{E_1,E_2}\Big[(U^{\eta}_{A_1E_1\to B_1E_1} \otimes U^{\eta}_{A_2E_2\to B_2E_2})(\rho_{A_1,A_2} \otimes |0,0\rangle\langle 0,0|_{E_1,E_2}) \\ \times(U^{\eta}_{A_1E_1\to B_1E_1} \otimes U^{\eta}_{A_2E_2\to B_2E_2})^{\dagger}\Big].\end{aligned} \tag{4.5.23}$$

Although this has a similar form to (4.5.4), which defines the amplitude damping channel, our particular realization of the qubit system in terms of dual-rail single photons results in a completely different output from that of the amplitude damping channel. In particular, using (4.5.22) along with (4.5.21), it is straightforward to show that

$$\begin{aligned}&\mathrm{Tr}_{E_1,E_2}\Big[(U^{\eta}_{A_1E_1\to B_1E_1} \otimes U^{\eta}_{A_2E_2\to B_2E_2})(\rho_{A_1,A_2} \otimes |0,0\rangle\langle 0,0|_{E_1,E_2}) \\ &\qquad \times(U^{\eta}_{A_1E_1\to B_1E_1} \otimes U^{\eta}_{A_2E_2\to B_2E_2})^{\dagger}\Big] \\ &= \eta\rho_{B_1,B_2} + (1-\eta)|0,0\rangle\langle 0,0|_{B_1,B_2} \\ &= \mathcal{E}_{1-\eta}(\rho_{A_1,A_2}).\end{aligned} \tag{4.5.24}$$

In other words, the pure-loss bosonic channel on a dual-rail qubit system is simply the qubit erasure channel $\mathcal{E}_{1-\eta}$ with erasure probability $1-\eta$ and erasure state $|0,0\rangle\langle 0,0|_{B_1,B_2}$. This means that a dual-rail single-photonic qubit sent through a pure-loss bosonic channel is transmitted to the receiver unchanged with probability η, or it is lost, and replaced by the vacuum state $|0,0\rangle\langle 0,0|$, with probability $1-\eta$.

4.5.3 Pauli Channels

The *Pauli channels* constitute an important class of channels on qubit systems. They are based on the qubit *Pauli operators* X, Y, Z, which we first introduced in (2.2.50) and (2.2.51) and have the following matrix representation in the standard basis $\{|0\rangle, |1\rangle\}$:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.5.25}$$

Recall that the Pauli operators are Hermitian, satisfy $X^2 = Y^2 = Z^2 = \mathbb{1}$, and mutually anti-commute.

A general Pauli channel is one whose Kraus operators are proportional to the Pauli operators, i.e.,

$$\rho \mapsto p_I \rho + p_X X\rho X + p_Y Y\rho Y + p_Z Z\rho Z, \tag{4.5.26}$$

where $p_I, p_X, p_Y, p_Z \geq 0$, $p_I + p_X + p_Y + p_Z = 1$.

Here we highlight two particular Pauli channels of interest.

1. *Dephasing channel*: We let $p_I = 1 - p$ and $p_Z = p$ for $0 \leq p \leq 1$, and $p_X = p_Y = 0$. The action of the dephasing channel is thus

$$\rho \mapsto (1-p)\rho + pZ\rho Z. \tag{4.5.27}$$

   To see why this is called the dephasing channel, consider again a general state ρ of the form (4.5.2) and let $p = \frac{1}{2}$. In this case, we call the channel the *completely dephasing channel*, and it is straightforward to see that

$$\rho = \begin{pmatrix} 1-\lambda & \alpha \\ \overline{\alpha} & \lambda \end{pmatrix} \mapsto \frac{1}{2}\rho + \frac{1}{2}Z\rho Z = \begin{pmatrix} 1-\lambda & 0 \\ 0 & \lambda \end{pmatrix}. \tag{4.5.28}$$

   In other words, the completely dephasing channel eliminates the off-diagonal elements of the input state (when represented in the standard basis, which is the same basis in which Z is diagonal), so that the relative phase between the $|0\rangle$ state and the $|1\rangle$ state vanishes and the state becomes effectively classical. In the more general case when $p \neq \frac{1}{2}$, the effect of the dephasing channel is to reduce the magnitude of the off-diagonal elements:

$$\rho = \begin{pmatrix} 1-\lambda & \alpha \\ \overline{\alpha} & \lambda \end{pmatrix} \mapsto (1-p)\rho + pZ\rho Z \tag{4.5.29}$$

$$= \begin{pmatrix} 1-\lambda & (1-2p)\alpha \\ (1-2p)\overline{\alpha} & \lambda \end{pmatrix}. \tag{4.5.30}$$

2. *Depolarizing Channel*: For $p \in [0,1]$, the depolarizing channel is defined by letting $p_I = 1 - p$ and $p_X = p_Y = p_Z = \frac{p}{3}$, so that

$$\rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{4.5.31}$$

By using the identity

$$\frac{1}{4}\rho + \frac{1}{4}X\rho X + \frac{1}{4}Y\rho Y + \frac{1}{4}Z\rho Z = \mathrm{Tr}[\rho]\pi, \tag{4.5.32}$$

(see Lemma 3.15 and (3.2.100)) we can equivalently write the action of the depolarizing channel as

$$\rho \mapsto \left(1 - \frac{4p}{3}\right)\rho + \frac{4p}{3}\mathrm{Tr}[\rho]\pi, \tag{4.5.33}$$

which has the interpretation that the state of the system is replaced by the maximally mixed state π with probability $\frac{4p}{3}$. Observe, however, that $\frac{4p}{3}$ can be interpreted as a probability only for $0 \leq p \leq \frac{3}{4}$.

4.5.4 Generalized Pauli Channels

Using the Heisenberg–Weyl operators defined in (3.2.47), we can generalize Pauli channels to the qudit case. For all $d \geq 2$, a generalized Pauli channel is defined as

$$\rho \mapsto \sum_{z,x=0}^{d-1} p(z,x) W_{z,x} \rho W_{z,x}^\dagger, \tag{4.5.34}$$

where $p : \{0, 1, \ldots, d-1\}^2 \to [0,1]$ is a probability distribution, so that $0 \leq p(z,x) \leq 1$ for all $z, x \in \{0, 1, \ldots, d-1\}$ and $\sum_{z,x=0}^{d-1} p(z,x) = 1$. In other words, a generalized Pauli channel randomly applies one of the Heisenberg–Weyl operators to the input. The Kraus operators of a generalized Pauli channel are therefore $\left\{\sqrt{p(z,x)}W_{z,x}\right\}_{z,x=0}^{d-1}$.

Exercise 4.28

Show that the Choi state of a generalized Pauli channel is a Bell-diagonal state. (Recall the definition of a Bell-diagonal state in (3.2.59).)

A special case of a generalized Pauli channel is a *d-dimensional dephasing channel*, which is obtained by letting $p(z,x) = 0$ for all $x \in \{1, 2, \ldots, d-1\}$:

$$\rho \mapsto \sum_{z=0}^{d-1} p(z,0) Z(z) \rho Z(z)^\dagger. \tag{4.5.35}$$

For this special case, only the phase operators $Z(z)$ are applied randomly to the input ρ.

Exercise 4.29

For the d-dimensional dephasing channel defined in (4.5.35), prove that, in the standard basis, only the off-diagonal elements of the input state ρ are affected by the channel.

The *d-dimensional depolarizing channel* is defined analogously to the qubit case as follows:

$$\mathcal{D}_p(\rho) := (1-p)\rho + \frac{p}{d^2-1} \sum_{(z,x)\neq(0,0)} W_{z,x}\rho W_{z,x}^{\dagger}, \tag{4.5.36}$$

for all $p \in [0,1]$.

Exercise 4.30

Using (3.2.97), prove that for all $p \in [0,1]$, the action of the depolarizing channel $\mathcal{D}_p$ can be written as

$$\mathcal{D}_p(\rho) = (1-q)\,\rho + q\mathrm{Tr}[\rho]\frac{\mathbb{1}_d}{d} \tag{4.5.37}$$

for every linear operator ρ, where $q = \frac{pd^2}{d^2-1}$.

Exercise 4.31

Prove that the Choi state of the depolarizing channel $\mathcal{D}_p$ is the isotropic state $\rho^{\mathrm{iso};1-p}$ (recall (3.2.130)). Conversely, using (4.2.16), prove that every isotropic state is the Choi state of a depolarizing channel. In other words, prove that for all $p \in [0,1]$,

$$\mathcal{D}_p(X) = d\mathrm{Tr}[(X^{\mathsf{T}} \otimes \mathbb{1})\rho^{\mathrm{iso};1-p}] \tag{4.5.38}$$

for all $X \in \mathrm{L}(\mathbb{C}^d)$.

Exercise 4.32

1. Using the result of Exercise 4.31, along with (3.2.131), show that the depolarizing channel has the following covariance property:

$$\mathcal{D}_p = \mathcal{U}^\dagger \circ \mathcal{D}_p \circ \mathcal{U}, \tag{4.5.39}$$

   for all $p \in [0,1]$ and every unitary U.

2. Using the result of Exercise 4.31, along with (3.2.133), prove that for every quantum channel $\mathcal{N}_{A\to B}$, with $d_A = d_B = d$,

$$\int_U \mathcal{U} \circ \mathcal{N} \circ \mathcal{U}^\dagger \, \mathrm{d}U = \mathcal{D}_{1-p}, \tag{4.5.40}$$

   where $p = \langle \Phi | \Phi^{\mathcal{N}} | \Phi \rangle$ is the *entanglement fidelity* of $\mathcal{N}$, which we define formally in Section 6.4.

4.6 Special Types of Channels

4.6.1 Petz Recovery Map

The following channel plays an important role in variations of the data-processing inequality for the quantum relative entropy and other entropic quantities that we define in the next chapter.

Definition 4.21 Petz Recovery Map

Let $\sigma \in \mathrm{L}(\mathcal{H})$ be a positive semi-definite operator and let $\mathcal{N} : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H}')$ be a quantum channel. The *Petz recovery map for σ and $\mathcal{N}$* is the completely positive and trace-non-increasing map $\mathcal{P}_{\sigma,\mathcal{N}} : \mathrm{L}(\mathcal{H}') \to \mathrm{L}(\mathcal{H}')$ defined as

$$\mathcal{P}_{\sigma,\mathcal{N}}(X) := \sigma^{\frac{1}{2}} \mathcal{N}^\dagger \left(\mathcal{N}(\sigma)^{-\frac{1}{2}} X \mathcal{N}(\sigma)^{-\frac{1}{2}} \right) \sigma^{\frac{1}{2}} \tag{4.6.1}$$

for all $X \in \mathrm{L}(\mathcal{H}')$.

REMARK: If the operator $\mathcal{N}(\sigma)$ is invertible, then the Petz recovery map $\mathcal{P}_{\sigma,\mathcal{N}}$ is a channel. If the operator $\mathcal{N}(\sigma)$ is not invertible, then the inverse $\mathcal{N}(\sigma)^{-\frac{1}{2}}$ is taken on the support on $\mathcal{N}(\sigma)$, following the convention from Section 2.2.8.1. In this latter case, the Petz recovery map $\mathcal{P}_{\sigma,\mathcal{N}}$ is a trace non-increasing map.

The Petz recovery map is indeed completely positive because it is the composition of the following completely positive maps:

1. Sandwiching by the positive semi-definite operator $\mathcal{N}(\sigma)^{-\frac{1}{2}}$.
2. The adjoint $\mathcal{N}^\dagger$ of $\mathcal{N}$.
3. Sandwiching by the positive semi-definite operator $\sigma^{\frac{1}{2}}$.

The Petz recovery map is also trace non-increasing, as we can readily verify. For every positive semi-definite operator X, the following holds

$$\operatorname{Tr}[\mathcal{P}_{\sigma,\mathcal{N}}(X)] = \operatorname{Tr}\left[\sigma^{\frac{1}{2}}\mathcal{N}^\dagger\left(\mathcal{N}(\sigma)^{-\frac{1}{2}}X\mathcal{N}(\sigma)^{-\frac{1}{2}}\right)\sigma^{\frac{1}{2}}\right] \tag{4.6.2}$$

$$= \operatorname{Tr}\left[\sigma\mathcal{N}^\dagger\left(\mathcal{N}(\sigma)^{-\frac{1}{2}}X\mathcal{N}(\sigma)^{-\frac{1}{2}}\right)\right] \tag{4.6.3}$$

$$= \operatorname{Tr}\left[\mathcal{N}(\sigma)\mathcal{N}(\sigma)^{-\frac{1}{2}}X\mathcal{N}(\sigma)^{-\frac{1}{2}}\right] \tag{4.6.4}$$

$$= \operatorname{Tr}\left[\mathcal{N}(\sigma)^{-\frac{1}{2}}\mathcal{N}(\sigma)\mathcal{N}(\sigma)^{-\frac{1}{2}}X\right] \tag{4.6.5}$$

$$= \operatorname{Tr}[\Pi_{\mathcal{N}(\sigma)}X] \tag{4.6.6}$$

$$\leq \operatorname{Tr}[X], \tag{4.6.7}$$

where $\Pi_{\mathcal{N}(\sigma)}$ is the projection onto the support of $\mathcal{N}(\sigma)$, which arises because $\mathcal{N}(\sigma)$ need not be invertible. If X is contained in the support of $\mathcal{N}(\sigma)$, then $\operatorname{Tr}[\Pi_{\mathcal{N}(\sigma)}X] = \operatorname{Tr}[X]$, which means that the Petz recovery channel is trace-preserving for all inputs with support contained in the support of $\mathcal{N}(\sigma)$.

One of the important properties of the Petz recovery channel $\mathcal{P}_{\sigma,\mathcal{N}}$ is that it reverses the action of $\mathcal{N}$ on σ whenever $\mathcal{N}(\sigma)$ is invertible. In particular,

$$\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) = \sigma^{\frac{1}{2}}\mathcal{N}^\dagger\left(\mathcal{N}(\sigma)^{-\frac{1}{2}}\mathcal{N}(\sigma)\mathcal{N}(\sigma)^{-\frac{1}{2}}\right)\sigma^{\frac{1}{2}} \tag{4.6.8}$$

$$= \sigma^{\frac{1}{2}}\mathcal{N}^\dagger(\mathbb{1}_{\mathcal{H}'})\sigma^{\frac{1}{2}} \tag{4.6.9}$$

$$= \sigma, \tag{4.6.10}$$

where we have used the fact that $\mathcal{N}(\sigma)$ is invertible, so that

$$\mathcal{N}(\sigma)^{-\frac{1}{2}}\mathcal{N}(\sigma)\mathcal{N}(\sigma)^{-\frac{1}{2}} = \mathbb{1}_{\mathcal{H}'}. \tag{4.6.11}$$

Then, since $\mathcal{N}$ is a channel, its adjoint is unital, which leads to the final equality.

Remark: The equality in (4.6.10) holds more generally; i.e., it holds even when $\mathcal{N}(\sigma)$ is not invertible. To see this, we use the fact that the projection $\Pi_{\mathcal{N}(\sigma)}$ onto the support of $\mathcal{N}(\sigma)$ satisfies $\Pi_{\mathcal{N}(\sigma)} \leq \mathbb{1}_{\mathcal{H}'}$. Then, we find that

$$\begin{aligned}
\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) &= \sigma^{\frac{1}{2}}\mathcal{N}^\dagger\left(\mathcal{N}(\sigma)^{-\frac{1}{2}}\mathcal{N}(\sigma)\mathcal{N}(\sigma)^{-\frac{1}{2}}\right)\sigma^{\frac{1}{2}} && (4.6.12)\\
&= \sigma^{\frac{1}{2}}\mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)})\sigma^{\frac{1}{2}} && (4.6.13)\\
&\leq \sigma^{\frac{1}{2}}\mathcal{N}^\dagger(\mathbb{1}_{\mathcal{H}'})\sigma^{\frac{1}{2}} && (4.6.14)\\
&= \sigma. && (4.6.15)
\end{aligned}$$

On the other hand, if we let $U : \mathcal{H} \to \mathcal{H}' \otimes \mathcal{H}_E$ be an isometric extension of $\mathcal{N}$, then we can use Lemma 3.3, which implies that $\text{supp}(U\sigma U^\dagger) \subseteq \text{supp}(\mathcal{N}(\sigma) \otimes \mathbb{1}_E)$, and in turn implies that $\Pi_{U\sigma U^\dagger} \leq \Pi_{\mathcal{N}(\sigma)\otimes\mathbb{1}_E}$. Then, for every vector $|\psi\rangle \in \mathcal{H}$, we obtain

$$\begin{aligned}
\langle\psi|\Pi_\sigma|\psi\rangle &= \langle\psi|U^\dagger\Pi_{U\sigma U^\dagger}U|\psi\rangle && (4.6.16)\\
&\leq \langle\psi|U^\dagger(\Pi_{\mathcal{N}(\sigma)} \otimes \mathbb{1}_E)U|\psi\rangle && (4.6.17)\\
&= \text{Tr}[U|\psi\rangle\!\langle\psi|U^\dagger(\Pi_{\mathcal{N}(\sigma)} \otimes \mathbb{1}_E)] && (4.6.18)\\
&= \text{Tr}[\text{Tr}_E[U|\psi\rangle\!\langle\psi|U^\dagger]\Pi_{\mathcal{N}(\sigma)}] && (4.6.19)\\
&= \text{Tr}[\mathcal{N}(|\psi\rangle\!\langle\psi|)\Pi_{\mathcal{N}(\sigma)}] && (4.6.20)\\
&= \text{Tr}[|\psi\rangle\!\langle\psi|\mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)})] && (4.6.21)\\
&= \langle\psi|\mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)})|\psi\rangle. && (4.6.22)
\end{aligned}$$

Since $|\psi\rangle$ is arbitrary, it holds that $\Pi_\sigma \leq \mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)})$. Using this, we find that

$$\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) = \sigma^{\frac{1}{2}}\mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)})\sigma^{\frac{1}{2}} \geq \sigma^{\frac{1}{2}}\Pi_\sigma\sigma^{\frac{1}{2}} = \sigma. \tag{4.6.23}$$

Having shown that $\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) \leq \sigma$ and $\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) \geq \sigma$, we conclude that

$$\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) = \sigma, \tag{4.6.24}$$

even when $\mathcal{N}(\sigma)$ is not invertible.

4.6.1.1 Petz Recovery Channel for Partial Trace

Let the input Hilbert space to the channel $\mathcal{N}$ in the definition of the Petz recovery map be $\mathcal{H} = \mathcal{H}_{AB}$. Then, let $\mathcal{N} = \text{Tr}_B$ be the partial trace over B, and note that (see

Exercise 4.16)

$$\mathcal{N}^\dagger(\sigma_B) = \sigma_A \otimes \mathbb{1}_B. \tag{4.6.25}$$

Indeed, using Definition 2.18 for the adjoint of a superoperator, we have

$$\langle \mathcal{N}^\dagger(\sigma_A), \sigma_{AB}\rangle = \langle \sigma_A \otimes \mathbb{1}_B, \sigma_{AB}\rangle \tag{4.6.26}$$
$$= \mathrm{Tr}[(\sigma_A \otimes \mathbb{1}_B)^\dagger \sigma_{AB}] \tag{4.6.27}$$
$$= \mathrm{Tr}[\sigma_A^\dagger \mathrm{Tr}_B(\sigma_{AB})] \tag{4.6.28}$$
$$= \langle \sigma_A, \mathcal{N}(\sigma_{AB})\rangle. \tag{4.6.29}$$

Therefore, the Petz recovery map corresponding to the partial trace over B is

$$\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}(X_A) = \sigma_{AB}^{\frac{1}{2}} \left(\sigma_A^{-\frac{1}{2}} X_A \sigma_A^{-\frac{1}{2}} \otimes \mathbb{1}_B\right) \sigma_{AB}^{\frac{1}{2}}. \tag{4.6.30}$$

By writing the identity on B as $\mathbb{1}_B = \sum_{j=0}^{d_B-1} |j\rangle\langle j|_B$, we can write the action $\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}$ as

$$\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}(X_A) = \sigma_{AB}^{\frac{1}{2}} \left(\sigma_A^{-\frac{1}{2}} X_A \sigma_A^{-\frac{1}{2}} \otimes \mathbb{1}_B\right) \sigma_{AB}^{\frac{1}{2}} \tag{4.6.31}$$
$$= \sum_{j=0}^{d_B-1} \sigma_{AB}^{\frac{1}{2}} \left(\sigma_A^{-\frac{1}{2}} X_A \sigma_A^{-\frac{1}{2}} \otimes |j\rangle\langle j|_B\right) \sigma_{AB}^{\frac{1}{2}} \tag{4.6.32}$$
$$= \sum_{j=0}^{d_B-1} \sigma_{AB}^{\frac{1}{2}} \left(\sigma_A^{-\frac{1}{2}} \otimes |j\rangle_B\right) X_A \left(\sigma_A^{-\frac{1}{2}} \otimes \langle j|_B\right) \sigma_{AB}^{\frac{1}{2}} \tag{4.6.33}$$
$$= \sum_{j=0}^{d_B-1} K_j X_A K_j^\dagger, \tag{4.6.34}$$

where

$$K_j := \sigma_{AB}^{\frac{1}{2}} \left(\sigma_A^{-\frac{1}{2}} \otimes |j\rangle_B\right). \tag{4.6.35}$$

The operators K_j, for $0 \leq j \leq d_B - 1$, are thus Kraus operators for the Petz recovery map for the partial trace. Using (4.3.9), and letting the environment E be denoted by $\hat{B}$ (since the dimension of the environment in the construction (4.3.9) is the same as the number of Kraus operators, which in this case is equal to the dimension of B), we find that

$$V_{A\to B\hat{B}} = \sum_{j=0}^{d_B-1} \sigma_{AB}^{\frac{1}{2}}(\sigma_A^{-\frac{1}{2}} \otimes |j\rangle_B) \otimes |j\rangle_{\hat{B}} \tag{4.6.36}$$

$$= \sum_{j=0}^{d_B-1} \sigma_{AB}^{\frac{1}{2}}(\sigma_A^{-\frac{1}{2}} \otimes |j\rangle_B \otimes |j\rangle_{\hat{B}}) \tag{4.6.37}$$

$$= (\sigma_{AB}^{\frac{1}{2}} \otimes \mathbb{1}_{\hat{B}})(\sigma_A^{-\frac{1}{2}} \otimes \mathbb{1}_B \otimes \mathbb{1}_{\hat{B}}) \left(\mathbb{1}_A \otimes \sum_{j=0}^{d_B-1} |j,j\rangle_{B\hat{B}} \right) \tag{4.6.38}$$

$$= (\sigma_{AB}^{\frac{1}{2}} \otimes \mathbb{1}_{\hat{B}})(\sigma_A^{-\frac{1}{2}} \otimes \mathbb{1}_{B\hat{B}})(\mathbb{1}_A \otimes |\Gamma\rangle_{B\hat{B}}), \tag{4.6.39}$$

which is an isometric extension of the Petz recovery map $\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}$. Omitting identity operators, this can be written more simply as follows:

$$V_{A\to B\hat{B}} = \sigma_{AB}^{\frac{1}{2}} \sigma_A^{-\frac{1}{2}} |\Gamma\rangle_{B\hat{B}}. \tag{4.6.40}$$

Exercise 4.33

Recall the Bayes theorem from probability theory:

$$p_{X|Y}(x|y) p_Y(y) = p_{Y|X}(y|x) p_X(x), \tag{4.6.41}$$

where X and Y are random variables with joint probability distribution $p_{XY}(x,y)$ over the alphabets $\mathcal{X}$ and $\mathcal{Y}$, and the distributions given above are derived from this joint distribution as

$$p_X(x) = \sum_{y\in\mathcal{Y}} p_{XY}(x,y), \qquad p_Y(y) = \sum_{x\in\mathcal{X}} p_{XY}(x,y), \tag{4.6.42}$$

$$p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}, \qquad p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}. \tag{4.6.43}$$

We now develop a connection between Bayes theorem and the Petz recovery map. Let $\{\rho_x\}_{x\in\mathcal{X}}$ be a set of states, let $\mathcal{N}$ be a channel, and let $\{M_y\}_{y\in\mathcal{Y}}$ be a POVM. Set

$$p_{Y|X}(y|x) = \mathrm{Tr}[M_y \mathcal{N}(\rho_x)]. \tag{4.6.44}$$

1. Show that (4.6.41) is satisfied with

$$p_{X|Y}(x|y) = \mathrm{Tr}[L_x \mathcal{P}(\sigma_y)], \tag{4.6.45}$$

for the set $\{\sigma_y\}_{y\in\mathcal{Y}}$ of states, channel $\mathcal{P}$, and POVM $\{L_x\}_{x\in\mathcal{X}}$ chosen as

$$\sigma_y = \frac{1}{p_Y(y)} [\mathcal{N}(\overline{\rho})]^{\frac{1}{2}} M_y [\mathcal{N}(\overline{\rho})]^{\frac{1}{2}}, \tag{4.6.46}$$

$$\mathcal{P}(\cdot) = \overline{\rho}^{\frac{1}{2}} \mathcal{N}^{\dagger} \left([\mathcal{N}(\overline{\rho})]^{-\frac{1}{2}} (\cdot) [\mathcal{N}(\overline{\rho})]^{-\frac{1}{2}} \right) \overline{\rho}^{\frac{1}{2}}, \tag{4.6.47}$$

$$L_x = p_X(x) [\overline{\rho}]^{-\frac{1}{2}} \rho_x [\overline{\rho}]^{-\frac{1}{2}}, \tag{4.6.48}$$

where

$$\overline{\rho} = \sum_{x \in \mathcal{X}} p_X(x) \rho_x. \tag{4.6.49}$$

2. Verify that $\{\sigma_y\}_{y \in \mathcal{Y}}$ is a set of states, $\mathcal{P}$ is a channel, and $\{L_x\}_{x \in \mathcal{X}}$ is a POVM. For simplicity, suppose that the states $\overline{\rho}$ and $\mathcal{N}(\overline{\rho})$ are positive definite.

4.6.2 LOCC Channels

A very common physical scenario encountered in quantum information theory is one in which two parties, Alice and Bob, who are distantly separated, perform local quantum operations (consisting of channels and/or measurements) at their respective locations and communicate classically with each other in order to transform some initially shared state to some final desired state. This sequence of *local operations and classical communication* is abbreviated *LOCC* and is an important element of many quantum communication protocols, such as entanglement distillation and secret key distillation.

As with every other transformation in quantum theory, an LOCC operation corresponds mathematically to a channel, which we call an *LOCC channel*. Formally, an LOCC channel is defined as follows:

Definition 4.22 LOCC Channel

Let $\mathcal{X}$ be a finite alphabet, let $\{\mathcal{M}^x\}_{x \in \mathcal{X}}$ be a quantum instrument (a set of completely positive trace-non-increasing maps such that $\sum_{x \in \mathcal{X}} \mathcal{M}^x$ is a channel), and let $\{\mathcal{N}^x\}_{x \in \mathcal{X}}$ be a set of quantum channels. Then, a *one-way LOCC channel from Alice to Bob* is the channel $\mathcal{L}^{\rightarrow}_{AB \to A'B'}$ from Alice's initial and final systems A and A' to Bob's initial and final systems B and B', defined as

$$\mathcal{L}^{\rightarrow}_{AB \to A'B'} = \sum_{x \in \mathcal{X}} \mathcal{M}^x_{A \to A'} \otimes \mathcal{N}^x_{B \to B'}. \tag{4.6.50}$$

A *one-way LOCC channel from Bob to Alice* is the channel $\mathcal{L}^{\leftarrow}_{AB\to A'B'}$ defined as

$$\mathcal{L}^{\leftarrow}_{AB\to A'B'} = \sum_{x\in\mathcal{X}} \mathcal{N}^x_{A\to A'} \otimes \mathcal{M}^x_{B\to B'}. \tag{4.6.51}$$

An *LOCC channel* is a composition of a finite, but arbitrarily large number of one-way LOCC channels from Alice to Bob and from Bob to Alice and can be written as

$$\mathcal{L}^{\leftrightarrow}_{AB\to A'B'} = \sum_{y\in\mathcal{Y}} \mathcal{S}^y_A \otimes \mathcal{W}^y_B \tag{4.6.52}$$

for some finite alphabet $\mathcal{Y}$ and sets $\{\mathcal{S}^y\}_{y\in\mathcal{Y}}$, $\{\mathcal{W}^y\}_{y\in\mathcal{Y}}$ of completely positive trace-non-increasing maps such that $\sum_{y\in\mathcal{Y}} \mathcal{S}^y \otimes \mathcal{W}^y$ is trace preserving.

Consider the one-way LOCC channel $\mathcal{L}^{\rightarrow}_{AB\to A'B'}$ from Alice to Bob defined as in (4.6.50). The values in the set $\mathcal{X}$ form the possible messages that can be communicated from Alice to Bob and constitute the "classical communication" part of LOCC. The set $\{\mathcal{M}^x\}_{x\in\mathcal{X}}$ of completely positive trace-non-increasing maps and the set $\{\mathcal{N}^x\}_{x\in\mathcal{X}}$ of quantum channels specify the actions of Alice and Bob for each value $x \in \mathcal{X}$ and constitute the "local operations" part of LOCC. The set $\{\mathcal{M}^x\}_{x\in\mathcal{X}}$ of completely positive trace-non-increasing maps that sum to a channel essentially specifies a quantum instrument. The operations corresponding to these maps can thus be considered probabilisitic since the maps are not trace preserving. In general, the party that performs the quantum instrument determines the direction of classical communication and thus the direction of the LOCC channel. In this case, Alice performs the classical communication since she performs the quantum instrument. The values in the set $\mathcal{X}$ specify the outcomes of the instrument, and Alice communicates the outcome to Bob, who performs the corresponding channel selected from his set $\{\mathcal{N}^x\}_{x\in\mathcal{X}}$ of channels.

In more detail, let ρ_{AB} be the initial state shared by Alice and Bob. Using the definition in (4.4.53) for the channel corresponding to the quantum instrument $\{\mathcal{M}^x\}_{x\in\mathcal{X}}$, the state after applying the quantum instrument is

$$\sum_{x\in\mathcal{X}} \mathcal{M}^x_{A\to A'}(\rho_{AB}) \otimes |x\rangle\langle x|_{X_A}, \tag{4.6.53}$$

where the system X_A stores the classical information corresponding to the outcome of the instrument. Alice then communicates the outcome of the instrument to Bob. This classical communication can be understood via the noiseless classical channel from X_A to X_B defined by

$$\theta_{X_A} \mapsto \sum_{x\in\mathcal{X}} \langle x|_{X_A}\theta_{X_A}|x\rangle_{X_A}|x\rangle\langle x|_{X_B}. \tag{4.6.54}$$

The state in (4.6.53) thus gets transformed to

$$\sum_{x\in\mathcal{X}} \mathcal{M}^x_{A\to A'}(\rho_{AB}) \otimes |x\rangle\langle x|_{X_B}. \tag{4.6.55}$$

Finally, Bob applies the channel specified by

$$\tau_B \otimes \omega_{X_B} \mapsto \sum_{x\in\mathcal{X}} \mathcal{N}^x_{B\to B'}(\tau_B) \otimes \langle x|_{X_B}\omega_{X_B}|x\rangle_{X_B}, \tag{4.6.56}$$

which corresponds to Bob measuring his system X_B in the basis $\{|x\rangle_{X_B}\}_{x\in\mathcal{X}}$ and applying a quantum channel from the set $\{\mathcal{N}^x_{B\to B'}\}_{x\in\mathcal{X}}$ to the system B based on the outcome. The final state is then

$$\sum_{x\in\mathcal{X}} (\mathcal{M}^x_{A\to A'} \otimes \mathcal{N}^x_{B\to B'})(\rho_{AB}), \tag{4.6.57}$$

which is precisely the output of the one-way LOCC channel $\mathcal{L}^{\to}_{AB\to A'B'}$ defined in (4.6.50). We can succinctly write the steps in (4.6.53)–(4.6.56) as

$$\mathcal{L}_{AB\to A'B'}(\rho_{AB}) = (\mathcal{N}_{X_B B\to B'} \circ \mathcal{C}_{X_A\to X_B} \circ \mathcal{M}_{A\to A'X_A})(\rho_{AB}), \tag{4.6.58}$$

where the channel $\mathcal{M}_{A\to A'X_A}$ is defined as

$$\mathcal{M}_{A\to A'X_A}(\xi_A) := \sum_{x\in\mathcal{X}} \mathcal{M}^x_{A\to A'}(\xi_A) \otimes |x\rangle\langle x|_{X_A}, \tag{4.6.59}$$

and the channel $\mathcal{N}_{BX_B\to B'}$ is defined as

$$\mathcal{N}_{BX_B\to B'}(\tau_B \otimes |x\rangle\langle x|_{X_B}) := \mathcal{N}^x_{B\to B'}(\tau_B) \tag{4.6.60}$$

for all $x \in \mathcal{X}$. As indicated above, the channel $\mathcal{C}_{X_A\to X_B}$, defined in (4.6.54), is a noiseless classical channel that transforms the classical register X_A, held by Alice, to the classical register X_B (which is simply a copy of X), held by Bob.

An example of an LOCC channel is illustrated in Figure 4.7. This is an LOCC channel consisting of k rounds of alternating Alice-to-Bob and Bob-to-Alice one-way LOCC channels and is of the form

$$\begin{aligned}\mathcal{L}^{\leftrightarrow}_{A_0B_0\to A_kB_k} = \mathcal{L}^{k,\to}_{A_{k-1}B_{k-1}\to A_kB_k} \circ \mathcal{L}^{k-1,\leftarrow}_{A_{k-2}B_{k-2}\to A_{k-1}B_{k-1}} \circ \cdots \\ \circ\, \mathcal{L}^{2,\leftarrow}_{A_1B_1\to A_2B_2} \circ \mathcal{L}^{1,\to}_{A_0B_0\to A_1B_1}.\end{aligned} \tag{4.6.61}$$

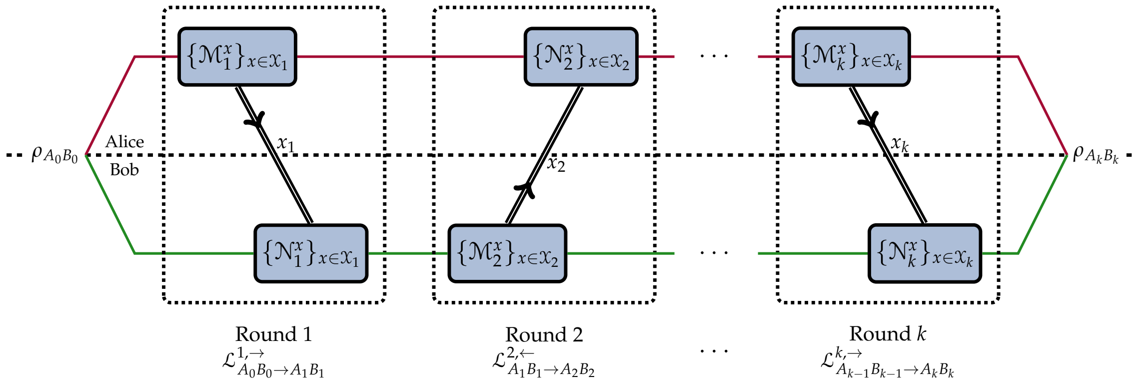


FIGURE 4.7: Illustration of an LOCC channel with k rounds of alternating one-way LOCC channels from Alice to Bob and from Bob to Alice. In each round i, there is a finite alphabet $\mathcal{X}_i$, a set $\{\mathcal{M}_i^x\}_{x\in\mathcal{X}_i}$ of completely positive trace-non-increasing maps that sum to a channel, and a set $\{\mathcal{N}_i^x\}_{x\in\mathcal{X}_i}$ of quantum channels.

For each round i, there is a finite alphabet $\mathcal{X}_i$ consisting of the messages communicated in the round, along with a set $\{\mathcal{M}_i^x\}_{x\in\mathcal{X}_i}$ of completely positive trace-non-increasing maps that sum to a quantum channel and a set $\{\mathcal{N}_i^x\}_{x\in\mathcal{X}_i}$ of quantum channels. In a multi-round LOCC channel such as this one, it possible for the operation sets $\{\mathcal{M}_i^x\}_{x\in\mathcal{X}_i}$ and $\{\mathcal{N}_i^x\}_{x\in\mathcal{X}_i}$ on the ith round to depend on the outcomes and actions taken in previous rounds. The quantum teleportation protocol in Section 5.1 provides a concrete example of a one-way LOCC protocol from Alice to Bob.

Definition 4.23 Separable Channel

A *separable channel* is a quantum channel $\mathcal{S}_{AB\to A'B'}$ such that

$$\mathcal{S}_{AB\to A'B'} = \sum_{x\in\mathcal{X}} \mathcal{C}^x_{A\to A'} \otimes \mathcal{D}^x_{B\to B'} \tag{4.6.62}$$

for some finite alphabet $\mathcal{X}$ and sets of completely positive and trace-non-increasing maps $\{\mathcal{C}^x\}_{x\in\mathcal{X}}$ and $\{\mathcal{D}^x\}_{x\in\mathcal{X}}$ such that $\mathcal{S}$ is trace preserving.

Every separable channel has a set of Kraus operators in product form; i.e., for every separable channel $\mathcal{S}_{AB\to A'B'}$ as in (4.6.62) there exists a finite alphabet $\mathcal{Y}$ and sets $\{C^y_{A\to A'}\}_{y\in\mathcal{Y}}$ and $\{D^y_{B\to B'}\}_{y\in\mathcal{Y}}$ such that

$$\mathcal{S}_{AB\to A'B'}(\rho_{AB}) = \sum_{y\in\mathcal{Y}}(C^y_{A\to A'}\otimes D^y_{B\to B'})\rho_{AB}(C^y_{A\to A'}\otimes D^y_{B\to B'})^\dagger \tag{4.6.63}$$

for all ρ_{AB}.

A key property of a separable channel is that it outputs a separable state if the input state is separable. To see this, consider the separable state $\sigma_{AB} = \sum_{z\in\mathcal{Z}} p(z)\tau^z_A\otimes\omega^z_B$. Then the output state $\mathcal{S}_{AB\to A'B'}(\sigma_{AB})$ is given by

$$\begin{aligned}
&\mathcal{S}_{AB\to A'B'}(\sigma_{AB})\\
&= \sum_{y\in\mathcal{Y}}(C^y_{A\to A'}\otimes D^y_{B\to B'})\left(\sum_{z\in\mathcal{Z}} p(z)\tau^z_A\otimes\omega^z_B\right)(C^y_{A\to A'}\otimes D^y_{B\to B'})^\dagger \qquad (4.6.64)\\
&= \sum_{y\in\mathcal{Y},z\in\mathcal{Z}} p(z)C^y_{A\to A'}\tau^z_A(C^y_{A\to A'})^\dagger\otimes D^y_{B\to B'}\omega^z_B(D^y_{B\to B'})^\dagger, \qquad (4.6.65)
\end{aligned}$$

and is manifestly separable.

Proposition 4.24 LOCC Channels are Separable Channels

From (4.6.52) and (4.6.62), we conclude that every LOCC channel is a separable channel.

The converse of Proposition 4.24 is not true. For example, let us define the following operators:

$$K_1 \coloneqq \frac{1}{\sqrt{2}}|0\rangle\langle 0| + |1\rangle\langle 1|, \quad K_2 \coloneqq |0\rangle\langle 0|, \quad K_3 \coloneqq |1\rangle\langle 1|. \tag{4.6.66}$$

Then, following the notation in (4.6.62), let

$$\mathcal{C}^1_{A\to A'}(\cdot) = K_1(\cdot)K_1^\dagger, \tag{4.6.67}$$
$$\mathcal{C}^2_{A\to A'}(\cdot) = K_2(\cdot)K_2^\dagger, \tag{4.6.68}$$
$$\mathcal{C}^3_{A\to A'}(\cdot) = K_3(\cdot)K_3^\dagger, \tag{4.6.69}$$

and

$$\mathcal{D}^1_{B\to B'}(\cdot) = K_2(\cdot)K_2^\dagger, \tag{4.6.70}$$

$$\mathcal{D}^2_{B\to B'}(\cdot) = K_1(\cdot)K_1^\dagger, \tag{4.6.71}$$
$$\mathcal{D}^3_{B\to B'}(\cdot) = K_3(\cdot)K_3^\dagger. \tag{4.6.72}$$

Then, the map

$$\mathcal{S}_{AB\to A'B'}(\cdot) := \sum_{x=1}^{3}(\mathcal{C}^x_{A\to A'}\otimes\mathcal{D}^x_{B\to B'})(\cdot) \tag{4.6.73}$$
$$= (K_1\otimes K_2)(\cdot)(K_1\otimes K_2)^\dagger + (K_2\otimes K_1)(\cdot)(K_2\otimes K_1)^\dagger \tag{4.6.74}$$
$$+ (K_3\otimes K_3)(\cdot)(K_3\otimes K_3)^\dagger \tag{4.6.75}$$

is a separable channel, but it can be shown that it is not an LOCC channel; please consult the Bibliographic Notes in Section 3.4.

4.6.2.1 LOCC and Separable Simulation of Channels

Given a bipartite quantum channel $\mathcal{N}_{AB\to A'B'}$, an important question related to the physical realization of the channel is whether the channel is an LOCC channel. This amounts to determining whether the channel can be decomposed as in (4.6.52). More generally, one can ask whether a given channel $\mathcal{N}_{A\to B}$ can be *simulated* by an LOCC channel acting on the input and a resource state, a notion that is illustrated in Figure 4.8 and defined below.

Definition 4.25 LOCC-Simulable Channel

A channel $\mathcal{N}_{A\to B}$ is called *LOCC-simulable* with associated resource state $\omega_{RB'}$ if there exists an auxiliary system R and an LOCC channel $\mathcal{L}^{\leftrightarrow}_{RAB'\to B}$ such that, for every state ρ_A,

$$\mathcal{N}(\rho_A) = \mathcal{L}^{\leftrightarrow}_{RAB'\to B}(\rho_A\otimes\omega_{RB'}). \tag{4.6.76}$$

For the LOCC channel $\mathcal{L}^{\leftrightarrow}_{RAB'\to B}$, the input systems RA are Alice's, the input system B' is Bob's, Alice's output system is trivial, and Bob's output system is B.

If a channel $\mathcal{N}_{A\to B}$ is LOCC-simulable with associated resource state $\omega_{RB'}$, it means that Alice and Bob, the sender and receiver of the channel, respectively, can execute an LOCC channel of the form as depicted in Figure 4.7, with the

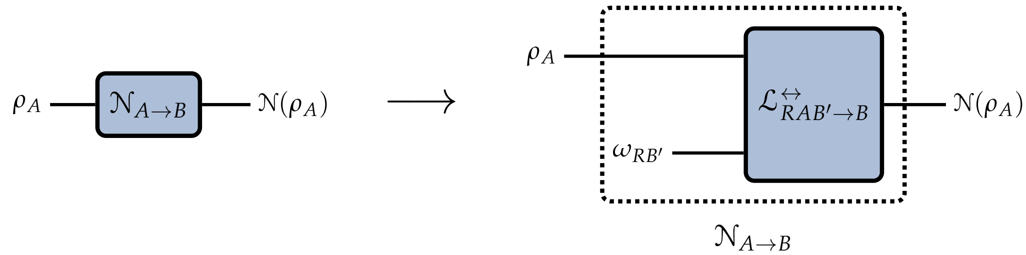


FIGURE 4.8: Depiction of an LOCC-simulable channel $\mathcal{N}_{A\to B}$ with associated resource state $\omega_{RB'}$. The channel $\mathcal{N}_{A\to B}$ can be realized via the LOCC channel $\mathcal{L}^{\leftrightarrow}_{RAB'\to B}$ and the resource state $\omega_{RB'}$, for some auxiliary system R. The LOCC channel $\mathcal{L}^{\leftrightarrow}_{RAB'\to B}$ could in principle consist of a sequence of several rounds of one-way LOCC channels, as depicted in Figure 4.7.

assistance of the auxiliary system R, such that the output on Bob's system at the end is $\mathcal{N}(\rho_A)$. The resource state $\omega_{RB'}$ is fixed, being such that the same resource state can be used for every input state ρ_A on Alice's system. A concrete example of an LOCC simulation of a channel is shown in the context of quantum teleportation in Section 5.1 below.

Due to the fact that separable channels strictly contain LOCC channels, it is sensible to generalize the notion of teleportation simulation even further to this case:

Definition 4.26 Separable-Simulable Channel

A channel $\mathcal{N}_{A\to B}$ is called *separable-simulable* with associated resource state $\omega_{RB'}$ if there exists an auxiliary system R and a separable channel $\mathcal{S}_{RAB'\to B}$ such that, for every state ρ_A,

$$\mathcal{N}(\rho_A) = \mathcal{S}_{RAB'\to B}(\rho_A \otimes \omega_{RB'}). \tag{4.6.77}$$

For the separable channel $\mathcal{S}^{\leftrightarrow}_{RAB'\to B}$, the input systems RA are Alice's, the input system B' is Bob's, Alice's output system is trivial, and Bob's output system is B.

4.6.3 Completely PPT-Preserving Channels

In Definition 3.17, we introduced PPT states as bipartite states ρ_{AB} such that the partial transpose $\mathrm{T}_B(\rho_{AB})$ is positive semi-definite. A class of channels that preserve PPT states are called *completely PPT-preserving channels*, abbreviated as C-PPT-P channels, and we define them formally as follows.

Definition 4.27 Completely PPT-Preserving Channel

A channel $\mathcal{P}_{AB\to A'B'}$ is called *completely PPT-preserving* if the following map is a channel:

$$\mathrm{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \mathrm{T}_B. \tag{4.6.78}$$

Proposition 4.28

Completely PPT-preserving channels preserve the set of PPT states.

PROOF: Suppose that ρ_{AB} is a PPT state and that $\mathcal{P}_{AB\to A'B'}$ is a completely PPT-preserving channel. If we take the partial transpose $\mathrm{T}_{B'}$ on the output state $\sigma_{A'B'} = \mathcal{P}_{AB\to A'B'}(\rho_{AB})$, then we find that

$$\mathrm{T}_{B'}(\sigma_{A'B'}) = (\mathrm{T}_{B'} \circ \mathcal{P}_{AB\to A'B'})(\rho_{AB}) \tag{4.6.79}$$

$$= (\mathrm{T}_B \circ \mathcal{P}_{AB\to A'B'} \circ \mathrm{T}_B)(\mathrm{T}_B(\rho_{AB})). \tag{4.6.80}$$

Since $\mathcal{P}_{AB\to A'B'}$ is completely PPT-preserving, by definition $\mathrm{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \mathrm{T}_B$ is completely positive. Since $\mathrm{T}_B(\rho_{AB})$ is positive, this implies that $\mathrm{T}_{B'}(\sigma_{A'B'})$ is positive, which means that the output state is a PPT state. ■

Proposition 4.29

Every separable channel is a completely PPT-preserving channel.

PROOF: Let $\mathcal{S}_{AB\to A'B'}$ be a separable channel. By definition, it has the form

$$\mathcal{S}_{AB\to A'B'} = \sum_{x\in\mathcal{X}} \mathcal{R}^x_{A\to A'} \otimes \mathcal{W}^x_{B\to B'}, \tag{4.6.81}$$

where $\mathcal{X}$ is a finite alphabet and $\{\mathcal{R}^x\}_{x\in\mathcal{X}}$ and $\{\mathcal{W}^x\}_{x\in\mathcal{X}}$ are sets of completely positive trace non-increasing maps such that $\sum_{x\in\mathcal{X}} \mathcal{R}^x \otimes \mathcal{W}^x$ is trace preserving.

Then,

$$\mathrm{T}_{B'} \circ \mathcal{S}_{AB\to A'B'} \circ \mathrm{T}_B = \sum_{x\in\mathcal{X}} \mathcal{R}^x_{A\to A'} \otimes (\mathrm{T}_{B'} \circ \mathcal{W}^x_{B\to B'} \circ \mathrm{T}_B). \tag{4.6.82}$$

By applying Lemma 4.30 below, we conclude that the maps $\mathrm{T}_{B'} \circ \mathcal{W}^x_{B\to B'} \circ \mathrm{T}_B$ are completely positive for all $x \in \mathcal{X}$, which means that $\mathrm{T}_{B'} \circ \mathcal{S}_{AB\to A'B'} \circ \mathrm{T}_B$ is completely positive. Therefore, $\mathcal{S}_{AB\to A'B'}$ is completely PPT-preserving. ■

As a consequence of Proposition 4.29, it follows that every LOCC channel is a completely PPT-preserving channel, because every LOCC channel is a separable channel.

The next lemma applies to channels that do not have a bipartite structure, i.e., there is no input or output system for Alice.

Lemma 4.30

Let $\mathcal{N}_{B\to B'}$ be a completely positive map. Then the map $\mathrm{T}_{B'} \circ \mathcal{N}_{B\to B'} \circ \mathrm{T}_B$ is completely positive, and its Choi operator is given by the full transpose of the Choi operator for $\mathcal{N}_{B\to B'}$, i.e.,

$$\Gamma^{\mathrm{T}_{B'}\circ\mathcal{N}_{B\to B'}\circ\mathrm{T}_B}_{BB'} = \mathrm{T}(\Gamma^{\mathcal{N}}_{BB'}). \tag{4.6.83}$$

Proof: Let $\Gamma_{B\hat{B}} = |\Gamma\rangle\langle\Gamma|_{B\hat{B}}$, where $|\Gamma\rangle$ is defined in (2.2.36) and $d_{\hat{B}} = d_B$. Observe that $\mathrm{T}_B(\Gamma_{\hat{B}B}) = \mathrm{T}_{\hat{B}}(\Gamma_{\hat{B}B})$, since

$$\begin{aligned}
\mathrm{T}_B(\Gamma_{\hat{B}B}) &= \sum_{i,j=0}^{d_B-1} |i\rangle\langle j|_{\hat{B}} \otimes (|i\rangle\langle j|_B)^{\mathsf{T}} && (4.6.84)\\
&= \sum_{i,j=0}^{d_B-1} |i\rangle\langle j|_{\hat{B}} \otimes |j\rangle\langle i|_B && (4.6.85)\\
&= \sum_{i,j=0}^{d_B-1} (|j\rangle\langle i|_{\hat{B}})^{\mathsf{T}} \otimes |j\rangle\langle i|_B && (4.6.86)\\
&= \mathrm{T}_{\hat{B}}(\Gamma_{\hat{B}B}). && (4.6.87)
\end{aligned}$$

Then the following holds for the Choi representation of $\mathrm{T}_{B'} \circ \mathcal{N}_{B\to B'} \circ \mathrm{T}_B$:

$$\Gamma^{\mathrm{T}_{B'}\circ\mathcal{N}_{B\to B'}\circ\mathrm{T}_B}_{BB'} = (\mathrm{T}_{B'} \circ \mathcal{N}_{B\to B'} \circ \mathrm{T}_B)(\Gamma_{\hat{B}B}) \tag{4.6.88}$$

$$= ((\mathrm{T}_{\hat{B}} \otimes \mathrm{T}_{B'}) \circ \mathcal{N}_{B\to B'})(\Gamma_{\hat{B}B}) \tag{4.6.89}$$
$$= \mathrm{T}(\Gamma^{\mathcal{N}}_{BB'}). \tag{4.6.90}$$

Since the map $\mathcal{N}_{B\to B'}$ is completely positive, its Choi representation $\Gamma^{\mathcal{N}}_{BB'}$ is positive semi-definite. Since positive semi-definiteness is preserved under transposition, we find that $\mathrm{T}(\Gamma^{\mathcal{N}}_{BB'})$ is positive semi-definite, which means that the map $\mathrm{T}_{B'} \circ \mathcal{N}_{B\to B'} \circ \mathrm{T}_B$ is completely positive (by applying Theorem 4.3). ■

As a generalization of the lemma above, we have the following:

Proposition 4.31 Choi States of C-PPT-P Channels

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. The channel $\mathcal{N}_{AB\to A'B'}$ is completely PPT-preserving if and only if its Choi state $\Phi^{\mathcal{N}}_{ABA'B'}$ is a PPT state with respect to the bipartite cut $AA'|BB'$.

PROOF: We begin by proving the only-if part. Suppose that $\mathcal{N}_{AB\to A'B'}$ is completely PPT-preserving. By definition, this implies that $\mathrm{T}_{B'} \circ \mathcal{N}_{AB\to A'B'} \circ \mathrm{T}_B$ is completely positive. By the definition of complete positivity, we conclude that $(\mathrm{T}_{B'} \circ \mathcal{N}_{AB\to A'B'} \circ \mathrm{T}_B)(\Phi_{\bar{A}\bar{B}AB})$ is a positive semi-definite operator, where the Hilbert spaces corresponding to systems $\bar{A}$ and $\bar{B}$ are isomorphic to the Hilbert spaces corresponding to the input systems A and B, respectively. By employing a calculation similar to that in (4.6.84)–(4.6.90), we conclude that

$$(\mathrm{T}_{B'} \circ \mathcal{N}_{AB\to A'B'} \circ \mathrm{T}_B)(\Phi_{\bar{A}\bar{B}A'B'}) = ((\mathrm{T}_{\bar{B}} \otimes \mathrm{T}_{B'}) \circ \mathcal{N}_{AB\to A'B'})(\Phi_{\bar{A}\bar{B}AB}), \tag{4.6.91}$$

which implies that the Choi state $\mathcal{N}_{AB\to A'B'})(\Phi_{\bar{A}\bar{B}AB})$ is a PPT state with respect to the bipartite cut $AA'|BB'$. Running this calculation backwards and making use of Theorem 4.3 establishes the if-part of the proposition. ■

4.6.3.1 PPT Simulation of Channels

Just as we asked whether a given channel $\mathcal{N}_{A\to B}$ can be simulated by an LOCC channel, we can ask whether the channel $\mathcal{N}_{A\to B}$ can be simulated by a completely PPT-preserving channel. This leads to the following definition:

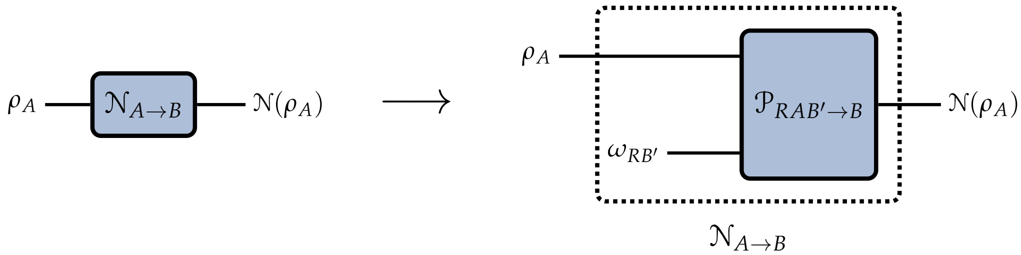


FIGURE 4.9: Depiction of a PPT-simulable channel $\mathcal{N}_{A\to B}$ with associated resource state $\omega_{RB'}$. The channel $\mathcal{N}_{A\to B}$ can be realized via the completely PPT-preserving channel $\mathcal{P}_{RAB'\to B}$ and the resource state $\omega_{RB'}$, for some auxiliary system R.

Definition 4.32 PPT-Simulable Channel

A channel $\mathcal{N}_{A\to B}$ is called *PPT-simulable* with associated resource state $\omega_{RB'}$ if there exists an auxiliary system R and a completely PPT-preserving channel $\mathcal{P}_{RAB'\to B}$ such that for every state ρ_A

$$\mathcal{N}(\rho_A) = \mathcal{P}_{RAB'\to B}(\rho_A \otimes \omega_{RB'}). \tag{4.6.92}$$

For the C-PPT-P channel $\mathcal{P}_{RAB'\to B}$, the input systems RA are Alice's, the input system B' is Bob's, Alice's output system is trivial, and Bob's output system is B.

The definition of a PPT-simulable channel is illustrated in Figure 4.9.

4.6.4 Non-Signaling Channels

One of the main applications considered in this book is communication and, more specifically, when communication is possible or impossible. To this end, suppose that Alice and Bob are connected by means of a bipartite channel $\mathcal{N}_{AB\to A'B'}$. Such a channel is said to be non-signaling from Alice to Bob if it is impossible for Alice and Bob to make use of it for the purpose of Alice to communicate a message to Bob. We give a precise definition as follows:

Definition 4.33 Non-Signaling Channel

A bipartite channel $\mathcal{N}_{AB\to A'B'}$ is non-signaling from Alice to Bob if the following condition holds

$$\mathrm{Tr}_{A'} \circ \mathcal{N}_{AB\to A'B'} = \mathrm{Tr}_{A'} \circ \mathcal{N}_{AB\to A'B'} \circ \mathcal{R}^{\pi}_{A}, \tag{4.6.93}$$

where $\mathcal{R}^{\pi}_{A}$ is a replacer channel, defined as $\mathcal{R}^{\pi}_{A}(\cdot) := \mathrm{Tr}_A[\cdot]\pi_A$, with $\pi_A := \mathbb{1}_A/d_A$ the maximally mixed state on system A.

To interpret the condition in (4.6.93), consider the following. For Bob, the reduced state of his output system B' is obtained by tracing out Alice's output system A'. Note that the reduced state on B' is all that Bob can access at the output in this scenario. If the condition in (4.6.93) holds, then the reduced state on Bob's output system B' has no dependence on Alice's input system. Thus, if (4.6.93) holds, then Alice cannot use $\mathcal{N}_{AB\to A'B'}$ to send a signal to Bob.

Proposition 4.34 Choi Operator of a Non-Signaling Channel

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Let

$$\Gamma^{\mathcal{N}}_{ABA'B'} := \mathcal{N}_{\bar{A}\bar{B}\to A'B'}(\Gamma_{A\bar{A}} \otimes \Gamma_{B\bar{B}}) \tag{4.6.94}$$

be the Choi operator of $\mathcal{N}_{AB\to A'B'}$, where $\bar{A}$ is isomorphic to A and $\bar{B}$ is isomorphic to B. The channel $\mathcal{N}_{AB\to A'B'}$ is non-signaling from Alice to Bob if and only if its Choi operator $\Gamma^{\mathcal{N}}_{ABA'B'}$ satisfies the following condition:

$$\mathrm{Tr}_{A'}[\Gamma^{\mathcal{N}}_{ABA'B'}] = \pi_A \otimes \mathrm{Tr}_{A'A}[\Gamma^{\mathcal{N}}_{ABA'B'}]. \tag{4.6.95}$$

Proof: We begin by proving the if-part. Consider that

$$\mathrm{Tr}_{A'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'}(\Gamma_{A\bar{A}} \otimes \Gamma_{B\bar{B}}) = \mathrm{Tr}_{A'}[\Gamma^{\mathcal{N}}_{ABA'B'}] \tag{4.6.96}$$

Also, consider that

$$\begin{aligned}
&(\mathrm{Tr}_{A'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'} \circ \mathcal{R}^{\pi}_{\bar{A}})(\Gamma_{A\bar{A}} \otimes \Gamma_{B\bar{B}}) \\
&= (\mathrm{Tr}_{A'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'})(\mathbb{1}_A \otimes \pi_{\bar{A}} \otimes \Gamma_{B\bar{B}}) && (4.6.97)\\
&= (\mathrm{Tr}_{A'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'})(\pi_A \otimes \mathbb{1}_{\bar{A}} \otimes \Gamma_{B\bar{B}}) && (4.6.98)
\end{aligned}$$

$$= \pi_A \otimes (\mathrm{Tr}_{A'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'})(\mathbb{1}_{\bar{A}} \otimes \Gamma_{B\bar{B}}) \tag{4.6.99}$$
$$= \pi_A \otimes (\mathrm{Tr}_{AA'} \circ \mathcal{N}_{\bar{A}\bar{B}\to A'B'})(\Gamma_{A\bar{A}} \otimes \Gamma_{B\bar{B}}) \tag{4.6.100}$$
$$= \pi_A \otimes \mathrm{Tr}_{A'A}[\Gamma^{\mathcal{N}}_{ABA'B'}]. \tag{4.6.101}$$

Thus, we conclude that (4.6.93) implies (4.6.95).

To see the other implication, we simply run the reasoning given above backwards and note that two channels are equal if and only if their Choi operators are equal (see Theorem 4.3). ■

A one-way LOCC channel from Bob to Alice is an interesting example of a bipartite channel that is non-signaling from Alice to Bob. Indeed, consider a bipartite channel of the form in (4.6.51), and let us check that the condition in (4.6.93) holds for such a channel. By tracing over the output system A', we find that

$$\mathrm{Tr}_{A'}[\mathcal{L}^{\leftarrow}_{AB\to A'B'}(\rho_{AB})] = \sum_{x\in\mathcal{X}} \mathrm{Tr}_{A'}[(\mathcal{N}^x_{A\to A'} \otimes \mathcal{M}^x_{B\to B'})(\rho_{AB})] \tag{4.6.102}$$
$$= \sum_{x\in\mathcal{X}} \mathcal{M}^x_{B\to B'}(\mathrm{Tr}_A[\rho_{AB}]) \tag{4.6.103}$$
$$= \sum_{x\in\mathcal{X}} \mathcal{M}^x_{B\to B'}(\rho_B). \tag{4.6.104}$$

The second equality follows because $\mathcal{N}^x_{A\to A'}$ is trace preserving for all x. Also, consider that

$$(\mathrm{Tr}_{A'} \circ \mathcal{L}^{\leftarrow}_{AB\to A'B'} \circ \mathcal{R}^{\pi}_A)(\rho_{AB})$$
$$= (\mathrm{Tr}_{A'} \circ \mathcal{L}^{\leftarrow}_{AB\to A'B'})(\pi_A \otimes \rho_B) \tag{4.6.105}$$
$$= \sum_{x\in\mathcal{X}} \mathrm{Tr}_{A'}[(\mathcal{N}^x_{A\to A'} \otimes \mathcal{M}^x_{B\to B'})(\pi_A \otimes \rho_B)] \tag{4.6.106}$$
$$= \sum_{x\in\mathcal{X}} \mathcal{M}^x_{B\to B'}(\mathrm{Tr}_A[\pi_A \otimes \rho_B]) \tag{4.6.107}$$
$$= \sum_{x\in\mathcal{X}} \mathcal{M}^x_{B\to B'}(\rho_B). \tag{4.6.108}$$

Thus, the condition in (4.6.93) holds, and as expected, $\mathcal{L}^{\leftarrow}_{AB\to A'B'}$ is non-signaling from Alice to Bob.

4.7 Summary

4.8 Bibliographic Notes

For an introduction to quantum dynamics, and in particular to the Schrödinger equation and explicit forms for the unitaries $U(t)$ describing the evolution of closed quantum systems, please see the book of Sakurai (1994). The book of Breuer and Petruccione (2002) provides a general introduction to the evolution of open quantum systems as a quantum generalization of classical stochastic processes, and shows how completely positive trace-preserving maps (quantum channels) arise from extensions of the Schrödinger and von Neumann equations to "master equations". We also refer to the tutorial article of Milz and Modi (2021) for a similar exposition on quantum channels from the perspective of stochastic processes and master equations.

The Choi representation of a linear map is named after Choi (1975), with his paper establishing the equivalence between complete positivity of a map and positivity of its Choi representation. Naimark's dilation theorem for positive operator-valued measures was established by Naimark (1940) (see also Gelfand and Naimark (1943)). Stinespring's dilation theorem for completely positive maps was established by Stinespring (1955). An early exposition of the theory of quantum channels was presented by Kraus (1983). Leifer (2007) proposed the concept of conditional states (see also (Leifer and Spekkens, 2013)). Proposition 4.4 was found by Christandl and Winter (2004). The channel in (4.4.13) for reversing the action of an isometric channel was presented by Wilde (2017a). The notions of a complementary channel and a degradable channel were defined by Devetak and Shor (2005). Anti-degradable channels were defined by Caruso and Giovannetti (2006). The quantum erasure channel was defined by Grassl et al. (1997). The connection of the amplitude damping channel to the bosonic pure-loss channel was realized by Giovannetti and Fazio (2005). The quantum instrument formalism was developed by Davies and Lewis (1970) and further developed by Ozawa (1984). Entanglement-breaking channels were defined by Horodecki et al. (2003) and several of their properties were established therein. Hadamard channels were defined by King et al. (2007).

The Petz recovery map was established by Petz (1986b) and Petz (1988) in the context of proving the conditions under which equality holds in the data-processing inequality for the quantum relative entropy. The results therein are stated for von Neumann algebras. A more accessible exposition that considers operators acting only on finite-dimensional Hilbert spaces can be found in Petz (2003) and Mosonyi and Petz (2004) (see also Hayden et al. (2004)).

The paradigm of local operations and classical communication was defined by Bennett et al. (1996c), and its mathematical properties were explored in more detail by Chitambar et al. (2014). See Section 4.3 of Chitambar et al. (2014) for a justification of the example provided in Section 4.6.2 of a separable channel that is not an LOCC channel. Separable channels were defined by Vedral et al. (1997); Barnum et al. (1998). The existence of a separable channel that is not LOCC was found by Bennett et al. (1999a). Completely PPT-preserving channels were defined by Rains (1999a) and further developed by Chitambar et al. (2020). Non-signaling channels were introduced by Beckman et al. (2001) and further considered by Eggeling et al. (2002); Piani et al. (2006).

4.9 Problems

1. Let $\mathcal{N}$ be a quantum channel, and let $\{K_i\}_{i=1}^{r}$ and $\{K'_i\}_{i=1}^{s}$ be two sets of Kraus operators for $\mathcal{N}$. Prove that these sets of Kraus operators are related by an isometry as in (4.3.3).

2. Let $F_{AA'} = \sum_{i,j=0}^{d_A-1} |j,i\rangle\langle i,j|_{AA'}$ be the swap operator, as defined in (3.2.82), and let $\mathcal{N}_{A\to B}$ be a superoperator. Consider the operator

$$F_{AB}^{\mathcal{N}} \coloneqq \mathcal{N}_{A'\to B}(F_{AA'}) = \sum_{i,j=0}^{d_A-1} |j\rangle\langle i|_A \otimes \mathcal{N}(|i\rangle\langle j|_{A'}). \tag{4.9.1}$$

   (a) Prove that $F_{AB}^{\mathcal{N}} = \mathrm{T}_A(\Gamma_{AB}^{\mathcal{N}})$.

   (b) Prove that $\mathrm{Tr}_A[F_{AB}^{\mathcal{N}}] = \mathcal{N}_{A\to B}(\mathbb{1}_A)$.

   (c) If $\mathcal{N}_{A\to B}$ is trace preserving, then prove that $\mathrm{Tr}_B[F_{AB}^{\mathcal{N}}] = \mathbb{1}_A$.

   (d) Prove that $\left\langle X_A^\dagger \otimes Y_B, F_{AB}^{\mathcal{N}} \right\rangle = \langle Y_B, \mathcal{N}_{A\to B}(X_A)\rangle$ for all $X_A \in \mathrm{L}(\mathcal{H}_A)$ and $Y_B \in \mathrm{L}(\mathcal{H}_B)$.

(e) Prove that $F^{\mathcal{N}}_{AB}$ uniquely characterizes $\mathcal{N}$, just as the Choi representation, by showing that

$$\mathcal{N}_{A\to B}(X_A) = \mathrm{Tr}_A[(X_A \otimes \mathbb{1}_B)F^{\mathcal{N}}_{AB}] \tag{4.9.2}$$

for every linear operator X_A.

(f) Prove that $F^{\mathcal{N}}_{AB}$ can be expressed in terms of the adjoint $\mathcal{N}^\dagger$ as follows:

$$F^{\mathcal{N}}_{AB} = \sum_{k,\ell=0}^{d_B-1} \mathcal{N}^\dagger(|k\rangle\langle\ell|_B)^\dagger \otimes |k\rangle\langle\ell|_B. \tag{4.9.3}$$

Conclude that $\mathrm{Tr}_B[F^{\mathcal{N}}_{AB}] = \mathcal{N}^\dagger(\mathbb{1}_B)^\dagger$. Furthermore, if $\mathcal{N}$ is Hermiticity preserving, then conclude that $\mathrm{Tr}_B[F^{\mathcal{N}}_{AB}] = \mathcal{N}^\dagger(\mathbb{1}_B)$ and that $F^{\mathcal{N}}_{AB} = (\mathcal{N}^\dagger)_{B'\to A}(F_{B'B})$, where $F_{B'B} = \sum_{k,\ell=0}^{d_B-1} |\ell, k\rangle\langle k, \ell|_{B'B}$ is the swap operator for system B.

(g) Prove that, for every unitary operator U_A,

$$F^{\mathcal{N}}_{AB} = (\mathcal{U}_A \otimes \mathcal{N}_{A'\to B} \circ \mathcal{U}_{A'})(F_{AA'}). \tag{4.9.4}$$

(h) Prove that $\mathcal{N}_{A\to B}$ is a positive map if and only if $(\langle\psi|_A \otimes \langle\phi|_B)F^{\mathcal{N}}_{AB}(|\psi\rangle_A \otimes |\phi\rangle_B) \geq 0$ for all $|\psi\rangle \in \mathcal{H}_A$ and $|\phi\rangle_B \in \mathcal{H}_B$.

(*Bibliographic Note*: The representation in (4.9.1) was defined by de Pillis (1967). It is sometimes called the *Jamiołkowski representation* of $\mathcal{N}$ due to the work of Jamiołkowski (1972), who proved the necessary and sufficient condition on $F^{\mathcal{N}}_{AB}$ in (h) such that $\mathcal{N}_{A\to B}$ is positive.)

3. Consider the states ζ_{AB} and $\zeta^{\perp}_{AB}$ defined in (3.2.136) and (3.2.137), respectively, and recall the channel $\mathcal{N}^{\rho}_{A\to B}$ defined in (4.2.16) for a given bipartite state ρ_{AB}.

(a) Show that

$$\mathcal{N}^{\zeta}_{A\to B}(X) = \frac{1}{d-1}\left(\mathrm{Tr}[X]\mathbb{1}_d - X^{\mathsf{T}}\right), \tag{4.9.5}$$

$$\mathcal{N}^{\zeta^\perp}_{A\to B}(X) = \frac{1}{d+1}\left(\mathrm{Tr}[X]\mathbb{1}_d + X^{\mathsf{T}}\right) \tag{4.9.6}$$

for every linear operator $X \in \mathrm{L}(\mathcal{H}_A)$.

(*Bibliographic Note*: The quantum channels $\mathcal{N}^{\zeta}$ and $\mathcal{N}^{\zeta^\perp}$ are sometimes called *Werner–Holevo channels*, after Werner and Holevo (2002).)

Chapter 5

Fundamental Quantum Information Processing Tasks

Having studied quantum states, measurements, and channels in detail in the previous two chapters, we are now ready to study three fundamental tasks in quantum information processing: quantum teleportation, quantum super-dense coding, and quantum hypothesis testing. Quantum hypothesis testing has been studied since the late 1960s, with the aim of generalizing (classical) statistical hypothesis testing to the quantum setting. The discovery of quantum teleportation and super-dense coding in the early 1990s demonstrated the practical advantages that entanglement could allow for with respect to communication, and it contributed to the rise of quantum information science as a prominent field of study in both theoretical and experimental physics.

All of the tasks that we study in this chapter provide us with prototypes of some of the quantum communication scenarios that we consider in Parts II and III of this book. In particular, listed below are the tasks and protocols that we study in this chapter and how they are connected to the communication tasks that we study later.

- Quantum teleportation (Section 5.1) is connected to the task of quantum communication (Chapter 14), and in particular to LOCC-assisted quantum communication (Chapter 19).
- Quantum super-dense coding (Section 5.2) is connected to the task of entanglement-assisted classical communication (Chapter 11).
- Quantum hypothesis testing (Section 5.3), in particular state discrimination

in Section 5.3.1 and Section 5.3.2, is connected to classical communication (Chapter 12). Furthermore, asymmetric hypothesis testing in Section 5.3.3 is fundamental to the analysis of every communication scenario that we consider in this book, as it provides us with a method for placing an upper bound on the rate of communication with a finite number of uses of a quantum channel.

Several fundamental quantities used in quantum information theory, particularly in the analysis of quantum communication protocols, arise naturally in the context of quantum hypothesis testing. For example, the *trace distance* arises in terms of the optimal success probability for discriminating between two quantum states in the task of symmetric hypothesis testing (Section 5.3.1), and similarly the *diamond distance* arises in terms of the optimal success probability for discriminating between two quantum channels in the task of symmetric quantum channel hypothesis testing (Section 5.4). The Chernoff divergence quantifies the optimal error exponent for symmetric hypothesis testing of two quantum states in the asymptotic setting (Section 5.3.1.1), and this quantity is related to the *Petz–Rényi relative entropy*, which we define later in Section 7.4. With respect to asymmetric hypothesis testing of two quantum states, the optimal error probability defines the so-called *hypothesis testing relative entropy* (Section 7.9), and in the asymptotic setting the optimal error exponent is given by the *quantum relative entropy* (Section 7.2). The task of hypothesis testing thus provides several fundamental quantities in quantum information theory with an operational meaning, and we devote Chapters 6 and 7 to the detailed study of these and other quantities.

5.1 Quantum Teleportation

Quantum teleportation is a remarkable and fundamental protocol in quantum information theory. The simplest version of the protocol allows two parties, Alice and Bob, to transfer the state of a qubit from Alice to Bob while making use of one shared pair of qubits in a maximally entangled state, along with two bits of classical communication.

5.1.1 Qubit Teleportation Protocol

Before stating the basic teleportation protocol, let us start by introducing a key element of the protocol, the *Bell measurement*.

The Bell measurement is a measurement on two qubits defined by the POVM $\{|\Phi_{z,x}\rangle\langle\Phi_{z,x}| : z, x \in \{0, 1\}\}$, where we recall from (3.2.42) that the two-qubit *Bell states* are defined as

$$|\Phi_{z,x}\rangle_{AB} = (\mathbb{1}_A \otimes X_B^x Z_B^z)|\Phi\rangle_{AB} = (Z_A^z X_A^x \otimes \mathbb{1}_B)|\Phi\rangle_{AB} \tag{5.1.1}$$

for all $x, z \in \{0, 1\}$, so that

$$|\Phi_{0,0}\rangle \coloneqq \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle), \tag{5.1.2}$$

$$|\Phi_{1,0}\rangle \coloneqq \frac{1}{\sqrt{2}}(|0,0\rangle - |1,1\rangle), \tag{5.1.3}$$

$$|\Phi_{0,1}\rangle \coloneqq \frac{1}{\sqrt{2}}(|0,1\rangle + |1,0\rangle), \tag{5.1.4}$$

$$|\Phi_{1,1}\rangle \coloneqq \frac{1}{\sqrt{2}}(|0,1\rangle - |1,0\rangle). \tag{5.1.5}$$

The Bell states are all maximally entangled, and they form an orthonormal basis for the Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$ of two qubits. As such, we have that $\sum_{z,x=0}^{1} |\Phi_{z,x}\rangle\langle\Phi_{z,x}| = \mathbb{1}_2 \otimes \mathbb{1}_2$, so that the set $\{|\Phi_{z,x}\rangle\langle\Phi_{z,x}| : z, x \in \{0, 1\}\}$ is indeed a POVM. Furthermore, the classical bits x and z can be viewed as being the outcomes of the measurement.

We can write the usual computational basis states for two qubits in terms of the Bell states as

$$|0,0\rangle = \frac{1}{\sqrt{2}}(|\Phi_{0,0}\rangle + |\Phi_{1,0}\rangle), \tag{5.1.6}$$

$$|0,1\rangle = \frac{1}{\sqrt{2}}(|\Phi_{0,1}\rangle + |\Phi_{1,1}\rangle), \tag{5.1.7}$$

$$|1,0\rangle = \frac{1}{\sqrt{2}}(|\Phi_{0,1}\rangle - |\Phi_{1,1}\rangle), \tag{5.1.8}$$

$$|1,1\rangle = \frac{1}{\sqrt{2}}(|\Phi_{0,0}\rangle - |\Phi_{1,0}\rangle). \tag{5.1.9}$$

We now detail the teleportation protocol; see Figure 5.1 for a circuit diagram depicting the protocol. The protocol starts with Alice and Bob sharing two qubits in the state $|\Phi\rangle_{AB}$. Alice has an additional qubit, which is in the state $|\psi\rangle_{A'}$, that she wishes to teleport to Bob, where

$$|\psi\rangle_{A'} = \alpha|0\rangle_{A'} + \beta|1\rangle_{A'}, \quad \alpha, \beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1. \tag{5.1.10}$$

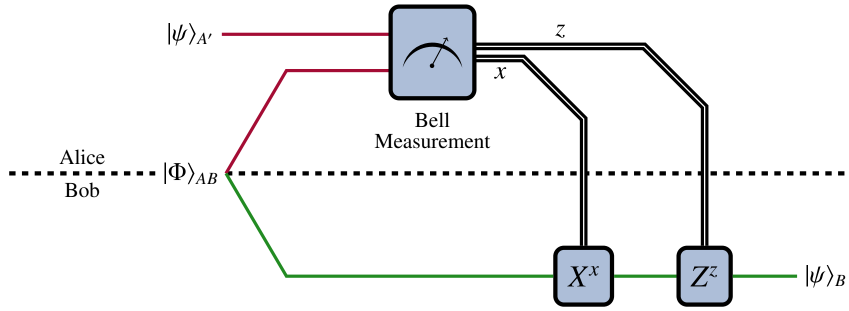


FIGURE 5.1: Circuit diagram for the quantum teleportation protocol. The protocol accomplishes the task of sending a quantum state ψ from Alice to Bob using a shared entangled state and two bits of classical communication. The outcomes (x,z) of Alice's Bell measurement on her qubits A and A' are communicated to Bob, who applies the unitary Z^zX^x on his qubit to transform it to the state ψ that Alice wished to send.

The state $|\psi\rangle_{A'}$ is arbitrary and need not be known to either Alice or Bob. The overall joint state between Alice and Bob at the start of the protocol is therefore

$$|\psi\rangle_{A'} \otimes |\Phi\rangle_{AB} = \frac{1}{\sqrt{2}}(\alpha|0,0,0\rangle_{A'AB} + \alpha|0,1,1\rangle_{A'AB} + \beta|1,0,0\rangle_{A'AB} + \beta|1,1,1\rangle_{A'AB}). \tag{5.1.11}$$

Alice and Bob then proceed as follows.

1. Alice performs a Bell measurement on her two qubits A' and A. To determine the measurement outcomes and their probabilities, it is helpful to write down the initial state (5.1.11) in the Bell basis on Alice's systems. Using (5.1.6)–(5.1.9), we find that

$$\begin{aligned}
&|\psi\rangle_{A'} \otimes |\Phi\rangle_{AB} \\
&= \frac{1}{2}\big(|\Phi_{0,0}\rangle_{A'A} \otimes (\alpha|0\rangle_B + \beta|1\rangle_B) + |\Phi_{1,0}\rangle_{A'A} \otimes (\alpha|0\rangle_B - \beta|1\rangle_B) \\
&\qquad + |\Phi_{0,1}\rangle_{A'A} \otimes (\alpha|1\rangle_B + \beta|0\rangle_B) + |\Phi_{1,1}\rangle_{A'A} \otimes (\alpha|1\rangle_B - \beta|0\rangle_B)\big) \qquad (5.1.12) \\
&= \frac{1}{2}\big(|\Phi_{0,0}\rangle_{A'A} \otimes |\psi\rangle_B + |\Phi_{1,0}\rangle_{A'A} \otimes Z_B|\psi\rangle_B \\
&\qquad + |\Phi_{0,1}\rangle_{A'A} \otimes X_B|\psi\rangle_B + |\Phi_{1,1}\rangle_{A'A} \otimes X_BZ_B|\psi\rangle_B\big). \qquad (5.1.13)
\end{aligned}$$

From this, it is clear that each outcome $(x,z) \in \{0,1\}^2$ of the Bell measurement occurs with equal probablity $\frac{1}{4}$ and that the state of Bob's qubit after the measurement is $X_B^xZ_B^z|\psi\rangle_B$.

Exercise 5.1

Verify (5.1.12) and (5.1.13).

2. Alice communicates to Bob the two classical bits x and z resulting from the Bell measurement.

3. Upon receiving the measurement outcomes, Bob performs X^x and then Z^z on his qubit. The resulting state of Bob's qubit is $|\psi\rangle$.

Although we have described the teleportation protocol using a pure state $|\psi\rangle_{A'}$ as the state being teleported, the protocol applies just as well if the state to be teleported is a mixed state $\rho_{A'}$.

5.1.2 Qudit Teleportation Protocol

The teleportation protocol for qubits described above can be easily generalized to qudits using the Heisenberg–Weyl operators $\{W_{z,x} : 0 \leq z, x \leq d-1\}$ introduced in Definition 3.7. Specifically, recall from (3.2.57) that we define the *two-qudit Bell states* in terms of the Heisenberg–Weyl operators as follows:

$$|\Phi_{z,x}\rangle_{AB} \coloneqq (W_A^{z,x} \otimes \mathbb{1}_B)|\Phi\rangle_{AB}, \tag{5.1.14}$$

which is a direct generalization of (5.1.1), where $|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i,i\rangle_{AB}$. Just like the qubit Bell states, the qudit Bell states form an orthonormal basis for $\mathbb{C}^d \otimes \mathbb{C}^d$ (see Exercise 3.12). This means that the set of operators

$$\{|\Phi_{z,x}\rangle\langle\Phi_{z,x}| : 0 \leq z, x \leq d-1\} \tag{5.1.15}$$

constitutes a POVM, which is the POVM corresponding to the Bell measurement on two qudits.

Now we start, like before, with Alice holding two qudits, one shared with Bob and in the joint state $|\Phi\rangle_{AB}$, and the other in the state $|\psi\rangle_{A'}$, where

$$|\psi\rangle_{A'} = \sum_{i=0}^{d-1} c_i |i\rangle_{A'}, \quad \sum_{i=0}^{d-1} |c_i|^2 = 1. \tag{5.1.16}$$

The state $|\psi\rangle_{A'}$ is the one to be teleported to Bob's system. The starting joint state on the three qudits A', A, and B is

$$|\psi\rangle_{A'} \otimes |\Phi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{i,j=0}^{d-1} c_i |i,j,j\rangle_{A'AB}. \tag{5.1.17}$$

Alice then performs a Bell measurement on her two qudits. By writing the Bell states $|\Phi_{z,x}\rangle_{A'A}$ as

$$|\Phi_{z,x}\rangle_{A'A} = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \mathrm{e}^{\frac{2\pi \mathrm{i}(k+x)z}{d}} |k+x,k\rangle_{A'A}, \tag{5.1.18}$$

for each outcome $(z,x) \in \{0,\dots,d-1\}^2$, we use (3.3.22) to find that the corresponding (unnormalized) post-measurement state of Bob's qudit is

$$\begin{aligned}
&(\langle\Phi_{z,x}|_{A'A} \otimes \mathbb{1}_B)(|\psi\rangle_{A'} \otimes |\Phi\rangle_{AB}) \\
&\quad = \frac{1}{d} \sum_{j,k,\ell=0}^{d-1} c_\ell \mathrm{e}^{-\frac{2\pi \mathrm{i}(k+x)z}{d}} \langle k+x|\ell\rangle\langle k|j\rangle|j\rangle_B && (5.1.19)\\
&\quad = \frac{1}{d} \sum_{k=0}^{d-1} c_{k+x} \mathrm{e}^{-\frac{2\pi \mathrm{i}(k+x)z}{d}} |k\rangle_B && (5.1.20)\\
&\quad = \frac{1}{d} \sum_{k'=0}^{d-1} c_{k'} \mathrm{e}^{-\frac{2\pi \mathrm{i}k'z}{d}} |k'-x\rangle_B && (5.1.21)\\
&\quad = \frac{1}{d} \sum_{k=0}^{d-1} c_{k'} X(-x)Z(-z)|k'\rangle_B && (5.1.22)\\
&\quad = \frac{1}{d} X(-x)Z(-z)|\psi\rangle_B. && (5.1.23)
\end{aligned}$$

Therefore, each outcome occurs with probability $\frac{1}{d^2}$ and the corresponding post-measurement state of Bob's qudit is $X(-x)Z(-z)|\psi\rangle_B$. This means that Bob, upon receiving the two classical values corresponding to the outcome of the Bell measurement, can apply the unitary $Z(z)X(x) = W_{z,x}$ in order to transform the state of his qudit to $|\psi\rangle_B$, completing the teleportation protocol.

Of course, the teleportation protocol works just as well if the state to be teleported is mixed. Also, as shown in Figure 5.2, we can write down the entire

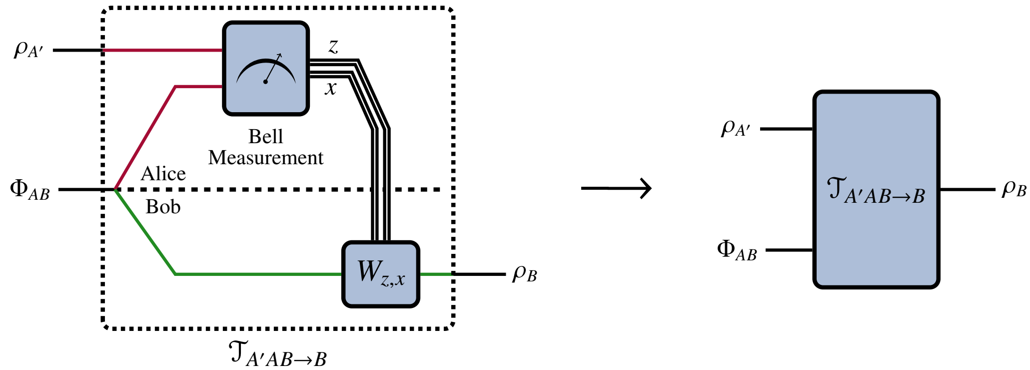


FIGURE 5.2: The qudit teleportation protocol, depicted on the left, can be regarded as an LOCC channel $\mathcal{T}_{A'AB\to B}$ with the input states $\rho_{A'}$ and $|\Phi\rangle\langle\Phi|_{AB}$ and the output state ρ_B, as shown on the right.

teleportation protocol as a one-way LOCC channel $\mathcal{T}_{A'AB\to B}$ from Alice to Bob, with the state $\rho_{A'}$ to be teleported and the state $|\Phi\rangle\langle\Phi|_{AB}$ as the inputs. By the analysis above, the output state received by Bob is exactly the same as the original input state:

$$\mathcal{T}_{A'AB\to B}(\rho_{A'} \otimes |\Phi\rangle\langle\Phi|_{AB}) = \rho_B. \tag{5.1.24}$$

This equation can also be interpreted as follows:

$$\mathcal{T}_{A'AB\to B}((\cdot)_{A'} \otimes |\Phi\rangle\langle\Phi|_{AB}) = \mathrm{id}_{A'\to B}(\cdot), \tag{5.1.25}$$

i.e., that the teleportation protocol simulates the identity channel.

To see that the teleportation protocol can indeed be viewed as a one-way LOCC channel from Alice to Bob, let us explicitly write down the quantum channel $\mathcal{T}_{A'AB\to B}$ defined above in the form of (4.6.58), i.e.,

$$\mathcal{T}_{A'AB\to B} = \mathcal{D}_{BY_1Y_2\to B} \circ \mathcal{C}_{X_1X_2\to Y_1Y_2} \circ \mathcal{E}_{A'A\to X_1X_2} \tag{5.1.26}$$

which we recall is equivalent to the form in (4.6.50). We have that

$$\mathcal{E}_{A'A\to X_1X_2} = \sum_{z,x=0}^{d-1} \mathcal{E}^{z,x}_{A'A\to\varnothing} \otimes |z,x\rangle\langle z,x|_{X_1X_2}, \tag{5.1.27}$$

$$\mathcal{E}^{z,x}_{A'A\to\varnothing}(\cdot) \coloneqq \mathrm{Tr}_{A'A}[|\Phi_{z,x}\rangle\langle\Phi_{z,x}|_{A'A}(\cdot)], \tag{5.1.28}$$

$$\mathcal{C}_{X_1X_2\to Y_1Y_2}(|z,x\rangle\langle z,x|_{X_1X_2}) = |z,x\rangle\langle z,x|_{Y_1Y_2}, \tag{5.1.29}$$

$$\mathcal{D}_{BY_1Y_2\to B}((\cdot)_B \otimes |z,x\rangle\langle z,x|_{Y_1Y_2}) = \mathcal{D}^{z,x}_{B\to B}(\cdot), \tag{5.1.30}$$

$$\mathcal{D}^{z,x}_{B\to B}(\cdot) \coloneqq W_{z,x}(\cdot)W^\dagger_{z,x}. \tag{5.1.31}$$

Exercise 5.2

Combine the quantum channels in (5.1.27)–(5.1.31) according to (5.1.26) and conclude that the channel $\mathcal{T}_{A'AB\to B}$ can be written as

$$\mathcal{T}_{A'AB\to B}(\sigma_{A'AB}) = \sum_{z,x=0}^{d-1} \mathrm{Tr}_{A'A}\left[\Phi^{z,x}_{A'A} W^{z,x}_{B}(\sigma_{A'AB})(W^{z,x}_{B})^{\dagger}\right] \tag{5.1.32}$$

for every state $\sigma_{A'AB}$. Verify that, for the input state $\sigma_{A'AB} = \rho_{A'} \otimes \Phi_{AB}$, we get $\mathcal{T}_{A'AB\to B}(\rho_{A'} \otimes \Phi_{AB}) = \rho_B$, as expected.

We can also connect with the previously defined notion of LOCC simulation of a quantum channel (Definition 4.25). That is, we can understand the teleportation protocol and (5.1.24) as demonstrating that the identity channel is LOCC simulable with associated resource state given by the maximally entangled state. Now, by using the teleportation protocol in conjunction with a quantum channel $\mathcal{N}_{A\to B}$, we find that every quantum channel $\mathcal{N}_{A\to B}$ is LOCC simulable with associated resource state given by the maximally entangled state of an appropriate Schmidt rank. To see this, observe that the channel $\mathcal{N}_{A\to B}$ can be trivially written as $\mathcal{N}_{A\to B} = \mathcal{N}_{B'\to B} \circ \mathrm{id}_{A\to B'}$, where B' is an auxiliary system with the same dimension as A. Then, by (5.1.25), we can simulate the identity channel $\mathrm{id}_{A\to B'}$ using the usual teleportation protocol, so that the overall LOCC channel $\mathcal{L}$ is $\mathcal{N}_{B'\to B} \circ \mathcal{T}_{AA'B'\to B'}$ and the resource state is $|\Phi\rangle\!\langle\Phi|_{A'B'}$, with the dimension of A' equal to the dimension of A. This is illustrated in Figure 5.3. We can also simulate $\mathcal{N}$ via teleportation in a different manner, in which Alice locally applies the channel $\mathcal{N}$ to her input state ρ_A, then teleports the resulting state to Bob. Mathematically, we write this as $\mathcal{N}_{A\to B} = \mathrm{id}_{\hat{A}\to B} \circ \mathcal{N}_{A\to\hat{A}} = \mathcal{T}_{\hat{A}\tilde{A}B\to B} \circ \mathcal{N}_{A\to\hat{A}}$, where $\hat{A}, \tilde{A}$ are auxiliary systems with the same dimension as B. We thus have have the following two ways to represent the action of the channel $\mathcal{N}$ using teleportation:

$$\mathcal{N}_{A\to B}(\rho_A) = \mathcal{N}_{B'\to B}\left(\mathcal{T}_{AA'B'\to B'}(\rho_A \otimes |\Phi\rangle\!\langle\Phi|_{A'B'})\right) \tag{5.1.33}$$

$$= \mathcal{T}_{\hat{A}\tilde{A}B\to B}\left(\mathcal{N}_{A\to\hat{A}}(\rho_A) \otimes |\Phi\rangle\!\langle\Phi|_{\tilde{A}B}\right). \tag{5.1.34}$$

Depending on whether the input dimension is smaller than the output dimension of the channel, there can be a more economical way to perform the simulation. If the channel's output dimension is smaller than its input dimension, then the more economical way to simulate the channel is for Alice to apply $\mathcal{N}_{A\to B}$ first and then for Alice and Bob to perform the teleportation protocol. In this way, they

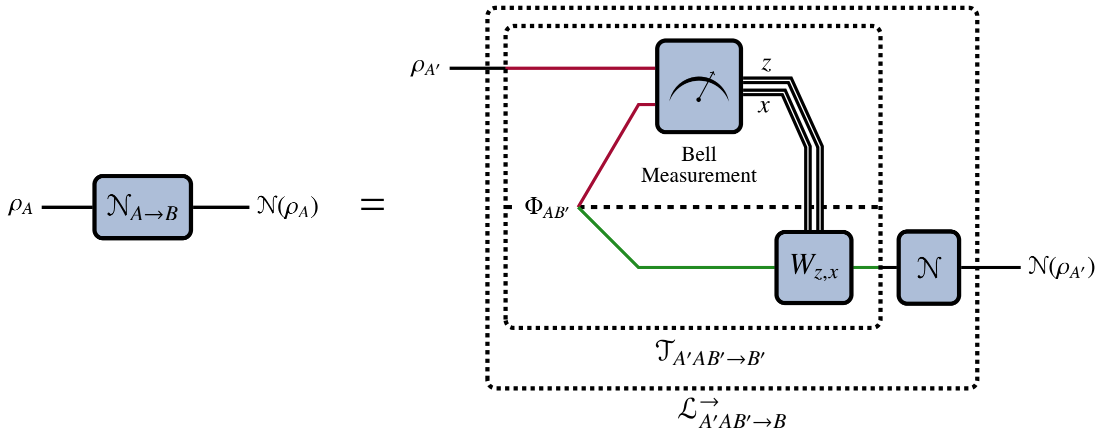


FIGURE 5.3: Every quantum channel $\mathcal{N}$ is teleportation simulable, because Alice and Bob can first perform the usual teleportation protocol, and then Bob can apply the channel $\mathcal{N}$ to the state teleported by Alice. The combination of these two steps can be taken as the LOCC channel $\mathcal{L}^{\to}_{A'AB'\to B}$.

exploit a maximally entangled state of Schmidt rank d_B in order to accomplish the simulation. If the channel's input dimension is smaller than its output dimension, then the more economical way to simulate the channel is for Alice and Bob to perform the teleportation protocol first, and then for Bob to apply the channel locally. In this way, they exploit a maximally entangled state of Schmidt rank d_A in order to accomplish the simulation. As we see in Section 5.1.4, depending on the channel and its symmetries, there can be even more economical methods to simulate a quantum channel via teleportation.

5.1.2.1 Qudit Teleportation Protocol With Respect to a Finite Group

The qudit teleportation protocol outlined above is based on the Heisenberg–Weyl operators $\{W_{z,x} : 0 \leq z, x \leq d-1\}$ acting on a d-dimensional system, which are used to form the qudit Bell states and thus the Bell measurement.

More generally, we can consider an arbitrary finite group G and an irreducible unitary representation $\{U^g\}_{g\in G}$ of G acting on a d-dimensional Hilbert space, where $d^2 \leq |G|$. It then follows from Schur's lemma (see Bibliographic Notes in Section 3.4) that

$$\frac{1}{|G|}\sum_{g\in G} U^g \rho U^{g\dagger} = \mathrm{Tr}[\rho]\frac{\mathbb{1}_d}{d} \tag{5.1.35}$$

for every state ρ. In particular, for bipartite states ρ_{AB} in which the systems A and

B are both d-dimensional, we find that

$$\frac{1}{|G|}\sum_{g\in G}(U_A^g\otimes\mathbb{1}_B)\rho_{AB}(U_A^{g\dagger}\otimes\mathbb{1}_B)=\frac{\mathbb{1}_A}{d}\otimes\operatorname{Tr}_A[\rho_{AB}]. \tag{5.1.36}$$

Now, let us take the maximally entangled state $|\Phi\rangle_{AB}$ and define the states

$$|\Phi^g\rangle_{AB}\coloneqq(U_A^g\otimes\mathbb{1}_B)|\Phi\rangle_{AB}. \tag{5.1.37}$$

We call these states the *generalized Bell states*. We see that these states are a direct generalization of the usual qudit Bell states in (3.2.57).

Exercise 5.3

Using (5.1.36), prove that

$$\frac{1}{|G|}\sum_{g\in G}|\Phi^g\rangle\langle\Phi^g|_{AB}=\frac{\mathbb{1}_{AB}}{d^2}. \tag{5.1.38}$$

By defining the operators

$$M_{AB}^g\coloneqq\frac{d^2}{|G|}|\Phi^g\rangle\langle\Phi^g|_{AB}, \tag{5.1.39}$$

we find that

$$\sum_{g\in G}M_{AB}^g=\mathbb{1}_{AB}. \tag{5.1.40}$$

Since the operators M_{AB}^g satisfy $0\leq M_{AB}^g\leq\mathbb{1}_{AB}$ for all $g\in G$ (the right-most inequality due to the assumption $d^2\leq|G|$), we conclude that the set $\{M_{AB}^g\}_{g\in G}$ constitutes a POVM. This POVM defines the *G-Bell measurement*.

We use the G-Bell measurement defined by the POVM $\{M^g\}_{g\in G}$ in order to construct the generalized teleportation protocol. The protocol proceeds as follows. As before, Alice and Bob start by sharing two qudits in the state $|\Phi\rangle_{AB}$, with Alice holding an extra qudit, i.e., in the state $|\psi\rangle_{A'}$, to be teleported to Bob.

1. Alice performs, on her qudits A and A', the generalized Bell measurement given by the POVM $\{M_{AA'}^g\}_{g\in G}$. For each outcome $g\in G$ of the measurement,

according to (3.3.22) the (unnormalized) post-measurement state of Bob's qudit is

$$\left(\frac{d}{\sqrt{|G|}}\langle\Phi_g|_{A'A}\otimes\mathbb{1}_B\right)(|\psi\rangle_{A'}\otimes|\Phi\rangle_{AB})$$
$$=\frac{d}{\sqrt{|G|}}\left(\langle\Phi|_{A'A}(U_{A'}^{g\dagger}\otimes\mathbb{1}_A)\otimes\mathbb{1}_B\right)(|\psi\rangle_{A'}\otimes|\Phi\rangle_{AB}) \tag{5.1.41}$$
$$=\frac{d}{\sqrt{|G|}}\left(\frac{1}{\sqrt{d}}\sum_{k=1}^{d}\langle k,k|_{A'A}(U_{A'}^{g\dagger}\otimes\mathbb{1}_A)\otimes\mathbb{1}_B\right)$$
$$\times\left(|\psi\rangle_{A'}\otimes\frac{1}{\sqrt{d}}\sum_{k'=1}^{d}|k',k'\rangle_{AB}\right) \tag{5.1.42}$$
$$=\frac{1}{\sqrt{|G|}}\sum_{k,k'=1}^{d}\langle k|_{A'}U_{A'}^{g\dagger}|\psi\rangle_{A'}|k\rangle_B\langle k|k'\rangle \tag{5.1.43}$$
$$=\frac{1}{\sqrt{|G|}}\sum_{k=1}^{d}\langle k|_{A'}U_{A'}^{g\dagger}|\psi\rangle_{A'}|k\rangle_B \tag{5.1.44}$$
$$=\frac{1}{\sqrt{|G|}}U_B^{g\dagger}|\psi\rangle_B \tag{5.1.45}$$

We see that each outcome occurs with probability $\frac{1}{|G|}$, and the post-measurement state of Bob's qudit is $U_B^{g\dagger}|\psi\rangle_B$.

2. Alice communicates the outcome g resulting from the measurement to Bob.

3. Upon receiving the measurement outcome, Bob applies U^g on his qudit. The resulting state of Bob's qudit is $|\psi\rangle_B$.

Observe that the original qudit teleportation protocol is a special case of the generalized teleportation protocol outlined above, in which the group G is $\mathbb{Z}_d\times\mathbb{Z}_d$ and its irreducible projective unitary representation $\{U^g\}_{g\in G}$ is taken to be the set of Heisenberg–Weyl operators. Then, the generalized Bell states Φ^g are precisely the qudit Bell states $\Phi_{z,x}$ defined in (3.2.57). Furthermore, since $|G|=d^2$, the POVM elements $M^g=\frac{d^2}{|G|}|\Phi^g\rangle\langle\Phi^g|$ are the projections on to the qudit Bell states, exactly as in the qudit teleportation protocol.

Exercise 5.4

By following a development similar to that in (5.1.26)–(5.1.31) and Exercise 5.2, verify that the one-way LOCC channel corresponding to the generalized teleportation protocol presented above has the following form analogous to (5.1.32):

$$\mathcal{T}^{G}_{A'AB\to B}(\sigma_{A'AB}) = \sum_{g\in G} \mathrm{Tr}_{A'A}\left[M^{g}_{A'A} U^{g}_{B}(\sigma_{A'AB}) U^{g\dagger}_{B}\right] \tag{5.1.46}$$

for every state $\sigma_{A'AB}$. Conclude that $\mathcal{T}^{G}_{A'AB\to B}(\rho_{A'} \otimes \Phi_{AB}) = \rho_B$, as expected.

5.1.3 Post-Selected Teleportation

Throughout this section, we have described teleportation protocols that involve performing a particular kind of Bell measurement between a system A', whose state is to be teleported, and a system A that is one share of a bipartite system AB in the joint resource state $|\Phi\rangle\langle\Phi|_{AB}$. Based on the outcome of the Bell measurement performed on $A'A$, Bob applies a particular correction operation in order to obtain the initial state of A' in his system B. Thus, although the individual outcomes of the Bell measurement occur with some probability, the overall teleportation protocol is deterministic, due to the correction operations; i.e., it succeeds with probability one.

If Bob does not have the ability to apply correction operations to his system based on the Bell measurement outcomes, then Alice and Bob can perform what is called *post-selected teleportation*. Post-selected teleportation is based on the fact that, in the teleportation protocols that we have considered, Bob does not need to apply a correction operation on his system if the outcome corresponding to $|\Phi\rangle\langle\Phi|_{A'A}$ occurs in the Bell measurement performed on $A'A$. This is due to the fact that the post-measurement state on Bob's system, conditioned on the $|\Phi\rangle\langle\Phi|_{A'A}$ outcome, is precisely the initial state $\rho_{A'}$ to be teleported:

$$\langle\Phi|_{A'A}(\rho_{A'} \otimes |\Phi\rangle\langle\Phi|_{AB})|\Phi\rangle_{A'A} = \frac{1}{d^2}\rho_B, \tag{5.1.47}$$

where $d = d_A = d_B$. This is a special case of (4.2.14) in which $\mathcal{N}_{A'\to B} = \mathrm{id}_{A'\to B}$ and $X_{RA'} = \rho_{A'}$. If we thus modify the teleportation protocol such that we consider

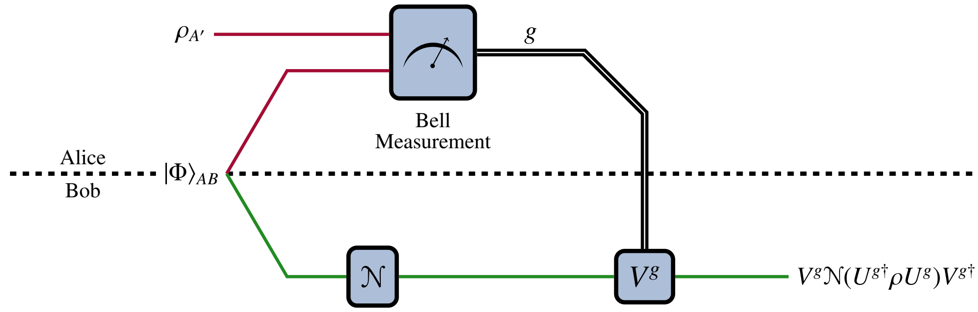


FIGURE 5.4: Circuit diagram of a modified teleportation protocol in which Bob applies a channel $\mathcal{N}$ to his qudit before performing a correction operation V^g, which is a unitary operation from the set $\{V^g : g \in G\}$ of unitary operators.

the $|\Phi\rangle\langle\Phi|_{A'A}$ outcome of the Bell measurement as a "success" and the rest of the outcomes as a "failure", then we obtain post-selected teleportation. Post-selected teleportation is probabilistic by definition. In particular, from (5.1.47), we see that it succeeds with probability $\frac{1}{d^2}$.

5.1.4 Teleportation-Simulable Channels

Let us now consider the even more general protocol depicted in Figure 5.4. Let G be a finite group. As before, Alice and Bob start with a shared pair of qudits in the state $|\Phi\rangle\langle\Phi|_{AB}$, while Alice holds an extra qudit in the state $\rho_{A'}$ to be teleported to Bob. Unlike the teleportation protocol above, however, Bob applies the channel $\mathcal{N}$ to his qudit before he receives the results of the Bell measurement. Once he receives the measurement results, he applies the unitary operation V^g from the set $\{V^g : g \in G\}$ of pre-determined unitary operators constituting a projective unitary representation of G.

The initial tripartite joint state of the protocol is

$$\rho_{A'} \otimes (\mathbb{1}_A \otimes \mathcal{N}_B)(|\Phi\rangle\langle\Phi|_{AB}). \tag{5.1.48}$$

Alice performs the same generalized Bell measurement as before on A and A', which we recall has the POVM $\{\Pi^g_{AA'}\}_{g\in G}$ with elements $\Pi^g_{AA'}$ defined in (5.1.39). Recall that this POVM corresponds to an irreducible projective unitary representation of G given by $\{U^g\}_{g\in G}$. Since the Bell measurement operates only on the systems A' and A, we can bring them inside the action of $\mathcal{N}$ on Bob's share of the state

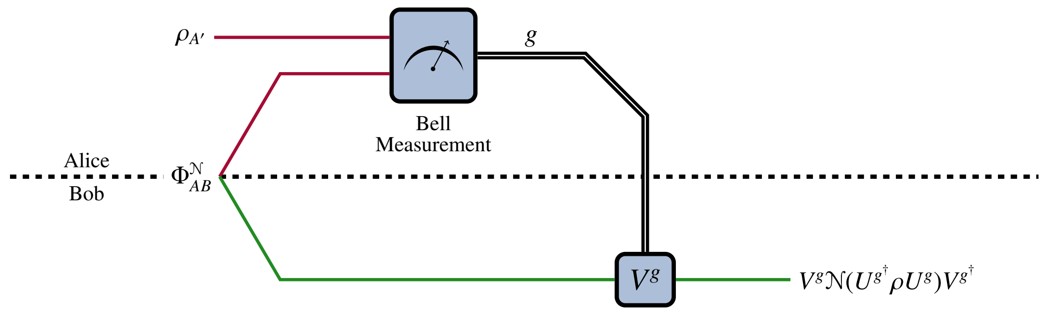


FIGURE 5.5: A mathematically equivalent way of describing the protocol in Figure 5.4. In this case, Alice and Bob start with the Choi state $\Phi^{\mathcal{N}}_{AB}$ of the channel $\mathcal{N}$ instead of the maximally entangled state $|\Phi\rangle\langle\Phi|_{AB}$.

$|\Phi\rangle\langle\Phi|_{AB}$. This means that the analysis for the qudit teleportation protocol from Section 5.1.2.1 carries over exactly in this case. In other words, each outcome $g \in G$ occurs with an equal probability of $\frac{1}{|G|}$, and the post-measurement state on Bob's qudit corresponding to the outcome g is

$$\mathcal{N}(U^{g^\dagger}\rho U^g). \tag{5.1.49}$$

After Bob applies the unitary V^g, the state of Bob's qudit at the end of the protocol is

$$V^g\mathcal{N}(U^{g^\dagger}\rho U^g)V^{g^\dagger}. \tag{5.1.50}$$

This occurs with probability $\frac{1}{|G|}$ for all $g \in G$.

Now, observe that the state (5.1.48) can be written as

$$\rho_{A'} \otimes \Phi^{\mathcal{N}}_{AB}, \tag{5.1.51}$$

where we recall that $\Phi^{\mathcal{N}}_{AB} = (\mathrm{id}_A \otimes \mathcal{N}_B)(|\Phi\rangle\langle\Phi|_{AB})$ is the Choi state of the channel $\mathcal{N}$. In other words, the protocol depicted in Figure 5.4 is mathematically equivalent to the teleportation protocol over a group G outlined above, except that instead of starting with the shared maximally entangled state $|\Phi\rangle_{AB}$, Alice and Bob start with the shared state $\Phi^{\mathcal{N}}_{AB}$. This equivalent protocol is depicted in Figure 5.5.

If Bob discards the classical message g at the end of the protocol, then the state of his system is given by

$$\frac{1}{|G|}\sum_{g\in G} V^g\mathcal{N}(U^{g^\dagger}\rho U^g)V^{g^\dagger}. \tag{5.1.52}$$

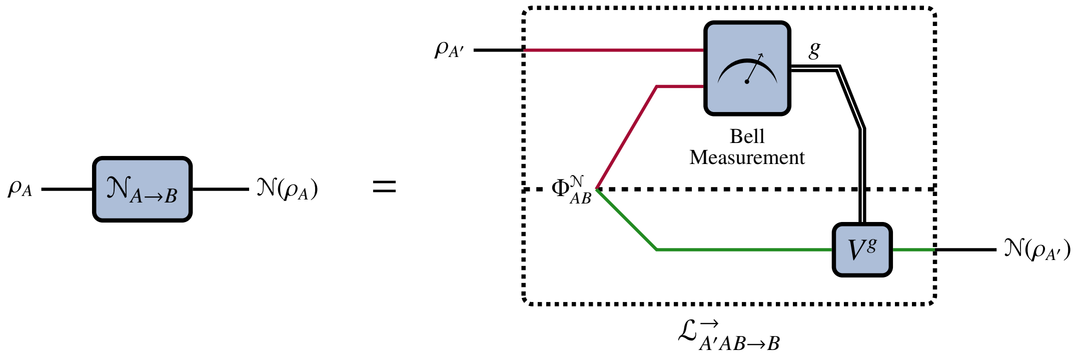


FIGURE 5.6: Teleportation simulation of a G-covariant channel $\mathcal{N}$, where the operators $\{V^g : g \in G\}$ form a unitary representation of G on the output space of $\mathcal{N}$.

Recall from (4.4.127) that this state is simply the output state of the twirl of $\mathcal{N}$ with respect to the unitary representations $\{U^g\}_{g\in G}$ and $\{V^g\}_{g\in G}$, because the twirled channel $\overline{\mathcal{N}}$ is a symmetrized version of the original channel $\mathcal{N}$. Thus, the generalized teleportation protocol gives an explicit procedure for implementing a channel twirl by implementing the teleportation protocol using the Choi state of the channel as the resource state.

Suppose now that the channel $\mathcal{N}$ satisfies the group covariance property from Definition 4.18 for all $g \in G$. In this case, we see that $\mathcal{N}(U^{g\dagger}\rho U^g) = V^{g\dagger}\mathcal{N}(\rho)V^g$ for every outcome g of Alice's generalized Bell measurement. Therefore, after Bob applies V^g, the state of his qudit is $\mathcal{N}(\rho)$. This generalized teleportation protocol therefore effectively applies the channel $\mathcal{N}$ to the state $\rho_{A'}$ and transfers the resulting state to Bob's qudit; see Figure 5.6. We say that the teleportation protocol *simulates* the action of the channel $\mathcal{N}$ on the input state $\rho_{A'}$. As stated earlier, in this sense, the original teleportation protocol can be regarded as a way to simulate the identity channel.

The notion of simulation of a channel by a teleportation protocol can be extended to a one-way LOCC channel $\mathcal{L}^{\to}$, as introduced in Definition 4.22, to obtain the following definition.

Definition 5.1 Teleportation-Simulable Channel

A channel $\mathcal{N}_{A\to B}$ is called *teleportation-simulable* with associated resource state ω_{RB} if there exists a one-way LOCC channel $\mathcal{L}^{\to}_{RAB'\to B}$ such that, for every

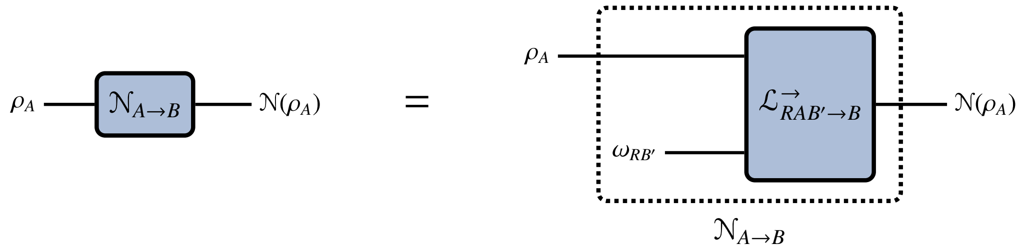


FIGURE 5.7: Depiction of a teleportation-simulable channel with associated resource state $\omega_{RB'}$. The teleportation-simulable channel $\mathcal{N}_{A\to B}$ can be realized via the interaction LOCC channel $\mathcal{L}^{\to}_{RAB\to B}$ and the resource state $\omega_{RB'}$.

input state ρ_A,

$$\mathcal{N}(\rho_A) = \mathcal{L}^{\to}_{RAB'\to B}(\rho_A \otimes \omega_{RB'}). \tag{5.1.53}$$

Figure 5.7 illustrates the concept of a teleportation-simulable channel. Note that in (5.1.53) the resource state $\omega_{RB'}$ is fixed, as well as the LOCC channel $\mathcal{L}^{\to}$. Both sides of the equation should thus be regarded as functions of ρ_A.

From the discussions above, we conclude that every group-covariant channel is teleportation-simulable, where the one-way LOCC channel $\mathcal{L}$ is simply the teleportation protocol with respect to the group, and the resource state is the Choi state of the channel.

5.2 Quantum Super-Dense Coding

We now discuss the quantum super-dense coding protocol. This protocol can be viewed as a "dual" to the quantum teleportation protocol in the following sense: while in the basic quantum teleportation protocol, Alice and Bob make use of two qubits in the entangled state vector $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$ and two bits of classical information to simulate a noiseless qubit channel, in quantum super-dense coding they make use of the shared entangled state $|\Phi^+\rangle$ along with one use of a noiseless qubit channel to communicate *two* bits of classical information. This is remarkable because, without the shared entanglement and only one use of the noiseless qubit channel, they can communicate at most only one bit of classical information. The quantum super-dense coding protocol thus represents one of the simplest examples in which prior shared entanglement provides an advantage for

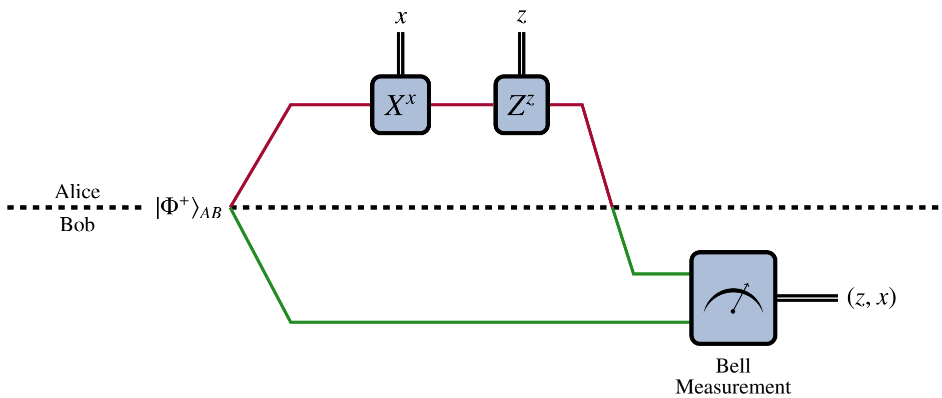


FIGURE 5.8: Circuit diagram for the super-dense coding protocol. Using the bits (z, x) that she wishes to send, Alice applies the appropriate Pauli X and/or Z operators to her share A of the maximally entangled qubits that are in the state $|\Phi^+\rangle_{AB}$ and sends it through a noiseless qubit channel to Bob. Bob then performs a Bell measurement on the two qubits to recover the encoded bits (z, x).

classical communication.

Let us now go through the quantum super-dense coding protocol. See Figure 5.8 for a depiction of the protocol. Alice wishes to send two classical bits $(z, x) \in \{0, 1\}^2$ to Bob by making use of a shared pair of qubits in the maximally entangled state $|\Phi^+\rangle_{AB}$ and one use of a noiseless qubit channel. Depending on the bits she wishes to send, she performs the following operations on her share of the entangled qubits:

- To send the bits $(0, 0)$, she does nothing.
- To send the bits $(0, 1)$, she applies the Pauli X operator to her qubit, transforming the joint state $|\Phi^+\rangle_{AB}$ to $|\Psi^+\rangle_{AB}$.
- To send the bits $(1, 0)$, she applies the Pauli Z operator to her qubit, transforming the joint state $|\Phi^+\rangle_{AB}$ to $|\Phi^-\rangle_{AB}$.
- To send the bits $(1, 1)$, she applies the X operator followed by the Z operator, so that the joint state becomes $|\Psi^-\rangle_{AB}$.

After applying the appropriate operation, Alice sends her qubit to Bob with the one allowed use of a noiseless qubit channel.

Bob now holds both qubits, and they are in one of the four Bell states

$$|\Phi_{z,x}\rangle_{AB} = (Z_A^z X_A^x \otimes \mathbb{1}_B)|\Phi^+\rangle_{AB} \tag{5.2.1}$$

depending on the bits (z, x) Alice sent. Bob then performs a Bell measurement on his two qubits, and the outcome of this measurement consists precisely of the bits (z, x) that Alice wished to send.

The super-dense coding protocol has a simple generalization to the qudit case. In this case, Alice and Bob share the qudit Bell state $|\Phi\rangle_{AB}$ before communication begins, and by applying one of the d^2 Heisenberg–Weyl operators $W_{z,x}$ from (3.2.47) on her share of the state, Alice can rotate the global state to one of the d^2 qudit Bell states in (3.2.57). After Alice sends her share of the encoded state over a noiseless qudit channel to Bob, Bob can then perform the qudit Bell measurement to decode which of the d^2 messages Alice transmitted.

5.3 Quantum Hypothesis Testing

We now consider the task of quantum hypothesis testing, which is a generalization of classical statistical hypothesis testing to the quantum setting. In the quantum setting, the statistical hypotheses are represented by the states of a particular quantum system[1], and the task is to determine which of the hypotheses is "true", i.e., to determine the state of the quantum system.

To be more specific, consider the following scenario. Bob is given a quantum system by Alice, which is either in the state ρ or in the state σ, and his task is to determine in which state the system has been prepared. Bob's strategy consists of performing a measurement of the system, described by the POVM $\{M_\rho, M_\sigma\}$ (so that $M_\rho, M_\sigma \geq 0$ and $M_\rho + M_\sigma = \mathbb{1}$), and then guessing "$\rho$" if the outcome corresponds to M_ρ and guessing "σ" if the outcome corresponds to M_σ; see Figure 5.9. Of course, Bob's guess might not always be correct, and there are two types of errors that can occur:

1. *Type-I Error*: Bob guesses "σ", but the system is in the state ρ. The probability of this occurring is $\text{Tr}[M_\sigma \rho]$.

[1]The hypotheses can be represented by quantum channels more generally, as we detail in Section 5.4.

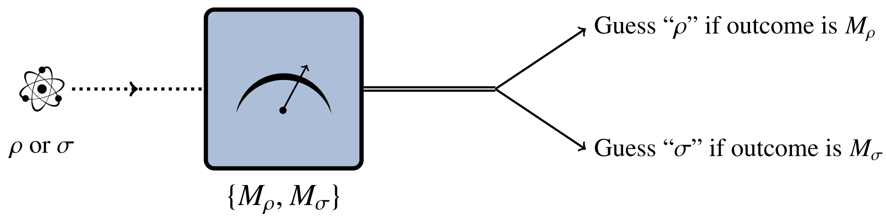


FIGURE 5.9: In quantum hypothesis testing, a given quantum system is known to be either in the state ρ or the state σ. The most general strategy to determine the state of the system consists of measuring it according to a two-outcome POVM $\{M_\rho, M_\sigma\}$. If the outcome corresponding to M_ρ occurs, then we guess that the system is in the state ρ, and if the outcome corresponding to M_σ occurs, then we guess that the system is in the state σ. The goal is to minimize the probability of error of this general strategy.

2. *Type-II Error*: Bob guesses "ρ", but the system is in the state σ. The probability of this occurring is $\text{Tr}[M_\rho \sigma]$.

In order to obtain an optimal strategy for Bob, there are two cases that are typically considered.

- *Symmetric Case*: Also called *quantum state discrimination*, in this setting, Bob has some prior knowledge about the state he is given. Specifically, he knows that the state is ρ with probability $\lambda \in [0, 1]$ and σ with probability $1 - \lambda$. The goal is then to minimize the average of the type-I and type-II error probabilities with respect to this probability distribution. In other words, letting $M \equiv M_\rho$, the goal is to minimize the function

$$p_{\text{err}}(\lambda, \rho, \sigma, M) := \lambda \text{Tr}[(\mathbb{1} - M)\rho] + (1 - \lambda)\text{Tr}[M\sigma]. \tag{5.3.1}$$

  The optimization problem we are interested in is thus:

$$\begin{aligned} \text{minimize} \quad & p_{\text{err}}(\lambda, \rho, \sigma, M) \\ \text{subject to} \quad & 0 \leq M \leq \mathbb{1}, \end{aligned} \tag{5.3.2}$$

  where the minimization is with respect to operators $0 \leq M \leq \mathbb{1}$ representing the two-outcome POVM $\{M, \mathbb{1} - M\}$. We discuss symmetric hypothesis testing, and the optimization problem in (5.3.2), in detail in Section 5.3.1.

- *Asymmetric Case*: In this setting, the goal is to minimize the type-II error probability, given an upper bound on the type-I error probability. In other words, letting $M \equiv M_\rho$, the optimal measurement is given by solving the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \text{Tr}[M\sigma] \\ \text{subject to} \quad & \text{Tr}[(\mathbb{1} - M)\rho] \leq \varepsilon, \\ & 0 \leq M \leq \mathbb{1}, \end{aligned} \tag{5.3.3}$$

where $\varepsilon \in [0,1]$ is the upper bound on the type-I error probability, and the optimization is with respect to every operator M satisfying $0 \leq M \leq \mathbb{1}$, representing the two-outcome POVM $\{M, \mathbb{1} - M\}$. We discuss asymmetric hypothesis testing, and the optimization problem in (5.3.3), in detail in Section 5.3.3.

Exercise 5.5

Consider a very simple hypothesis testing strategy in which Bob discards the state of the quantum system and simply guesses "ρ" with some probability $q \in [0,1]$ and "σ" with probability $1 - q$.

1. What is the POVM corresponding to this strategy?
2. Evaluate the type-I and type-II error probabilities for this strategy.
3. If, in the symmetric setting, the prior probability for the state ρ is $\lambda \in [0,1]$, then evaluate the error probability in (5.3.1) for this strategy.

Now, suppose that Bob is given several copies, say $n \geq 1$, of a quantum system, each one of which is either in the state ρ or the state σ. His strategy to determine the state can now make use of these multiple copies in an adaptive manner, for example, and could allow the error probabilities to go below the "single-shot" ($n = 1$) error probabilities defined above. Since Bob ultimately has to make a decision between ρ and σ, his strategy is still described by a two-outcome POVM, which we denote by $\{M_\rho^{(n)}, M_\sigma^{(n)}\}$. This setting of hypothesis testing with multiple copies is depicted in Figure 5.10. The type-I and type-II error probabilities are defined in an analogous manner as before. Specifically, the type-I error is $\text{Tr}[M_\sigma^{(n)}\rho^{\otimes n}]$ and the type-II error is $\text{Tr}[M_\rho^{(n)}\sigma^{\otimes n}]$. In the symmetric case, if $\lambda \in [0,1]$ is the probability that each system is in the state ρ, then the error probability is

$$\begin{aligned} &\lambda \text{Tr}[M_\sigma^{(n)}\rho^{\otimes n}] + (1 - \lambda)\text{Tr}[M_\rho^{(n)}\sigma^{\otimes n}] \\ &\quad = \lambda \text{Tr}[(\mathbb{1}^{\otimes n} - M^{(n)})\rho^{\otimes n}] + (1 - \lambda)\text{Tr}[M^{(n)}\sigma^{\otimes n}] \end{aligned} \tag{5.3.4}$$

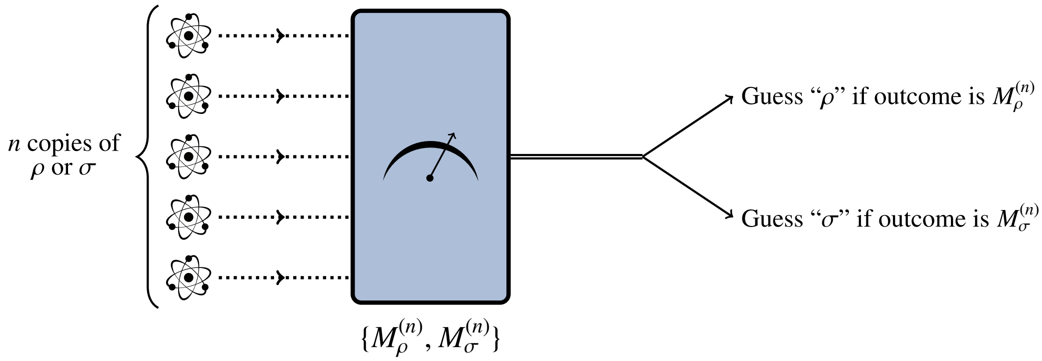


FIGURE 5.10: Quantum hypothesis testing with $n \geq 1$ copies of the state. As in the case of one copy shown in Figure 5.9 ($n = 1$), the most general decision strategy in this case consists of a measurement of all n copies of the system according to a POVM $\{M_\rho^{(n)}, M_\sigma^{(n)}\}$. If the outcome corresponding to $M_\rho^{(n)}$ occurs, then we guess that each copy of the system is in the state ρ, and if the outcome corresponding to $M_\sigma^{(n)}$ occurs, then we guess that each copy of the system is in the state σ.

$$= p_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}, M^{(n)}), \tag{5.3.5}$$

where we have let $M_\rho^{(n)} \equiv M^{(n)}$.

Exercise 5.6

Consider states ρ and σ along with a POVM $\{M_0, M_1\}$ representing a strategy for hypothesis testing of a single copy of the quantum system, where the outcome "0" corresponds to ρ and the outcome "1" corresponds to σ. Let $\lambda \in [0, 1]$ be the prior probability for ρ, and let $n \geq 2$. Construct the POVM $\{M_\rho^{(n)}, M_\sigma^{(n)}\}$, and evaluate the type-I and type-II error probabilities for the following two decision strategies for hypothesis testing of n copies of the quantum system.

1. The *majority-vote* decision strategy: (1) Measure each system according to the POVM $\{M_0, M_1\}$, and let N_x be the number of times the outcome x occurs. (2) If $N_0 > N_1$, guess "ρ", and if $N_1 > N_0$, guess "σ". If n is even and $N_0 = N_1$, then guess "ρ" with probability $q \in [0, 1]$ and guess "σ" with probability $1 - q$.

2. The *unanimous-vote* decision strategy: (1) Measure each system according to the POVM $\{M_0, M_1\}$, and let N_x be the number of times the outcome x

occurs. (2) If $N_0 = n$, then guess "ρ"; otherwise, guess "σ".

5.3.1 Symmetric Case (State Discrimination)

Given quantum states ρ and σ, the goal of symmetric hypothesis testing, also known as quantum state discrimination, is to devise a measurement strategy that minimizes the error probability defined in (5.3.1), where $\lambda \in [0, 1]$ is the probability that the state is ρ and $1 - \lambda$ is the probability that the state is σ. The value of the corresponding optimization problem in (5.3.2) is

$$p^*_{\text{err}}(\lambda, \rho, \sigma) := \inf_{M: 0 \le M \le \mathbb{1}} \{\text{Tr}[(\mathbb{1} - M)(\lambda\rho)] + \text{Tr}[M(1-\lambda)\sigma]\}\,. \tag{5.3.6}$$

Exercise 5.7

Show that $p^*_{\text{err}}(\lambda, \rho, \sigma)$ can be evaluated using a semi-definite program. Then, using strong duality, prove that an alternate expression for $p^*_{\text{err}}(\lambda, \rho, \sigma)$ is

$$p^*_{\text{err}}(\lambda, \rho, \sigma) = \sup_{W \text{ Hermitian}} \{\text{Tr}[W] : W \le \lambda\rho,\ W \le (1-\lambda)\sigma\}\,. \tag{5.3.7}$$

Finally, evaluate the complementary slackness conditions from Proposition 2.29. An optimal operator W is known as the "greatest lower bound operator".

Exercise 5.8

Prove that $p^*_{\text{err}}(\lambda, \rho, \sigma)$ is isometrically invariant: for every isometry V, $p^*_{\text{err}}(\lambda, \rho, \sigma) = p^*_{\text{err}}(\lambda, V\rho V^\dagger, V\sigma V^\dagger)$.

More generally than isometric invariance, the following *data-processing inequality* holds for the error probability for discriminating two quantum states. The intuition behind the proof of Proposition 5.2 is as follows: Suppose that we are given a quantum system in an unknown state. Before applying a measurement to determine the state, we could perform a quantum channel $\mathcal{N}$. However, if we do so, this strategy is not necessarily an optimal strategy, and the error probability is never smaller than if we simply apply an optimal measurement to distinguish the states.

Proposition 5.2 Data-Processing Inequality for State Discrimination

Consider states ρ and σ, $\lambda \in [0,1]$, and let $\mathcal{N}$ be a positive and trace preserving superoperator. Then,

$$p^*_{\text{err}}(\lambda,\rho,\sigma) \leq p^*_{\text{err}}(\lambda,\mathcal{N}(\rho),\mathcal{N}(\sigma)) \tag{5.3.8}$$

Proof: Let M' be an operator satisfying $0 \leq M' \leq \mathbb{1}$, and consider the operator $\mathcal{N}^\dagger(M')$ (this is the measurement operator corresponding to performing the channel $\mathcal{N}$ first and then applying the measurement operator M'). Due to the positivity of $\mathcal{N}$, and thus of $\mathcal{N}^\dagger$, we have $\mathcal{N}^\dagger(M') \geq 0$. Now, the condition $M' \leq \mathbb{1}$ implies $\mathbb{1} - M' \geq 0$. Thus, by the positivity of $\mathcal{N}^\dagger$, it follows that $\mathcal{N}^\dagger(\mathbb{1} - M') \geq 0$, which implies $\mathcal{N}^\dagger(M') \leq \mathcal{N}^\dagger(\mathbb{1})$. Now, $\mathcal{N}^\dagger$ is unital, because $\mathcal{N}$ is trace preserving (see Exercise 4.10), which means that $\mathcal{N}^\dagger(\mathbb{1}) = \mathbb{1}$. We thus conclude that $0 \leq \mathcal{N}^\dagger(M') \leq \mathbb{1}$. Therefore, $\mathcal{N}^\dagger(M')$ is a measurement operator and thus a feasible point in the optimization problem for $p^*_{\text{err}}(\lambda,\rho,\sigma)$, so that

$$\begin{aligned} p^*_{\text{err}}(\lambda,\rho,\sigma) &= \inf_{M:0\leq M\leq \mathbb{1}} \{\text{Tr}[(\mathbb{1}-M)(\lambda\rho)] + \text{Tr}[M(1-\lambda)\sigma]\} && (5.3.9)\\ &\leq \text{Tr}[(\mathbb{1}-\mathcal{N}^\dagger(M'))(\lambda\rho)] + \text{Tr}[\mathcal{N}^\dagger(M')(1-\lambda)\sigma] && (5.3.10)\\ &= \text{Tr}[(\mathbb{1}-M')(\lambda\mathcal{N}(\rho))] + \text{Tr}[M'(1-\lambda)\mathcal{N}(\sigma)], && (5.3.11) \end{aligned}$$

where the last line follows from the definition of the adjoint of a superoperator. Finally, because the inequality

$$p^*_{\text{err}}(\lambda,\rho,\sigma) \leq \text{Tr}[(\mathbb{1}-M')(\lambda\mathcal{N}(\rho))] + \text{Tr}[M'(1-\lambda)\mathcal{N}(\sigma)] \tag{5.3.12}$$

holds for all M' satisfying $0 \leq M' \leq \mathbb{1}$, we conclude that

$$\begin{aligned} p^*_{\text{err}}(\lambda,\rho,\sigma) &\leq \inf_{M':0\leq M'\leq \mathbb{1}} \text{Tr}[(\mathbb{1}-M')(\lambda\mathcal{N}(\rho))] + \text{Tr}[M'(1-\lambda)\mathcal{N}(\sigma)] && (5.3.13)\\ &= p^*_{\text{err}}(\lambda,\mathcal{N}(\rho),\mathcal{N}(\sigma)), && (5.3.14) \end{aligned}$$

as required. ■

It turns out that $p^*_{\text{err}}(\lambda,\rho,\sigma)$ can be written in terms of the trace norm (Section 2.2.9.2) as

$$p^*_{\text{err}}(\lambda,\rho,\sigma) = \frac{1}{2}\left(1 - \|\lambda\rho - (1-\lambda)\sigma\|_1\right), \tag{5.3.15}$$

which is an immediate consequence of the following theorem.

Theorem 5.3 Helstrom–Holevo Theorem

For all positive semi-definite operators A and B,

$$\inf_{M:0\leq M\leq \mathbb{1}} \{\mathrm{Tr}[(\mathbb{1}-M)A]+\mathrm{Tr}[MB]\} = \frac{1}{2}\left(\mathrm{Tr}[A+B]-\|A-B\|_1\right). \tag{5.3.16}$$

A measurement operator M is optimal if and only if $M = \Pi_+ + \Lambda_0$, where Π_+ is the projection onto the strictly positive part of $A - B$, the operator Π_0 is the projection onto the zero eigenspace of $A - B$, and $0 \leq \Lambda_0 \leq \Pi_0$. Furthermore,

$$\sup_{M:0\leq M\leq \mathbb{1}} \mathrm{Tr}[M(A-B)] = \frac{1}{2}\left(\mathrm{Tr}[A-B]+\|A-B\|_1\right), \tag{5.3.17}$$

and the conditions for an optimal M are the same as given above.

Remark: Letting $A = \lambda\rho$ and $B = (1-\lambda)\sigma$ in the statement of Theorem 5.3, we recognize that the objective function on the left-hand side of (5.3.16) is equal to $p_{\mathrm{err}}(\lambda, \rho, \sigma, M)$ as defined in (5.3.1). We thus obtain (5.3.15). Note that Theorem 5.3 also gives us a measurement that achieves the minimal error probability.

Proof: Let M be an arbitrary operator satisfying $0 \leq M \leq \mathbb{1}$. Let $\Delta \coloneqq A - B$ and let Δ_+ and Δ_- be the positive and negative parts, respectively, of Δ, so that $A - B = \Delta_+ - \Delta_-$ and $\Delta_+\Delta_- = 0$ (recall (2.2.68)). We can then write the objective function in (5.3.16) as

$$\mathrm{Tr}[(\mathbb{1}-M)A]+\mathrm{Tr}[MB] = \mathrm{Tr}[A] - (\mathrm{Tr}[M\Delta_+]-\mathrm{Tr}[M\Delta_-]). \tag{5.3.18}$$

Now, since $\mathrm{Tr}[M\Delta_-] \geq 0$, on account of both M and Δ_- being positive semi-definite, we find that

$$\mathrm{Tr}[M\Delta_+]-\mathrm{Tr}[M\Delta_-] \leq \mathrm{Tr}[M\Delta_+] \leq \mathrm{Tr}[\Delta_+], \tag{5.3.19}$$

where the last inequality follows because $M \leq \mathbb{1}$. The equality in (5.3.18) and the inequality in (5.3.19) imply that

$$\mathrm{Tr}[(\mathbb{1}-M)A]+\mathrm{Tr}[MB] \geq \mathrm{Tr}[A]-\mathrm{Tr}[\Delta_+]. \tag{5.3.20}$$

Since

$$\|A-B\|_1 = \mathrm{Tr}[|A-B|] = \mathrm{Tr}[\Delta_+]+\mathrm{Tr}[\Delta_-] \tag{5.3.21}$$

and

$$\Delta_- = \Delta_+ + B - A, \tag{5.3.22}$$

we can write $\mathrm{Tr}[\Delta_+]$ as

$$\mathrm{Tr}[\Delta_+] = \frac{1}{2}\left(\|A-B\|_1 - \mathrm{Tr}[B-A]\right). \tag{5.3.23}$$

This means that the objective function on the left-hand side of (5.3.18) can be bounded from below by $\frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right)$. We have thus shown that

$$\mathrm{Tr}[(\mathbb{1}-M)A] + \mathrm{Tr}[MB] \geq \frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right). \tag{5.3.24}$$

for all M such that $0 \leq M \leq \mathbb{1}$, which implies that

$$\inf_{M:0\leq M\leq \mathbb{1}} \mathrm{Tr}[(\mathbb{1}-M)A] + \mathrm{Tr}[MB] \geq \frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right). \tag{5.3.25}$$

To see the reverse inequality, let $M = \Pi_+ + \Lambda_0$, where Π_+ is the projection onto the strictly positive part of $A-B$ and Λ_0 satisfies $0 \leq \Lambda_0 \leq \Pi_0$, with Π_0 the projection onto the zero eigenspace of $A-B$. Then,

$$\begin{aligned}
\mathrm{Tr}[M(A-B)] &= \mathrm{Tr}[(\Pi_+ + \Lambda_0)(\Delta_+ - \Delta_-)] && (5.3.26)\\
&= \mathrm{Tr}[(\Pi_+ + \Lambda_0)\Delta_+] - \mathrm{Tr}[(\Pi_+ + \Lambda_0)\Delta_-] && (5.3.27)\\
&= \mathrm{Tr}[\Delta_+], && (5.3.28)
\end{aligned}$$

where the last equality follows because $\mathrm{Tr}[\Pi_+\Delta_+] = \mathrm{Tr}[\Delta_+]$ and $\mathrm{Tr}[\Pi_+\Delta_-] = 0$, since Π_+ and Δ_- are by definition orthogonal. We also used $\mathrm{Tr}[\Lambda_0\Delta_+] = \mathrm{Tr}[\Lambda_0\Delta_-] = 0$, with these latter equalities following because $0 \leq \mathrm{Tr}[\Lambda_0\Delta_\pm] \leq \mathrm{Tr}[\Pi_0\Delta_\pm] = 0$. Therefore, using (5.3.23), we find that

$$\mathrm{Tr}[A] - \mathrm{Tr}[(\Pi_+ + \Lambda_0)(A-B)] = \frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right). \tag{5.3.29}$$

The operator $\Pi_+ + \Lambda_0$ thus achieves the bound in (5.3.24), which means that

$$\inf_{M:0\leq M\leq \mathbb{1}} \mathrm{Tr}[(\mathbb{1}-M)A] + \mathrm{Tr}[MB] = \frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right), \tag{5.3.30}$$

so that (5.3.16) is established.

To see that $\Pi_+ + \Lambda_0$ is the only form for an optimal measurement operator, suppose that M is optimal, i.e., satisfies $0 \leq M \leq \mathbb{1}$ and saturates (5.3.16) with equality. Then it follows that the two inequalities in (5.3.19) are saturated with equality, so that

$$\mathrm{Tr}[M(\Delta_+ - \Delta_-)] = \mathrm{Tr}[M\Delta_+] = \mathrm{Tr}[\Delta_+] = \mathrm{Tr}[\Pi_+\Delta_+]. \tag{5.3.31}$$

The leftmost equality implies that $\mathrm{Tr}[M\Delta_-] = 0$, where Δ_- is the strictly negative part of $A - B$. Since both M and Δ_- are positive semi-definite and Π_- is the projection onto the strictly negative part of Δ_-, we conclude that $M\Pi_- = \Pi_- M = 0$. This in turn implies that

$$M(\Pi_+ + \Pi_0) = (\Pi_+ + \Pi_0)M = M, \tag{5.3.32}$$

which, after sandwiching $M \leq \mathbb{1}$ on the left and right by $\Pi_+ + \Pi_0$, implies that $M \leq \Pi_+ + \Pi_0$. Since

$$\begin{aligned} 0 &= \mathrm{Tr}[\Delta_+(\Pi_+ - M)] && (5.3.33)\\ &= \mathrm{Tr}[\Delta_+(\Pi_+ - \Pi_+ M \Pi_+)] && (5.3.34)\\ &= \mathrm{Tr}[\Delta_+\Pi_+(\mathbb{1} - M)\Pi_+], && (5.3.35) \end{aligned}$$

we find that $\Pi_+(\mathbb{1} - M)\Pi_+ = 0$. Now consider that $\Pi_-(\mathbb{1} - M)\Pi_+ = 0$ because $\Pi_-\Pi_+ = 0$ and $\Pi_- M\Pi_+ = \Pi_-(\Pi_+ + \Pi_0)M\Pi_+ = 0$. So then $\Pi_+(\mathbb{1} - M)\Pi_+ = 0$ and $\Pi_-(\mathbb{1} - M)\Pi_+ = 0$ imply that $(\mathbb{1} - M)\Pi_+ = 0$. From this equation, we conclude that $\Pi_+ = \Pi_+ M = M\Pi_+$. By sandwiching $\Pi_+ \leq \mathbb{1}$ by M and applying operator monotonicity of the square-root function (see Section 2.2.8.1), we conclude that $\Pi_+ \leq M$. Combining this operator inequality with the previous one, we conclude that an optimal M satisfies $\Pi_+ \leq M \leq \Pi_+ + \Pi_0$, which is equivalent to M decomposing as $M = \Pi_+ + \Lambda_0$ for $0 \leq \Lambda_0 \leq \Pi_0$.

The equality in (5.3.17) follows as a rewrite of (5.3.16):

$$\begin{aligned} \frac{1}{2}\left\|A - B\right\|_1 &= \frac{1}{2}\mathrm{Tr}[A + B] - \inf_{M:0\leq M\leq \mathbb{1}} \mathrm{Tr}[(\mathbb{1} - M)A] + \mathrm{Tr}[MB] && (5.3.36)\\ &= \sup_{M:0\leq M\leq \mathbb{1}} \frac{1}{2}\mathrm{Tr}[A + B] - (\mathrm{Tr}[(\mathbb{1} - M)A] + \mathrm{Tr}[MB]) && (5.3.37)\\ &= \sup_{M:0\leq M\leq \mathbb{1}} \frac{1}{2}\mathrm{Tr}[B - A] + \mathrm{Tr}[M(A - B)] && (5.3.38)\\ &= \frac{1}{2}\mathrm{Tr}[B - A] + \sup_{M:0\leq M\leq \mathbb{1}} \mathrm{Tr}[M(A - B)]. && (5.3.39) \end{aligned}$$

Rearranging this equality, we arrive at (5.3.17). An optimal M having the form $\Pi_+ + \Lambda_0$ again follows from (5.3.19), (5.3.26)–(5.3.28), and the reasoning given above. ■

Exercise 5.9

Let $\rho = |\psi\rangle\langle\psi| \equiv \psi$ and $\sigma = |\phi\rangle\langle\phi| \equiv \phi$ be pure states, and let $\lambda \in [0,1]$. Show that

$$p_{\text{err}}^*(\lambda, \psi, \phi) = \frac{1}{2}\left(1 - \sqrt{1 - 4\lambda(1-\lambda)\,|\langle\psi|\phi\rangle|^2}\right). \tag{5.3.40}$$

What is a measurement that achieves this optimal error probability?

Observe from (5.3.40) that if $|\psi\rangle$ and $|\phi\rangle$ are orthogonal, then $p_{\text{err}}^*(\lambda, \psi, \phi) = 0$.

Exercise 5.10

Let ρ and σ be quantum states that are orthogonal, in the sense that $\Pi_\rho \Pi_\sigma = \Pi_\sigma \Pi_\rho = 0$, where Π_ρ and Π_σ are the projections onto the support of ρ and σ, respectively (recall (2.2.67)). Prove that the optimal error probability for discriminating ρ and σ vanishes, i.e., that $p_{\text{err}}^*(\lambda, \rho, \sigma) = 0$. What is a measurement achieving this optimal error probability?

Exercise 5.11

Evaluate the optimal error probability $p_{\text{err}}^*\left(\lambda, \rho_{AB}^{\text{iso};p_1}, \rho_{AB}^{\text{W};p_2}\right)$ for discriminating between the isotropic state $\rho_{AB}^{\text{iso};p_1}$, $p_1 \in [0,1]$, and the Werner state $\rho_{AB}^{\text{W};p_2}$, $p_2 \in [0,1]$, where $\lambda \in [0,1]$.

The Helstrom–Holevo theorem gives us the lowest possible error probability in distinguishing between two states ρ and σ, given just one copy of either state. Suppose now that, instead of just one copy, Alice sends Bob n copies of either ρ or σ. Bob's task is then to discriminate between the states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$. The Helstrom–Holevo theorem still applies in this case, so that

$$p_{\text{err}}^*(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) = \frac{1}{2}\left(1 - \left\|\lambda\rho^{\otimes n} - (1-\lambda)\sigma^{\otimes n}\right\|_1\right) \tag{5.3.41}$$

is the lowest possible error probability. However, because Bob now has n copies of either ρ or σ, he can perform a discrimination strategy that involves a collective measurement acting on the n copies of the state. This means that the optimal *error exponent* can generally be lower with $n \geq 2$ copies than with just one copy.

Exercise 5.12

Prove that the optimal error probability $p^*_{\text{err}}(\lambda, \rho, \sigma)$ for quantum state discrimination is monotonically non-increasing with n, i.e., prove that

$$p^*_{\text{err}}(\lambda, \rho^{\otimes n+1}, \sigma^{\otimes n+1}) \leq p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) \tag{5.3.42}$$

for all $n \geq 1$.

5.3.1.1 Asymptotic Setting

Given states ρ and σ and $\lambda \in (0,1)$, how does the optimal error probability $p_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n})$ behave as the number n of copies of the state increases? If $\rho \equiv \psi$ and $\sigma \equiv \phi$ are pure states, then because $\psi^{\otimes n}$ are $\phi^{\otimes n}$ are both pure states, we can use (5.3.40) and the expansion $\frac{1}{2}\left(1 - \sqrt{1-4x}\right) = x + O(x^2)$ to see that the following approximation holds as n becomes large:

$$p^*_{\text{err}}(\lambda, \psi^{\otimes n}, \phi^{\otimes n}) \approx \lambda(1-\lambda) \left|\langle\psi|\phi\rangle\right|^{2n} = \lambda(1-\lambda) 2^{-n(-\log_2 |\langle\psi|\phi\rangle|^2)}. \tag{5.3.43}$$

Now, because $|\langle\psi|\phi\rangle|^2 \in [0,1]$, we have that $-\log_2 |\langle\phi|\psi\rangle|^2 \geq 0$, which means that, as n becomes large, the optimal error probability decays exponentially to zero. Does the exponential decay hold more generally? In other words, for arbitrary states ρ and σ, is it true that $p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) \approx 2^{-n\xi(\lambda,\rho,\sigma)}$ as n becomes large, where $\xi(\lambda, \rho, \sigma) = -\frac{1}{n}\log_2 p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n})$ is a non-negative asymptotic *error exponent* that is independent of n? To be more precise, does the limit $\lim_{n\to\infty} -\frac{1}{n}\log_2 p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n})$ exist, and if so, what is its value?

The following theorem provides positive answers to both questions. The characterization given below is useful because the quantity $p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n})$ becomes more and more difficult to calculate as n increases, so that the asymptotic error exponent is a helpful characterization of $p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n})$.

Theorem 5.4 Quantum Chernoff Bound

For all quantum states ρ and σ, and $\lambda \in (0,1)$, the following limit exists and is equal to the the *quantum Chernoff divergence* of ρ and σ:

$$\lim_{n\to\infty} -\frac{1}{n}\log_2 p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) = C(\rho\|\sigma), \tag{5.3.44}$$

where

$$C(\rho\|\sigma) := \sup_{s\in(0,1)} \left(-\log_2 \mathrm{Tr}[\rho^s\sigma^{1-s}]\right). \tag{5.3.45}$$

That is, $C(\rho\|\sigma)$ is the optimal asymptotic error exponent for symmetric hypothesis testing of ρ and σ.

REMARK: Theorem 5.4 tells us that, as n becomes large, the following approximation holds

$$p^*_{\mathrm{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) \approx 2^{-n\,C(\rho\|\sigma)}, \tag{5.3.46}$$

so that the optimal error probability does indeed decay exponentially with the number n of copies of the state. In particular, the quantum Chernoff divergence is the *optimal asymptotic error exponent* for the exponential decay of the error probability as the number n of copies increases.

Note that the optimal error exponent in (5.3.44) is independent of the prior probability distribution. This means that, in the asymptotic limit (i.e., in the limit $n \to \infty$), the prior probability distribution is irrelevant for the optimal error exponent.

We call Theorem 5.4 the quantum Chernoff bound because it is analogous to the Chernoff bound from (classical) symmetric hypothesis testing, which is the task of discriminating between two hypotheses, each of which has a corresponding probability distribution (please consult the Bibliographic Notes in Section 3.4 for details).

An important element of the proof of the quantum Chernoff bound is Lemma 5.5 below.

Lemma 5.5

Let A and B be positive semi-definite operators. Then, for all $s \in (0,1)$,

$$\frac{1}{2}\left(\mathrm{Tr}[A+B] - \|A-B\|_1\right) \leq \mathrm{Tr}[A^s B^{1-s}]. \tag{5.3.47}$$

PROOF: Let $\Delta := A - B$, and let Δ_+ and Δ_- be the positive and negative parts, respectively, of Δ, so that $\Delta = \Delta_+ - \Delta_-$ (recall (2.2.68)). Then,

$$|\Delta| = |\Delta_+ - \Delta_-| = \Delta_+ + \Delta_-. \tag{5.3.48}$$

Therefore,

$$\|A-B\|_1 = \|\Delta\|_1 = \mathrm{Tr}[|\Delta|] = \mathrm{Tr}[\Delta_+] + \mathrm{Tr}[\Delta_-]. \tag{5.3.49}$$

In addition, we write

$$A + B = A - B + 2B = \Delta_+ - \Delta_- + 2B, \tag{5.3.50}$$

so that

$$\begin{aligned}
&\frac{1}{2}\left(\operatorname{Tr}[A + B] - \|A - B\|_1\right) \\
&= \frac{1}{2}\left(\operatorname{Tr}[\Delta_+] - \operatorname{Tr}[\Delta_-] + 2\operatorname{Tr}[B] - \operatorname{Tr}[\Delta_+] - \operatorname{Tr}[\Delta_-]\right) && (5.3.51)\\
&= \operatorname{Tr}[B] - \operatorname{Tr}[\Delta_-]. && (5.3.52)
\end{aligned}$$

So it suffices to prove that the following inequality holds for all $s \in (0, 1)$:

$$\operatorname{Tr}[B] - \operatorname{Tr}[\Delta_-] \leq \operatorname{Tr}[A^s B^{1-s}]. \tag{5.3.53}$$

Using the fact that $\Delta_+ \geq 0$, by definition of the positive part of Δ, we obtain

$$B + \Delta_+ \geq B. \tag{5.3.54}$$

Similarly, using $A - B = \Delta_+ - \Delta_-$, we obtain

$$A + \Delta_- = B + \Delta_+ \geq B. \tag{5.3.55}$$

By the operator monotonicity of $x \mapsto x^s$ for $s \in (0, 1)$ (recall Definition 2.13 and thereafter), the inequality in (5.3.55) implies that

$$B^s \leq (A + \Delta_-)^s. \tag{5.3.56}$$

Using this, along with the fact that $\operatorname{Tr}[B] = \operatorname{Tr}[B^s B^{1-s}]$, we find that

$$\begin{aligned}
\operatorname{Tr}[B] - \operatorname{Tr}[A^s B^{1-s}] &= \operatorname{Tr}[(B^s - A^s)B^{1-s}] && (5.3.57)\\
&\leq \operatorname{Tr}[((A + \Delta_-)^s - A^s)B^{1-s}] && (5.3.58)\\
&\leq \operatorname{Tr}[((A + \Delta_-)^s - A^s)(A + \Delta_-)^{1-s}] && (5.3.59)\\
&= \operatorname{Tr}[A] + \operatorname{Tr}[\Delta_-] - \operatorname{Tr}[A^s(A + \Delta_-)^{1-s}] && (5.3.60)\\
&\leq \operatorname{Tr}[A] + \operatorname{Tr}[\Delta_-] - \operatorname{Tr}[A] && (5.3.61)\\
&= \operatorname{Tr}[\Delta_-]. && (5.3.62)
\end{aligned}$$

The first inequality follows because $B^{1-s} \geq 0$ and from (5.3.56). The second inequality follows because $(A + \Delta_-)^s \geq A^s$ and from (5.3.56) with the substitution $s \to 1 - s$. The last inequality follows because

$$\operatorname{Tr}[A^s(A + \Delta_-)^{1-s}] \geq \operatorname{Tr}[A^s A^{1-s}] = \operatorname{Tr}[A], \tag{5.3.63}$$

which in turn follows because $A^s \geq 0$ and $(A + \Delta_-)^{1-s} \geq A^{1-s}$. Therefore, we conclude that

$$\mathrm{Tr}[B] - \mathrm{Tr}[A^s B^{1-s}] \leq \mathrm{Tr}[\Delta_-], \tag{5.3.64}$$

establishing the desired inequality in (5.3.53). ■

Exercise 5.13

Let A and B be positive semi-definite operators and $s \in (0, 1)$. Starting with Lemma 5.5, prove that

$$\frac{1}{2} \|A - B\|_1 \geq \frac{1}{2}\mathrm{Tr}[A + B] - \left\|A^s B^{1-s}\right\|_1 . \tag{5.3.65}$$

(*Hint*: Recall Theorem 2.10.)

Remark: In the case $s = \frac{1}{2}$, the inequality in (5.3.65) becomes

$$\frac{1}{2} \|A - B\|_1 \geq \frac{1}{2}\mathrm{Tr}[A + B] - \left\|\sqrt{A}\sqrt{B}\right\|_1 \tag{5.3.66}$$

$$= \frac{1}{2}\mathrm{Tr}[A + B] - \sqrt{F(A, B)}, \tag{5.3.67}$$

where in the second line we let

$$F(A, B) \coloneqq \left\|\sqrt{A}\sqrt{B}\right\|_1^2 . \tag{5.3.68}$$

The quantity $F(A, B)$ is called the *fidelity* between A and B, and it plays a critical role in the analysis of quantum communication protocols. We study the fidelity function in detail in Chapter 6.

We now proceed to the proof of the quantum Chernoff bound.

Proof of the Quantum Chernoff Bound (Theorem 5.4): Since the limit on the left-hand side of (5.3.44) need not exist *a priori*, let us define the following quantities based on the discussion in Section 2.3.1:

$$\underline{\xi}(\rho, \sigma) \coloneqq \liminf_{n\to\infty} -\frac{1}{n} \log_2 p^*_{\mathrm{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}), \tag{5.3.69}$$

$$\overline{\xi}(\rho, \sigma) \coloneqq \limsup_{n\to\infty} -\frac{1}{n} \log_2 p^*_{\mathrm{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}). \tag{5.3.70}$$

Note that, by definition, we always have $\underline{\xi}(\rho,\sigma) \leq \overline{\xi}(\rho,\sigma)$; see (2.3.5). Our goal is to prove that $\underline{\xi}(\rho,\sigma) = \overline{\xi}(\rho,\sigma) = \lim_{n\to\infty} -\frac{1}{n}\log_2 p_{\text{err}}^*(\lambda,\rho^{\otimes n},\sigma^{\otimes n}) = C(\rho\|\sigma)$.

Now, if λ is the probability with which ρ is chosen, and $1-\lambda$ the probability with which σ is chosen, then an application of Lemma 5.5, with $A = \lambda\rho^{\otimes n}$, $B = (1-\lambda)\sigma^{\otimes n}$, and $s \in (0,1)$, gives the following:

$$\begin{aligned} p_{\text{err}}^*(\lambda,\rho^{\otimes n},\sigma^{\otimes n}) &= \frac{1}{2}\left(1 - \left\|\lambda\rho^{\otimes n} - (1-\lambda)\sigma^{\otimes n}\right\|_1\right) && (5.3.71)\\ &\leq \text{Tr}\left[\lambda^s(\rho^{\otimes n})^s(1-\lambda)^{1-s}(\sigma^{\otimes n})^{1-s}\right] && (5.3.72)\\ &= \lambda^s(1-\lambda)^{1-s}\text{Tr}[(\rho^s)^{\otimes n}(\sigma^{1-s})^{\otimes n}] && (5.3.73)\\ &= \lambda^s(1-\lambda)^{1-s}\left(\text{Tr}[\rho^s\sigma^{1-s}]\right)^n && (5.3.74)\\ &\leq \left(\text{Tr}[\rho^s\sigma^{1-s}]\right)^n, && (5.3.75) \end{aligned}$$

where the last line follows from the fact that $\lambda^s(1-\lambda)^{1-s} \leq 1$. By taking a negative logarithm and dividing by n, this bound becomes

$$-\frac{1}{n}\log_2 p_{\text{err}}^*(\lambda,\rho^{\otimes n},\sigma^{\otimes n}) \geq -\log_2 \text{Tr}[\rho^s\sigma^{1-s}]. \quad (5.3.76)$$

Since the bound above holds for all $s \in (0,1)$, we obtain the following bound:

$$\begin{aligned} \underline{\xi}(\rho,\sigma) &= \liminf_{n\to\infty} -\frac{1}{n}\log_2 p_{\text{err}}^*(\lambda,\rho^{\otimes n},\sigma^{\otimes n}) && (5.3.77)\\ &\geq \sup_{s\in(0,1)} -\log_2 \text{Tr}[\rho^s\sigma^{1-s}] && (5.3.78)\\ &= C(\rho\|\sigma). && (5.3.79) \end{aligned}$$

For the opposite inequality, we start by observing that it suffices to optimize over projectors when determining the optimal error probability $p_{\text{err}}^*(\lambda,\rho^{\otimes n},\sigma^{\otimes n})$. (This is true because one choice of an optimal measurement is a projective measurement, as shown in the proof of Theorem 5.3, where we can set $\Lambda_0 = 0$.) Next, suppose that ρ and σ have the following spectral decompositions:

$$\rho = \sum_{i=0}^{d-1} \lambda_i |\psi_i\rangle\langle\psi_i|, \quad \sigma = \sum_{j=0}^{d-1} \mu_j |\phi_j\rangle\langle\phi_j|, \quad (5.3.80)$$

where d is the dimension of the underlying Hilbert space. Then, for every projection Π,

$$\begin{aligned}
\mathrm{Tr}[(\mathbb{1}-\Pi)\rho] &= \sum_{i=0}^{d-1} \lambda_i \mathrm{Tr}[(\mathbb{1}-\Pi)|\psi_i\rangle\langle\psi_i|] && (5.3.81)\\
&= \sum_{i=0}^{d-1} \lambda_i \mathrm{Tr}[(\mathbb{1}-\Pi)|\psi_i\rangle\langle\psi_i|(\mathbb{1}-\Pi)] && (5.3.82)\\
&= \sum_{i,j=0}^{d-1} \lambda_i \mathrm{Tr}[|\phi_j\rangle\langle\phi_j|(\mathbb{1}-\Pi)|\psi_i\rangle\langle\psi_i|(\mathbb{1}-\Pi)] && (5.3.83)\\
&= \sum_{i,j=0}^{d-1} \lambda_i \left|\langle\psi_i|(\mathbb{1}-\Pi)|\phi_j\rangle\right|^2, && (5.3.84)
\end{aligned}$$

where we have used the fact that $\mathbb{1} = \sum_{j=0}^{d-1} |\phi_j\rangle\langle\phi_j|$. Similarly, using the fact that $\mathbb{1} = \sum_{i=0}^{d-1} |\psi_i\rangle\langle\psi_i|$, we have

$$\begin{aligned}
\mathrm{Tr}[\Pi\sigma] &= \sum_{j=0}^{d-1} \mu_j \mathrm{Tr}[\Pi|\phi_j\rangle\langle\phi_j|] && (5.3.85)\\
&= \sum_{j=0}^{d-1} \mu_j \mathrm{Tr}[\Pi|\phi_j\rangle\langle\phi_j|\Pi] && (5.3.86)\\
&= \sum_{i,j=0}^{d-1} \mu_j \mathrm{Tr}[|\psi_i\rangle\langle\psi_i|\Pi|\phi_j\rangle\langle\phi_j|\Pi] && (5.3.87)\\
&= \sum_{i,j=0}^{d-1} \mu_j \left|\langle\psi_i|\Pi|\phi_j\rangle\right|^2. && (5.3.88)
\end{aligned}$$

Then, using the fact that $\lambda_i \geq \min\{\lambda_i, \mu_j\}$ and $\mu_j \geq \min\{\lambda_i, \mu_j\}$ for all $0 \leq i, j \leq d-1$, the error probability under the measurement given by Π is

$$\begin{aligned}
&p_{\mathrm{err}}(\lambda, \rho, \sigma, \Pi)\\
&\quad = \mathrm{Tr}[(\mathbb{1}-\Pi)(\lambda\rho)] + \mathrm{Tr}[\Pi(1-\lambda)\sigma] && (5.3.89)\\
&\quad = \sum_{i,j=0}^{d-1} \left(\lambda\lambda_i \left|\langle\psi_i|(\mathbb{1}-\Pi)|\phi_j\rangle\right|^2 + (1-\lambda)\mu_j \left|\langle\psi_i|\Pi|\phi_j\rangle\right|^2\right) && (5.3.90)
\end{aligned}$$

$$\geq \frac{1}{2}\min\{\lambda\lambda_i, (1-\lambda)\mu_j\}\left(\left|x_{i,j}\right|^2 + \left|y_{i,j}\right|^2\right), \tag{5.3.91}$$

where we have defined $x_{i,j} \coloneqq \langle\psi_i|(\mathbb{1}-\Pi)|\phi_j\rangle$ and $y_{i,j} \coloneqq \langle\psi_i|\Pi|\phi_j\rangle$ in the last line. Using the identity $|a|^2 + |b|^2 \geq \frac{1}{2}|a+b|^2$, which holds for all $a, b \in \mathbb{C}$, we obtain

$$p_{\text{err}}(\lambda, \rho, \sigma, \Pi) \geq \frac{1}{2}\sum_{i,j=0}^{d-1}\min\{\lambda\lambda_i, (1-\lambda)\mu_j\}\left|\langle\psi_i|\phi_j\rangle\right|^2. \tag{5.3.92}$$

Now, let us define two probability distributions, $p, q : \{0, 1, \ldots, d-1\}^2 \to [0, 1]$ as follows:

$$p(i, j) \coloneqq \lambda_i\left|\langle\psi_i|\phi_j\rangle\right|^2, \quad q(i, j) \coloneqq \mu_j\left|\langle\psi_i|\phi_j\rangle\right|^2, \quad \forall\, 0 \leq i, j \leq d-1. \tag{5.3.93}$$

It is straightforward to verify that p and q are indeed probability distributions. Then, since the projector Π is arbitrary, and it suffices to optimize over projective measurements, as explained above, the following inequality holds

$$p^*_{\text{err}}(\lambda, \rho, \sigma) \geq \frac{1}{2}\sum_{i,j=0}^{d-1}\min\{\lambda p(i, j), (1-\lambda)q(i, j)\}. \tag{5.3.94}$$

The expression on the right-hand side of the above inequality is precisely half the optimal error probability for discriminating the two probability distributions p and q, with a prior probability of λ. Indeed, we can see this by an application of Theorem 5.3 to the case of commutative A and B. To this end, letting $A = \sum_{i=0}^{d-1} a_i|i\rangle\langle i|$ and $B = \sum_{i=0}^{d-1} b_i|i\rangle\langle i|$ where $a_i, b_i \geq 0$ for all $i \in \{0, \ldots, d-1\}$, it follows that an optimal measurement operator $\Pi_+ = \sum_{i:a_i \geq b_i}|i\rangle\langle i|$ and its complement $\Pi_- = \sum_{i:a_i<b_i}|i\rangle\langle i|$, so that

$$\inf_{M:0\leq M\leq\mathbb{1}}\{\text{Tr}[(\mathbb{1}-M)A] + \text{Tr}[MB]\} = \text{Tr}[\Pi_- A] + \text{Tr}[\Pi_+ B] \tag{5.3.95}$$

$$= \sum_{i:a_i<b_i} a_i + \sum_{i:a_i\geq b_i} b_i \tag{5.3.96}$$

$$= \sum_{i=0}^{d-1}\min\{a_i, b_i\}. \tag{5.3.97}$$

We denote this optimal error probability by $p_{\text{err}}(\lambda, p, q)$. Therefore,

$$p^*_{\text{err}}(\lambda, \rho, \sigma) \geq \frac{1}{2}p^*_{\text{err}}(\lambda, p, q). \tag{5.3.98}$$

Now, for $n \geq 2$, by following the same arguments as above, we obtain

$$p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) \geq \frac{1}{2} p^*_{\text{err}}(\lambda, p^{(n)}, q^{(n)}), \tag{5.3.99}$$

where $p^{(n)}, q^{(n)}$ are the n-fold i.i.d. probability distributions corresponding to p and q, respectively. Then, by the classical Chernoff bound (please consult the Bibliographic Notes in Section 3.4), we find that

$$\overline{\xi}(\rho, \sigma) = \limsup_{n \to \infty} -\frac{1}{n} \log_2 p^*_{\text{err}}(\lambda, \rho^{\otimes n}, \sigma^{\otimes n}) \tag{5.3.100}$$

$$\leq \limsup_{n \to \infty} -\frac{1}{n} \log_2 p^*_{\text{err}}(\lambda, p^{(n)}, q^{(n)}) \tag{5.3.101}$$

$$= \sup_{s \in (0,1)} -\log_2 \sum_{i,j=0}^{d-1} p(i,j)^s q(i,j)^{1-s}, \tag{5.3.102}$$

where the factor of $\frac{1}{2}$ in (5.3.99) vanishes in the asymptotic limit. Finally, observe that

$$\sum_{i,j=0}^{d-1} p(i,j)^s q(i,j)^{1-s} = \sum_{i,j=0}^{d-1} \lambda_i^s \left(|\langle \psi_i | \phi_j \rangle|^2 \right)^s \mu_j^{1-s} \left(|\langle \psi_i | \phi_j \rangle|^2 \right)^{1-s} \tag{5.3.103}$$

$$= \sum_{i,j=0}^{d-1} \lambda_i^s \mu_j^{1-s} |\langle \psi_i | \phi_j \rangle|^2 \tag{5.3.104}$$

$$= \sum_{i,j=0}^{d-1} \lambda_i^s \mu_j^{1-s} \text{Tr}[|\psi_i\rangle\langle\psi_i|\phi_j\rangle\langle\phi_j|] \tag{5.3.105}$$

$$= \text{Tr}\left[\left(\sum_{i=0}^{d-1} \lambda_i^s |\psi_i\rangle\langle\psi_i| \right) \left(\sum_{j=0}^{d-1} \mu_j^{1-s} |\phi_j\rangle\langle\phi_j| \right) \right] \tag{5.3.106}$$

$$= \text{Tr}[\rho^s \sigma^{1-s}]. \tag{5.3.107}$$

Therefore,

$$\overline{\xi}(\rho, \sigma) \leq \sup_{s \in (0,1)} -\log_2 \text{Tr}[\rho^s \sigma^{1-s}] = C(\rho \| \sigma), \tag{5.3.108}$$

which, combined with (5.3.79) and $\underline{\xi}(\rho, \sigma) \leq \overline{\xi}(\rho, \sigma)$, implies that

$$\underline{\xi}(\rho, \sigma) = \overline{\xi}(\rho, \sigma) = \sup_{s \in (0,1)} -\log_2 \text{Tr}[\rho^s \sigma^{1-s}] = C(\rho \| \sigma), \tag{5.3.109}$$

which is what we set out to prove. ■

5.3.2 Multiple State Discrimination

We now briefly consider state discrimination when there are more than two states. Suppose that Alice prepares a quantum system in a state chosen randomly from a set $\{\rho^x\}_{x\in\mathcal{X}}$ of states. We assume that $\mathcal{X}$ is some finite alphabet with size $|\mathcal{X}| \geq 2$ and that the state ρ^x is chosen with probability $p(x)$, where $p : \mathcal{X} \to [0,1]$ is a probability distribution. Alice sends her chosen state to Bob, whose task is to guess the value of x, i.e., which state Alice sent. Bob's knowledge of the system can be described by the ensemble $\{(p(x), \rho_B^x)\}_{x\in\mathcal{X}}$, which has the following associated classical–quantum state:

$$\rho_{X_A B} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_A} \otimes \rho_B^x, \tag{5.3.110}$$

where X_A is a classical $|\mathcal{X}|$-dimensional register that contains Alice's choice of state. Note that while Bob knows both the prior probability distribution p and the association $x \leftrightarrow \rho^x$ between the labels x and the states ρ^x, he does not have access to the register X_A. (If Bob did have access to the classical register X_A, he could simply measure it in the basis $\{|x\rangle\}_{x\in\mathcal{X}}$ and figure out what state Alice sent.) Therefore, as before, Bob must make a measurement. His strategy is to choose a POVM $\{M_B^x\}_{x\in\mathcal{X}}$ with elements indexed by the elements of $\mathcal{X}$. If he obtains the outcome corresponding to $x \in \mathcal{X}$, then he guesses that the state sent was ρ^x.

The scenario of multiple state discrimination is very similar to the task of classical communication over a quantum channel $\mathcal{N}$ that we consider in Chapter 12. The classical messages to be sent correspond to the labels $x \in \mathcal{X}$, while the states ρ^x correspond to an encoding of the messages into quantum states, which are then sent through the quantum channel, and p corresponds to the prior probability over the set of messages. The measurement performed, in order to guess the state, corresponds to a decoding channel that is applied at the receiving end of the quantum channel in order to determine the message that was sent. The quantity in (5.3.119), evaluated for the ensemble $\{(p(x), \mathcal{N}(\rho^x))\}_{x\in\mathcal{X}}$, is then the optimal average probability of correctly guessing the message sent, where the optimization is over all POVMs indexed by the messages.

Let $\mathcal{M}_{B\to X_B}(\cdot) := \sum_{x\in\mathcal{X}} \text{Tr}[M_B^x(\cdot)]|x\rangle\langle x|_{X_B}$ be the measurement channel corresponding to the POVM $\{M_B^x\}_{x\in\mathcal{X}}$, where X_B is a $|\mathcal{X}|$-dimensional classical register. (Recall the definition of a measurement channel from Definition 4.10.) After the measurement, the classical–quantum state in (5.3.110) transforms to

$$\omega_{X_A X_B} := \mathcal{M}_{B\to X_B}(\rho_{X_A B}) \tag{5.3.111}$$

$$= \sum_{x,x' \in \mathcal{X}} p(x)\mathrm{Tr}[M_B^{x'} \rho_B^x] |x\rangle\langle x|_{X_A} \otimes |x'\rangle\langle x'|_{X_B}. \tag{5.3.112}$$

Let

$$\Pi^{\mathrm{succ}}_{X_A X_B} \coloneqq \sum_{x \in \mathcal{X}} |x\rangle\langle x|_{X_A} \otimes |x\rangle\langle x|_{X_B}, \tag{5.3.113}$$

which is a projector corresponding to the registers X_A and X_B having the same value; this is what we want for state discrimination to be successful. The expected success probability of the strategy given by the POVM $\{M^x\}_{x\in\mathcal{X}}$ is thus

$$p_{\mathrm{succ}}(\{(p(x), \rho^x)\}_x, \{M^x\}_x) \coloneqq \mathrm{Tr}[\Pi_{X_A X_B} \omega_{X_A X_B}] \tag{5.3.114}$$

$$= \sum_{x \in \mathcal{X}} p(x)\mathrm{Tr}[M^x \rho^x]. \tag{5.3.115}$$

Exercise 5.14

Show that the error probability for discriminating the states in the ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ using the POVM $\{M^x\}_{x\in\mathcal{X}}$ is

$$p_{\mathrm{err}}(\{(p(x), \rho^x)\}_x, \{M^x\}_x) \coloneqq 1 - p_{\mathrm{succ}}(\{(p(x), \rho^x)\}_x, \{M^x\}_x) \tag{5.3.116}$$

$$= \mathrm{Tr}[(\mathbb{1}_{X_A X_B} - \Pi_{X_A X_B})\omega_{X_A X_B}] \tag{5.3.117}$$

$$= \sum_{x \in \mathcal{X}} p(x)\mathrm{Tr}[(\mathbb{1} - M^x)\rho^x]. \tag{5.3.118}$$

The optimal success probability for discriminating the states in the ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ is

$$p^*_{\mathrm{succ}}(\{(p(x), \rho^x)\}_x)$$

$$\coloneqq \sup_{\{M^x\}_x} \left\{ \sum_{x \in \mathcal{X}} p(x)\mathrm{Tr}[M^x \rho^x] : 0 \leq M^x \leq \mathbb{1}\ \forall x, \sum_{x \in \mathcal{X}} M^x = \mathbb{1} \right\}. \tag{5.3.119}$$

The optimal error probability is then $p^*_{\mathrm{err}}(\{(p(x), \rho^x)\}_x) \coloneqq 1 - p^*_{\mathrm{succ}}(\{(p(x), \rho^x)\}_x)$.

Exercise 5.15

Given an ensemble $\{(p(x), \rho^x)\}_{x\in\mathcal{X}}$ of quantum states with associated classical–quantum state $\rho_{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_B^x$, show that the optimal success probability $p^*_{\text{succ}}(\{(p(x),\rho^x)\}_x)$ can be evaluated using the following semi-definite program:

$$\begin{aligned} \text{maximize} \quad & \text{Tr}[M_{XB}\rho_{XB}] \\ \text{subject to} \quad & \text{Tr}_X[M_{XB}] = \mathbb{1}_B, \\ & M_{XB} \geq 0. \end{aligned} \tag{5.3.120}$$

In other words, show that

$$p^*_{\text{succ}}(\{(p(x),\rho^x)\}_x) = \sup_{M_{XB}\geq 0} \{\text{Tr}[M_{XB}\rho_{XB}] : \text{Tr}_X[M_{XB}] = \mathbb{1}_B\}. \tag{5.3.121}$$

Then, using strong duality, prove that

$$p^*_{\text{succ}}(\{(p(x),\rho^x)\}_x) = \inf_{Y_B\geq 0}\{\text{Tr}[Y_B] : \mathbb{1}_X \otimes Y_B \geq \rho_{XB}\}, \tag{5.3.122}$$

$$= \inf_{Y_B\geq 0}\{\text{Tr}[Y_B] : Y_B \geq p(x)\rho^x_B \ \forall x \in \mathcal{X}\}. \tag{5.3.123}$$

Finally, evaluate the complementary slackness conditions from Proposition 2.29.

Exercise 5.16

Let $\{|\psi^x\rangle\}_{x\in\mathcal{X}}$ be a finite set of mutually orthogonal vectors, let $\rho^x = |\psi^x\rangle\langle\psi^x|$ for all $x \in \mathcal{X}$, and consider the ensemble $\{(p(x),\rho^x)\}_{x\in\mathcal{X}}$, where $p : \mathcal{X} \to [0,1]$ is a probability distribution. Prove that $p^*_{\text{succ}}(\{(p(x),\rho^x)\}_x) = 1$ and construct the corresponding optimal measurement.

Exercise 5.17

Generalize Proposition 5.2 to the case of multiple state discrimination. Specifically, for every finite ensemble $\{(p(x),\rho^x)\}_{x\in\mathcal{X}}$ of states and every positive, trace preserving map $\mathcal{N}$, prove that

$$p^*_{\text{succ}}(\{(p(x),\rho^x)\}_x) \geq p^*_{\text{succ}}(\{(p(x),\mathcal{N}(\rho^x))\}_x). \tag{5.3.124}$$

5.3.3 Asymmetric Case

Given quantum states ρ and σ, the goal of asymmetric quantum hypothesis testing is to minimize the type-II error probability given an upper bound on the type-I error probability, as per the optimization problem in (5.3.3). The value of that optimization problem is given by

$$\beta_\varepsilon(\rho\|\sigma) := \inf\{\text{Tr}[M\sigma] : 0 \leq M \leq \mathbb{1}, \text{Tr}[M\rho] \geq 1-\varepsilon\}, \tag{5.3.125}$$

with $\varepsilon \in [0,1]$ being an upper bound on the type-I error probability. Intuitively, we might expect a trade-off between the type-I and type-II error probabilities. In particular, we might expect that we can achieve a lower minimum type-II error probability by increasing our tolerance on the type-I error probability. This is indeed true. Observe that every measurement operator M that satisfies $\text{Tr}[M\rho] \geq 1-\varepsilon$ also satisfies $\text{Tr}[M\rho] \geq 1-\varepsilon'$ for all ε' greater than ε. All such measurement operators are thus feasible points in the optimization for $\beta_{\varepsilon'}(\rho\|\sigma)$. Therefore,

$$\beta_{\varepsilon'}(\rho\|\sigma) \leq \beta_\varepsilon(\rho\|\sigma) \quad \forall\, \varepsilon' \in [\varepsilon, 1]. \tag{5.3.126}$$

Exercise 5.18

Show that $\beta_\varepsilon(\rho\|\sigma)$ can be evaluated using a semi-definite program. Then, using strong duality, prove that an alternate expression for $\beta_\varepsilon(\rho\|\sigma)$ is

$$\beta_\varepsilon(\rho\|\sigma) = \sup_{\mu\geq 0, Z\geq 0} \{\mu(1-\varepsilon) - \text{Tr}[Z] : \mu\rho \leq \sigma + Z\}. \tag{5.3.127}$$

Finally, evaluate the complementary slackness conditions from Proposition 2.29 and conclude that

$$M(\sigma+Z) = \mu M\rho, \qquad \text{Tr}[M\rho]\mu = (1-\varepsilon)\mu, \qquad MZ = Z. \tag{5.3.128}$$

Exercise 5.19

Prove that the minimum type-II error probability for asymmetric hypothesis testing of states ρ and σ, with $\varepsilon \in [0,1]$, is isometrically invariant: for every isometry V, we have that $\beta_\varepsilon(\rho\|\sigma) = \beta_\varepsilon(V\rho V^\dagger\|V\sigma V^\dagger)$.

As with the minimum error probability for symmetric hypothesis testing, the minimum type-II error probability for asymmetric hypothesis testing obeys the following data-processing inequality.

Proposition 5.6 Data-Processing Inequality for Asymmetric Hypothesis Testing

Consider states ρ and σ, $\varepsilon \in [0,1]$, and let $\mathcal{N}$ be a positive and trace preserving map. Then,

$$\beta_\varepsilon(\rho\|\sigma) \leq \beta_\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{5.3.129}$$

Proof: The proof is analogous to the proof of Proposition 5.2, and the intuition behind it is as well. Let M' be an operator satisfying $0 \leq M' \leq \mathbb{1}$ and $\text{Tr}[M'\mathcal{N}(\rho)] \geq 1-\varepsilon$. Then, due to the positivity of $\mathcal{N}$, and thus of $\mathcal{N}^\dagger$, we have that $\mathcal{N}^\dagger(M') \geq 0$ and $\mathcal{N}^\dagger(\mathbb{1}-M') \geq 0 \Rightarrow \mathcal{N}^\dagger(M') \leq \mathcal{N}^\dagger(\mathbb{1})$. Since $\mathcal{N}$ is trace preserving, the adjoint $\mathcal{N}^\dagger$ is unital (see Exercise 4.10), which means that $\mathcal{N}^\dagger(\mathbb{1}) = \mathbb{1}$, so that $0 \leq \mathcal{N}^\dagger(M') \leq \mathbb{1}$. Furthermore, by definition of the adjoint of a superoperator, the inequality $\text{Tr}[\mathcal{N}^\dagger(M')\rho] \geq 1-\varepsilon$ holds. Therefore, $\mathcal{N}^\dagger(M')$ is a feasible point in the optimization for $\beta_\varepsilon(\rho\|\sigma)$, so that

$$\beta_\varepsilon(\rho\|\sigma) = \inf\{\text{Tr}[M\sigma] : 0 \leq M \leq \mathbb{1},\ \text{Tr}[M\rho] \geq 1-\varepsilon\} \tag{5.3.130}$$
$$\leq \text{Tr}[\mathcal{N}^\dagger(M')\sigma] \tag{5.3.131}$$
$$= \text{Tr}[M'\mathcal{N}(\sigma)], \tag{5.3.132}$$

where the last line follows by definition of the adjoint of a superoperator. Finally, because the inequality $\beta_\varepsilon(\rho\|\sigma) \leq \text{Tr}[M'\mathcal{N}(\sigma)]$ holds for every operator M' satisfying $0 \leq M' \leq \mathbb{1}$ and $\text{Tr}[M'\mathcal{N}(\rho)] \geq 1-\varepsilon$, we conclude that

$$\beta_\varepsilon(\rho\|\sigma) \leq \inf\{\text{Tr}[M'\mathcal{N}(\sigma)] : 0 \leq M' \leq \mathbb{1},\ \text{Tr}[M'\mathcal{N}(\rho)] \geq 1-\varepsilon\} \tag{5.3.133}$$
$$= \beta_\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \tag{5.3.134}$$

as required. ■

Proposition 5.7 Optimal Measurement for Asymmetric Hypothesis Testing

For every state ρ, positive semi-definite operator σ, and $\varepsilon \in [0,1]$, the minimum type-II error probability[a] $\beta_\varepsilon(\rho\|\sigma)$ is achieved by the following measurement operator:

$$M(\mu^*, p^*) := \Pi_{\mu^*\rho>\sigma} + p^*\Pi_{\mu^*\rho=\sigma}, \tag{5.3.135}$$

where $\Pi_{\mu^*\rho>\sigma}$ is the projection onto the strictly positive part of $\mu^*\rho - \sigma$, the projection $\Pi_{\mu^*\rho=\sigma}$ projects onto the zero eigenspace of $\mu^*\rho - \sigma$, and $\mu^* \geq 0$ and $p^* \in [0, 1]$ are chosen as follows:

$$\mu^* := \sup\left\{\mu : \text{Tr}[\Pi_{\mu\rho>\sigma}\rho] \leq 1 - \varepsilon\right\}, \tag{5.3.136}$$

$$p^* := \frac{1 - \varepsilon - \text{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]}{\text{Tr}[\Pi_{\mu^*\rho=\sigma}\rho]}. \tag{5.3.137}$$

[a] Even though this quantity need not be an error probability when σ is a general positive semi-definite operator, we still refer to it as such.

Proof: First, observe that it suffices to optimize with respect to every measurement operator M that meets the constraint $\text{Tr}[M\rho] \geq 1 - \varepsilon$ with equality. This follows because for every measurement operator M such that $\text{Tr}[M\rho] > 1 - \varepsilon$, there exists a positive number $\lambda \in [0, 1)$ such that $\text{Tr}[(\lambda M)\rho] = 1 - \varepsilon$. Note that $0 \leq \lambda M \leq \mathbb{1}$, so that λM is a measurement operator, and because $\text{Tr}[(\lambda M)\sigma] < \text{Tr}[M\sigma]$, we conclude that

$$\beta_\varepsilon(\rho\|\sigma) = \inf\{\text{Tr}[M\sigma] : 0 \leq M \leq \mathbb{1}, \text{Tr}[M\rho] = 1 - \varepsilon\}. \tag{5.3.138}$$

Based on this, let M be a measurement operator satisfying $\text{Tr}[M\rho] = 1 - \varepsilon$ and let $\mu \geq 0$. Then,

$$\begin{aligned}
\text{Tr}[M\sigma] &= \text{Tr}[M\sigma] + \mu\left(1 - \varepsilon - \text{Tr}[M\rho]\right) &(5.3.139)\\
&= -\mu\varepsilon + \text{Tr}[(\mathbb{1} - M)\mu\rho] + \text{Tr}[M\sigma] &(5.3.140)\\
&\geq -\mu\varepsilon + \frac{1}{2}\left(\text{Tr}[\mu\rho + \sigma] - \|\mu\rho - \sigma\|_1\right) &(5.3.141)\\
&= -\mu\varepsilon + \frac{1}{2}\left(\mu + \text{Tr}[\sigma] - \|\mu\rho - \sigma\|_1\right). &(5.3.142)
\end{aligned}$$

The sole inequality follows as an application of Theorem 5.3, with $B = \sigma$ and $A = \mu\rho$. Observe that the final expression is a universal bound independent of M.

To determine an optimal measurement operator, we can look to Theorem 5.3. There, it was established that the following measurement operator is an optimal one for $\inf_{M:0\leq M\leq\mathbb{1}}\{\text{Tr}[(\mathbb{1} - M)\mu\rho] + \text{Tr}[M\sigma]\}$:

$$M(\mu, p) := \Pi_{\mu\rho>\sigma} + p\Pi_{\mu\rho=\sigma}, \tag{5.3.143}$$

where $\Pi_{\mu\rho>\sigma}$ is the projection onto the strictly positive part of $\mu\rho - \sigma$, the projection $\Pi_{\mu\rho=\sigma}$ projects onto the zero eigenspace of $\mu\rho - \sigma$, and $p \in [0,1]$. The measurement operator $M(\mu, p)$ is called a *quantum Neyman–Pearson test*. We still need to choose the parameters $\mu \geq 0$ and $p \in [0,1]$. Let us pick μ according to the following optimization:

$$\mu^* \coloneqq \sup\left\{\mu : \mathrm{Tr}[\Pi_{\mu\rho>\sigma}\rho] \leq 1 - \varepsilon\right\}. \tag{5.3.144}$$

If it happens that μ^* is such that $\mathrm{Tr}[\Pi_{\mu^*\rho>\sigma}\rho] = 1 - \varepsilon$, then we are done; we can pick $p = 0$. However, if μ^* is such that $\mathrm{Tr}[\Pi_{\mu^*\rho>\sigma}\rho] < 1 - \varepsilon$, then we pick $p^* \in [0,1]$ such that

$$p^* \coloneqq \frac{1 - \varepsilon - \mathrm{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]}{\mathrm{Tr}[\Pi_{\mu^*\rho=\sigma}\rho]}, \tag{5.3.145}$$

with it following that $p^* \in [0,1]$ because

$$\mathrm{Tr}[\Pi_{\mu^*\rho>\sigma}\rho] < 1 - \varepsilon \leq \mathrm{Tr}[\Pi_{\mu^*\rho\geq\sigma}\rho]. \tag{5.3.146}$$

With these choices, we then find that

$$\mathrm{Tr}[M(\mu^*, p^*)\rho] = 1 - \varepsilon. \tag{5.3.147}$$

By the analysis in (5.3.139)–(5.3.142), it then follows that

$$\mathrm{Tr}[M\sigma] \geq \mathrm{Tr}[M(\mu^*, p^*)\sigma] \tag{5.3.148}$$

for every measurement operator M satisfying $0 \leq M \leq \mathbb{1}$ and $\mathrm{Tr}[M\rho] = 1 - \varepsilon$. ■

5.3.3.1 Asymptotic Setting

Now, suppose that multiple copies, say n, of the state (ρ or σ) are available. Based on the discussion at the beginning of Section 5.3, the optimal type-II error probability, given an upper bound of ε on the type-I error probability, is $\beta_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$ for all $n \geq 1$. Then, as in the symmetric case, we are interested in the behaviour of this type-II error probability as n becomes large. Furthermore, based on the earlier discussion of the trade-off between the type-I and type-II error probabilities, we might imagine that as n becomes large it is possible to bring the type-I error probability all the way down to zero, because the states become more distinguishable as n increases. In order to investigate this possibility, we consider the *optimal type-II error exponent*, by analogy with the error exponent for state discrimination that we considered in Section 5.3.1.1. To be precise, we consider the following quantities:

$$E(\rho,\sigma) \coloneqq \inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} -\frac{1}{n}\log_2 \beta_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}), \tag{5.3.149}$$

$$\widetilde{E}(\rho,\sigma) \coloneqq \sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} -\frac{1}{n}\log_2 \beta_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}). \tag{5.3.150}$$

We refer to $E(\rho,\sigma)$ as the optimal achievable type-II error exponent and $\widetilde{E}(\rho,\sigma)$ as the optimal strong converse type-II error exponent. Note that the following inequality is a direct consequence of definitions:

$$E(\rho,\sigma) \le \widetilde{E}(\rho,\sigma). \tag{5.3.151}$$

The following result, known as the *quantum Stein's lemma*, provides us with a tractable expression for $E(\rho,\sigma)$ and $\widetilde{E}(\rho,\sigma)$ in terms of the *quantum relative entropy*, an important quantity in quantum information theory that we introduce formally in Chapter 7. We also delay the proof of the result to Chapter 7, when all of the required elements of the proof become available to us.

Theorem 5.8 Quantum Stein's Lemma

For all states ρ and σ, the optimal achievable and strong converse rates are equal to the quantum relative entropy of ρ and σ, i.e.,

$$E(\rho,\sigma) = \widetilde{E}(\rho,\sigma) = D(\rho\|\sigma), \tag{5.3.152}$$

where

$$D(\rho\|\sigma) \coloneqq \begin{cases} \mathrm{Tr}[\rho(\log_2\rho - \log_2\sigma)] & \text{if } \mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma), \\ +\infty & \text{otherwise} \end{cases} \tag{5.3.153}$$

is the *quantum relative entropy*.

PROOF: See Section 7.10. ■

REMARK: See Section 2.2.8.1 for the definition of the logarithm of a Hermitian operator. See Section 7.2 for a more detailed explanation of the support conditions in the definition of the quantum relative entropy.

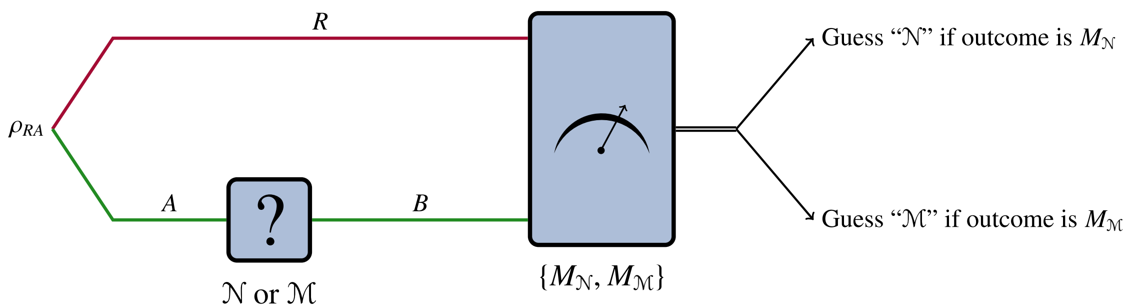


FIGURE 5.11: The most general strategy for discriminating two quantum channels $\mathcal{N}$ and $\mathcal{M}$ is to prepare a bipartite state ρ_{RA}, with the reference system R having arbitrary dimension, sending the system A through the unknown quantum channel, and then measuring both systems R and A according to a two-outcome POVM $\{M_{\mathcal{N}}, M_{\mathcal{M}}\}$. If the outcome corresponding to $M_{\mathcal{N}}$ occurs, then we guess that the unknown channel is $\mathcal{N}$; otherwise, we guess that it is $\mathcal{M}$. The minimum error probability among all such strategies is given by Theorem 5.9.

5.4 Quantum Channel Discrimination

In Sections 5.3.1 and 5.3.2, we considered symmetric hypothesis testing with respect to quantum states, also known as quantum state discrimination, in which the hypotheses are modelled as quantum states, and the task is to devise a measurement strategy that maximizes the probability of success of correctly identifying the state of a given quantum system. Let us now consider the analogous scenario with quantum channels, which we refer to as *quantum channel discrimination*. We do not discuss asymmetric hypothesis testing with respect to quantum channels in this book; please see the Bibliographic Notes in Section 5.6 for references.

The task of quantum channel discrimination is as follows. Consider that a quantum system A undergoes an evolution according to one of two quantum channels, $\mathcal{N}_{A\to B}$ or $\mathcal{M}_{A\to B}$. Furthermore, we suppose that the quantum channel is chosen to be $\mathcal{N}$ with probability $\lambda \in [0,1]$ and $\mathcal{M}$ with probability $1-\lambda$. The goal is to devise a strategy that correctly guesses the unknown quantum channel with the highest probability.

The most general strategy for quantum channel discrimination is illustrated in Figure 5.11. The strategy consists of a state ρ_{RA} and a two-outcome measurement described by the POVM $\{M_{\mathcal{N}}, M_{\mathcal{M}}\}$. The bipartite state ρ_{RA} is such that the system R can be arbitrarily large in principle, in order try to achieve the lowest error probability. The measurement acts on both the system R and the system A after system A has passed through the unknown channel. The expected error probability of this strategy is analogous to the expected error probability of a strategy for quantum state discrimination: it is the expectation, with respect to the prior probability distribution given by λ, of the probabilities of the two types of errors that can occur: guessing "$\mathcal{M}$" when the channel is $\mathcal{N}$, and guessing "$\mathcal{N}$" when the channel is $\mathcal{M}$. In other words, the expected error probability is

$$\lambda \mathrm{Tr}[M_{\mathcal{M}} \mathcal{N}_{A\to B}(\rho_{RA})] + (1-\lambda)\mathrm{Tr}[M_{\mathcal{N}} \mathcal{M}_{A\to B}(\rho_{RA})] \\ = p_{\mathrm{err}}(\lambda, \mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA}), M), \quad (5.4.1)$$

where the last line follows by letting $M_{\mathcal{N}} \equiv M$ and from the definition of p_{err} in (5.3.1). We see that, given a state ρ_{RA}, the task of discriminating $\mathcal{N}$ and $\mathcal{M}$ reduces to the task of discriminating the states $\mathcal{N}_{A\to B}(\rho_{RA})$ and $\mathcal{M}_{A\to B}(\rho_{RA})$. The optimal error probability is obtained by optimizing with respect to every state ρ_{RA} and measurement operator M, so that

$$p^*_{\mathrm{err}}(\lambda, \mathcal{N}, \mathcal{M}) := \inf_{\substack{\rho_{RA} \\ M:0\le M\le \mathbb{1}}} p_{\mathrm{err}}(\lambda, \mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA}), M) \quad (5.4.2)$$

$$= \inf_{\rho_{RA}} p^*_{\mathrm{err}}(\lambda, \mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA})), \quad (5.4.3)$$

where the optimization is with respect to every quantum state ρ_{RA}, and there is an implicit optimization with respect to the dimension of the system R. This gives us a first look into how a quantity defined initially for quantum states can be "lifted" to a quantity defined for quantum channels. In particular, the quantity p^*_{err}, initially defined for quantum states as in (5.3.6), has been extended to quantum channels by evaluating the state quantity with respect to the states $\mathcal{N}_{A\to B}(\rho_{RA})$ and $\mathcal{M}_{A\to B}(\rho_{RA})$ and then optimizing with respect to both to every state ρ_{RA} and the dimension of R. Such constructions of channel quantities from state quantities arise throughout the rest of the book.

Just as the optimal error probability for discriminating two states can be expressed using the trace norm (recall (5.3.15)), we now show that, analogously, the optimal error probability for discriminating two quantum channels can be expressed in terms of the diamond norm.

Theorem 5.9 Minimum Error Probability for Quantum Channel Discrimination

Let $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ be quantum channels, and $\lambda \in [0,1]$. The optimal error probability for discriminating $\mathcal{N}$ and $\mathcal{M}$ is given by

$$p^*_{\mathrm{err}}(\lambda, \mathcal{N}, \mathcal{M}) = \inf_{\psi_{A'A}} p^*_{\mathrm{err}}(\lambda, \mathcal{N}_{A\to B}(\psi_{A'A}), \mathcal{M}_{A\to B}(\psi_{A'A})) \tag{5.4.4}$$

$$= \frac{1}{2}(1 - \|\lambda\mathcal{N} - (1-\lambda)\mathcal{M}\|_\diamond), \tag{5.4.5}$$

where, in the first line, the optimization is with respect to every pure state $\psi_{A'A}$ with $d_{A'} = d_A$.

PROOF: Using (5.3.15), we have

$$p^*_{\mathrm{err}}(\lambda, \mathcal{N}, \mathcal{M}) = \frac{1}{2}\left(1 - \sup_{\rho_{RA}} \|\lambda\mathcal{N}_{A\to B}(\rho_{RA}) - (1-\lambda)\mathcal{M}_{A\to B}(\rho_{RA})\|_1\right). \tag{5.4.6}$$

Now, consider a state ρ_{RA} in the optimization above, and let $\psi_{R'RA}$ be a purification of ρ_{RA}. Then,

$$\begin{aligned}
&\|\lambda\mathcal{N}_{A\to B}(\rho_{RA}) - (1-\lambda)\mathcal{M}_{A\to B}(\rho_{RA})\|_1 \\
&\quad = \|\lambda\mathrm{Tr}_{R'}[\mathcal{N}_{A\to B}(\psi_{R'RA})] - (1-\lambda)\mathrm{Tr}_{R'}[\mathcal{M}_{A\to B}(\psi_{R'RA})]\|_1 && (5.4.7)\\
&\quad \le \|\lambda\mathcal{N}_{A\to B}(\psi_{R'RA}) - (1-\lambda)\mathcal{M}_{A\to B}(\psi_{R'RA})\|_1, && (5.4.8)
\end{aligned}$$

where the inequality follows from (4.1.7), with respect to the partial trace channel $\mathrm{Tr}_{R'}$. Now, without loss of generality, we can let $d_{R'} \ge d_R d_A$; see Section 3.2.5. Then, by the Schmidt decomposition theorem (Theorem 2.2), in particular (2.2.61), the state vector $|\psi\rangle_{R'RA}$ can be expressed according to the $R'R|A$ bipartition as $|\psi\rangle_{R'RA} = \sum_{k=1}^{d_A} \sqrt{p_k}|u_k\rangle_{R'R} \otimes |v_k\rangle_A$, where p_k is a probability and the vectors $|u_k\rangle_{R'R}$ and $|v_k\rangle_A$ form orthonormal bases for a d_A-dimensional vector space. In other words, only a d_A-dimensional subspace of $\mathcal{H}_{R'R}$, call it $\mathcal{H}_{A'}$, is relevant for calculating the trace norm in (5.4.8), and there exists an isometry $V_{A'\to R'R}$ such that $V_{A'\to R'R}|\psi\rangle_{A'A} = |\psi\rangle_{R'RA}$. Adopting the shorthand $V \equiv V_{A'\to R'R}$, it follows that

$$\begin{aligned}
&\|\lambda\mathcal{N}_{A\to B}(\rho_{RA}) - (1-\lambda)\mathcal{M}_{A\to B}(\rho_{RA})\|_1 \\
&\quad \le \left\|\lambda\mathcal{N}_{A\to B}(V\psi_{A'A}V^\dagger) - (1-\lambda)\mathcal{M}_{A\to B}(V\psi_{A'A}V^\dagger)\right\|_1 && (5.4.9)\\
&\quad = \|\lambda\mathcal{N}_{A\to B}(\psi_{A'A}) - (1-\lambda)\mathcal{M}_{A\to B}(\psi_{A'A})\|_1 && (5.4.10)
\end{aligned}$$

$$\leq \sup_{\psi_{A'A}} \|\lambda \mathcal{N}_{A\to B}(\psi_{A'A}) - (1-\lambda)\mathcal{M}_{A\to B}(\psi_{A'A})\|_1 \tag{5.4.11}$$

$$= \|\lambda \mathcal{N} - (1-\lambda)\mathcal{M}\|_\diamond , \tag{5.4.12}$$

where the last line follows from (2.2.175), because the map $\lambda\mathcal{N} - (1-\lambda)\mathcal{M}$ is Hermiticity preserving. Therefore,

$$p^*_{\text{err}}(\lambda, \mathcal{N}, \mathcal{M}) \geq \frac{1}{2}\left(1 - \|\lambda\mathcal{N} - (1-\lambda)\mathcal{M}\|_\diamond\right). \tag{5.4.13}$$

The opposite inequality holds simply by restricting the optimization in (5.4.3) to every pure state $\psi_{A'A}$ satisfying $d_{A'} = d_A$. We thus obtain (5.4.5), as required. ■

Exercise 5.20

Consider quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, and let $\lambda \in [0,1]$. Using (5.4.4) and (5.4.1), show that the optimal error probability $p^*_{\text{err}}(\lambda, \mathcal{N}, \mathcal{M})$ can be evaluated using a semi-definite program. Then, using strong duality, prove that an alternate expression for $p^*_{\text{err}}(\lambda, \mathcal{N}, \mathcal{M})$ is

$$p^*_{\text{err}}(\lambda, \mathcal{N}, \mathcal{M}) = \sup_{\substack{W_{AB} \\ \text{Hermitian}}} \left\{\lambda_{\min}(\text{Tr}_B[W_{AB}]) : W_{AB} \leq \lambda\Gamma^{\mathcal{N}}_{AB},\ W_{AB} \leq (1-\lambda)\Gamma^{\mathcal{M}}_{AB}\right\}, \tag{5.4.14}$$

where $\lambda_{\min}(\text{Tr}_B[W_{AB}])$ is the smallest eigenvalue of $\text{Tr}_B[W_{AB}]$; see Exercise 2.31.

Exercise 5.21

Consider quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, and let $\lambda \in [0,1]$. Prove the following bounds on the optimal error probability for discriminating $\mathcal{N}$ and $\mathcal{M}$ in terms of the optimal error probability for discriminating the Choi states of $\mathcal{N}$ and $\mathcal{M}$:

$$d_A p^*_{\text{err}}(\lambda, \Phi^{\mathcal{N}}_{AB}, \Phi^{\mathcal{M}}_{AB}) \leq p^*_{\text{err}}(\lambda, \mathcal{N}, \mathcal{M}) \leq p^*_{\text{err}}(\lambda, \Phi^{\mathcal{N}}_{AB}, \Phi^{\mathcal{M}}_{AB}). \tag{5.4.15}$$

(*Hint*: See Exercise 4.7.)

The upper bound in (5.4.15) corresponds to the strategy that consists of letting the state ρ_{RA} in Figure 5.11 be the maximally-entangled state $\Phi_{A'A} = |\Phi\rangle\langle\Phi|_{A'A}$, with $|\Phi\rangle_{A'A} = \frac{1}{\sqrt{d_A}} \sum_{i=0}^{d_A-1} |i,i\rangle_{A'A}$. The following exercise tells us when this strategy is optimal, i.e., when the upper bound in (5.4.15) is achieved.

Exercise 5.22

Let the quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ be jointly covariant with respect to a group G, so that

$$\mathcal{N}_{A\to B}(U_A^g \rho_A U_A^{g\dagger}) = V_B^g \mathcal{N}_{A\to B}(\rho_A) V_B^{g\dagger}, \tag{5.4.16}$$

$$\mathcal{M}_{A\to B}(U_A^g \rho_A U_A^{g\dagger}) = V_B^g \mathcal{M}_{A\to B}(\rho_A) V_B^{g\dagger}, \tag{5.4.17}$$

for every $g \in G$ and every state ρ_A, where $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$ are unitary representations of G acting on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Furthermore, let $\{U_A^g\}_{g\in G}$ be such that

$$\mathcal{T}_A^G(\cdot) := \frac{1}{|G|} \sum_{g\in G} U_A^g (\cdot) U_A^{g\dagger} = \mathrm{Tr}[\cdot] \frac{\mathbb{1}_A}{d_A}. \tag{5.4.18}$$

Prove that $p_{\text{err}}^*(\lambda, \mathcal{N}, \mathcal{M}) = p_{\text{err}}^*(\lambda, \Phi_{AB}^{\mathcal{N}}, \Phi_{AB}^{\mathcal{M}})$. (*Hint*: Use Proposition 4.20.)

5.5 Summary

5.6 Bibliographic Notes

The quantum teleportation protocol for qubits and qudits presented in Section 5.1 was devised by Bennett et al. (1993). The generalization to groups that act irreducibly on the input state was presented by Braunstein et al. (2000); Werner (2001). Covariant channels were considered by Holevo (2002b). For an overview of group- and representation-theoretic concepts for finite groups, see Steinberg (2011).

The notion of teleportation simulation of a quantum channel was introduced by Bennett et al. (1996c) (see Section V therein). The LOCC simulation of a quantum channel was given by Horodecki et al. (1999) (see Eq. (10) therein), and the related PPT simulation of a quantum channel was given by Kaur and Wilde (2017). The generalized teleportation simulation of a quantum channel was developed for groups that act irreducibly on the channel input space by Chiribella et al. (2009) (see Section VII therein). Pirandola et al. (2017) played a role in developing the tool of LOCC/teleportation simulation more recently. The fact that the channel twirl can be realized by generalized teleportation simulation was observed by Kaur and Wilde (2017) (see Appendix B therein). The teleportation protocol was extended to states of bosonic systems by Braunstein and Kimble (1998). Teleportation simulation of bosonic Gaussian channels was given by Giedke and Ignacio Cirac (2002); Niset et al. (2009), and a detailed analysis of convergence in this scenario was presented by Wilde (2018a).

The quantum super-dense coding protocol was discovered by Bennett and Wiesner (1992).

The problem of state discrimination, as described in Section 5.3.1, was considered by Helstrom (1967) (see also Helstrom (1969)) and Holevo (1972a), who determined the optimal success probability in the case of projective measurements and general POVMs, respectively. The proof of the optimal measurement operators in Theorem 5.3 is due to Jencova (2010). Lemma 5.5 is due to Audenaert et al. (2007), with the simple proof presented here attributed by Jaksic et al. (2012) and Audenaert (2014) as being due to N. Ozawa. The Chernoff bound for probability distributions was given by Chernoff (1952). The corresponding bound for quantum states in Theorem 5.4 was established in two works: Audenaert et al. (2007) determined the upper bound on the optimal error exponent, while Nussbaum and Szkoła (2009) established the lower bound. The semi-definite program and complementary slackness conditions for multiple state discrimination in Section 5.3.2 are due to Yuen et al. (1975). For work on measurement strategies in two-state discrimination, including adaptive strategies and strategies that achieve the Chernoff bound, we refer to the work of Brody and Meister (1996); Ban et al. (1997); Acín et al. (2005); Higgins et al. (2011); Brandsen et al. (2020).

The operational interpretation of diamond distance in terms of symmetric hypothesis testing of quantum channels (specifically, Theorem 5.9) was given by Rosgen and Watrous (2005); Sacchi (2005). The work of Kretschmann and Werner (2004); Gilchrist et al. (2005) provides a different operational interpretation of diamond distance in terms of quantifying the error between an ideal channel and an experimental approximation of it, which we elaborate upon in Chapter 6. ...(references for multiple channel discrimination...references for asymmetric hypothesis testing for channels...)

5.7 Problems

1. Let ρ_{AB} be a quantum state with $d_A = d_B = d \geq 2$, and consider the quantity

$$q_{\text{corr}}(A|B)_\rho \equiv q_{\text{corr}}(\rho_{AB}) := d \sup_{\mathcal{N}} \langle \Phi|_{AB} (\text{id}_A \otimes \mathcal{N}_B)(\rho_{AB}) |\Phi\rangle_{AB}, \tag{5.7.1}$$

where the optimization is with respect to quantum channels $\mathcal{N}_B$ acting on system B.

(a) Show that $q_{\text{corr}}(\rho_{AB})$ can be evaluated using a semi-definite program, and determine the corresponding dual program.

(b) Suppose that ρ_{AB} is a classical–quantum state, i.e., $\rho_{AB} \equiv \rho_{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_B^x$, where $\mathcal{X}$ is a set of size d, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\rho_B^x\}_{x\in\mathcal{X}}$ is a set of states. Prove that

$$q_{\text{corr}}(\rho_{XB}) = p^*_{\text{succ}}(\{(p(x), \rho^x)\}_x). \tag{5.7.2}$$

In other words, for classical–quantum states, the quantity q_{corr} reduces to the optimal success probability for multiple state discrimination of the set $\{\rho_B^x\}_{x\in\mathcal{X}}$. (*Hint*: See Exercise 4.11.)

(*Bibliographic Note*: The function q_{corr} was defined by Koenig et al. (2009) within the context of the min-entropy (a quantity that we encounter in Chapter 7) and its operational meaning.)

Chapter 6

Distinguishibility Measures for Quantum States and Channels

[in progress]

6.1 Trace Distance

The trace distance is a distance measure based on the trace norm; see Section 2.2.9.2. For two quantum states ρ and σ, we define the *normalized trace distance between* ρ *and* σ as $\frac{1}{2}\left\|\rho-\sigma\right\|_1$.

Using (2.23), observe that the normalized trace distance between pure states $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$ is given by

$$\frac{1}{2}\left\||\psi\rangle\langle\psi|-|\phi\rangle\langle\phi|\right\|_1=\sqrt{1-|\langle\psi|\phi\rangle|^2}. \tag{6.1.1}$$

In Section 5.3.1, we showed that the trace distance arises in terms of the optimal error probability for discriminating two quantum states. Specifically, for two quantum states ρ and σ, and for $\lambda=\frac{1}{2}$, we have

$$p_{\text{err}}^*(1/2,\rho,\sigma)=\frac{1}{2}\left(1-\frac{1}{2}\left\|\rho-\sigma\right\|_1\right). \tag{6.1.2}$$

The trace distance thus has an operational meaning as quantifying the optimal error probability for distinguishing two quantum states with equal prior probability.

An alternative operational meaning of the trace distance is in terms of assessing the performance of a quantum information processing protocol in which the ideal state to be generated is ρ but the actual state generated is σ. To see that this is the case, suppose that a third party is trying to assess how distinguishable the actual state σ is from the ideal state ρ. Such an individual can do so by performing a quantum measurement described by the POVM $\{M_x\}_{x\in\mathcal{X}}$ whose elements are indexed by some finite set $\mathcal{X}$. In the case that ρ was prepared, the probability of obtaining the outcome corresponding to x is $\text{Tr}[M_x\rho]$, and in the case that σ was prepared, this probability is $\text{Tr}[M_x\sigma]$. In order for ρ and σ to be considered "close," what we demand is that the absolute deviation between the actual probability $\text{Tr}[M_x\sigma]$ and the ideal probability $\text{Tr}[M_x\rho]$ be no larger than some desired tolerance $\varepsilon > 0$, so that $|\text{Tr}[M_x\rho] - \text{Tr}[M_x\sigma]| \leq \varepsilon$. We want this condition to hold for *all* possible measurements that one could perform, so what we demand mathematically is that

$$\sup_{0\leq M\leq \mathbb{1}} |\text{Tr}[M\rho] - \text{Tr}[M\sigma]| \leq \varepsilon. \tag{6.1.3}$$

As stated in Theorem 6.1 below, the following identity holds

$$\sup_{0\leq M\leq \mathbb{1}} |\text{Tr}[M\rho] - \text{Tr}[M\sigma]| = \frac{1}{2}\left\|\rho - \sigma\right\|_1, \tag{6.1.4}$$

indicating that if $\frac{1}{2}\left\|\rho - \sigma\right\|_1 \leq \varepsilon$, then the deviation between probabilities for every possible measurement operator never exceeds ε, so that the approximation between states ρ and σ is naturally quantified by the trace distance $\frac{1}{2}\left\|\rho - \sigma\right\|_1$.

Let us now prove the variational characterization of the trace distance stated in (6.1.4).

Theorem 6.1 Trace Distance via Measurement

For two states ρ and σ, the following equality holds

$$\frac{1}{2}\left\|\rho - \sigma\right\|_1 = \sup_{0\leq M\leq \mathbb{1}} |\text{Tr}[M(\rho - \sigma)]|. \tag{6.1.5}$$

An optimal measurement operator M is equal to $\Pi_+ + \Lambda_0$, where Π_+ is the projection onto the strictly positive part of $\rho - \sigma$ and Λ_0 is an operator satisfying $0 \leq \Lambda_0 \leq \Pi_0$, with Π_0 the projection onto the zero eigenspace of $\rho - \sigma$.

PROOF: Consider that

$$\sup_{0\leq\Lambda\leq\mathbb{1}} |\text{Tr}[M(\rho-\sigma)]| = \sup_{0\leq M\leq\mathbb{1}} \text{Tr}[M(\rho-\sigma)], \tag{6.1.6}$$

because there are always choices for M such that $\text{Tr}[M(\rho-\sigma)] \geq 0$. Then the equality follows as a direct application of (5.3.17) and the optimality statement following it, with $A = \rho$ and $B = \sigma$. ■

From Exercise 2.30, we know that, for Hermitian operators, the trace norm can be evaluated using semi-definite programming. The normalized trace distance $\frac{1}{2}\|\rho-\sigma\|_1$ can therefore be evaluated using semi-definite programming, because $\rho-\sigma$ is Hermitian. Since ρ and σ are positive semi-definite, we obtain the following simpler semi-definite programs for their normalized trace distance.

Proposition 6.2 SDPs for Normalized Trace Distance

The trace distance between every two quantum states ρ and σ can be written as the following semi-definite programs:

$$\frac{1}{2}\|\rho-\sigma\|_1 = \sup_{M\geq 0}\{\text{Tr}[M(\rho-\sigma)] : \Lambda \leq \mathbb{1}\} \tag{6.1.7}$$

$$= \inf_{Z\geq 0}\{\text{Tr}[Z] : Z \geq \rho-\sigma\}. \tag{6.1.8}$$

PROOF: The expression in (6.1.7) is immediate from (5.3.17), which already provides an expression for $\frac{1}{2}\|\rho-\sigma\|_1$ as a semi-definite program in the primal form as in (2.4.3). Obtaining the expression in (6.1.8) is then straightforward; see Exercise 6.1. ■

Exercise 6.1

Prove (6.1.8).

The trace distance obeys the following data-processing inequality.

Theorem 6.3 Data-Processing Inequality for Trace Distance

Let ρ and σ be states, and let $\mathcal{N}$ be a positive, trace-non-increasing map. Then,

$$\|\rho - \sigma\|_1 \geq \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 . \tag{6.1.9}$$

PROOF: This is immediate from (4.1.7), which tells us that the trace norm is monotone non-increasing under the action of every positive trace-non-increasing superoperator for every linear operator. It is also possible to provide a direct proof using the expression in (6.1.7); see Exercise 6.2. ■

Exercise 6.2

Provide a direct proof of (6.1.9) using the expression in (6.1.7). (*Hint*: See the proof of Proposition 5.2.)

By combining the results of Theorems 6.1 and 6.3, we find that the trace distance is achieved by a measurement channel:

Theorem 6.4 Trace Distance Achieved by Measurement Channel

For two states ρ and σ, the following equality holds

$$\|\rho - \sigma\|_1 = \max_{\{\Lambda_x\}_x} \sum_{x \in \mathcal{X}} |\mathrm{Tr}[\Lambda_x \rho] - \mathrm{Tr}[\Lambda_x \sigma]| , \tag{6.1.10}$$

where the optimization is performed over POVMs $\{\Lambda_x\}_{x\in\mathcal{X}}$ defined with respect to a finite alphabet $\mathcal{X}$, and an optimal POVM is given by $\{\Lambda^*, \mathbb{1} - \Lambda^*\}$, where $\Lambda^* = \Pi_+ + \Lambda_0$, the projection Π_+ is the projection onto the strictly positive part of $\rho - \sigma$, and Λ_0 satisfies $0 \leq \Lambda_0 \leq \Pi_0$, with Π_0 the projection onto the zero eigenspace of $\rho - \sigma$.

PROOF: The inequality

$$\|\rho - \sigma\|_1 \geq \max_{\{\Lambda_x\}_x} \sum_{x \in \mathcal{X}} |\mathrm{Tr}[\Lambda_x \rho] - \mathrm{Tr}[\Lambda_x \sigma]| \tag{6.1.11}$$

follows from Theorem 6.3 by taking the channel $\mathcal{N}$ to be the quantum–classical channel defined as $\mathcal{N}(\omega) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[\Lambda_x \omega] |x\rangle\!\langle x|$. Then, from Theorem 6.1, we have the equality

$$\frac{1}{2}\|\rho-\sigma\|_1 = |\mathrm{Tr}[\Lambda^*(\rho-\sigma)]|. \tag{6.1.12}$$

Since $|\mathrm{Tr}[\Lambda^*(\rho-\sigma)]| = |\mathrm{Tr}[(\mathbb{1}-\Lambda^*)(\rho-\sigma)]|$, we conclude that

$$\|\rho-\sigma\|_1 = |\mathrm{Tr}[\Lambda^*(\rho-\sigma)]| + |\mathrm{Tr}[(\mathbb{1}-\Lambda^*)(\rho-\sigma)]|, \tag{6.1.13}$$

so that an optimal POVM is given by $\{\Lambda^*, \mathbb{1}-\Lambda^*\}$. ■

6.2 Fidelity

In addition to the trace distance, another distinguishability measure for states that we consider in this book is the fidelity (also called Uhlmann fidelity).

Definition 6.5 Fidelity

For two quantum states ρ and σ, the *fidelity between ρ and σ*, denoted by $F(\rho,\sigma)$, is defined as

$$F(\rho,\sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2 = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]\right)^2. \tag{6.2.1}$$

Observe that the fidelity is symmetric in its arguments. We also have that $F(\rho,\sigma)\in[0,1]$ for all states ρ and σ, a fact that we prove below.

For a pure state $|\psi\rangle$ and mixed state ρ, the fidelity between them is equal to

$$F(\rho,|\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle = \mathrm{Tr}[|\psi\rangle\langle\psi|\rho]. \tag{6.2.2}$$

Also, for two pure states $|\psi\rangle$ and $|\phi\rangle$, the fidelity is simply

$$F(|\psi\rangle\langle\psi|,|\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle|^2. \tag{6.2.3}$$

The formula in (6.2.2) gives the fidelity an operational meaning that we employ in later chapters. Suppose that the goal of a quantum information processing protocol is to produce the pure state $|\psi\rangle\langle\psi|$, but it instead produces a mixed state ρ. Then the fidelity $F(\rho,|\psi\rangle\langle\psi|)$ is equal to the probability that the actual state ρ passes a test for being the ideal state $|\psi\rangle\langle\psi|$, with the test being given by the POVM

$\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$. That is, the probability of obtaining the first outcome of the measurement (i.e., "success") is equal to $F(\rho, |\psi\rangle\langle\psi|)$. In this way, the fidelity provides another natural way for assessing the performance of quantum information processing protocols.

Like the trace distance, the fidelity can be computed via a semi-definite program, as stated in the following proposition:

Proposition 6.6 SDPs for Root Fidelity of States

The root fidelity $\sqrt{F}(\rho,\sigma) = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$ of quantum states ρ and σ is characterized by the following primal and dual semi-definite programs:

$$\sqrt{F}(\rho,\sigma) = \frac{1}{2}\sup_{X\in\mathcal{L}(\mathcal{H})}\left\{\operatorname{Tr}[X] + \operatorname{Tr}[X^\dagger] : \begin{pmatrix} \rho & X \\ X^\dagger & \sigma \end{pmatrix} \geq 0\right\} \tag{6.2.4}$$

$$= \frac{1}{2}\inf_{Y,Z\geq 0}\left\{\operatorname{Tr}[Y\rho] + \operatorname{Tr}[Z\sigma] : \begin{pmatrix} Y & \mathbb{1} \\ \mathbb{1} & Z \end{pmatrix} \geq 0\right\} \tag{6.2.5}$$

PROOF: See Appendix 6.B.1. ■

Theorem 6.7 Basic Properties of Fidelity

1. For two states ρ and σ, the inequalities $0 \leq F(\rho,\sigma) \leq 1$ hold. Furthermore, $F(\rho,\sigma) = 1$ if and only if $\rho = \sigma$, and $F(\rho,\sigma) = 0$ if and only if ρ and σ are supported on orthogonal subspaces.
2. *Isometric invariance*: For all states ρ and σ, and for every isometry V,
$$F(\rho,\sigma) = F(V\rho V^\dagger, V\sigma V^\dagger). \tag{6.2.6}$$
3. *Multiplicativity*: The fidelity is multiplicative with respect to tensor-product states, meaning that for all states $\rho_1, \sigma_1, \rho_2, \sigma_2$, we have
$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1,\sigma_1)F(\rho_2,\sigma_2). \tag{6.2.7}$$

PROOF:

1. The fact that $F(\rho,\sigma) \geq 0$ for all states ρ and σ follows from the definition

of the fidelity as the squared trace norm and the fact that the trace norm is always non-negative. If ρ and σ are supported on orthogonal subspaces, then $\sqrt{\rho}\sqrt{\sigma} = 0$, which means that $F(\rho, \sigma) = 0$. Conversely, if $F(\rho, \sigma) = 0$, then $\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 = 0$, which implies (by definition of a norm) that $\sqrt{\rho}\sqrt{\sigma} = 0$, which in turn implies that ρ and σ are supported on orthogonal subspaces.

Now, using (2.2.115), there exists a unitary U such that

$$F(\rho, \sigma) = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2 = \left|\mathrm{Tr}[U\sqrt{\rho}\sqrt{\sigma}]\right|^2. \tag{6.2.8}$$

Then, using the Cauchy–Schwarz inequality for the Hilbert–Schmidt inner product (see (2.2.32)), we find that

$$\begin{aligned} F(\rho, \sigma) &= \left|\mathrm{Tr}[U\sqrt{\rho}\sqrt{\sigma}]\right|^2 && (6.2.9)\\ &\leq \mathrm{Tr}[U\sqrt{\rho}\sqrt{\rho}U^\dagger]\mathrm{Tr}[\sqrt{\sigma}\sqrt{\sigma}] && (6.2.10)\\ &= \mathrm{Tr}[\sqrt{\rho}\sqrt{\rho}]\mathrm{Tr}[\sqrt{\sigma}\sqrt{\sigma}] && (6.2.11)\\ &= \mathrm{Tr}[\rho]\mathrm{Tr}[\sigma] && (6.2.12)\\ &= 1. && (6.2.13) \end{aligned}$$

If $\rho = \sigma$, then $F(\rho, \sigma) = \|\rho\|_1^2 = \mathrm{Tr}[\rho]^2 = 1$. On the other hand, if $F(\rho, \sigma) = 1$, then the inequality in the Cauchy–Schwarz inquality is saturated. The Cauchy–Schwarz inequality is saturated if and only if the two operators involved are proportional to each other. This means that $\rho = \alpha\sigma$ for some $\alpha > 0$. But since both ρ and σ are states, it must be the case that $\alpha = 1$, which means that $\rho = \sigma$.

2. *Proof of isometric invariance*: For every isometry V and every two states ρ and σ, since the action of an isometry does not change the eigenvalues, we have that $\sqrt{V\rho V^\dagger} = V\sqrt{\rho}V^\dagger$ and $\sqrt{V\sigma V^\dagger} = V\sqrt{\sigma}V^\dagger$. Therefore,

$$\begin{aligned} F(V\rho V^\dagger, V\sigma V^\dagger) &= \left\|\sqrt{V\rho V^\dagger}\sqrt{V\sigma V^\dagger}\right\|_1^2 && (6.2.14)\\ &= \left\|V\sqrt{\rho}V^\dagger V\sqrt{\sigma}V^\dagger\right\|_1^2 && (6.2.15)\\ &= \left\|V\sqrt{\rho}\sqrt{\sigma}V^\dagger\right\|_1^2 && (6.2.16)\\ &= \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2, && (6.2.17) \end{aligned}$$

as required, where the last line is due to the isometric invariance of the Schatten norms, as stated in (2.2.92).

3. *Proof of multiplicativity*: Using the fact that $\sqrt{\rho_1 \otimes \rho_2} = \sqrt{\rho_1} \otimes \sqrt{\rho_2}$, and similarly for $\sqrt{\sigma_1 \otimes \sigma_2}$, and using the multiplicativity of the trace norm with respect to the tensor product (see (2.2.95)), we find that

$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = \left\| \sqrt{\rho_1 \otimes \rho_2}\sqrt{\sigma_1 \otimes \sigma_2} \right\|_1^2 \tag{6.2.18}$$
$$= \left\| (\sqrt{\rho_1} \otimes \sqrt{\rho_2})(\sqrt{\sigma_1} \otimes \sqrt{\sigma_2}) \right\|_1^2 \tag{6.2.19}$$
$$= \left\| \sqrt{\rho_1}\sqrt{\sigma_1} \otimes \sqrt{\rho_2}\sqrt{\sigma_2} \right\|_1^2 \tag{6.2.20}$$
$$= \left\| \sqrt{\rho_1}\sqrt{\sigma_1} \right\|_1^2 \left\| \sqrt{\rho_2}\sqrt{\sigma_2} \right\|_1^2 \tag{6.2.21}$$
$$= F(\rho_1, \sigma_1) F(\rho_2, \sigma_2), \tag{6.2.22}$$

as required. ■

Theorem 6.8 Uhlmann's Theorem

For two quantum states ρ_A and σ_A, let $|\psi^\rho\rangle := (\mathbb{1}_R \otimes \sqrt{\rho_A})|\Gamma\rangle_{RA}$ and $|\psi^\sigma\rangle := (\mathbb{1}_R \otimes \sqrt{\sigma_A})|\Gamma\rangle_{RA}$ be purifications of ρ and σ, respectively, with the dimension of R equal to the dimension of A. Then,

$$F(\rho, \sigma) = \max_{U_R} |\langle \psi^\rho|_{RA}(U_R \otimes \mathbb{1}_A)|\psi^\sigma\rangle_{RA}|^2, \tag{6.2.23}$$

where the optimization is over unitaries on R.

Remark: Since all purifications are related to each other by isometries on the purifying system (which is the system R as in the statement of the theorem), Uhlmann's theorem tells us that the fidelity between two quantum states is equal to the maximum overlap between their purifications.

Furthermore, it is straightforward to show that it suffices to take the dimension of R the same as the dimension of A, as we have done in the statement of the theorem. In other words, performing an optimization over the dimension of R leads to the same result as in (6.2.23).

Proof: Using the definitions of $|\psi^\rho\rangle_{RA}$ and $|\psi^\sigma\rangle_{RA}$, we find for every unitary U_R that

$$|\langle \psi^\rho|_{RA}(U_R \otimes \mathbb{1}_A)|\psi^\sigma\rangle_{RA}|^2$$
$$= \left|\langle \Gamma|_{RA}(\mathbb{1}_R \otimes \sqrt{\rho_A})(U_R \otimes \mathbb{1}_A)(\mathbb{1}_R \otimes \sqrt{\sigma_A})|\Gamma\rangle_{RA}\right|^2 \tag{6.2.24}$$
$$= \left|\langle \Gamma|_{RA}(U_R \otimes \sqrt{\rho_A}\sqrt{\sigma_A})|\Gamma\rangle_{RA}\right|^2 \tag{6.2.25}$$

$$= \left|\langle\Gamma|_{RA}(\mathbb{1}_R \otimes \sqrt{\rho_A}\sqrt{\sigma_A}U_A^{\mathsf{T}})|\Gamma_{RA}\rangle\right|^2, \tag{6.2.26}$$

where the last line follows from the transpose trick in (2.2.42). Then, using (2.2.43), we find that

$$|\langle\psi^\rho|_{RA}(U_R \otimes \mathbb{1}_A)|\psi^\sigma\rangle_{RA}|^2 = \left|\mathrm{Tr}[\sqrt{\rho_A}\sqrt{\sigma_A}U_A^{\mathsf{T}}]\right|^2. \tag{6.2.27}$$

Since U_A is arbitrary, and U_A^{T} is also a unitary, we use (2.2.115) to obtain

$$\max_U |\langle\psi^\rho|_{RA}(U_R \otimes \mathbb{1}_A)|\psi^\sigma\rangle_{RA}|^2 = \max_U \left|\mathrm{Tr}[\sqrt{\rho_A}\sqrt{\sigma_A}U_A^{\mathsf{T}}]\right|^2 \tag{6.2.28}$$

$$= \max_U \left|\mathrm{Tr}[\sqrt{\rho_A}\sqrt{\sigma_A}U_A]\right|^2 \tag{6.2.29}$$

$$= \left\|\sqrt{\rho_A}\sqrt{\sigma_A}\right\|_1^2 \tag{6.2.30}$$

as required. ■

Theorem 6.9 Data-Processing Inequality for Fidelity

Let ρ and σ be states, and let $\mathcal{N}$ be a quantum channel. Then,

$$F(\rho, \sigma) \leq F(\mathcal{N}(\rho), \mathcal{N}(\sigma)). \tag{6.2.31}$$

PROOF: Recall that every quantum channel $\mathcal{N}_{A\to B}$ can be written in the Stinespring form as $\mathcal{N}_{A\to B}(\rho_A) = \mathrm{Tr}_E[V\rho_A V^\dagger]$, where $V \equiv V_{A\to BE}$ is some isometric extension of $\mathcal{N}$ and $d_E \leq \mathrm{rank}(\Gamma_{AB}^{\mathcal{N}})$. Since we have shown that the fidelity is invariant under isometric channels, it remains to show that the fidelity is non-decreasing under the action of the partial trace. To this end, consider bipartite states ρ_{AB} and σ_{AB}, and let $|\psi\rangle_{RAB}$ be an arbitrary purification of ρ_{AB} and let $|\phi\rangle_{RAB}$ be an arbitrary purification of σ_{AB}, where $d_R = d_A d_B$. Observe that $|\psi\rangle_{RAB}$ and $|\phi\rangle_{RAB}$ are also purifications of $\rho_A = \mathrm{Tr}_B[\rho_{AB}]$ and $\sigma_A = \mathrm{Tr}_B[\sigma_{AB}]$, respectively. Then, by Uhlmann's theorem, we have that

$$F(\rho_A, \sigma_A) = \max_{U_{RB}} |\langle\psi|_{RAB}(U_{RB} \otimes \mathbb{1}_A)|\phi_{RAB}\rangle|^2. \tag{6.2.32}$$

By restricting the maximization above to unitaries of the form $U_R \otimes \mathbb{1}_B$, we have that

$$F(\rho_A, \sigma_A) \geq |\langle\psi|_{RAB}(U_R \otimes \mathbb{1}_{AB})|\phi\rangle_{RAB}|^2 \tag{6.2.33}$$

for all unitaries U_R. Therefore,

$$\max_{U_R} |\langle\psi|_{RAB}(U_R \otimes \mathbb{1}_{AB})|\phi\rangle_{RAB}|^2 \leq \max_{U_{RB}} |\langle\psi|_{RAB}(U_{RB} \otimes \mathbb{1}_A)|\phi\rangle_{RAB}|^2 . \quad (6.2.34)$$

But, by Uhlmann's theorem,

$$\max_{U_R} |\langle\psi|_{RAB}(U_R \otimes \mathbb{1}_{AB})|\phi\rangle_{RAB}|^2 = F(\rho_{AB}, \sigma_{AB}). \quad (6.2.35)$$

Therefore,

$$F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A) = F(\mathrm{Tr}_B[\rho_{AB}], \mathrm{Tr}_B[\sigma_{AB}]). \quad (6.2.36)$$

The fidelity thus satisfies the data-processing inequality with respect to the partial trace.

Using the data-processing inequality for the fidelity with respect to the partial trace, along with its invariance under isometries, we conclude that

$$F(\mathcal{N}(\rho), \mathcal{N}(\sigma)) = F\left(\mathrm{Tr}_E[V\rho V^\dagger], \mathrm{Tr}_E[V\rho V^\dagger]\right) \quad (6.2.37)$$

$$\geq F(V\rho V^\dagger, V\sigma V^\dagger) \quad (6.2.38)$$

$$= F(\rho, \sigma), \quad (6.2.39)$$

as required. ■

With Uhlmann's theorem and the data-processing inequality for the fidelity in hand, we can now establish two more properties of the fidelity.

Theorem 6.10 Concavity of Fidelity

The fidelity is concave in either one of its arguments:

$$F\left(\sum_{x\in\mathcal{X}} p(x)\rho^x, \sigma\right) \geq \sum_{x\in\mathcal{X}} p(x) F(\rho^x, \sigma), \quad (6.2.40)$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0, 1]$ is a probability distribution, and σ and $\{\rho^x\}_{x\in\mathcal{X}}$ are states.

Proof: By Uhlmann's theorem, we know that the fidelity is given by the maximum overlap between the purifications of the two states under consideration. Based on this, let $|\psi^\sigma\rangle_{RA}$ be a purification of σ_A. Then, for $x \in \mathcal{X}$, let $|\phi^x\rangle_{RA}$ be a purification of ρ_A^x such that $F(\rho_A^x, \sigma) = |\langle\phi^x|\psi^\sigma\rangle|^2$. Then,

$$\sum_{x\in\mathcal{X}} p(x)F(\rho_A^x, \sigma_A) = \sum_{x\in\mathcal{X}} p(x)\,|\langle\phi^x|\psi^\sigma\rangle|^2 \tag{6.2.41}$$

$$= \langle\psi^\sigma|_{RA}\left(\sum_{x\in\mathcal{X}} p(x)|\phi^x\rangle\!\langle\phi^x|_{RA}\right)|\psi^\sigma\rangle_{RA} \tag{6.2.42}$$

$$= F\left(\sum_{x\in\mathcal{X}} p(x)|\phi^x\rangle\!\langle\phi^x|_{RA}, |\psi^\sigma\rangle\!\langle\psi^\sigma|_{RA}\right), \tag{6.2.43}$$

where the last line follows from the formula in (6.2.2) for the fidelity between a pure state and a mixed state. Then, using the data-processing inequality for the fidelity with respect to the partial trace, we obtain

$$\sum_{x\in\mathcal{X}} p(x)F(\rho_A^x, \sigma_A) = F\left(\sum_{x\in\mathcal{X}} p(x)|\phi^x\rangle\!\langle\phi^x|_{RA}, |\psi^\sigma\rangle\!\langle\psi^\sigma|_{RA}\right) \tag{6.2.44}$$

$$\leq F\left(\sum_{x\in\mathcal{X}} p(x)\mathrm{Tr}_R[|\phi^x\rangle\!\langle\phi^x|_{RA}], \mathrm{Tr}_R[|\psi^\sigma\rangle\!\langle\psi^\sigma|_{RA}]\right) \tag{6.2.45}$$

$$= F\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x, \sigma_A\right), \tag{6.2.46}$$

which is the required result. ■

A more general result than the concavity result proved above, namely *joint concavity*, can be obtained if we consider instead the square root of the fidelity, which we call the "root fidelity" and denote by

$$\sqrt{F}(\rho, \sigma) := \sqrt{F(\rho, \sigma)} = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1. \tag{6.2.47}$$

Theorem 6.11 Joint Concavity of Root Fidelity

The root fidelity is jointly concave:

$$\sqrt{F}\left(\sum_{x\in\mathcal{X}} p(x)\rho^x, \sum_{x\in\mathcal{X}} p(x)\sigma^x\right) \geq \sum_{x\in\mathcal{X}} p(x)\sqrt{F}(\rho^x, \sigma^x), \tag{6.2.48}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0, 1]$ is a probability distribution, and $\{\rho^x\}_{x\in\mathcal{X}}$ and $\{\sigma^x\}_{x\in\mathcal{X}}$ are sets of states.

Proof: This result follows by defining the classical–quantum states

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{6.2.49}$$

$$\sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \sigma_A^x, \tag{6.2.50}$$

and observing that

$$\begin{aligned}
&\sqrt{F}(\rho_{XA}, \sigma_{XA}) \\
&= \left\|\sqrt{\rho_{XA}}\sqrt{\sigma_{XA}}\right\|_1 && (6.2.51)\\
&= \left\|\left(\sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \sqrt{p(x)\rho_A^x}\right)\left(\sum_{x'\in\mathcal{X}} |x'\rangle\langle x'|_X \otimes \sqrt{p(x')\sigma_A^{x'}}\right)\right\|_1 && (6.2.52)\\
&= \left\|\sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes p(x)\sqrt{\rho_A^x}\sqrt{\sigma_A^x}\right\|_1 && (6.2.53)\\
&= \sum_{x\in\mathcal{X}} \left\|p(x)\sqrt{\rho_A^x}\sqrt{\sigma_A^x}\right\|_1 && (6.2.54)\\
&= \sum_{x\in\mathcal{X}} p(x)\left\|\sqrt{\rho_A^x}\sqrt{\sigma_A^x}\right\|_1 && (6.2.55)\\
&= \sum_{x\in\mathcal{X}} p(x)\sqrt{F}(\rho_A^x, \sigma_A^x). && (6.2.56)
\end{aligned}$$

From the above and the data-processing inequality for the fidelity under partial trace (Theorem 6.9), we conclude that

$$\sum_{x\in\mathcal{X}} p(x)\sqrt{F}(\rho_A^x, \sigma_A^x) = \sqrt{F}(\rho_{XA}, \sigma_{XA}) \le \sqrt{F}(\rho_A, \sigma_A), \tag{6.2.57}$$

which is equivalent to (6.2.48). ■

The steps in (6.2.51)–(6.2.56) demonstrate that the root fidelity satisfies the direct-sum property: for every finite alphabet $\mathcal{X}$, probability distributions $p, q : \mathcal{X} \to [0, 1]$, and sets $\{\rho_A^x\}_{x\in\mathcal{X}}$, $\{\sigma_A^x\}_{x\in\mathcal{X}}$ of states, we have

$$\sqrt{F}\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x\right) = \sum_{x\in\mathcal{X}} \sqrt{p(x)q(x)}\sqrt{F}(\rho_A^x, \sigma_A^x). \quad (6.2.58)$$

Just as the trace distance can be achieved with a measurement, so it holds that the fidelity can also be achieved with a measurement, as we now show.

Theorem 6.12 Fidelity via Measurement

For two states ρ and σ, the following equality holds

$$F(\rho, \sigma) = \min_{\{\Lambda_x\}_x} \left(\sum_{x\in\mathcal{X}} \sqrt{\mathrm{Tr}[\Lambda_x\rho]}\sqrt{\mathrm{Tr}[\Lambda_x\sigma]}\right)^2, \quad (6.2.59)$$

where the optimization is with respect to all POVMs $\{\Lambda_x\}_{x\in\mathcal{X}}$ defined with respect to a finite alphabet $\mathcal{X}$.

Proof: Let $\mathcal{M}$ be a measurement channel defined as

$$\mathcal{M}(\rho) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[\Lambda_x\rho]|x\rangle\langle x|, \quad (6.2.60)$$

where $\{\Lambda_x\}_{x\in\mathcal{X}}$ is a POVM and $\mathcal{X}$ is a finite alphabet. Then, by the data-processing inequality for the fidelity with respect to the channel $\mathcal{M}$, and since the action of $\mathcal{M}$ leads to a state that is diagonal in the orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$, we obtain

$$F(\rho, \sigma) \leq F(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \quad (6.2.61)$$

$$= \left\|\sqrt{\mathcal{M}(\rho)}\sqrt{\mathcal{M}(\sigma)}\right\|_1^2 \quad (6.2.62)$$

$$= \left\|\sqrt{\sum_{x\in\mathcal{X}} \mathrm{Tr}[\Lambda_x\rho]|x\rangle\langle x|}\sqrt{\sum_{x\in\mathcal{X}} \mathrm{Tr}[\Lambda_x\sigma]|x\rangle\langle x|}\right\|_1^2 \quad (6.2.63)$$

$$= \left\|\left(\sum_{x\in\mathcal{X}} \sqrt{\mathrm{Tr}[\Lambda_x\rho]}|x\rangle\langle x|\right)\left(\sum_{x\in\mathcal{X}} \sqrt{\mathrm{Tr}[\Lambda_x\sigma]}|x\rangle\langle x|\right)\right\|_1^2 \quad (6.2.64)$$

$$= \left\|\sum_{x\in\mathcal{X}} \sqrt{\mathrm{Tr}[\Lambda_x\rho]}\sqrt{\mathrm{Tr}[\Lambda_x\sigma]}|x\rangle\langle x|\right\|_1^2 \quad (6.2.65)$$

$$= \left(\sum_{x \in \mathcal{X}} \sqrt{\text{Tr}[\Lambda_x \rho]} \sqrt{\text{Tr}[\Lambda_x \sigma]} \right)^2 . \tag{6.2.66}$$

Since the POVM $\{\Lambda_x\}_{x \in \mathcal{X}}$ is arbitrary, we find that

$$F(\rho, \sigma) \leq \min_{\{\Lambda_x\}_x} \left(\sum_{x \in \mathcal{X}} \sqrt{\text{Tr}[\Lambda_x \rho]} \sqrt{\text{Tr}[\Lambda_x \sigma]} \right)^2 . \tag{6.2.67}$$

We now prove the reverse inequality by explicitly constructing a POVM that achieves the fidelity. First, observe that we can write the fidelity $F(\rho, \sigma)$ as

$$F(\rho, \sigma) = \left\| \sqrt{\rho} \sqrt{\sigma} \right\|_1^2 = \text{Tr}\left[\sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}} \right]^2 = \text{Tr}[A\sigma]^2, \tag{6.2.68}$$

where

$$A := \sigma^{-\frac{1}{2}} \left(\sigma^{\frac{1}{2}} \rho \sigma^{\frac{1}{2}} \right)^{\frac{1}{2}} \sigma^{-\frac{1}{2}}. \tag{6.2.69}$$

If σ is not invertible, then the inverse is understood to be on the support of σ, in which case A is supported on the support of σ. So the fidelity is simply equal to the squared expectation value of the Hermitian operator A with respect to the state σ. Observe that we can also write

$$F(\rho, \sigma) = \left(\text{Tr}\left[\sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}} \right] \right)^2 \tag{6.2.70}$$

$$= \left(\text{Tr}\left[\sqrt{\sqrt{\sigma} \Pi_\sigma \rho \Pi_\sigma \sqrt{\sigma}} \right] \right)^2 \tag{6.2.71}$$

$$= F(\Pi_\sigma \rho \Pi_\sigma, \sigma), \tag{6.2.72}$$

where Π_σ is the projection onto the support of σ. This holds because $\sqrt{\sigma} = \Pi_\sigma \sqrt{\sigma} = \sqrt{\sigma} \Pi_\sigma$.

Now, let us consider a measurement in the eigenbasis of A. Let $\{|\psi_i\rangle\}_{i=0}^{r-1}$ be the eigenvectors of A, where $r = \text{rank}(A)$. If σ is not invertible, then we can always add a set $\{|\psi_i\rangle\}_{i=r}^{d-1}$ of linearly independent pure states orthogonal to the eigenbasis of A in order to obtain a POVM on the full d-dimensional space. Therefore, suppose that

$$A = \sum_{i=0}^{d-1} \lambda_i |\psi_i\rangle\langle\psi_i|, \tag{6.2.73}$$

where we have included the vectors $\{|\psi_i\rangle\}_{i=r}^{d-1}$ that have corresponding eigenvalues equal to zero. Then,

$$\operatorname{Tr}[A\sigma] = \operatorname{Tr}\left[\sum_{i=0}^{d-1} \lambda_i |\psi_i\rangle\langle\psi_i|\sigma\right] \tag{6.2.74}$$

$$= \sum_{i=0}^{d-1} \lambda_i \langle\psi_i|\sigma|\psi_i\rangle \tag{6.2.75}$$

$$= \sum_{i=0}^{d-1} \sqrt{\langle\psi_i|\lambda_i\sigma\lambda_i|\psi_i\rangle}\sqrt{\langle\psi_i|\sigma|\psi_i\rangle} \tag{6.2.76}$$

$$= \sum_{i=0}^{r-1} \sqrt{\langle\psi_i|A\sigma A|\psi_i\rangle}\sqrt{\langle\psi_i|\sigma|\psi_i\rangle}, \tag{6.2.77}$$

where the last line follows because $A|\psi_i\rangle = \lambda_i|\psi_i\rangle$ for all $0 \leq i \leq r-1$, and we have used the fact that $\lambda_i = 0$ for all $r \leq i \leq d-1$. Now, it is straightforward to show that $A\sigma A = \Pi_\sigma\rho\Pi_\sigma$. Therefore,

$$F(\rho,\sigma) = \operatorname{Tr}[A\sigma]^2 = \left(\sum_{i=0}^{r-1} \sqrt{\langle\psi_i|\Pi_\sigma\rho\Pi_\sigma|\psi_i\rangle}\sqrt{\langle\psi_i|\sigma|\psi_i\rangle}\right)^2 \tag{6.2.78}$$

$$= \left(\sum_{i=0}^{r-1} \sqrt{\langle\psi_i|\rho|\psi_i\rangle}\sqrt{\langle\psi_i|\sigma|\psi_i\rangle}\right)^2, \tag{6.2.79}$$

where the last line follows because A is defined on the support of σ, which means that $\Pi_\sigma|\psi_i\rangle = |\psi_i\rangle$ for all $0 \leq i \leq r-1$. We thus have

$$\min_{\{\Lambda_x\}_x} \left(\sum_{x\in\mathcal{X}} \sqrt{\operatorname{Tr}[\Lambda_x\rho]}\sqrt{\operatorname{Tr}[\Lambda_x\sigma]}\right)^2 \leq \left(\sum_{i=0}^{r-1} \sqrt{\langle\psi_i|\rho|\psi_i\rangle}\sqrt{\langle\psi_i|\sigma|\psi_i\rangle}\right)^2 \tag{6.2.80}$$

$$= F(\rho,\sigma), \tag{6.2.81}$$

which is precisely the reverse inequality, as desired. ■

By employing Theorem 6.12, we conclude the following stronger data-processing inequality for the fidelity, which strengthens the statement of Theorem 6.9 considerably:

Proposition 6.13 Improved Data Processing for Fidelity

Let ρ and σ be quantum states, and let $\mathcal{N}$ be a positive, trace-preserving map. Then the following inequality holds:

$$F(\rho,\sigma) \leq F(\mathcal{N}(\rho),\mathcal{N}(\sigma)). \tag{6.2.82}$$

PROOF: The reasoning here follows the reasoning of the proof of Theorem 6.3 closely. Let $\{\Lambda'_x\}_{x\in\mathcal{X}}$ be a POVM. Then consider that

$$\begin{aligned}\sum_{x\in\mathcal{X}}\sqrt{\operatorname{Tr}[\Lambda'_x\mathcal{N}(\rho)]\operatorname{Tr}[\Lambda'_x\mathcal{N}(\sigma)]} &= \sum_{x\in\mathcal{X}}\sqrt{\operatorname{Tr}[\mathcal{N}^\dagger(\Lambda'_x)\rho]\operatorname{Tr}[\mathcal{N}^\dagger(\Lambda'_x)\sigma]} &(6.2.83)\\ &\geq \min_{\{\Lambda_x\}_{x\in\mathcal{X}}}\sum_{x\in\mathcal{X}}\sqrt{\operatorname{Tr}[\Lambda_x\rho]\operatorname{Tr}[\Lambda_x\sigma]} &(6.2.84)\\ &= \sqrt{F}(\rho,\sigma). &(6.2.85)\end{aligned}$$

The inequality follows because $\{\mathcal{N}^\dagger(\Lambda'_x)\}_{x\in\mathcal{X}}$ is a POVM since $\{\Lambda'_x\}_{x\in\mathcal{X}}$ is and $\mathcal{N}$ is a positive, trace-preserving map, so that $\mathcal{N}^\dagger(\Lambda'_x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x\in\mathcal{X}}\mathcal{N}^\dagger(\Lambda'_x) = \mathcal{N}^\dagger\left(\sum_{x\in\mathcal{X}}\Lambda'_x\right) = \mathcal{N}^\dagger(\mathbb{1}) = \mathbb{1}$. The last equality follows from Theorem 6.12. Since the inequality holds for all POVMs $\{\Lambda'_x\}_{x\in\mathcal{X}}$, we conclude that

$$\begin{aligned}\sqrt{F}(\mathcal{N}(\rho),\mathcal{N}(\sigma)) &= \min_{\{\Lambda'_x\}_{x\in\mathcal{X}}}\sum_{x\in\mathcal{X}}\sqrt{\operatorname{Tr}[\Lambda'_x\mathcal{N}(\rho)]\operatorname{Tr}[\Lambda'_x\mathcal{N}(\sigma)]} &(6.2.86)\\ &\geq \sqrt{F}(\rho,\sigma), &(6.2.87)\end{aligned}$$

where we have again employed Theorem 6.12 for the equality. ■

We now establish a useful relation between trace distance and fidelity.

Theorem 6.14 Relation Between Trace Distance and Fidelity

For two states ρ and σ, the following chain of inequalities relates their trace distance with the fidelity between them:

$$1-\sqrt{F(\rho,\sigma)} \leq \frac{1}{2}\|\rho-\sigma\|_1 \leq \sqrt{1-F(\rho,\sigma)}. \tag{6.2.88}$$

Proof: We first prove the upper bound. To do so, recall the formula in (6.1.1) for the trace distance between two pure states. If we let $|\psi^\rho\rangle_{RA}$ and $|\psi^\sigma\rangle_{RA}$ be purifications of ρ_A and σ_A, respectively, such that $F(\rho_A, \sigma_A) = |\langle\psi^\rho|\psi^\sigma\rangle|^2$, and if we use the data-processing inequality for the trace distance with respect to the partial trace channel Tr_R, then we obtain

$$\frac{1}{2}\|\rho_A - \sigma_A\|_1 = \frac{1}{2}\|\mathrm{Tr}_R[|\psi^\rho\rangle\langle\psi^\rho|_{RA} - |\psi^\sigma\rangle\langle\psi^\sigma|_{RA}]\|_1 \tag{6.2.89}$$

$$\leq \frac{1}{2}\||\psi^\rho\rangle\langle\psi^\rho|_{RA} - |\psi^\sigma\rangle\langle\psi^\sigma|_{RA}\|_1 \tag{6.2.90}$$

$$= \sqrt{1 - |\langle\psi^\rho|\psi^\sigma\rangle|^2} \tag{6.2.91}$$

$$= \sqrt{1 - F(\rho_A, \sigma_A)}, \tag{6.2.92}$$

as required.

For the lower bound, we use the results of Theorems 6.12 and 6.4. Theorem 6.12 tells us that there exists a POVM $\{\Lambda_x\}_{x\in\mathcal{X}}$ such that

$$F(\rho, \sigma) = \left(\sum_{x\in\mathcal{X}} \sqrt{\mathrm{Tr}[\Lambda_x\rho]}\sqrt{\mathrm{Tr}[\Lambda_x\sigma]}\right)^2 \tag{6.2.93}$$

$$\equiv \left(\sum_{x\in\mathcal{X}} \sqrt{p(x)q(x)}\right)^2, \tag{6.2.94}$$

where we have let $p(x) \coloneqq \mathrm{Tr}[\Lambda_x\rho]$ and $q(x) \coloneqq \mathrm{Tr}[\Lambda_x\sigma]$. Using this, observe that

$$\sum_{x\in\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 = \sum_{x\in\mathcal{X}} \left(p(x) - 2\sqrt{p(x)q(x)} + q(x)\right) \tag{6.2.95}$$

$$= 2 - 2\sum_{x\in\mathcal{X}} \sqrt{p(x)q(x)} \tag{6.2.96}$$

$$= 2 - 2\sqrt{F(\rho, \sigma)}. \tag{6.2.97}$$

Now, Theorem 6.4 tells us that

$$\|\rho - \sigma\|_1 = \max_{\{\Omega_y\}_y} \sum_{x\in\mathcal{X}} |r(y) - s(y)|, \tag{6.2.98}$$

where $r(y) \coloneqq \mathrm{Tr}[\Omega_y\rho]$ and $s(y) \coloneqq \mathrm{Tr}[\Omega_y\sigma]$. In particular, for the POVM $\{\Lambda_x\}_{x\in\mathcal{X}}$ that achieves the fidelity, we have

$$\sum_{x\in\mathcal{X}} |p(x) - q(x)| \leq \|\rho - \sigma\|_1\,. \tag{6.2.99}$$

Using this, and the fact that $|\sqrt{p(x)} - \sqrt{q(x)}| \leq |\sqrt{p(x)} + \sqrt{q(x)}|$, we obtain

$$\sum_{x\in\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \leq \sum_{x\in\mathcal{X}} \left|\sqrt{p(x)} - \sqrt{q(x)}\right| \left|\sqrt{p(x)} + \sqrt{q(x)}\right| \tag{6.2.100}$$

$$= \sum_{x\in\mathcal{X}} |p(x) - q(x)| \tag{6.2.101}$$

$$\leq \|\rho - \sigma\|_1 . \tag{6.2.102}$$

So we have

$$2 - 2\sqrt{F(\rho,\sigma)} = \sum_{x\in\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \leq \|\rho - \sigma\|_1 , \tag{6.2.103}$$

which is the required lower bound. ■

Lemma 6.15 Gentle Measurement

Let ρ be a density operator and Λ a measurement operator, satisfying $0 \leq \Lambda \leq I$ and $\text{Tr}[\Lambda\rho] \geq 1 - \varepsilon$, for $\varepsilon \in [0, 1]$. Then the post-measurement state

$$\rho' := \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\text{Tr}[\Lambda\rho]} \tag{6.2.104}$$

satisfies

$$F(\rho,\rho') \geq 1 - \varepsilon, \tag{6.2.105}$$

$$\frac{1}{2} \|\rho - \rho'\|_1 \leq \sqrt{\varepsilon}. \tag{6.2.106}$$

PROOF: Suppose first that ρ is a pure state $|\psi\rangle\langle\psi|$. The post-measurement state is then

$$\frac{\sqrt{\Lambda}|\psi\rangle\langle\psi|\sqrt{\Lambda}}{\langle\psi|\Lambda|\psi\rangle}. \tag{6.2.107}$$

The fidelity between the original state $|\psi\rangle$ and the post-measurement state above is as follows:

$$\langle\psi| \left(\frac{\sqrt{\Lambda}|\psi\rangle\langle\psi|\sqrt{\Lambda}}{\langle\psi|\Lambda|\psi\rangle}\right) |\psi\rangle = \frac{\left|\langle\psi|\sqrt{\Lambda}|\psi\rangle\right|^2}{\langle\psi|\Lambda|\psi\rangle} \geq \frac{|\langle\psi|\Lambda|\psi\rangle|^2}{\langle\psi|\Lambda|\psi\rangle} \tag{6.2.108}$$

$$= \langle\psi|\Lambda|\psi\rangle \geq 1 - \varepsilon. \tag{6.2.109}$$

The first inequality follows because $\sqrt{\Lambda} \geq \Lambda$ when $\Lambda \leq I$. The second inequality follows from the hypothesis of the lemma. Now let us consider when we have mixed states ρ_A and ρ'_A. Suppose $|\psi\rangle_{RA}$ and $|\psi'\rangle_{RA}$ are respective purifications of ρ_A and ρ'_A, where

$$|\psi'\rangle_{RA} \equiv \frac{I_R \otimes \sqrt{\Lambda_A}|\psi\rangle_{RA}}{\sqrt{\langle\psi|I_R \otimes \Lambda_A|\psi\rangle_{RA}}}. \tag{6.2.110}$$

Then we can apply the data-processing inequality for fidelity (Proposition 6.13) and the result above for pure states to conclude that

$$F(\rho_A, \rho'_A) \geq F(\psi_{RA}, \psi'_{RA}) \geq 1 - \varepsilon. \tag{6.2.111}$$

We obtain the bound on the normalized trace distance $\frac{1}{2}\left\|\rho_A - \rho'_A\right\|_1$ by exploiting Theorem 6.14. ■

6.2.1 Sine Distance

Unlike the trace distance, the fidelity is not a distance measure in the mathematical sense because it does not satisfy the triangle inequality. The following distance measure based on the fidelity, however, does satisfy the triangle inequality, along with the other properties that define a distance measure.

Definition 6.16 Sine Distance

For two states ρ and σ, we define the *sine distance* between ρ and σ as

$$P(\rho, \sigma) := \sqrt{1 - F(\rho, \sigma)}. \tag{6.2.112}$$

The measure $P(\rho, \sigma)$ is known as the sine distance due to the fact that $F(\rho, \sigma)$ has the interpretation as the largest value of the squared cosine of the angle between two arbitrary purifications of ρ and σ (see Theorem 6.8), which means that $\sqrt{1 - F(\rho, \sigma)}$ has the interpretation as the sine of the same angle. Related to this interpretation, the measure $P(\rho, \sigma)$ is equal to the minimum trace distance between purifications of ρ and σ:

$$\inf_{|\psi^\rho\rangle_{RA},|\psi^\sigma\rangle_{RA}} \frac{1}{2} \left\| |\psi^\rho\rangle\!\langle\psi^\rho|_{RA} - |\psi^\sigma\rangle\!\langle\psi^\sigma|_{RA} \right\|_1 \\ = \inf_{|\psi^\rho\rangle_{RA},|\psi^\sigma\rangle_{RA}} \sqrt{1 - |\langle\psi^\rho|\psi^\sigma\rangle_{RA}|^2} = P(\rho,\sigma), \quad (6.2.113)$$

where the optimization is over all purifications $|\psi^\rho\rangle_{RA}$ and $|\psi^\sigma\rangle_{RA}$ of ρ and σ, respectively. This follows by applying (6.1.1), as well as Uhlmann's theorem (Theorem 6.8).

Since the fidelity satisfies the data-processing inequality with respect to positive, trace-preserving maps (Proposition 6.13), so does the sine distance: for two states ρ and σ and a positive, trace-preserving map $\mathcal{N}$, we have that

$$P(\rho,\sigma) \geq P(\mathcal{N}(\rho),\mathcal{N}(\sigma)). \quad (6.2.114)$$

Lemma 6.17 Triangle Inequality for Sine Distance

Let ρ, σ, and ω be quantum states. Then the triangle inequality holds for the sine distance:

$$P(\rho,\sigma) \leq P(\rho,\omega) + P(\omega,\sigma). \quad (6.2.115)$$

Proof: Define the canonical purifications as

$$|\psi^\rho\rangle_{RA} = (\mathbb{1}_R \otimes \sqrt{\rho_A})|\Gamma\rangle_{RA}, \quad (6.2.116)$$
$$|\psi^\sigma\rangle_{RA} = (\mathbb{1}_R \otimes \sqrt{\sigma_A})|\Gamma\rangle_{RA}, \quad (6.2.117)$$
$$|\psi^\omega\rangle_{RA} = (\mathbb{1}_R \otimes \sqrt{\omega_A})|\Gamma\rangle_{RA}, \quad (6.2.118)$$

where $|\Gamma\rangle_{RA}$ is the maximally entangled vector from (2.2.36). Recalling (6.1.1), for pure states $|\phi\rangle$ and $|\varphi\rangle$, we have that

$$\frac{1}{2} \left\| |\phi\rangle\!\langle\phi| - |\varphi\rangle\!\langle\varphi| \right\|_1 = \sqrt{1 - F(\phi,\varphi)} = \sqrt{1 - |\langle\phi|\varphi\rangle|^2}. \quad (6.2.119)$$

Let U_R and V_R be arbitrary unitaries acting on the reference system R. From the fact that trace distance obeys the triangle inequality and the equality given above, we find that

$$\sqrt{1 - |\langle\psi^\sigma|_{RA}(W_R \otimes \mathbb{1}_A)|\psi^\rho\rangle_{RA}|^2} \leq \sqrt{1 - |\langle\psi^\sigma|_{RA}(U_R^\dagger \otimes \mathbb{1}_A)|\psi^\omega\rangle_{RA}|^2} \\ + \sqrt{1 - |\langle\psi^\omega|_{RA}(V_R \otimes \mathbb{1}_A)|\psi^\rho\rangle_{RA}|^2}, \quad (6.2.120)$$

where $W_R := U_R^\dagger V_R$. By minimizing the left-hand side with respect to all unitaries W_R, and applying Uhlmann's theorem, we find that

$$\sqrt{1 - F(\sigma_A, \rho_A)} \leq \sqrt{1 - |\langle\psi^\sigma|_{RA}(U_R^\dagger \otimes \mathbb{1}_A)|\psi^\omega\rangle_{RA}|^2} + \sqrt{1 - |\langle\psi^\omega|_{RA}(V_R \otimes \mathbb{1}_A)|\psi^\rho\rangle_{RA}|^2}. \quad (6.2.121)$$

Since the inequality holds for arbitrary unitaries U and V, it holds for the minimum of each term on the right, and so this, combined with Uhlmann's theorem (Theorem 6.8), implies the desired result:

$$\sqrt{1 - F(\sigma_A, \rho_A)} \leq \sqrt{1 - F(\sigma_A, \omega_A)} + \sqrt{1 - F(\omega_A, \rho_A)}, \quad (6.2.122)$$

concluding the proof. ■

6.3 Diamond Distance

Just as there are measures of distinguishability for quantum states, it is important to develop measures of distinguishability for quantum channels, in order to assess the performance of quantum information-processing protocols that attempt to simulate an ideal process. The measures that we introduce in this section are again motivated by operational concerns, stemming from the ability of an experimenter to distinguish one quantum channel from another when given access to a single use of the channel. In what follows, our discussion mirrors and generalizes the discussion in Section 6.1 that motivated trace distance as a measure of distinguishability for quantum states.

How should we measure the distance between two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$? Related, how should we assess the performance of a quantum information-processing protocol in which the ideal channel to be simulated is $\mathcal{N}_{A\to B}$ but the channel realized in practice is $\mathcal{M}_{A\to B}$? Suppose that a third party is trying to assess how distinguishable the actual channel $\mathcal{M}_{A\to B}$ is from the ideal channel $\mathcal{N}_{A\to B}$. Such an individual has access to both the input and output ports of the channel, and so the most general strategy for the distinguisher to employ is to prepare a state ρ_{RA} of a reference system R and the channel input system A. The distinguisher transmits the A system of ρ_{RA} into the unknown channel. After that, the distinguisher receives the channel output system B and then performs a measurement described by the POVM $\{\Lambda^x_{RB}\}_x$ on the reference system R and the channel output system B. The probability of obtaining a particular outcome Λ^x_{RB} is given by the Born rule. In the case that the unknown channel is $\mathcal{N}_{A\to B}$, this probability is $\mathrm{Tr}[\Lambda^x_{RB}\mathcal{N}_{A\to B}(\rho_{RA})]$, and in the case that the unknown channel is $\mathcal{M}_{A\to B}$, this probability is $\mathrm{Tr}[\Lambda^x_{RB}\mathcal{M}_{A\to B}(\rho_{RA})]$. What we demand is that the absolute deviation between the two probabilities $\mathrm{Tr}[\Lambda^x_{RB}\mathcal{N}_{A\to B}(\rho_{RA})]$ and $\mathrm{Tr}[\Lambda^x_{RB}\mathcal{M}_{A\to B}(\rho_{RA})]$ is no larger than some tolerance ε. Since this should be the case for all possible input states and measurement outcomes, what we demand mathematically is that

$$\sup_{\rho_{RA},\, 0\leq\Lambda_{RB}\leq\mathbb{1}_{RB}} |\mathrm{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\rho_{RA})] - \mathrm{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\rho_{RA})]| \leq \varepsilon. \tag{6.3.1}$$

As a consequence of the characterization of trace distance from Theorem 6.1 we have

$$\sup_{\rho_{RA},0\leq\Lambda_{RB}\leq\mathbb{1}_{RB}} |\mathrm{Tr}[\Lambda_{RB}(\mathcal{N}_{A\to B} - \mathcal{M}_{A\to B})(\rho_{RA})]|$$
$$= \sup_{\rho_{RA}} \frac{1}{2}\left\|\mathcal{N}_{A\to B}(\rho_{RA}) - \mathcal{M}_{A\to B}(\rho_{RA})\right\|_1 =: \frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_\diamond, \tag{6.3.2}$$

where $\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_\diamond$ is defined to be the *normalized diamond distance* between $\mathcal{N}$ and $\mathcal{M}$. This indicates that if $\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_\diamond \leq \varepsilon$, then the absolute deviation between probabilities for every possible input state and measurement operator never exceeds ε, so that the approximation between quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ is naturally quantified by the normalized diamond distance $\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_\diamond$.

With the above in mind, we now define the diamond norm for Hermiticity-preserving maps, from which the diamond distance measure for channels arises.

Definition 6.18 Diamond Norm

The *diamond norm* of a Hermiticity-preserving map $\mathcal{P}_{A\to B}$ is defined as

$$\left\|\mathcal{P}\right\|_\diamond := \sup_{\rho_{RA}} \left\|\mathcal{P}_{A\to B}(\rho_{RA})\right\|_1, \tag{6.3.3}$$

where the optimization is over all states ρ_{RA}, with the dimension of R unbounded.

For two quantum channels $\mathcal{N}$ and $\mathcal{M}$, note that the difference $\mathcal{N}-\mathcal{M}$ is a Hermiticity-preserving map. An important simplification of the diamond norm of a Hermiticity-preserving map is given by the following proposition:

Proposition 6.19

The diamond norm of a Hermiticity-preserving map $\mathcal{P}_{A\to B}$ can be calculated as

$$\|\mathcal{P}\|_\diamond = \sup_{\psi_{RA}} \|\mathcal{P}_{A\to B}(\psi_{RA})\|_1 , \tag{6.3.4}$$

where the optimization is over all pure states ψ_{RA}, such that the dimension of R is equal to the dimension of the system A.

PROOF: Let ρ_{RA} be an arbitrary state. It has a spectral decomposition as follows:

$$\rho_{RA} = \sum_x p(x)\psi^x_{RA}, \tag{6.3.5}$$

where $\{p(x)\}_x$ is a probability distribution and $\{\psi^x_{RA}\}_x$ is a set of pure states. From the convexity of the trace norm (see Section 2.2.9), it follows that

$$\|\mathcal{P}_{A\to B}(\rho_{RA})\|_1 \leq \sum_x p(x) \left\|\mathcal{P}_{A\to B}(\psi^x_{RA})\right\|_1 \tag{6.3.6}$$

$$\leq \sup_x \left\|\mathcal{P}_{A\to B}(\psi^x_{RA})\right\|_1 \tag{6.3.7}$$

$$\leq \sup_{\psi_{RA}} \|\mathcal{P}_{A\to B}(\psi_{RA})\|_1 \tag{6.3.8}$$

From the Schmidt decomposition (Theorem 2.2), it follows that the rank of the reduced state ψ_R is no larger than the dimension of system A. So then it suffices to optimize with respect to all pure states ψ_{RA}, such that the dimension of R is equal to the dimension of the system A. ■

The normalized diamond distance between two quantum channels can be computed via a semi-definite program (SDP). The following proposition states this fact formally and also states the dual optimization problem. Appendix 6.A provides a proof.

Proposition 6.20 SDPs for Normalized Diamond Distance

The diamond distance between two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ can be written as the following semi-definite programs:

$$\frac{1}{2}\left\|\mathcal{N}-\mathcal{M}\right\|_{\diamond}$$
$$= \sup_{\substack{\rho_R\geq 0,\\ \Omega_{RB}\geq 0}} \{\mathrm{Tr}[\Omega_{RB}(\Gamma^{\mathcal{N}}_{RB}-\Gamma^{\mathcal{M}}_{RB})] : \Omega_{RB} \leq \rho_R \otimes \mathbb{1}_B, \mathrm{Tr}[\rho_R] = 1\} \tag{6.3.9}$$
$$= \inf_{\substack{\mu\geq 0,\\ Z_{RB}\geq 0}} \{\mu : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB}-\Gamma^{\mathcal{M}}_{RB}, \mu\mathbb{1}_R \geq \mathrm{Tr}_B[Z_{RB}]\}. \tag{6.3.10}$$

The latter expression is equal to

$$\inf_{Z_{RB}\geq 0}\{\left\|\mathrm{Tr}_B[Z_{RB}]\right\|_\infty : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB}-\Gamma^{\mathcal{M}}_{RB}\}. \tag{6.3.11}$$

As described at the beginning of this section, the diamond distance has an operational meaning in terms of the task of *channel discrimination*, which is a generalization of state discrimination (see Section 5.3.1). Let us now analyze the task of channel discrimination in more detail.

Suppose that Alice gives Bob a device that implements either the channel $\mathcal{N}$ or the channel $\mathcal{M}$, but she does not tell him which channel it implements. Bob's task is to decide which channel the device implements. Suppose that the channels $\mathcal{N}$ and $\mathcal{M}$ have prior probabilities λ and $1-\lambda$ of being selected, respectively. The only way for Bob to determine which channel the device implements (without guessing randomly) is to pass a quantum system, say in the state ρ, through it. He can then perform a measurement on the resulting output state and make a guess as to which channel was implemented. Therefore, in addition to having the freedom to choose any binary measurement (which is the case in state discrimination), in channel discrimination Bob also has the freedom to prepare a system A and a reference system R in any state ρ_{RA} of his choosing, with the system A being passed through the device.

For every fixed input state ρ_{RA}, there are two possible output states, depending on which channel was implemented. This means that, for every fixed input state, the task of channel discrimination reduces to the task of state discrimination. Using the result of Theorem 5.3, for the input state ρ_{RA} the corresponding optimal error probability (i.e., the error probability obtained by optimizing over all measurements) is

$$p^*_{\text{err}}(\rho_{RA}) = \frac{1}{2}\left(1 - \|\lambda\mathcal{N}_{A\to B}(\rho_{RA}) - (1-\lambda)\mathcal{M}_{A\to B}(\rho_{RA})\|_1\right). \tag{6.3.12}$$

Then, optimizing over all input states ρ_{RA} in order to minimize the error probability, we find that

$$\inf_{\rho_{RA}} p^*_{\text{err}}(\rho_{RA}) = \frac{1}{2}\left(1 - \sup_{\rho_{RA}} \|(\lambda\mathcal{N}_{A\to B} - (1-\lambda)\mathcal{M}_{A\to B})(\rho_{RA})\|_1\right) \tag{6.3.13}$$

$$= \frac{1}{2}\left(1 - \|\lambda\mathcal{N} - (1-\lambda)\mathcal{M}\|_\diamond\right), \tag{6.3.14}$$

where the last line follows from the definition of the diamond norm. The optimal error probability for the task of channel discrimination is thus a simple function of the diamond norm.

6.4 Fidelity Measures for Channels

The diamond distance is a distance measure for channels that is based on the trace distance for states. We now define a fidelity-based quantity for channels that can be used to assess its ability to preserve entanglement.

Definition 6.21 Entanglement Fidelity of a Channel

For a quantum channel $\mathcal{N}_A$ with input and output systems of equal dimension d, we define its *entanglement fidelity* as

$$F_e(\mathcal{N}) := \langle\Phi|_{RA}(\text{id}_R \otimes \mathcal{N}_A)(\Phi_{RA})|\Phi\rangle_{RA}, \tag{6.4.1}$$

where

$$|\Phi\rangle_{RA} = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1} |i,i\rangle_{RA}. \tag{6.4.2}$$

Notice that the entanglement fidelity of a channel is the fidelity of the maximally entangled state with the Choi state of the channel. Intuitively, then, the entanglement fidelity quantifies how good a channel is at preserving the entanglement between two systems when it acts on one of the two systems.

It turns out that the entanglement fidelity is very closely related to another fidelity-based measure on quantum channels called the *average fidelity*:

$$\overline{F}(\mathcal{N}) := \int_{\psi} \langle\psi|\mathcal{N}(\psi)|\psi\rangle \, \mathrm{d}\psi, \tag{6.4.3}$$

where we integrate over all pure states acting on the input Hilbert space of $\mathcal{N}$ with respect to the Haar measure. The Haar measure is the uniform probability measure on pure quantum states (see the remark after (2.5.18)). For a quantum channel $\mathcal{N}$ with input system dimension d, the following identity holds

$$\overline{F}(\mathcal{N}) = \frac{dF_e(\mathcal{N}) + 1}{d + 1}. \tag{6.4.4}$$

Instead of taking the average as in (6.4.3), we can take the minimum over all input states to obtain the *minimum fidelity*:

$$F_{\min}(\mathcal{N}) := \inf_{\psi} \langle\psi|\mathcal{N}(\psi)|\psi\rangle, \tag{6.4.5}$$

where the optimization is over all pure states ψ acting on the input Hilbert space of the channel $\mathcal{N}$. By introducing a reference system R and optimizing over all joint states $|\psi\rangle_{RA}$ of R and the input system A of the channel $\mathcal{N}$, we obtain a fidelity measure that generalizes the entanglement fidelity.

Definition 6.22 Fidelity of a Quantum Channel

For a quantum channel $\mathcal{N}_A$ with equal input and output system dimension, we define the *fidelity of* $\mathcal{N}$ as

$$F(\mathcal{N}) := \inf_{\psi_{RA}} \langle\psi|_{RA}(\mathrm{id}_R \otimes \mathcal{N}_A)(\psi_{RA})|\psi\rangle_{RA}, \tag{6.4.6}$$

where we take the infimum over all pure states $|\psi\rangle_{RA}$, with the dimension of R equal to the dimension of A.

Note that the state $|\psi\rangle_{RA} = |\Phi\rangle_{RA}$ is a special case in the optimization in (6.4.6). This implies that, for a channel $\mathcal{N}$, $F(\mathcal{N}) \leq F_e(\mathcal{N})$.

More generally, we define the fidelity between two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ as follows:

Definition 6.23 Fidelity of Quantum Channels

Let $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ be quantum channels. Their channel fidelity is defined as

$$F(\mathcal{N},\mathcal{M}) = \inf_{\rho_{RA}} F(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})), \tag{6.4.7}$$

where the infimum is taken over all bipartite states ρ_{RA}, with the dimension of R arbitrarily large.

Remark: Similar to the diamond distance, we define the channel fidelity as above in order to indicate its operational meaning with an infimum over all possible input states, but it is not necessary to take the infimum over all bipartite states. One can instead restrict the infimum to be over pure bipartite states where the reference system R is isomorphic to the channel input system A, so that

$$F(\mathcal{N},\mathcal{M}) = \inf_{\psi_{RA}} F(\mathcal{N}_{A\to B}(\psi_{RA}),\mathcal{M}_{A\to B}(\psi_{RA})), \tag{6.4.8}$$

where ψ_{RA} is a pure bipartite state with system R is isomorphic to system A. The same statement thus applies to (6.4.6). An argument for this is similar to that given in the proof of Proposition 6.19, except using the joint concavity of root fidelity rather than convexity of the trace norm.

Here, we provide a different argument for this fact. First, we have that

$$\inf_{\rho_{RA}} F(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})) \le \inf_{\psi_{RA}} F(\mathcal{N}_{A\to B}(\psi_{RA}),\mathcal{M}_{A\to B}(\psi_{RA})) \tag{6.4.9}$$

which holds simply by restricting the optimization on the left-hand side to pure states.

Next, given a state ρ_{RA}, with the dimension of R not necessarily equal to the dimension of A, we can purify it to a state $\psi_{R'RA}$. Then, using the data-processing inequality for the fidelity with respect to the partial trace channel $\text{Tr}_{R'}$ (Proposition 6.13), we find that

$$\begin{aligned} F(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})) &= F\left(\mathcal{N}_{A\to B}(\text{Tr}_{R'}[\psi_{R'RA}]),\mathcal{M}_{A\to B}(\text{Tr}_{R'}[\psi_{R'RA}])\right) &(6.4.10)\\ &= F\left(\text{Tr}_{R'}[\mathcal{N}_{A\to B}(\psi_{R'RA})],\text{Tr}_{R'}[\mathcal{M}_{A\to B}(\psi_{R'RA})]\right) &(6.4.11)\\ &\ge F(\mathcal{N}_{A\to B}(\psi_{R'RA}),\mathcal{M}_{A\to B}(\psi_{R'RA})) &(6.4.12)\\ &\ge \inf_{\psi_{R'RA}} F(\mathcal{N}_{A\to B}(\psi_{R'RA}),\mathcal{M}_{A\to B}(\psi_{R'RA})). &(6.4.13)\end{aligned}$$

Since the state ρ_{RA} is arbitrary, we obtain

$$\inf_{\rho_{RA}} F(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})) \ge \inf_{\psi_{RA}} F(\mathcal{N}_{A\to B}(\psi_{RA}),\mathcal{M}_{A\to B}(\psi_{RA})). \tag{6.4.14}$$

Finally, by the Schmidt decomposition theorem (Theorem 2.2), for every pure state ψ_{RA}, the rank of the reduced state ψ_R need not exceed the dimension of A, implying that it suffices to optimize over pure states for which the system R has the same dimension as the system A. We thus obtain

$$\inf_{\rho_{RA}} F(\mathcal{N}_{A\to B}(\rho_{RA}),\mathcal{M}_{A\to B}(\rho_{RA})) = \inf_{\psi_{RA}} F(\mathcal{N}_{A\to B}(\psi_{RA}),\mathcal{M}_{A\to B}(\psi_{RA}) \tag{6.4.15}$$

$$= F(\mathcal{N}, \mathcal{M}). \tag{6.4.16}$$

We then have that $F(\mathcal{N}) = F(\mathcal{N}, \mathrm{id})$. In other words, the fidelity $F(\mathcal{N})$ of a quantum channel $\mathcal{N}$ can be viewed as the fidelity between $\mathcal{N}$ and the identity channel id.

Similar to the diamond distance, the fidelity of quantum channels can be computed by means of primal and dual semi-definite programs:

Proposition 6.24 SDP for Root Fidelity of Channels

Let $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ be quantum channels with respective Choi operators $\Gamma^{\mathcal{N}}_{RB}$ and $\Gamma^{\mathcal{M}}_{RB}$. Then their root channel fidelity

$$\sqrt{F}(\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}) := \inf_{\psi_{RA}} \sqrt{F}(\mathcal{N}_{A\to B}(\psi_{RA}), \mathcal{M}_{A\to B}(\psi_{RA})) \tag{6.4.17}$$

can be calculated by means of the following semi-definite program:

$$\sqrt{F}(\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B})$$
$$= \sup_{\lambda\geq 0, Q_{RB}} \left\{\lambda : \lambda I_R \leq \mathrm{Re}[\mathrm{Tr}_B[Q_{RB}]], \quad \begin{pmatrix} \Gamma^{\mathcal{N}}_{RB} & Q^\dagger_{RB} \\ Q_{RB} & \Gamma^{\mathcal{M}}_{RB} \end{pmatrix} \geq 0\right\} \tag{6.4.18}$$
$$= \frac{1}{2} \inf_{\rho_R\geq 0, W_{RB}, Z_{RB}} \left\{ \mathrm{Tr}[\Gamma^{\mathcal{N}}_{RB} W_{RB}] + \mathrm{Tr}[\Gamma^{\mathcal{M}}_{RB} Z_{RB}] : \mathrm{Tr}[\rho_R] = 1, \right.$$
$$\left. \begin{pmatrix} W_{RB} & \rho_R \otimes I_B \\ \rho_R \otimes I_B & Z_{RB} \end{pmatrix} \geq 0 \right\}. \tag{6.4.19}$$

The expression in (6.4.18) is equal to

$$\sup_{Q_{RB}} \left\{ \lambda_{\min} \left(\mathrm{Re}[\mathrm{Tr}_B[Q_{RB}]]\right) : \begin{pmatrix} \Gamma^{\mathcal{N}}_{RB} & Q^\dagger_{RB} \\ Q_{RB} & \Gamma^{\mathcal{M}}_{RB} \end{pmatrix} \geq 0 \right\}, \tag{6.4.20}$$

where $\lambda_{\min}$ denotes the minimum eigenvalue of its argument.

Proof: See Appendix 6.B.2. ■

The inequality in (6.2.88) relating the fidelity between two states ρ and σ and their trace distance can be used to relate the fidelity-based distance measure $F(\mathcal{N},\mathcal{M})$ on channels and the diamond distance $\frac{1}{2}\|\mathcal{N}-\mathcal{M}\|_\diamond$. It is straightfoward to show that

$$1-\sqrt{F(\mathcal{N},\mathcal{M})} \leq \frac{1}{2}\|\mathcal{N}-\mathcal{M}\|_\diamond \leq \sqrt{1-F(\mathcal{N},\mathcal{M})}. \tag{6.4.21}$$

The following proposition relates the fidelity $F(\mathcal{N})$ of a channel $\mathcal{N}$ to its minimum fidelity $F_{\min}(\mathcal{N})$ in (6.4.5), telling us that if $F_{\min}(\mathcal{N})$ is large then so is $F(\mathcal{N})$.

Proposition 6.25

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in [0,1]$, if $F_{\min}(\mathcal{N}) \geq 1-\varepsilon$, then $F(\mathcal{N}) \geq 1-2\sqrt{\varepsilon}$.

Proof: The inequality in $F_{\min}(\mathcal{N}) \geq 1-\varepsilon$ implies that the following inequality holds for all state vectors $|\phi\rangle \in \mathcal{H}$:

$$\langle\phi|\left[|\phi\rangle\langle\phi| - \mathcal{N}(|\phi\rangle\langle\phi|)\right]|\phi\rangle \leq \varepsilon. \tag{6.4.22}$$

By (6.2.88), this implies that

$$\left\||\phi\rangle\langle\phi| - \mathcal{N}(|\phi\rangle\langle\phi|)\right\|_1 \leq 2\sqrt{\varepsilon}, \tag{6.4.23}$$

for all state vectors $|\phi\rangle \in \mathcal{H}$. We will show that

$$\left|\langle\phi|\left[|\phi\rangle\langle\phi^\perp| - \mathcal{N}(|\phi\rangle\langle\phi^\perp|)\right]|\phi^\perp\rangle\right| \leq 2\sqrt{\varepsilon}, \tag{6.4.24}$$

for every orthonormal pair $\{|\phi\rangle, |\phi^\perp\rangle\}$ of state vectors in $\mathcal{H}$. Set

$$|w_k\rangle \coloneqq \frac{|\phi\rangle + \mathrm{i}^k|\phi^\perp\rangle}{\sqrt{2}}, \tag{6.4.25}$$

for $k \in \{0,1,2,3\}$. Then, it follows that

$$|\phi\rangle\langle\phi^\perp| = \frac{1}{2}\sum_{k=0}^{3} \mathrm{i}^k |w_k\rangle\langle w_k|. \tag{6.4.26}$$

Consider now that

$$\begin{aligned}
&\left|\langle\phi|\left[|\phi\rangle\!\langle\phi^\perp| - \mathcal{N}(|\phi\rangle\!\langle\phi^\perp|)\right]|\phi^\perp\rangle\right| \\
&\leq \left\||\phi\rangle\!\langle\phi^\perp| - \mathcal{N}(|\phi\rangle\!\langle\phi^\perp|)\right\|_\infty && (6.4.27)\\
&\leq \frac{1}{2}\sum_{k=0}^{3}\||w_k\rangle\!\langle w_k| - \mathcal{N}(|w_k\rangle\!\langle w_k|)\|_\infty && (6.4.28)\\
&\leq \frac{1}{4}\sum_{k=0}^{3}\||w_k\rangle\!\langle w_k| - \mathcal{N}(|w_k\rangle\!\langle w_k|)\|_1 && (6.4.29)\\
&\leq 2\sqrt{\varepsilon}. && (6.4.30)
\end{aligned}$$

The first inequality follows from the characterization of the operator norm in (2.2.106) as $\|X\|_\infty = \sup_{|\phi\rangle,|\psi\rangle}|\langle\psi|X|\phi\rangle|$, where the optimization is with respect to pure states. The second inequality follows from substituting (6.4.26) and applying the triangle inequality and homogeneity of the ∞-norm. The third inequality follows because the ∞-norm of a traceless Hermitian operator is bounded from above by half of its trace norm (see Lemma 2.11 below). The final inequality follows from applying (6.4.23). Let $|\psi\rangle \in \mathcal{H}'\otimes\mathcal{H}$ be an arbitrary state. All such states have a Schmidt decomposition of the following form:

$$|\psi\rangle = \sum_x \sqrt{p(x)}|\zeta_x\rangle\otimes|\varphi_x\rangle, \tag{6.4.31}$$

where $\{p(x)\}_x$ is a probability distribution and $\{|\zeta_x\rangle\}_x$ and $\{|\varphi_x\rangle\}_x$ are sets of states. Then, consider that

$$\begin{aligned}
&1 - \langle\psi|(\mathrm{id}_{\mathcal{H}'}\otimes\mathcal{N})(|\psi\rangle\!\langle\psi|)|\psi\rangle \\
&= \langle\psi|(\mathrm{id}_{\mathcal{H}'}\otimes\mathrm{id}_{\mathcal{H}} - \mathrm{id}_{\mathcal{H}'}\otimes\mathcal{N})(|\psi\rangle\!\langle\psi|)|\psi\rangle && (6.4.32)\\
&= \langle\psi|(\mathrm{id}_{\mathcal{H}'}\otimes[\mathrm{id}_{\mathcal{H}} - \mathcal{N}])(|\psi\rangle\!\langle\psi|)|\psi\rangle && (6.4.33)\\
&= \sum_{x,y} p(x)p(y)\langle\varphi_x|\left[|\varphi_x\rangle\!\langle\varphi_y| - \mathcal{N}(|\varphi_x\rangle\!\langle\varphi_y|)\right]|\varphi_y\rangle. && (6.4.34)
\end{aligned}$$

Now, applying the triangle inequality and (6.4.24), we find that the following holds for all $|\psi\rangle \in \mathcal{H}'\otimes\mathcal{H}$:

$$\begin{aligned}
&1 - \langle\psi|(\mathrm{id}_{\mathcal{H}'}\otimes\mathcal{N})(|\psi\rangle\!\langle\psi|)|\psi\rangle \\
&= \left|\sum_{x,y} p(x)p(y)\langle\varphi_x|\left[|\varphi_x\rangle\!\langle\varphi_y| - \mathcal{N}(|\varphi_x\rangle\!\langle\varphi_y|)\right]|\varphi_y\rangle\right| && (6.4.35)
\end{aligned}$$

$$\leq \sum_{x,y} p(x)p(y) \left| \langle \varphi_x | \left[|\varphi_x\rangle\langle\varphi_y| - \mathcal{N}(|\varphi_x\rangle\langle\varphi_y|) \right] |\varphi_y\rangle \right| \tag{6.4.36}$$

$$\leq 2\sqrt{\varepsilon}. \tag{6.4.37}$$

This concludes the proof. ■

6.5 Bibliographic Notes

The quantum fidelity was defined by Uhlmann (1976), and Theorem 6.8 is due to Uhlmann (1976). The semi-definite program for root fidelity in Proposition 6.6 was established by Watrous (2013), and we have followed the proof therein. The fact that the fidelity is achieved by a quantum measurement was realized by Fuchs and Caves (1995). The relation between trace distance and fidelity presented in Theorem 6.14 was proved by Fuchs and van de Graaf (1998). For very closely related inequalities, with the fidelity replaced by the "Holevo fidelity" $(\mathrm{Tr}[\sqrt{\rho}\sqrt{\sigma}])^2$, see Holevo (1972b). The sine distance was defined by Rastegin (2002, 2003); Gilchrist et al. (2005); Rastegin (2006), and its interpretation in terms of the minimal trace distance of purifications was given by Rastegin (2006). The sine distance was generalized to subnormalized states by Tomamichel et al. (2010), where it was given the name "purified distance."

The diamond norm was presented and studied by Kitaev (1997), who applied it to problems in quantum information theory and quantum computation. The operational interpretation of the diamond distance in terms of hypothesis testing of quantum channels was given by Kretschmann and Werner (2004); Rosgen and Watrous (2005); Gilchrist et al. (2005). More properties of the diamond norm can be found in Watrous (2018). The SDP in Proposition 6.20 for the normalized diamond distance of quantum channels was given by Watrous (2009).

Schumacher (1996) introduced the entanglement fidelity of a quantum channel, and Barnum et al. (2000) made further observations regarding it. Nielsen (2002) provided a simple proof for the relation between entanglement fidelity and average fidelity in (6.4.4). The fidelity of quantum channels was introduced by Gilchrist et al. (2005), and it can be understood as a special case of the generalized channel divergence (Leditzky et al., 2018). A semi-definite program for the root fidelity of channels was given by Yuan and Fung (2017). The particular semi-definite program in Proposition 6.24, for the root fidelity of channels, was presented by Katariya and Wilde (2021). Proposition 6.25 was established by Barnum et al. (2000) and reviewed by Kretschmann and Werner (2004). Here we followed the proof given by Watrous (2018, Theorem 3.56), which therein established a relation between trace distance and diamond distance between an arbitrary channel and the identity channel.

Appendix 6.A SDP for Normalized Diamond Distance

Here, we provide a proof of Proposition 6.20.

PROOF OF PROPOSITION 6.20: Employing (6.3.2), consider that

$$\frac{1}{2}\|\mathcal{N}-\mathcal{M}\|_\diamond = \sup_{\substack{\psi_{RA}\geq 0,\\ \Lambda_{RB}\geq 0}} \left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}(\mathcal{N}_{A\to B}-\mathcal{M}_{A\to B})(\psi_{RA})] : \Lambda_{RB}\leq \mathbb{1}_{RB}, \\ \operatorname{Tr}[\psi_{RA}]=1, \operatorname{Tr}[\psi_{RA}^2]=1 \end{array} \right\}, \tag{6.A.1}$$

where the constraints $\psi_{RA}\geq 0$, $\operatorname{Tr}[\psi_{RA}]=1$, and $\operatorname{Tr}[\psi_{RA}^2]=1$ correspond to ψ_{RA} being a pure bipartite state. Note that the above is equal to

$$\sup_{\substack{\psi_{RA}\geq 0,\\ \Lambda_{RB}\geq 0}} \left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}(\mathcal{N}_{A\to B}-\mathcal{M}_{A\to B})(\psi_{RA})] : \Lambda_{RB}\leq \mathbb{1}_{RB}, \psi_R>0, \\ \operatorname{Tr}[\psi_{RA}]=1, \operatorname{Tr}[\psi_{RA}^2]=1 \end{array} \right\}, \tag{6.A.2}$$

due to the fact that the set of pure states with reduced state ψ_R positive definite is dense in the set of all pure states. Now, recall from (2.2.40) that any such pure state can be written as $\psi_{RA}=X_R\Gamma_{RA}X_R^\dagger$ for some linear operator X_R such that $\operatorname{Tr}[X_R^\dagger X_R]=1$ and $|X_R|>0$, where Γ_{RA} defined in (2.2.36). Using this, we find that the objective function can be rewritten as

$$\operatorname{Tr}[\Lambda_{RB}(\mathcal{N}_{A\to B}-\mathcal{M}_{A\to B})(X_R\Gamma_{RA}X_R^\dagger)] \tag{6.A.3}$$

$$=\operatorname{Tr}[X_R^\dagger\Lambda_{RB}X_R(\mathcal{N}_{A\to B}-\mathcal{M}_{A\to B})(\Gamma_{RA})] \tag{6.A.4}$$

$$=\operatorname{Tr}[X_R^\dagger\Lambda_{RB}X_R(\Gamma_{RB}^{\mathcal{N}}-\Gamma_{RB}^{\mathcal{M}})]. \tag{6.A.5}$$

Now, observe the following equivalence:

$$0 \leq \Lambda_{RB} \leq \mathbb{1}_{RB} \Leftrightarrow 0 \leq X_R^\dagger \Lambda_{RB} X_R \leq X_R^\dagger X_R \otimes \mathbb{1}_B. \tag{6.A.6}$$

Thus, defining $\Omega_{RB} \coloneqq X_R^\dagger \Lambda_{RB} X_R$ and $\rho_R \coloneqq X_R^\dagger X_R$, the optimization in (6.A.2) is equivalent to the following one:

$$\sup_{\substack{\rho_R, \\ \Omega_{RB} \geq 0}} \{\mathrm{Tr}[\Omega_{RB}(\Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB})] : \Omega_{RB} \leq \rho_R \otimes \mathbb{1}_B, \rho_R > 0, \mathrm{Tr}[\rho_R] = 1\}, \tag{6.A.7}$$

giving the equality in (6.3.9). Finally, setting

$$Y \coloneqq \begin{pmatrix} \Omega_{RB} & 0 \\ 0 & \rho_R \end{pmatrix}, \tag{6.A.8}$$

$$D \coloneqq \begin{pmatrix} \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB} & 0 \\ 0 & 0 \end{pmatrix}, \tag{6.A.9}$$

$$\Phi(Y) \coloneqq \begin{pmatrix} \Omega_{RB} - \rho_R \otimes \mathbb{1}_B & 0 & 0 \\ 0 & \mathrm{Tr}[\rho_R] & 0 \\ 0 & 0 & -\mathrm{Tr}[\rho_R] \end{pmatrix}, \tag{6.A.10}$$

$$C \coloneqq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{6.A.11}$$

we find that (6.3.9) is now in the standard form from (2.4.3), namely,

$$\sup_{Y \geq 0} \{\mathrm{Tr}[DY] : \Phi(Y) \leq C\}. \tag{6.A.12}$$

Now, to establish the dual SDP in (6.3.10), we first determine the adjoint $\Phi^\dagger$ of Φ using

$$\mathrm{Tr}[\Phi(Y)Z] = \mathrm{Tr}[Y\Phi^\dagger(Z)], \tag{6.A.13}$$

where without loss of generality we can take Z to be

$$Z \coloneqq \begin{pmatrix} Z_{RB} & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_2 \end{pmatrix}. \tag{6.A.14}$$

Then, we find that

$$\mathrm{Tr}[\Phi(Y)Z] = \mathrm{Tr}[(\Omega_{RB} - \rho_R \otimes \mathbb{1}_B)Z_{RB}] + \mathrm{Tr}[\rho_R]\mu_1 - \mathrm{Tr}[\rho_R]\mu_2 \tag{6.A.15}$$

$$= \mathrm{Tr}[\Omega_{RB} Z_{RB}] + \mathrm{Tr}[\rho_R((\mu_1 - \mu_2)\mathbb{1}_R - \mathrm{Tr}_B[Z_{RB}])], \tag{6.A.16}$$

from which we conclude that

$$\Phi^\dagger(Z) = \begin{pmatrix} Z_{RB} & 0 \\ 0 & (\mu_1 - \mu_2)\mathbb{1}_R - \mathrm{Tr}_B[Z_{RB}] \end{pmatrix}. \tag{6.A.17}$$

The standard form of the dual SDP from (2.4.4), which is

$$\inf_{Z \geq 0} \{\mathrm{Tr}[CZ] : \Phi^\dagger(Z) \geq D\}, \tag{6.A.18}$$

then becomes

$$\inf_{\substack{\mu_1 \geq 0, \\ \mu_2 \geq 0, \\ Z_{RB} \geq 0}} \{\mu_1 - \mu_2 : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB}, (\mu_1 - \mu_2)\mathbb{1}_R \geq \mathrm{Tr}_B[Z_{RB}]\}. \tag{6.A.19}$$

Now, observe that the variables μ_1 and μ_2 always appear together in the above optimization as $\mu_1 - \mu_2$, and so can be reduced to the a single real variable $\mu \in \mathbb{R}$. Then, the condition $\mu\mathbb{1}_R \geq \mathrm{Tr}_B[Z_{RB}]$ implies that $\mu \geq 0$. Thus, the optimization in (6.A.19) can be simplified to

$$\inf_{\substack{\mu \geq 0, \\ Z_{RB} \geq 0}} \{\mu : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB}, \mu\mathbb{1}_R \geq \mathrm{Tr}_B[Z_{RB}]\}, \tag{6.A.20}$$

which is precisely (6.3.10). Equality of the primal and dual SDPs is due to strong duality, which holds for the SDP in (6.3.10) because $Z_{RB} = \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB} + \delta\mathbb{1}_{RB}$ and $\mu = \mathrm{Tr}_B[Z_{RB}] + \delta\mathbb{1}_R$ together form a strictly feasible point for all $\delta > 0$ and a feasible point for the primal is $\rho_R = \pi_R$ and $\Omega_{RB} = \pi_R \otimes \mathbb{1}_B$.

Finally, the equality

$$\inf_{\substack{\mu \geq 0, \\ Z_{RB} \geq 0}} \{\mu : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB}, \mu\mathbb{1}_R \geq \mathrm{Tr}_B[Z_{RB}]\}$$
$$= \inf_{Z_{RB} \geq 0} \{\|\mathrm{Tr}_B[Z_{RB}]\|_\infty : Z_{RB} \geq \Gamma^{\mathcal{N}}_{RB} - \Gamma^{\mathcal{M}}_{RB}\} \tag{6.A.21}$$

holds by the expression in (2.4.47) for the Schatten ∞-norm for positive semi-definite operators.

Appendix 6.B SDPs for Fidelity of States and Channels

6.B.1 Proof of Proposition 6.6

First, let us verify that strong duality holds for the primal and dual semi-definite programs in (6.2.4) and (6.2.5), respectively. Consider that $X = 0$ is a feasible choice for the primal program, while $Y = Z = 2\mathbb{1}$ is strictly feasible for the dual program. Thus, strong duality holds according to Theorem 2.28.

In order to prove the equality in (6.2.4), we start with the following lemma:

Lemma 6.26

Let P and Q be positive semi-definite operators in $\mathcal{L}(\mathcal{H})$, and let $X \in \mathcal{L}(\mathcal{H})$. Then the operator

$$\begin{pmatrix} P & X \\ X^\dagger & Q \end{pmatrix} \tag{6.B.1}$$

is positive semi-definite if and only if there exists $K \in \mathcal{L}(\mathcal{H})$ satisfying $\|K\|_\infty \leq 1$ and $X = \sqrt{P}K\sqrt{Q}$.

Proof: See Theorem IX.5.9 of Bhatia (1997). ■

It follows from this lemma that the operators X in the primal optimization in (6.2.4) can range over $X = \sqrt{P}K\sqrt{Q}$ such that $K \in \mathcal{L}(\mathcal{H})$ and $\|K\|_\infty \leq 1$, so that the optimization over X is reduced to an optimization over K. We then find that the primal optimal value is given by

$$\frac{1}{2} \sup_{X \in \mathcal{L}(\mathcal{H})} \left\{ \mathrm{Tr}[X] + \mathrm{Tr}[X^\dagger] : \begin{pmatrix} \rho & X \\ X^\dagger & \sigma \end{pmatrix} \geq 0 \right\}$$

$$= \frac{1}{2} \sup_{K : \|K\|_\infty \leq 1} \mathrm{Tr}[\sqrt{\rho}K\sqrt{\sigma}] + \mathrm{Tr}[\sqrt{\sigma}K^\dagger\sqrt{\rho}] \tag{6.B.2}$$

$$= \sup_{K : \|K\|_\infty \leq 1} \mathrm{Re}[\mathrm{Tr}[\sqrt{\rho}K\sqrt{\sigma}]] \tag{6.B.3}$$

$$= \sup_{K : \|K\|_\infty \leq 1} \left|\mathrm{Tr}[\sqrt{\rho}K\sqrt{\sigma}]\right| \tag{6.B.4}$$

$$= \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 = \sqrt{F}(\rho, \sigma). \tag{6.B.5}$$

The first equality follows from Lemma 6.26. The third equality follows because we can use the optimization over K to adjust a global phase such that the real part is equal to the absolute value (here, one should think of the fact that $\mathrm{Re}[z] = r\cos(\theta)$ for $z = re^{i\theta}$, and then one can optimize the value of θ so that $\mathrm{Re}[z] = r$). The final equality follows by a generalization of Proposition 2.10 (in fact the same proof given there implies that the optimization can be with respect to U satisfying $\|U\|_\infty \leq 1$, rather than just with respect to isometries).

We now prove that (6.2.5) is the dual program of (6.2.4). We can rewrite the primal SDP as

$$\frac{1}{2} \sup \mathrm{Tr}[X] + \mathrm{Tr}[X^\dagger] \tag{6.B.6}$$

subject to

$$\begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix} \geq \begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}, \qquad \begin{pmatrix} R & X \\ X^\dagger & S \end{pmatrix} \geq 0 \tag{6.B.7}$$

because R and S are not involved in the objective function and can always be chosen so that the second operator is PSD. Also, the following equivalences hold

$$\begin{pmatrix} \rho & X \\ X^\dagger & \sigma \end{pmatrix} \geq 0 \quad \Longleftrightarrow \quad \begin{pmatrix} \rho & -X \\ -X^\dagger & \sigma \end{pmatrix} \geq 0 \tag{6.B.8}$$

$$\Longleftrightarrow \quad \begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix} \geq \begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}. \tag{6.B.9}$$

As given in (2.4.3) and (2.4.4), the standard forms of primal and dual SDPs for Hermitian A and B and Hermiticity-preserving map Φ are as follows:

$$\sup_{G \geq 0} \{\mathrm{Tr}[AG] : \Phi(G) \leq B\}, \tag{6.B.10}$$

$$\inf_{Y \geq 0} \left\{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\right\}. \tag{6.B.11}$$

So the SDP above is in standard form with

$$G = \begin{pmatrix} R & X \\ X^\dagger & S \end{pmatrix}, \quad A = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \tag{6.B.12}$$

$$\Phi(G) = \begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix}. \tag{6.B.13}$$

Setting

$$Y = \begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix}, \tag{6.B.14}$$

the adjoint of Φ is given by

$$\begin{aligned} \mathrm{Tr}[Y\Phi(X)] &= \mathrm{Tr}\left[\begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix}\begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}\right] && (6.B.15)\\ &= \mathrm{Tr}\left[\begin{pmatrix} VX^\dagger & WX \\ ZX^\dagger & V^\dagger X \end{pmatrix}\right] && (6.B.16)\\ &= \mathrm{Tr}[VX^\dagger] + \mathrm{Tr}[XV^\dagger] && (6.B.17)\\ &= \mathrm{Tr}\left[\begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix}\begin{pmatrix} R & X \\ X^\dagger & S \end{pmatrix}\right], && (6.B.18) \end{aligned}$$

so that

$$\Phi^\dagger(Y) = \begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix}. \tag{6.B.19}$$

Then the dual is given by

$$\frac{1}{2}\inf \mathrm{Tr}\left[\begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix}\begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix}\right] \tag{6.B.20}$$

subject to

$$\begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix} \geq \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \qquad \begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix} \geq 0. \tag{6.B.21}$$

This simplifies to

$$\frac{1}{2}\inf \mathrm{Tr}[\rho W] + \mathrm{Tr}[\sigma Z], \tag{6.B.22}$$

subject to

$$\begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix} \geq \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \qquad \begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix} \geq 0 \tag{6.B.23}$$

Since

$$\begin{aligned} \begin{pmatrix} W & V \\ V^\dagger & Z \end{pmatrix} \geq 0 \quad &\Longleftrightarrow \quad \begin{pmatrix} W & -V \\ -V^\dagger & Z \end{pmatrix} \geq 0 && (6.B.24)\\ &\Longleftrightarrow \quad \begin{pmatrix} W & 0 \\ 0 & Z \end{pmatrix} \geq \begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix}, && (6.B.25) \end{aligned}$$

we find that there is a single condition

$$\begin{pmatrix} W & 0 \\ 0 & Z \end{pmatrix} \geq \begin{pmatrix} 0 & V \\ V^\dagger & 0 \end{pmatrix} \geq \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \tag{6.B.26}$$

and since V plays no role in the objective function, we can set $V = \mathbb{1}$. So the final SDP simplifies as follows:

$$\frac{1}{2} \inf_{W,Z} \operatorname{Tr}[\rho W] + \operatorname{Tr}[\sigma Z] \tag{6.B.27}$$

subject to

$$\begin{pmatrix} W & -\mathbb{1} \\ -\mathbb{1} & Z \end{pmatrix} \geq 0. \tag{6.B.28}$$

Using the fact that

$$\begin{pmatrix} W & -\mathbb{1} \\ -\mathbb{1} & Z \end{pmatrix} \geq 0 \quad \Longleftrightarrow \quad \begin{pmatrix} W & \mathbb{1} \\ \mathbb{1} & Z \end{pmatrix} \geq 0 \tag{6.B.29}$$

we can do one final rewriting as follows:

$$\frac{1}{2} \inf_{W,Z} \operatorname{Tr}[\rho W] + \operatorname{Tr}[\sigma Z] \tag{6.B.30}$$

subject to

$$\begin{pmatrix} W & \mathbb{1} \\ \mathbb{1} & Z \end{pmatrix} \geq 0. \tag{6.B.31}$$

6.B.2 Proof of Proposition 6.24

First, strong duality holds, according to Theorem 2.28, because $Q_{RB} = 0$ and $\lambda = 0$ is feasible for the primal program, while $\rho_R = \mathbb{1}/|R|$ and $W_{RB} = Z_{RB} = 2\mathbb{1}_{RB}$ is strictly feasible for the dual.

For a pure bipartite state ψ_{RA}, we use (2.2.40) to conclude that

$$\psi_{RA} = X_R \Gamma_{RA} X_R^\dagger, \tag{6.B.32}$$

where $\operatorname{Tr}[X_R^\dagger X_R] = 1$ to see that

$$\mathcal{N}_{A\to B}(\psi_{RA}) = X_R \Gamma_{RB}^{\mathcal{N}} X_R^\dagger, \quad \mathcal{M}_{A\to B}(\psi_{RA}) = X_R \Gamma_{RB}^{\mathcal{M}} X_R^\dagger, \tag{6.B.33}$$

and then plug in to (6.2.5) to get that

$$\sqrt{F}(\mathcal{N}_{A\to B},\mathcal{M}_{A\to B}) = \frac{1}{2}\inf_{W_{RB},Z_{RB}} \operatorname{Tr}[X_R\Gamma^{\mathcal{N}}_{RB}X_R^{\dagger}W_{RB}] + \operatorname{Tr}[X_R\Gamma^{\mathcal{M}}_{RB}X_R^{\dagger}Z_{RB}] \tag{6.B.34}$$

subject to

$$\begin{pmatrix} W_{RB} & \mathbb{1}_{RB} \\ \mathbb{1}_{RB} & Z_{RB} \end{pmatrix} \geq 0. \tag{6.B.35}$$

Consider that the objective function can be written as

$$\operatorname{Tr}[\Gamma^{\mathcal{N}}_{RB}W'_{RB}] + \operatorname{Tr}[\Gamma^{\mathcal{M}}_{RB}Z'_{RB}], \tag{6.B.36}$$

with

$$W'_{RB} := X_R^{\dagger}W_{RB}X_R, \quad Z'_{RB} := X_R^{\dagger}Z_{RB}X_R \tag{6.B.37}$$

Now consider that the inequality in (6.B.35) is equivalent to

$$\begin{pmatrix} X_R & 0 \\ 0 & X_R \end{pmatrix}^{\dagger} \begin{pmatrix} W_{RB} & \mathbb{1}_{RB} \\ \mathbb{1}_{RB} & Z_{RB} \end{pmatrix} \begin{pmatrix} X_R & 0 \\ 0 & X_R \end{pmatrix} \geq 0. \tag{6.B.38}$$

(Here we have assumed that X_R is invertible, but it suffices to do so for this optimization because the set of invertible X_R is dense in the set of all possible X_R.) Multiplying out the last matrix we find that

$$\begin{aligned} &\begin{pmatrix} X_R & 0 \\ 0 & X_R \end{pmatrix}^{\dagger} \begin{pmatrix} W_{RB} & \mathbb{1}_{RB} \\ \mathbb{1}_{RB} & Z_{RB} \end{pmatrix} \begin{pmatrix} X_R & 0 \\ 0 & X_R \end{pmatrix} \\ &= \begin{pmatrix} X_R^{\dagger}W_{RB}X_R & X_R^{\dagger}X_R \otimes \mathbb{1}_B \\ X_R^{\dagger}X_R \otimes \mathbb{1}_B & X_R^{\dagger}Z_{RB}X_R \end{pmatrix} &(6.B.39) \\ &= \begin{pmatrix} W'_{RB} & \rho_R \otimes \mathbb{1}_B \\ \rho_R \otimes \mathbb{1}_B & Z'_{RB} \end{pmatrix}, &(6.B.40) \end{aligned}$$

where we defined $\rho_R = X_R^{\dagger}X_R$. Observing that $\rho_R \geq 0$ and $\operatorname{Tr}[\rho_R] = 1$, we can write the final SDP as follows:

$$\sqrt{F}(\mathcal{N}_{A\to B},\mathcal{M}_{A\to B}) = \frac{1}{2}\inf_{\rho_R,W_{RB},Z_{RB}} \operatorname{Tr}[\Gamma^{\mathcal{N}}_{RB}W_{RB}] + \operatorname{Tr}[\Gamma^{\mathcal{M}}_{RB}Z_{RB}], \tag{6.B.41}$$

subject to

$$\rho_R \geq 0, \quad \operatorname{Tr}[\rho_R] = 1, \quad \begin{pmatrix} W_{RB} & \rho_R \otimes \mathbb{1}_B \\ \rho_R \otimes \mathbb{1}_B & Z_{RB} \end{pmatrix} \geq 0. \tag{6.B.42}$$

Now let us calculate the dual SDP to this, using the following standard forms for primal and dual SDPs, with Hermitian operators A and B and a Hermiticity-preserving map Φ (as given in (2.4.3) and (2.4.4)):

$$\sup_{X\geq 0}\left\{\mathrm{Tr}[AX]:\Phi(X)\leq B\right\},\qquad \inf_{Y\geq 0}\left\{\mathrm{Tr}[BY]:\Phi^\dagger(Y)\geq A\right\}. \tag{6.B.43}$$

Consider that the constraint in (6.B.42) implies $W_{RB}\geq 0$ and $Z_{RB}\geq 0$, so that we can set

$$Y=\begin{pmatrix} W_{RB} & 0 & 0\\ 0 & Z_{RB} & 0\\ 0 & 0 & \rho_R\end{pmatrix},\quad B=\begin{pmatrix}\Gamma^{\mathcal{N}}_{RB} & 0 & 0\\ 0 & \Gamma^{\mathcal{M}}_{RB} & 0\\ 0 & 0 & 0\end{pmatrix}, \tag{6.B.44}$$

$$\Phi^\dagger(Y)=\begin{pmatrix} W_{RB} & \rho_R\otimes\mathbb{1}_B & 0 & 0\\ \rho_R\otimes\mathbb{1}_B & Z_{RB} & 0 & 0\\ 0 & 0 & \mathrm{Tr}[\rho_R] & 0\\ 0 & 0 & 0 & -\mathrm{Tr}[\rho_R]\end{pmatrix}, \tag{6.B.45}$$

$$A=\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}. \tag{6.B.46}$$

Then with

$$X=\begin{pmatrix} P_{RB} & Q^\dagger_{RB} & 0 & 0\\ Q_{RB} & S_{RB} & 0 & 0\\ 0 & 0 & \lambda & 0\\ 0 & 0 & 0 & \mu\end{pmatrix} \tag{6.B.47}$$

the map Φ is given by

$$\begin{aligned}
&\mathrm{Tr}[X\Phi^\dagger(Y)]\\
&=\mathrm{Tr}\left[\begin{pmatrix} P_{RB} & Q^\dagger_{RB} & 0 & 0\\ Q_{RB} & S_{RB} & 0 & 0\\ 0 & 0 & \lambda & 0\\ 0 & 0 & 0 & \mu\end{pmatrix}\begin{pmatrix} W_{RB} & \rho_R\otimes\mathbb{1}_B & 0 & 0\\ \rho_R\otimes\mathbb{1}_B & Z_{RB} & 0 & 0\\ 0 & 0 & \mathrm{Tr}[\rho_R] & 0\\ 0 & 0 & 0 & -\mathrm{Tr}[\rho_R]\end{pmatrix}\right]\\
&=\mathrm{Tr}[P_{RB}W_{RB}]+\mathrm{Tr}[Q^\dagger_{RB}(\rho_R\otimes\mathbb{1}_B)]+\mathrm{Tr}[Q_{RB}(\rho_R\otimes\mathbb{1}_B)]\\
&\qquad+\mathrm{Tr}[S_{RB}Z_{RB}]+(\lambda-\mu)\,\mathrm{Tr}[\rho_R]\\
&=\mathrm{Tr}[P_{RB}W_{RB}]+\mathrm{Tr}[S_{RB}Z_{RB}]+\mathrm{Tr}[(\mathrm{Tr}_B[Q_{RB}+Q^\dagger_{RB}]+(\lambda-\mu)\,\mathbb{1}_R)\rho_R]
\end{aligned}$$

$$= \operatorname{Tr}\left[\begin{pmatrix} P_{RB} & 0 & 0 \\ 0 & S_{RB} & 0 \\ 0 & 0 & \operatorname{Tr}_B[Q_{RB} + Q_{RB}^\dagger] + (\lambda - \mu)\,\mathbb{1}_R \end{pmatrix}\begin{pmatrix} W_{RB} & 0 & 0 \\ 0 & Z_{RB} & 0 \\ 0 & 0 & \rho_R \end{pmatrix}\right]. \tag{6.B.48}$$

So then

$$\Phi(X) = \begin{pmatrix} P_{RB} & 0 & 0 \\ 0 & S_{RB} & 0 \\ 0 & 0 & \operatorname{Tr}_B[Q_{RB} + Q_{RB}^\dagger] + (\lambda - \mu)\,\mathbb{1}_R \end{pmatrix}. \tag{6.B.49}$$

The primal is then given by

$$\frac{1}{2}\sup \operatorname{Tr}\left[\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} P_{RB} & Q_{RB}^\dagger & 0 & 0 \\ Q_{RB} & S_{RB} & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix}\right], \tag{6.B.50}$$

subject to

$$\begin{pmatrix} P_{RB} & 0 & 0 \\ 0 & S_{RB} & 0 \\ 0 & 0 & \operatorname{Tr}_B[Q_{RB} + Q_{RB}^\dagger] + (\lambda - \mu)\,\mathbb{1}_R \end{pmatrix} \leq \begin{pmatrix} \Gamma_{RB}^{\mathcal{N}} & 0 & 0 \\ 0 & \Gamma_{RB}^{\mathcal{M}} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.B.51}$$

$$\begin{pmatrix} P_{RB} & Q_{RB}^\dagger & 0 & 0 \\ Q_{RB} & S_{RB} & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix} \geq 0, \tag{6.B.52}$$

which simplifies to

$$\frac{1}{2}\sup\,(\lambda - \mu) \tag{6.B.53}$$

subject to

$$P_{RB} \leq \Gamma_{RB}^{\mathcal{N}}, \tag{6.B.54}$$

$$S_{RB} \leq \Gamma_{RB}^{\mathcal{M}}, \tag{6.B.55}$$

$$\operatorname{Tr}_B[Q_{RB} + Q_{RB}^\dagger] + (\lambda - \mu)\,\mathbb{1}_R \leq 0, \tag{6.B.56}$$

$$\begin{pmatrix} P_{RB} & Q_{RB}^\dagger \\ Q_{RB} & S_{RB} \end{pmatrix} \geq 0, \tag{6.B.57}$$

$$\lambda, \mu \geq 0. \tag{6.B.58}$$

We can simplify this even more. We can set $\lambda' = \lambda - \mu \in \mathbb{R}$, and we can substitute Q_{RB} with $-Q_{RB}$ without changing the value, so then it becomes

$$\frac{1}{2}\sup \lambda' \tag{6.B.59}$$

subject to

$$P_{RB} \leq \Gamma^{\mathcal{N}}_{RB}, \tag{6.B.60}$$

$$S_{RB} \leq \Gamma^{\mathcal{M}}_{RB}, \tag{6.B.61}$$

$$\lambda' \mathbb{1}_R \leq \operatorname{Tr}_B[Q_{RB} + Q^\dagger_{RB}], \tag{6.B.62}$$

$$\begin{pmatrix} P_{RB} & -Q^\dagger_{RB} \\ -Q_{RB} & S_{RB} \end{pmatrix} \geq 0, \tag{6.B.63}$$

$$\lambda' \in \mathbb{R}. \tag{6.B.64}$$

We can rewrite

$$\begin{pmatrix} P_{RB} & -Q^\dagger_{RB} \\ -Q_{RB} & S_{RB} \end{pmatrix} \geq 0 \quad \Longleftrightarrow \quad \begin{pmatrix} P_{RB} & Q^\dagger_{RB} \\ Q_{RB} & S_{RB} \end{pmatrix} \geq 0 \tag{6.B.65}$$

$$\Longleftrightarrow \quad \begin{pmatrix} P_{RB} & 0 \\ 0 & S_{RB} \end{pmatrix} \geq \begin{pmatrix} 0 & -Q^\dagger_{RB} \\ -Q_{RB} & 0 \end{pmatrix} \tag{6.B.66}$$

We then have the simplified condition

$$\begin{pmatrix} 0 & -Q^\dagger_{RB} \\ -Q_{RB} & 0 \end{pmatrix} \leq \begin{pmatrix} P_{RB} & 0 \\ 0 & S_{RB} \end{pmatrix} \leq \begin{pmatrix} \Gamma^{\mathcal{N}}_{RB} & 0 \\ 0 & \Gamma^{\mathcal{M}}_{RB} \end{pmatrix}. \tag{6.B.67}$$

Since P_{RB} and S_{RB} do not appear in the objective function, we can set them to their largest value and obtain the following simplification

$$\frac{1}{2}\sup \lambda' \tag{6.B.68}$$

subject to

$$\lambda' \mathbb{1}_R \leq \operatorname{Tr}_B[Q_{RB} + Q^\dagger_{RB}], \quad \begin{pmatrix} \Gamma^{\mathcal{N}}_{RB} & Q^\dagger_{RB} \\ Q_{RB} & \Gamma^{\mathcal{M}}_{RB} \end{pmatrix} \geq 0, \quad \lambda' \in \mathbb{R} \tag{6.B.69}$$

Since a feasible solution is $\lambda' = 0$ and $Q_{RB} = 0$, it is clear that we can restrict to $\lambda' \geq 0$. After a relabeling, this becomes

$$\frac{1}{2} \sup_{\lambda \geq 0, Q_{RB}} \left\{ \lambda : \lambda \mathbb{1}_R \leq \operatorname{Tr}_B[Q_{RB} + Q_{RB}^\dagger], \quad \begin{pmatrix} \Gamma_{RB}^{\mathcal{N}} & Q_{RB}^\dagger \\ Q_{RB} & \Gamma_{RB}^{\mathcal{M}} \end{pmatrix} \geq 0 \right\}$$
$$= \sup_{\lambda \geq 0, Q_{RB}} \left\{ \lambda : \lambda \mathbb{1}_R \leq \operatorname{Re}[\operatorname{Tr}_B[Q_{RB}]], \quad \begin{pmatrix} \Gamma_{RB}^{\mathcal{N}} & Q_{RB}^\dagger \\ Q_{RB} & \Gamma_{RB}^{\mathcal{M}} \end{pmatrix} \geq 0 \right\}. \quad (6.B.70)$$

This is equivalent to

$$\sup_{Q_{RB}} \left\{ \lambda_{\min} \left(\operatorname{Re}[\operatorname{Tr}_B[Q_{RB}]] \right) : \begin{pmatrix} \Gamma_{RB}^{\mathcal{N}} & Q_{RB}^\dagger \\ Q_{RB} & \Gamma_{RB}^{\mathcal{M}} \end{pmatrix} \geq 0 \right\}. \quad (6.B.71)$$

This concludes the proof.

Chapter 7

Quantum Entropies and Information

In this chapter, we introduce various entropic and information quantities that play a fundamental role in the analysis of quantum communication protocols. Here we see that the notions of entropy and information take many forms. The most basic and fundamental entropic and information measures are members of the "von Neumann family," and these often correspond to optimal communication rates for information-theoretic tasks in the asymptotic regime of many uses of an independent and identically distributed (i.i.d.) resource. More refined entropy measures belong to the "Rényi family," and interestingly, in part due to the non-commutativity of quantum states, there are several interesting ways of generalizing the classical Rényi relative entropy that are meaningful for understanding information-theoretic tasks. The Rényi measures reduce to the von Neumann ones in a particular limit, and they are useful for characterizing optimal rates of information-theoretic tasks in the non-asymptotic regime, in particular when trying to determine how fast an error probability is converging to zero or one (so-called error exponents and strong converse exponents, respectively). Even more broadly, we define entropy measures from the meaningful "one-shot" information-theoretic task given by quantum hypothesis testing. Even though one might debate whether such an operationally defined quantity is truly an entropy, our opinion is that this perspective is quite powerful, and so we adopt it in this chapter and the rest of the book. Thus, this chapter explores quite broadly a variety of entropic measures and their mathematical properties, as they are the basis for analyzing a wide variety of quantum information-processing protocols.

We start our development with a brief preview of quantum entropies and information (Section 7.1) and then proceed to the well-known quantum relative entropy (Section 7.2), which plays a foundational role in understanding properties of other entropies. We then proceed to defining a generalized divergence, which is a concept that plays an important role in the proofs of strong converse theorems throughout this book (Section 7.3). Of particular focus are several prominent examples of generalized divergences, the Petz–Rényi relative entropy (Section 7.4), the sandwiched Rényi relative entropy (Section 7.5), the geometric Rényi relative entropy (Section 7.6), the Belavkin–Staszewski relative entropy (Section 7.7), the (smooth) max-relative entropy (Section 7.8), and the hypothesis testing relative entropy (Section 7.9). The first of these plays an important role in achievability proofs for quantum channel capacities, and the second plays an important role for strong converses. The hypothesis testing relative entropy is fundamental in establishing bounds on one-shot channel capacities.

7.1 Preview

Arguably one of the most important quantities in quantum information theory is the *von Neumann entropy*, which is the quantum generalization of the Shannon entropy. We also refer to it as the *quantum entropy* and do so from now on. For a quantum system A in the state $\rho_A \in \mathrm{D}(\mathcal{H}_A)$, the von Neumann entropy is defined as

$$H(\rho_A) \equiv H(A)_\rho := -\mathrm{Tr}[\rho_A \log_2 \rho_A]. \tag{7.1.1}$$

If ρ_A has a spectral decomposition of the form

$$\rho_A = \sum_{i=1}^{r} \lambda_i |\psi_i\rangle\langle\psi_i|_A, \tag{7.1.2}$$

where $r \equiv \mathrm{rank}(\rho_A)$, then we can write $H(\rho_A)$ in terms of the non-zero eigenvalues $\{\lambda_i\}_{i=1}^{r}$ of ρ_A as

$$H(\rho_A) = -\sum_{i=1}^{r} \lambda_i \log_2 \lambda_i. \tag{7.1.3}$$

Note that the zero eigenvalues of ρ_A do not contribute to the entropy due to the convention that $0 \log_2 0 = 0$, which is taken because $\lim_{x\to 0^+} x \log_2 x = 0$. By viewing ρ_A as a probabilistic mixture of the pure states $\{|\psi_i\rangle_A\}_{i=1}^{r}$ defined in (7.1.2), the quantum entropy quantifies the uncertainty about which of these pure states the system A is in. In particular, $H(\rho_A)$ is, in a rough sense, the expected information gain upon performing an experiment to determine the state of the system.

Note that the right-hand side of (7.1.3) is the formula for the *Shannon entropy* of the probability distribution $\{\lambda_i\}_i$ corresponding to the eigenvalues of ρ_A. The Shannon entropy of the probability distribution $\{p, 1-p\}$, for $p \in [0,1]$, shows up frequently and is denoted by $h_2(p)$, i.e.,

$$h_2(p) := -p \log_2 p - (1-p) \log_2(1-p). \tag{7.1.4}$$

It is called the binary entropy function.

Just as the Shannon entropy has an operational meaning as the optimal rate of (classical) data compression, the quantum entropy has an operational interpretation as the optimal rate of quantum data compression. More precisely, given the state $\rho_A^{\otimes n}$, the quantum entropy $H(\rho_A)$ is the minimum number of qubits per copy of the state ρ_A that are needed to faithfully represent $\rho_A^{\otimes n}$, when n becomes large. This task is also called Schumacher compression.

Other fundamental information-theoretic quantities, which are functions of the Shannon entropy, have straightforward generalizations to the quantum setting. Let ρ_{AB} be a bipartite state, and let σ_{ABC} be a tripartite state.

1. The *quantum conditional entropy* is defined as

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho. \tag{7.1.5}$$

   The quantum conditional entropy quantifies the uncertainty about the state of the system A in the presence of additional quantum side information in the form of the quantum system B.

2. The *coherent information* is defined as

$$I(A\rangle B)_\rho := H(B)_\rho - H(AB)_\rho = -H(A|B)_\rho, \tag{7.1.6}$$

   and it arises in the context of communication of quantum information over quantum channels (see Chapter 14). The coherent information is asymmetric and can be interpreted as having a directionality. We obtain a quantity called the *reverse coherent information* by swapping the systems A and B in (7.1.6):

$$I(B\rangle A)_\rho := H(A)_\rho - H(AB)_\rho = -H(B|A)_\rho. \tag{7.1.7}$$

   This quantity arises when studying feedback-assisted quantum communication.

3. The *quantum mutual information* is defined as

$$I(A;B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho, \tag{7.1.8}$$
$$= H(A)_\rho - H(A|B)_\rho \tag{7.1.9}$$
$$= H(B)_\rho - H(B|A)_\rho, \tag{7.1.10}$$

   and it arises in the context of communicating classical information over quantum channels (see Chapters 11 and 12).

4. The *quantum conditional mutual information* is defined as

$$I(A;B|C)_\sigma := H(A|C)_\sigma + H(B|C)_\sigma - H(AB|C)_\sigma, \tag{7.1.11}$$

   and it is the basis for an entanglement measure called squashed entanglement (see Chapter 9). An important result is that the quantum conditional mutual information is non-negative, i.e., $I(A;B|C)_\sigma \geq 0$ for every tripartite state σ_{ABC}. This inequality goes by the name *strong subadditivity of quantum entropy*, and we show at the end of Section 7.2 that it follows from the data-processing inequality for the quantum relative entropy (Theorem 7.4).

As it turns out, all of these quantities, including the entropy itself, can be derived from a single parent quantity, the quantum relative entropy, which we introduce in the next section.

7.2 Quantum Relative Entropy

We define the quantum relative entropy as follows:

Definition 7.1 Quantum Relative Entropy

For every state ρ and positive semi-definite operator σ, the *quantum relative entropy of ρ and σ*, denoted by $D(\rho\|\sigma)$, is defined as

$$D(\rho\|\sigma) = \begin{cases} \mathrm{Tr}[\rho(\log_2\rho - \log_2\sigma)] & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \tag{7.2.1}$$

> REMARK: More generally, we could define the quantum relative entropy exactly as above, but with both arguments being positive semi-definite operators. For our purposes in this book, however, it suffices to restrict the first argument to be a state.

The quantum relative entropy is a particular quantum generalization of the *Kullback–Leibler divergence* or classical relative entropy, which for two probability distributions $p, q : \mathcal{X} \to [0, 1]$ defined on a finite alphabet $\mathcal{X}$ is given by

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) \log_2\left(\frac{p(x)}{q(x)}\right). \tag{7.2.2}$$

The quantum relative entropy has an operational meaning in terms of the task of quantum hypothesis testing, as is shown in Section 7.10 on the quantum Stein's lemma (Theorem 7.78). The quantum relative entropy $D(\rho\|\sigma)$ can also be interpreted as a distinguishability measure for the quantum states ρ and σ, in part due to the facts that $D(\rho\|\sigma) \geq 0$ and $D(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$, which is shown in Proposition 7.3 below.

The support condition $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ in the definition of the quantum relative entropy essentially has to do with the term $\mathrm{Tr}[\rho \log_2 \sigma]$ and the fact that the logarithm of an operator is really only well defined for positive definite operators, while we allow for σ to be positive semi-definite, which means that it could have some eigenvalues equal to zero. Recall that the expression $\rho \log_2 \rho$ is well defined even for states with eigenvalues equal to zero since we set $0 \log_2 0 = 0$. We justified this with the fact that $\lim_{x \to 0^+} x \log_2 x = 0$. We can similarly make sense of the support condition in the definition of the quantum relative entropy by using the following fact.

Proposition 7.2

For every state ρ and positive semi-definite operator σ,

$$D(\rho\|\sigma) = \lim_{\varepsilon \to 0^+} \mathrm{Tr}[\rho(\log_2 \rho - \log_2(\sigma + \varepsilon \mathbb{1}))]. \tag{7.2.3}$$

Consequently, whenever σ does not have full support (i.e., it is positive semi-definite as opposed to positive definite), we can write $D(\rho\|\sigma)$ as the following limit:

$$D(\rho\|\sigma) = \lim_{\varepsilon \to 0^+} D(\rho\|\sigma + \varepsilon \mathbb{1}). \tag{7.2.4}$$

Proof: Observe that, for all $\varepsilon > 0$, the operator $\sigma + \varepsilon\mathbb{1}$ has full support; i.e., $\text{supp}(\sigma + \varepsilon\mathbb{1}) = \mathcal{H}$ for all $\varepsilon > 0$, where $\mathcal{H}$ is the underlying Hilbert space. This means that the quantity

$$\text{Tr}[\rho \log_2(\sigma + \varepsilon\mathbb{1})] \tag{7.2.5}$$

is finite for all $\varepsilon > 0$. Now, let us decompose the Hilbert space $\mathcal{H}$ into the direct sum of the orthogonal subspaces $\text{supp}(\sigma)$ and $\ker(\sigma)$, so that $\mathcal{H} = \text{supp}(\sigma) \oplus \ker(\sigma)$. Let Π_σ be the projection onto $\text{supp}(\sigma)$ and $\Pi_\sigma^\perp$ the projection onto $\ker(\sigma)$. Then with respect to this decomposition, the operators ρ and σ can be written as the block matrices

$$\begin{aligned} \rho &= \begin{pmatrix} \Pi_\sigma\rho\Pi_\sigma & \Pi_\sigma\rho\Pi_\sigma^\perp \\ \Pi_\sigma^\perp\rho\Pi_\sigma & \Pi_\sigma^\perp\rho\Pi_\sigma^\perp \end{pmatrix} \equiv \begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^\dagger & \rho_{1,1} \end{pmatrix}, \\ \sigma &= \begin{pmatrix} \Pi_\sigma\sigma\Pi_\sigma & \Pi_\sigma\sigma\Pi_\sigma^\perp \\ \Pi_\sigma^\perp\sigma\Pi_\sigma & \Pi_\sigma^\perp\sigma\Pi_\sigma^\perp \end{pmatrix} = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{7.2.6}$$

Now, let us first suppose that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. This means that $\rho_{0,0} = \rho$ and that $\rho_{0,1} = 0$ and $\rho_{1,1} = 0$. Using the fact that $\mathbb{1}_\mathcal{H} = \begin{pmatrix} \Pi_\sigma & 0 \\ 0 & \Pi_\sigma^\perp \end{pmatrix}$, we can write the second term in (7.2.3) as

$$\text{Tr}[\rho \log_2(\sigma + \varepsilon\mathbb{1})] = \text{Tr}\left[\begin{pmatrix} \rho_{0,0} & 0 \\ 0 & 0 \end{pmatrix} \log_2 \begin{pmatrix} \sigma + \varepsilon\Pi_\sigma & 0 \\ 0 & \varepsilon\Pi_\sigma^\perp \end{pmatrix}\right] \tag{7.2.7}$$

$$= \text{Tr}\left[\begin{pmatrix} \rho_{0,0} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \log_2(\sigma + \varepsilon\Pi_\sigma) & 0 \\ 0 & \log_2(\varepsilon\Pi_\sigma^\perp) \end{pmatrix}\right] \tag{7.2.8}$$

$$= \text{Tr}[\rho_{0,0} \log_2(\sigma + \varepsilon\Pi_\sigma)]. \tag{7.2.9}$$

Therefore,

$$\lim_{\varepsilon\to 0^+} \text{Tr}[\rho \log_2(\sigma + \varepsilon\mathbb{1})] = \text{Tr}[\rho_{0,0} \log_2(\sigma + \varepsilon\Pi_\sigma)] = \text{Tr}[\rho \log_2 \sigma], \tag{7.2.10}$$

which means that

$$\lim_{\varepsilon\to 0^+} \text{Tr}[\rho(\log_2 \rho - \log_2(\sigma + \varepsilon\mathbb{1}))] = \lim_{\varepsilon\to 0^+} \text{Tr}[\rho(\log_2 \rho - \log_2 \sigma)] \tag{7.2.11}$$

whenever $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$.

If $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$, then the block $\rho_{1,1}$ of ρ is non-zero (and the block $\rho_{0,1}$ could be non-zero), and we obtain

$$\text{Tr}[\rho \log_2(\sigma + \varepsilon\mathbb{1})] = \text{Tr}\left[\begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^\dagger & \rho_{1,1} \end{pmatrix} \begin{pmatrix} \log_2(\sigma + \varepsilon\Pi_\sigma) & 0 \\ 0 & \log_2(\varepsilon\Pi_\sigma^\perp) \end{pmatrix}\right] \tag{7.2.12}$$

$$= \text{Tr}[\rho_{0,0} \log_2(\sigma + \varepsilon\Pi_\sigma)] + \text{Tr}[\rho_{1,1} \log_2(\varepsilon\Pi_\sigma^\perp)] \quad (7.2.13)$$
$$= \text{Tr}[\rho_{0,0} \log_2(\sigma + \varepsilon\Pi_\sigma)] + \log_2(\varepsilon)\text{Tr}[\rho_{1,1}\Pi_\sigma^\perp]. \quad (7.2.14)$$

Then, using the fact that $\lim_{\varepsilon\to 0^+}(-\log_2 \varepsilon) = +\infty$, we find that

$$\lim_{\varepsilon\to 0^+} \text{Tr}[\rho(\log_2 \rho - \log_2(\sigma + \varepsilon\mathbb{1}))] = \text{Tr}[\rho \log_2 \rho] \\ - \lim_{\varepsilon\to 0^+} \text{Tr}[\rho_{0,0} \log_2(\sigma + \varepsilon\Pi_\sigma)] - \log_2(\varepsilon) \lim_{\varepsilon\to 0^+} \text{Tr}[\rho_{1,1}\Pi_\sigma^\perp] = +\infty. \quad (7.2.15)$$

We thus conclude that

$$\lim_{\varepsilon\to 0^+} \text{Tr}[\rho(\log_2 \rho - \log_2(\sigma + \varepsilon\mathbb{1}))] \\ = \begin{cases} \text{Tr}[\rho(\log_2 \rho - \log_2 \sigma)] & \text{if } \text{supp}(\rho) \subseteq \text{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \quad (7.2.16) \\ = D(\rho\|\sigma),$$

as required. ■

The proposition above, in particular the fact that we can write the quantum relative entropy as $D(\rho\|\sigma) = \lim_{\varepsilon\to 0+} D(\rho\|\sigma+\varepsilon\mathbb{1})$, allows us to take the logarithm $\log_2(\sigma)$ of σ on only the support of σ when determining the quantum relative entropy. We can use this fact to write another formula for the quantum relative entropy. We start by writing a spectral decomposition of the state ρ and positive semi-definite operator σ. In particular, setting $r_\rho \equiv \text{rank}(\rho)$ and $r_\sigma \equiv \text{rank}(\sigma)$, let

$$\rho = \sum_{j=1}^{d} p_j |\psi_j\rangle\langle\psi_j| = \sum_{j=1}^{r_\rho} p_j |\psi_j\rangle\langle\psi_j|, \quad (7.2.17)$$
$$\sigma = \sum_{k=1}^{d} q_k |\phi_k\rangle\langle\phi_k| = \sum_{k=1}^{r_\sigma} q_k |\phi_k\rangle\langle\phi_k|, \quad (7.2.18)$$

be spectral decompositions of ρ and σ, where $d = \dim(\mathcal{H})$ and in the second equality we have restricted the sum to only those eigenvalues that are non-zero. Then,

$$D(\rho\|\sigma) = \sum_{j=1}^{r_\rho} p_j \log_2 p_j$$

$$- \operatorname{Tr}\left[\left(\sum_{j=1}^{r_\rho} p_j |\psi_j\rangle\langle\psi_j|\right)\left(\sum_{k=1}^{r_\sigma} \log_2 q_k |\phi_k\rangle\langle\phi_k|\right)\right] \tag{7.2.19}$$

$$= \sum_{j=1}^{r_\rho} p_j \log_2 p_j - \sum_{j=1}^{r_\rho}\sum_{k=1}^{r_\sigma} \left|\langle\psi_j|\phi_k\rangle\right|^2 p_j \log_2 q_k \tag{7.2.20}$$

$$= \sum_{j=1}^{r_\rho}\left[p_j \log_2 p_j - \sum_{k=1}^{r_\sigma} \left|\langle\psi_j|\phi_k\rangle\right|^2 p_j \log_2 q_k\right]. \tag{7.2.21}$$

Now, using the fact that the eigenvectors $\{|\phi_k\rangle : 1 \leq k \leq d\}$ form a complete orthonormal basis for $\mathcal{H}$, so that $\mathbb{1}_{\mathcal{H}} = \sum_{k=1}^{d} |\phi_k\rangle\langle\phi_k|$, we conclude that

$$1 = \langle\psi_j|\psi_j\rangle = \sum_{k=1}^{d}\langle\psi_j|\phi_k\rangle\langle\phi_k|\psi_j\rangle = \sum_{k=1}^{r_\sigma}\left|\langle\psi_j|\phi_k\rangle\right|^2, \tag{7.2.22}$$

for $1 \leq j \leq r_\rho$, where the last equality follows from the assumption that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, which implies that the eigenvectors $\{|\psi_j\rangle : 1 \leq j \leq r_\rho\}$ of ρ can be expressed as a linear combination of the eigenvectors $\{|\phi_k\rangle : 1 \leq k \leq r_\sigma\}$ of σ. Therefore,

$$\boxed{D(\rho\|\sigma) = \sum_{j=1}^{r_\rho}\sum_{k=1}^{r_\sigma}\left|\langle\psi_j|\phi_k\rangle\right|^2 p_j \log_2\left(\frac{p_j}{q_k}\right),} \tag{7.2.23}$$

whenever $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$.

We now detail some important mathematical properties of the quantum relative entropy.

Proposition 7.3 Basic Properties of the Quantum Relative Entropy

The quantum relative entropy satisfies the following properties for all states ρ, ρ_1, ρ_2 and positive semi-definite operators $\sigma, \sigma_1, \sigma_2$:

1. *Isometric invariance*: For every isometry V,

$$D(V\rho V^\dagger \| V\sigma V^\dagger) = D(\rho\|\sigma). \tag{7.2.24}$$

2. (a) *Klein's inequality*: If $\operatorname{Tr}(\sigma) \leq 1$, then $D(\rho\|\sigma) \geq 0$.

   (b) *Faithfulness*: $D(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.

(c) If $\rho \leq \sigma$, then $D(\rho\|\sigma) \leq 0$.

(d) If $\sigma \leq \sigma'$, then $D(\rho\|\sigma) \geq D(\rho\|\sigma')$.

3. *Additivity*:

$$D(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) = D(\rho_1\|\sigma_1) + D(\rho_2\|\sigma_2). \tag{7.2.25}$$

As a special case, for all $\beta \in (0, \infty)$,

$$D(\rho\|\beta\sigma) = D(\rho\|\sigma) + \log_2\left(\frac{1}{\beta}\right). \tag{7.2.26}$$

4. *Direct-sum property*: Let $p : \mathcal{X} \to [0, 1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $q : \mathcal{X} \to [0, \infty)$ be a non-negative function on $\mathcal{X}$. Let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$D(\rho_{XA}\|\sigma_{XA}) = D(p\|q) + \sum_{x\in\mathcal{X}} p(x) D(\rho_A^x\|\sigma_A^x). \tag{7.2.27}$$

where

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.2.28}$$

$$\sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.2.29}$$

Remark: If we let the first argument of the relative entropy be a general positive semi-definite operator instead of just a state, then (7.2.26) can be generalized for every $\alpha, \beta \in (0, \infty)$ as

$$D(\alpha\rho\|\beta\sigma) = \alpha D(\rho\|\sigma) + \alpha \log_2\left(\frac{\alpha}{\beta}\right). \tag{7.2.30}$$

Proof:

1. *Proof of isometric invariance*: When $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$, there is nothing to prove because $\operatorname{supp}(V\rho V^\dagger) \not\subseteq \operatorname{supp}(V\sigma V^\dagger)$, which means that both

$D(V\rho V^\dagger \| V\sigma V^\dagger)$ and $D(\rho\|\sigma)$ are equal to $+\infty$.

Suppose that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, which implies that $\text{supp}(V\rho V^\dagger) \subseteq \text{supp}(V\sigma V^\dagger)$. Let ρ and σ have the spectral decompositions given in (7.2.17) and (7.2.18), respectively. Using the formula in (7.2.23), we find that

$$\begin{aligned} D(V\rho V^\dagger \| V\sigma V^\dagger) &= \sum_{j=1}^{r_\rho} \sum_{k=1}^{r_\sigma} \left|\langle\psi_j|V^\dagger V|\phi_k\rangle\right|^2 p_j \log_2\left(\frac{p_j}{q_k}\right) \\ &= \sum_{j=1}^{r_\rho} \sum_{k=1}^{r_\sigma} \left|\langle\psi_j|\phi_k\rangle\right|^2 p_j \log_2\left(\frac{p_j}{q_k}\right) \\ &= D(\rho\|\sigma). \end{aligned} \tag{7.2.31}$$

We used the fact that V is an isometry, i.e., satisfying $V^\dagger V = \mathbb{1}$.

2. (a) *Proof of Klein's inequality*: The result is trivial in the case that $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, and so we assume that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. We can write the relative entropy as in (7.2.21),

$$D(\rho\|\sigma) = \sum_{j=1}^{d}\left[p_j \log_2 p_j - \sum_{k=1}^{d} \left|\langle\psi_j|\phi_k\rangle\right|^2 p_j \log_2 q_k\right], \tag{7.2.32}$$

where we have extended the sums to include all terms up to $d = \dim(\mathcal{H})$. Now, let $c_{j,k} \equiv \left|\langle\psi_j|\phi_k\rangle\right|^2$, and observe that

$$c_{j,k} \geq 0 \quad \forall\, 1 \leq j,k \leq d, \quad \text{and} \quad \sum_{k=1}^{d} c_{j,k} = 1 \quad \forall\, 1 \leq j \leq d. \tag{7.2.33}$$

The latter indeed holds because

$$\begin{aligned} \sum_{k=1}^{d} c_{j,k} &= \sum_{k=1}^{d} \langle\psi_j|\phi_k\rangle\langle\phi_k|\psi_j\rangle && (7.2.34) \\ &= \langle\psi_j| \underbrace{\left(\sum_{k=1}^{d} |\phi_k\rangle\!\langle\phi_k|\right)}_{\mathbb{1}} |\psi_j\rangle && (7.2.35) \\ &= \langle\psi_j|\psi_j\rangle && (7.2.36) \\ &= 1. && (7.2.37) \end{aligned}$$

Therefore, for each j, the set $\{c_{j,k} : 1 \leq k \leq d\}$ constitutes a probability distribution over k. Using the concavity of the function $\log_2$, we thus obtain

$$\sum_{k=1}^{d} c_{j,k} \log_2(q_k) \leq \log_2\left(\sum_{k=1}^{d} c_{j,k} q_k\right) = \log_2(r_j) \tag{7.2.38}$$

for all $1 \leq j \leq d$, where

$$r_j \coloneqq \sum_{k=1}^{d} c_{j,k} q_k. \tag{7.2.39}$$

Therefore, we obtain

$$D(\rho\|\sigma) = \sum_{j=1}^{d} p_j \log_2(p_j) - \sum_{j=1}^{d} p_j \left(\sum_{k=1}^{d} c_{j,k} \log_2(q_k)\right) \tag{7.2.40}$$

$$\geq \sum_{j=1}^{d} p_j \log_2\left(\frac{p_j}{r_j}\right) \tag{7.2.41}$$

$$= -\sum_{j=1}^{d} p_j \log_2\left(\frac{r_j}{p_j}\right). \tag{7.2.42}$$

Now, we make use of the fact that

$$-\log_2(x) \geq \frac{1-x}{\ln(2)} \quad \forall x > 0, \tag{7.2.43}$$

with equality if and only if $x = 1$. This fact can be readily verified by elementary calculus. Using this and (7.2.40)–(7.2.42), we obtain

$$D(\rho\|\sigma) \geq \frac{1}{\ln(2)} \sum_{j=1}^{d} p_j \left(1 - \frac{r_j}{p_j}\right) \tag{7.2.44}$$

$$= \frac{1}{\ln(2)} \sum_{j=1}^{d} p_j - \frac{1}{\ln(2)} \sum_{j=1}^{d} r_j. \tag{7.2.45}$$

Now, $\sum_{j=1}^{d} p_j = \operatorname{Tr}(\rho) = 1$, and since

$$\sum_{j=1}^{d} c_{j,k} = \langle\phi_k| \underbrace{\left(\sum_{j=1}^{d} |\psi_j\rangle\!\langle\psi_j|\right)}_{\mathbb{1}} |\phi_k\rangle = 1, \tag{7.2.46}$$

we obtain

$$\sum_{j=1}^{d} r_j = \sum_{k=1}^{d} q_k = \mathrm{Tr}(\sigma). \tag{7.2.47}$$

Therefore,

$$D(\rho\|\sigma) \geq \frac{1}{\ln(2)}(1 - \mathrm{Tr}(\sigma)) \geq 0, \tag{7.2.48}$$

as required, where the last inequality holds by the assumption that $\mathrm{Tr}(\sigma) \leq 1$.

(b) We are now interested in the case of equality in the statement $D(\rho\|\sigma) \geq 0$ that we just proved in (a). In that proof, we made use of two inequalities. The first was in (7.2.38), where we made use of the concavity of the logarithm. Equality holds in (7.2.38) if and only if for each j there exists k such that $c_{j,k} = 1$. The second inequality we used was in (7.2.43), where equality holds if and only if $x = 1$. Therefore, equality holds in (7.2.44) if and only if $p_j = r_j$ for all j, and equality in (7.2.38) is true if and only if the eigenvectors of ρ and σ are, up to relabeling, the same. Therefore $D(\rho\|\sigma) = 0$ if and only if $p_j = q_j$ for all j and the corresponding eigenvectors are (up to relabeling) equal, which is true if and only if $\rho = \sigma$.

(c) Suppose that both ρ and σ are positive definite. Since the logarithm is operator monotone (see Section 2.2.8.1), the operator inequality $\rho \leq \sigma$ implies that $\log_2(\rho) \leq \log_2(\sigma)$. This implies the inequality $\rho^{\frac{1}{2}} \log_2(\rho)\rho^{\frac{1}{2}} \leq \rho^{\frac{1}{2}} \log_2(\sigma)\rho^{\frac{1}{2}}$, which implies that $\mathrm{Tr}[\rho \log_2(\rho)] \leq \mathrm{Tr}[\rho \log_2(\sigma)]$, proving the result. In the case that ρ and/or σ are not positive definite, we first apply the result to the positive definite state $(1-\delta)\rho+\delta\pi$ and the positive definite operator $\sigma + \varepsilon\mathbb{1}$, with $\delta, \varepsilon > 0$, so that $D((1-\delta)\rho + \delta\pi\|\sigma + \varepsilon\mathbb{1}) \leq 0$. Then, using

$$\lim_{\varepsilon\to 0^+} \lim_{\delta\to 0^+} D((1-\delta)\rho + \delta\pi\|\sigma + \varepsilon\mathbb{1}) = D(\rho\|\sigma), \tag{7.2.49}$$

we obtain the desired result.

(d) As in (c), first suppose that ρ, σ, and σ' are positive definite. Since the logarithm is operator monotone, the operator inequality $\sigma' \geq \sigma$ implies that $\log_2(\sigma') \geq \log_2(\sigma)$, which implies that $\mathrm{Tr}[\rho \log_2(\sigma')] \geq \mathrm{Tr}[\rho \log_2 \sigma]$. Therefore,

$$D(\rho\|\sigma) = \mathrm{Tr}[\rho(\log_2 \rho - \log_2 \sigma)] \tag{7.2.50}$$

$$\geq \text{Tr}[\rho(\log_2 \rho - \log_2 \sigma')] \tag{7.2.51}$$
$$= D(\rho\|\sigma'), \tag{7.2.52}$$

as required. In the general case that the operators are not positive definite, as in (c) we apply the result to the positive definite operators $(1-\delta)\rho+\delta\pi$, $\sigma+\varepsilon\mathbb{1}$, and $\sigma'+\varepsilon'\mathbb{1}$, for $\delta, \varepsilon, \varepsilon' > 0$, and then use (7.2.49) to obtain the result.

3. *Proof of additivity*: Since $\text{supp}(\rho_1\otimes\rho_2) = \text{supp}(\rho_1)\otimes\text{supp}(\rho_2)$ and $\text{supp}(\sigma_1\otimes\sigma_2) = \text{supp}(\sigma_1)\otimes\text{supp}(\sigma_2)$, the condition $\text{supp}(\rho_1\otimes\rho_2) \nsubseteq \text{supp}(\sigma_1\otimes\sigma_2)$ is equivalent to the condition $\text{supp}(\rho_1) \nsubseteq \text{supp}(\sigma_1)$ or $\text{supp}(\rho_2) \nsubseteq \text{supp}(\sigma_2)$. Therefore, $D(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) = +\infty$ and $D(\rho_1\|\sigma_1) = +\infty$ or $D(\rho_2\|\sigma_2) = +\infty$ if one of the support conditions is violated. Now suppose that $\text{supp}(\rho_1\otimes\rho_2) \subseteq \text{supp}(\sigma_1\otimes\sigma_2)$. Letting ρ_1 and ρ_2 have spectral decompositions

$$\rho_1 = \sum_{j=1}^{d} p_j^1 |\psi_j^1\rangle\langle\psi_j^1|, \quad \rho_2 = \sum_{k=1}^{d} p_k^2 |\psi_k^2\rangle\langle\psi_k^2|, \tag{7.2.53}$$

we find that

$$\log_2(\rho_1\otimes\rho_2) = \log_2\left(\sum_{j,k=1}^{d} p_j^1|\psi_j^1\rangle\langle\psi_j^1| \otimes p_k^2|\psi_k^2\rangle\langle\psi_k^2|\right) \tag{7.2.54}$$
$$= \sum_{j,k=1}^{d} \log_2(p_j^1 p_k^2)|\psi_j^1\rangle\langle\psi_j^1| \otimes |\psi_k^2\rangle\langle\psi_k^2| \tag{7.2.55}$$
$$= \sum_{j,k=1}^{d} \log_2(p_j^1)|\psi_j^1\rangle\langle\psi_j^1| \otimes |\psi_k^2\rangle\langle\psi_k^2| \tag{7.2.56}$$
$$+ \sum_{j,k=1}^{d} \log_2(p_k^2)|\psi_j^1\rangle\langle\psi_j^1| \otimes |\psi_k^2\rangle\langle\psi_k^2| \tag{7.2.57}$$
$$= \log_2(\rho_1)\otimes\mathbb{1} + \mathbb{1}\otimes\log_2(\rho_2). \tag{7.2.58}$$

Similarly, $\log_2(\sigma_1\otimes\sigma_2) = \log_2(\sigma_1)\otimes\mathbb{1} + \mathbb{1}\otimes\log_2(\sigma_2)$. Therefore,

$$\begin{aligned} &D(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) \\ &= \text{Tr}\big[(\rho_1\otimes\rho_2)(\log_2(\rho_1)\otimes\mathbb{1} + \mathbb{1}\otimes\log_2(\rho_2)\big] \\ &\qquad - \text{Tr}\big[(\rho_1\otimes\rho_2)(\log_2(\sigma_1)\otimes\mathbb{1} + \mathbb{1}\otimes\log_2(\sigma_2)\big] \end{aligned} \tag{7.2.59}$$

$$= \operatorname{Tr}(\rho_2)\left(\operatorname{Tr}[\rho_1 \log_2(\rho_1)] - \operatorname{Tr}[\rho_1 \log_2(\sigma_1)]\right)$$
$$- \operatorname{Tr}(\rho_1)\left(\operatorname{Tr}[\rho_2 \log_2(\rho_2)] - \operatorname{Tr}[\rho_2 \log_2(\sigma_2)]\right) \tag{7.2.60}$$
$$= D(\rho_1\|\sigma_1) + D(\rho_2\|\sigma_2). \tag{7.2.61}$$

Then, to see (7.2.26), let $\rho = \rho_1$, $\alpha = \rho_2 = 1$, $\sigma = \sigma_1$, and $\beta = \sigma_2$. Recognizing that the tensor product with a scalar is just multiplication by the scalar, we find that

$$\begin{aligned} D(\rho\|\beta\sigma) &= D(\rho\|\sigma) + D(1\|\beta) \\ &= D(\rho\|\sigma) + (\log_2(1) - \log_2\beta) \\ &= D(\rho\|\sigma) + \log_2\left(\frac{1}{\beta}\right). \end{aligned} \tag{7.2.62}$$

4. *Proof of the direct-sum property*: Define the classical–quantum operators

$$\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \quad \sigma_{XA} = \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.2.63}$$

Observe that

$$\log_2 \rho_{XA} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2(p(x)\rho_A^x) \tag{7.2.64}$$
$$= \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2 p(x)\mathbb{1}_A + \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2 \rho_A^x, \tag{7.2.65}$$
$$\log_2 \sigma_{XA} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2(q(x)\sigma_A^x) \tag{7.2.66}$$
$$= \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2 q(x)\mathbb{1}_A + \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \log_2 \sigma_A^x. \tag{7.2.67}$$

Then, in the case $\operatorname{supp}(\rho_{XA}) \subseteq \operatorname{supp}(\sigma_{XA})$, we obtain

$$D(\rho_{XA}\|\sigma_{XA}) = \operatorname{Tr}[\rho_{XA}\log_2\rho_{XA}] - \operatorname{Tr}[\rho_{XA}\log_2\sigma_{XA}] \tag{7.2.68}$$
$$= \operatorname{Tr}\left[\sum_{X\in\mathcal{X}} p(x)\log_2(p(x))|x\rangle\langle x|_X \otimes \rho_A^x \right.$$
$$\left. + \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \log_2 \rho_A^x\right]$$
$$- \operatorname{Tr}\left[\sum_{x\in\mathcal{X}} p(x)\log_2(q(x))|x\rangle\langle x|_X \otimes \rho_A^x\right.$$

$$+\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \log_2 \sigma_A^x \Bigg] \tag{7.2.69}$$

$$= \sum_{x\in\mathcal{X}} \left[p(x)\log_2 p(x) - p(x)\log_2 q(x)\right]$$
$$+ \sum_{x\in\mathcal{X}} p(x)\mathrm{Tr}\left[\rho_A^x \log_2 \rho_A^x - \rho_A^x \log_2 \sigma_A^x\right] \tag{7.2.70}$$

$$= D(p\|q) + \sum_{x\in\mathcal{X}} p(x) D(\rho_A^x\|\sigma_A^x), \tag{7.2.71}$$

as required. ■

An important consequence of Klein's inequality from the proposition above is that

$$\boxed{\begin{array}{l} D(\rho\|\sigma) \geq 0 \text{ for all states } \rho, \sigma, \text{ and} \\ D(\rho\|\sigma) = 0 \text{ if and only if } \rho = \sigma. \end{array}} \tag{7.2.72}$$

This allows us to use the quantum relative entropy as a distinguishability measure for quantum states. We emphasize, however, that the quantum relative entropy is not a metric in the mathematical sense since it is neither symmetric in its two arguments nor does it satisfy the triangle inequality.

We now come to one of the most important properties of the quantum relative entropy that is used frequently throughout this book: the *data-processing inequality*. It is also called the *monotonicity of the quantum relative entropy*.

Theorem 7.4 Data-Processing Inequality for Quantum Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then,

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.2.73}$$

In other words, the quantum relative entropy $D(\rho\|\sigma)$ can only decrease or stay the same if we apply the same quantum channel $\mathcal{N}$ to the states ρ and σ. When the quantum relative entropy is interpreted as a distinguishability measure on quantum states, the data-processing inequality tells us that the distinguishability of two quantum states cannot increase when we act on them with the same quantum channel; see Figure 7.1. We postpone the proof of the data-processing inequality

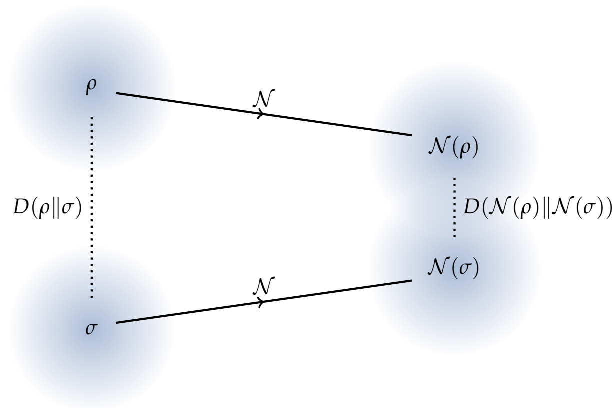


Figure 7.1: Illustration of the data-processing inequality for the quantum relative entropy (Theorem 7.4). The quantum states ρ, σ, $\mathcal{N}(\rho)$, and $\mathcal{N}(\sigma)$ are represented by spheres that signify the amount of "space" they occupy in the Hilbert space. While the states ρ and σ are nearly distinguishable as depicted, since their spheres do not overlap, after processing with the channel $\mathcal{N}$, the states become much less distinguishable because the spheres overlap significantly.

to later in the chapter, where it follows easily as a consequence of the data-processing inequality for the Petz–Rényi and sandwiched Rényi relative entropies (see Corollaries 7.25 and 7.34, respectively).

It is typically interesting and illuminating to investigate the conditions under which an important inequality is saturated. The data-processing inequality for quantum relative entropy is no exception. In the case that the action of the quantum channel can be reversed on ρ and σ, so that there exists a recovery channel $\mathcal{R}$ satisfying

$$\rho = (\mathcal{R} \circ \mathcal{N})(\rho), \qquad \sigma = (\mathcal{R} \circ \mathcal{N})(\sigma), \tag{7.2.74}$$

then it follows from an application of Theorem 7.4 that

$$\begin{aligned} D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) &\geq D((\mathcal{R} \circ \mathcal{N})(\rho)\|(\mathcal{R} \circ \mathcal{N})(\sigma)) && (7.2.75)\\ &= D(\rho\|\sigma). && (7.2.76) \end{aligned}$$

Thus, by combining with Theorem 7.4, we in fact have that

$$D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \tag{7.2.77}$$

so that the existence of a recovery channel implies saturation of the data-processing inequality in Theorem 7.4.

On the other hand, suppose that ρ, σ, and $\mathcal{N}$ are such that the inequality in Theorem 7.4 is saturated:

$$D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \tag{7.2.78}$$

Then it is a non-trivial result that there exists a recovery channel $\mathcal{R}$ such that the equality in (7.2.74) holds. In fact, this channel can be taken as the Petz recovery channel from Definition 4.21. We do not provide a proof here and instead point to the Bibliographic Notes in Section 7.13 for more details.

One of the remarkable aspects of the data-processing inequality for the quantum relative entropy is that it alone can be used to prove many of the properties of the quantum relative entropy stated in Proposition 7.3. For example, Klein's inequality follows by considering the trace channel Tr, so that for every state ρ and positive semi-definite operator σ such that $\mathrm{Tr}[\sigma] \leq 1$, we find that

$$D(\rho\|\sigma) \geq D(\mathrm{Tr}(\rho)\|\mathrm{Tr}(\sigma)) = \mathrm{Tr}(\rho)\log_2\left(\frac{\mathrm{Tr}(\rho)}{\mathrm{Tr}(\sigma)}\right) \tag{7.2.79}$$

$$= \log_2\left(\frac{1}{\mathrm{Tr}(\sigma)}\right) \geq 0. \tag{7.2.80}$$

Isometric invariance also follows from the data-processing inequality. The inequality $D(\rho\|\sigma) \geq D(V\rho V^\dagger\|V\sigma V^\dagger)$ follows from data processing because $(\cdot) \to V(\cdot)V^\dagger$ is a channel. The reverse inequality also follows from data processing because $D(V\rho V^\dagger\|V\sigma V^\dagger) \geq D(\mathcal{R}_V(V\rho V^\dagger)\|\mathcal{R}_V(V\sigma V^\dagger)) = D(\rho\|\sigma)$, where $\mathcal{R}_V$ is the reversal channel defined in (4.4.13) and we used (4.4.17)–(4.4.20).

Another important fact that follows from the data-processing inequality for quantum relative entropy is its *joint convexity*.

Proposition 7.5 Joint Convexity of Quantum Relative Entropy

Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$\sum_{x\in\mathcal{X}} p(x)D(\rho_A^x\|\sigma_A^x) \geq D\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right). \tag{7.2.81}$$

PROOF: Define the classical–quantum state and operator, respectively, as

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \quad \sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.2.82}$$

By the direct-sum property of the quantum relative entropy (Proposition 7.3), we find that

$$D(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)D(\rho_A^x\|\sigma_A^x). \tag{7.2.83}$$

Then, since we have $\rho_A = \text{Tr}_X[\rho_{XA}] = \sum_{x\in\mathcal{X}} p(x)\rho_A^x$ and $\sigma_A = \text{Tr}_X[\sigma_{XA}] = \sum_{x\in\mathcal{X}} p(x)\sigma_A^x$, we apply the data-processing inequality for quantum relative entropy with respect to the partial trace channel Tr_X to obtain

$$\sum_{x\in\mathcal{X}} p(x)D(\rho_A^x\|\sigma_A^x) = D(\rho_{XA}\|\sigma_{XA}) \tag{7.2.84}$$

$$\geq D(\text{Tr}_X(\rho_{XA})\|\text{Tr}_X(\sigma_{XA})) \tag{7.2.85}$$

$$= D\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right), \tag{7.2.86}$$

which is the desired joint convexity of quantum relative entropy. ■

Exercise 7.1

Let $\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$ and $\sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x$. Prove that

$$D(\rho_{XA}\|\sigma_{XA}) \geq D(p\|q) + D\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right). \tag{7.2.87}$$

7.2.1 Information Measures from Quantum Relative Entropy

As stated at the beginning of this chapter, the quantum relative entropy acts, as in the classical case, as a parent quantity for all of the fundamental information-theoretic quantities based on the quantum entropy. Indeed, using the properties of the quantum relative entropy stated previously, it is straightforward to verify the following:

1. The quantum entropy $H(\rho)$ of a state ρ is given by

$$H(\rho) = -D(\rho\|\mathbb{1}). \tag{7.2.88}$$

2. The quantum conditional entropy $H(A|B)_\rho$ of a bipartite state ρ_{AB} is given by

$$\begin{aligned} H(A|B)_\rho &= -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) && (7.2.89)\\ &= -\inf_{\sigma_B \in \mathrm{D}(\mathcal{H}_B)} D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B), && (7.2.90) \end{aligned}$$

and the coherent information $I(A\rangle B)_\rho$ of a bipartite state ρ_{AB} is given by

$$\begin{aligned} I(A\rangle B)_\rho &= D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) && (7.2.91)\\ &= \inf_{\sigma_B \in \mathrm{D}(\mathcal{H}_B)} D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B). && (7.2.92) \end{aligned}$$

Observe that

$$H(A|B)_\rho = -I(A\rangle B)_\rho \tag{7.2.93}$$

for every bipartite state ρ_{AB}. Similarly, we can write the reverse coherent information as

$$\begin{aligned} I(B\rangle A)_\rho &= D(\rho_{AB}\|\rho_A \otimes \mathbb{1}_B) && (7.2.94)\\ &= \inf_{\sigma_A \in \mathrm{D}(\mathcal{H}_A)} D(\rho_{AB}\|\sigma_A \otimes \mathbb{1}_B) && (7.2.95) \end{aligned}$$

3. The quantum mutual information $I(A;B)_\rho$ of a bipartite state ρ_{AB} is given by

$$\begin{aligned} I(A;B)_\rho &= D(\rho_{AB}\|\rho_A \otimes \rho_B) && (7.2.96)\\ &= \inf_{\sigma_B \in \mathrm{D}(\mathcal{H}_B)} D(\rho_{AB}\|\rho_A \otimes \sigma_B) && (7.2.97)\\ &= \inf_{\tau_A \in \mathrm{D}(\mathcal{H}_A)} D(\rho_{AB}\|\tau_A \otimes \rho_B) && (7.2.98)\\ &= \inf_{\substack{\tau_A \in \mathrm{D}(\mathcal{H}_A)\\ \sigma_B \in \mathrm{D}(\mathcal{H}_B)}} D(\rho_{AB}\|\tau_A \otimes \sigma_B). && (7.2.99) \end{aligned}$$

4. The quantum conditional mutual information $I(A;B|C)_\rho$ of a tripartite state ρ_{ABC} is given by

$$I(A;B|C)_\rho = D(\rho_{ABC}\|\sigma_{ABC}), \tag{7.2.100}$$

where

$$\sigma_{ABC} := 2^{\log_2(\rho_{AC}\otimes\mathbb{1}_B)+\log_2(\rho_{BC}\otimes\mathbb{1}_A)-\log_2(\mathbb{1}_{AB}\otimes\rho_C)}. \tag{7.2.101}$$

The expressions in (7.2.90), (7.2.92), and (7.2.97), in terms of an optimization of the conditional entropy, coherent information, and mutual information, respectively, are useful for defining information-theoretic quantities analogous to these ones in the context of generalized divergences, which we introduce in the next section.

Exercise 7.2

Verify the equalities in (7.2.90), (7.2.92), and (7.2.97). *Hint*: First prove that

$$D(\rho_{AB}\|\tau_A\otimes\rho_B)+D(\tau_A\otimes\rho_B\|\tau_A\otimes\sigma_B)=D(\rho_{AB}\|\tau_A\otimes\sigma_B), \quad (7.2.102)$$

for all states τ_A and σ_B. Then use the fact that

$$D(\tau_A\otimes\rho_B\|\tau_A\otimes\sigma_B)\geq 0 \quad (7.2.103)$$

which holds for all states τ_A and σ_B, by Klein's inequality as stated in (7.2.72). Set $\tau_A=\pi_A$ to prove (7.2.90) and (7.2.92), and set $\tau_A=\rho_A$ to prove (7.2.97).

The properties of the quantum relative entropy, such as the ones stated in Propositions 7.3 and 7.5, can be directly translated to properties of the derived information measures stated above. Some of these properties are used frequently throughout the book, and so we state them here for convenience. They are straightforward to verify using definitions and properties of the quantum relative entropy.

- *Additivity of the quantum entropy for product states ρ and τ*:

$$H(\rho\otimes\tau)=H(\rho)+H(\tau). \quad (7.2.104)$$

- *Isometric invariance of the quantum entropy for a state ρ and an isometry V*:

$$H(\rho)=H(V\rho V^\dagger). \quad (7.2.105)$$

- *Concavity of the quantum entropy*: The joint convexity of the quantum relative entropy, as stated in Proposition 7.5, and the identity in (7.2.88) imply that the quantum entropy is concave in its input: if $p:\mathcal{X}\to[0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$ and $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states on a system A, then

$$H\left(\sum_{x\in\mathcal{X}}p(x)\rho_A^x\right)\geq\sum_{x\in\mathcal{X}}p(x)H(\rho_A^x). \quad (7.2.106)$$

By taking a spectral decomposition of a state ρ, applying the concavity inequality above, and the fact that the entropy of a pure state is equal to zero, we conclude that the quantum entropy is non-negative for every state ρ:

$$H(\rho) \geq 0. \tag{7.2.107}$$

By employing the mixing property of the Heisenberg–Weyl unitaries from (3.2.97), the invariance of entropy under a unitary, its concavity, and the fact that the entropy of the maximally mixed state π_A is equal to $\log_2 d_A$, we conclude the following dimension bound for the entropy of a state of system A:

$$H(\rho) \leq \log_2 d_A. \tag{7.2.108}$$

- *Direct-sum property of the quantum entropy*: The direct-sum property of the quantum relative entropy (see Proposition 7.3) translates to the following for the quantum entropy: If $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X and $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states on a system A, then

$$H\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x\right) = H(p) + \sum_{x\in\mathcal{X}} p(x) H(\rho_A^x), \tag{7.2.109}$$

  where $H(p) \coloneqq -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)$ is the (classical) Shannon entropy of the probability distribution p.

- *Chain rule for conditional entropy*: For every state ρ_{ABC}, the following equality holds

$$H(AB|C)_\rho = H(A|C)_\rho + H(B|AC)_\rho. \tag{7.2.110}$$

  If the system C is trivial, so that the state is a bipartite state ρ_{AB}, then this equality reduces to the following:

$$H(AB)_\rho = H(A)_\rho + H(B|A)_\rho. \tag{7.2.111}$$

- *Chain rule for quantum mutual information*: For every state ρ_{ABC}, the following equality holds

$$I(A;BC)_\rho = I(A;B)_\rho + I(A;C|B)_\rho. \tag{7.2.112}$$

  We call this the *chain rule* because it can be interpreted as saying that the correlations between A and BC can be built up by first establishing correlations between A and B (signified by $I(A;B)_\rho$), then establishing correlations between A and C, given the correlations with B (signified by $I(A;C|B)_\rho$).

Exercise 7.3

Provide explicit proofs of the properties in (7.2.104)–(7.2.109), by following what is stated above.

Exercise 7.4

Verify the chain rules stated in (7.2.110) and (7.2.112). More generally, prove that $I(A;BC|D)_\rho = I(A;B|D)_\rho + I(A;C|BD)_\rho$ for a four-party state ρ_{ABCD}.

7.2.2 Quantum Conditional Mutual Information

We now develop in detail some properties of the quantum conditional mutual information that we use later in the book.

We start with the proof of the strong subadditivity property of the quantum entropy, which we recall from (7.1.11) is the statement that

$$I(A;B|C)_\rho = H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho \geq 0 \tag{7.2.113}$$

for every state ρ_{ABC}.

Theorem 7.6 Strong Subadditivity of Quantum Entropy

For every state ρ_{ABC}, the following inequality holds

$$I(A;B|C)_\rho \geq 0. \tag{7.2.114}$$

Proof: One way to prove this result is by means the data-processing inequality for the quantum relative entropy, along with the expression for the quantum conditional entropy in (7.2.89). We start by using the definition of the quantum conditional entropy to rewrite the quantum conditional mutual information defined in (7.1.11) as

$$I(A;B|C)_\rho = H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho \tag{7.2.115}$$
$$= H(B|C)_\rho - H(B|AC)_\rho. \tag{7.2.116}$$

Then, using the expression in (7.2.89) for the quantum conditional entropy in terms of the quantum relative entropy, we find that

$$I(A;B|C)_\rho = D(\rho_{ABC}\|\mathbb{1}_B\otimes\rho_{AC}) - D(\rho_{BC}\|\mathbb{1}_B\otimes\rho_C). \tag{7.2.117}$$

Finally, observe that, by the data-processing inequality for the quantum relative entropy with respect to the partial trace channel Tr_A, we obtain

$$D(\rho_{ABC}\|\mathbb{1}_B\otimes\rho_{AC}) \geq D(\mathrm{Tr}_A[\rho_{ABC}]\|\mathbb{1}_B\otimes\mathrm{Tr}_A[\rho_{AC}]) \tag{7.2.118}$$
$$= D(\rho_{BC}\|\mathbb{1}_B\otimes\rho_C), \tag{7.2.119}$$

which implies that $I(A;B|C)_\rho \geq 0$, as required. ■

Two direct consequences of strong subadditivity are that the conditional entropy is concave and non-negative for every separable state. We detail these properties below.

Corollary 7.7 Concavity of Conditional Entropy

The conditional entropy is concave, and the coherent information is convex:

$$H(A|B)_{\bar{\rho}} \geq \sum_{x\in\mathcal{X}} p(x)H(A|B)_{\rho^x}, \tag{7.2.120}$$
$$I(A\rangle B)_{\bar{\rho}} \leq \sum_{x\in\mathcal{X}} p(x)I(A\rangle B)_{\rho^x}, \tag{7.2.121}$$

where $\mathcal{X}$ is a finite alphabet, $p:\mathcal{X}\to[0,1]$ is a probability distribution, $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ is a set of bipartite states, and $\bar{\rho} \coloneqq \sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}$.

Proof: This follows directly by constructing the following classical–quantum state ρ_{XAB}:

$$\rho_{XAB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|\otimes\rho^x_{AB}, \tag{7.2.122}$$

applying strong subadditivity $H(A|B) - H(A|BX)_\rho = I(A;X|B)_\rho \geq 0$, and observing that

$$H(A|BX)_\rho = \sum_{x\in\mathcal{X}} p(x)H(A|B)_{\rho^x}. \tag{7.2.123}$$

By applying (7.2.93), we conclude from (7.2.120) that coherent information is convex. ■

Corollary 7.8 Non-Negativity of Conditional Entropy on Separable States

The conditional entropy $H(A|B)_\sigma$ is non-negative for every separable state σ_{AB}.

PROOF: Recall from Definition 3.5 that σ_{AB} is separable if it can be written as

$$\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\tau_A^x \otimes \omega_B^x, \tag{7.2.124}$$

where $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$ and $\{\tau_A^x\}_{x\in\mathcal{X}}$ and $\{\omega_B^x\}_{x\in\mathcal{X}}$ are sets of states. Defining the extension σ_{XAB} of σ_{AB} as

$$\sigma_{XAB} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \tau_A^x \otimes \omega_B^x, \tag{7.2.125}$$

we then conclude that $I(A;X|B)_\sigma \geq 0$, which implies the desired inequality:

$$H(A|B)_\sigma \geq H(A|BX)_\sigma \tag{7.2.126}$$

$$= \sum_{x\in\mathcal{X}} p(x)H(A|B)_{\tau^x\otimes\omega^x} \tag{7.2.127}$$

$$= \sum_{x\in\mathcal{X}} p(x)H(A)_{\tau^x} \geq 0. \tag{7.2.128}$$

This concludes the proof. ■

We now prove some other properties of the quantum conditional mutual information that recur throughout the book.

Proposition 7.9 Properties of Quantum Conditional Mutual Information

The quantum conditional mutual information has the following properties:

1. *Symmetry*: For every state ρ_{ABC}, we have $I(A;B|C)_\rho = I(B;A|C)_\rho$.
2. *Local entropy and dimension bounds*: For every state ρ_{ABC},

$$I(A;B|C)_\rho \leq 2\min\{H(A)_\rho, H(B)_\rho\} \leq 2\log_2(\min\{d_A, d_B\}). \tag{7.2.129}$$

Let σ_{XBC} be a classical–quantum state of the form

$$\sigma_{XBC} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{BC}, \tag{7.2.130}$$

where $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and $\{\rho^x_{BC}\}_{x\in\mathcal{X}}$ is a set of states. Then the following bounds hold

$$I(X;B|C)_\sigma \le H(X)_\sigma \le \log_2 |\mathcal{X}|. \tag{7.2.131}$$

3. *Product conditioning system*: For a state $\rho_{ABC} = \sigma_{AB} \otimes \tau_C$,

$$I(A;B|C)_\rho = I(A;B)_\sigma. \tag{7.2.132}$$

4. *Additivity*: For states $\rho_{A_1B_1C_1}$ and $\tau_{A_2B_2C_2}$, the following equality holds for the product state $\rho_{A_1B_1C_1} \otimes \tau_{A_2B_2C_2}$:

$$I(A_1A_2;B_1B_2|C_1C_2)_{\rho\otimes\tau} = I(A_1;B_1|C_1)_\rho + I(A_2;B_2|C_2)_\tau. \tag{7.2.133}$$

5. *Direct-sum property*: For the classical–quantum state

$$\sigma_{XABC} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{ABC} \tag{7.2.134}$$

the following equality holds

$$I(A;B|CX)_\sigma = \sum_{x\in\mathcal{X}} p(x) I(A;B|C)_{\rho^x}. \tag{7.2.135}$$

6. *Chain rule*: For every state $\rho_{AB_1B_2C}$,

$$I(A;B_1B_2|C)_\rho = I(A;B_1|C)_\rho + I(A;B_2|B_1C)_\rho \tag{7.2.136}$$

7. *Data-processing inequality for local channels*: For every state ρ_{ABC} and all local channels $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$, the following inequality holds

$$I(A;B|C)_\rho \ge I(A';B'|C)_\omega, \tag{7.2.137}$$

where $\omega_{A'B'C} := (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{ABC})$.

PROOF: We establish the various properties one by one.

1. Symmetry under exchange of A and B follows immediately from the definition.

2. Using the definition of conditional entropy, we can write $I(A;B|C)_\rho$ as

$$I(A;B|C)_\rho = H(AC)_\rho - H(C)_\rho + H(BC)_\rho - H(ABC)_\rho \tag{7.2.138}$$
$$= H(A|C)_\rho - H(A|BC)_\rho. \tag{7.2.139}$$

The inequality $H(A|C)_\rho \leq H(A)_\rho$ is a direct consequence of strong subadditivity, under a relabeling and with trivial conditioning system. Then, by the dimension bound from (7.2.108), it follows that $H(A)_\rho \leq \log_2 d_A$. Therefore,

$$H(A|C)_\rho \leq \log_2 d_A. \tag{7.2.140}$$

Now, let $|\psi\rangle_{ABCE}$ be a purification of ρ_{ABC}. Then, since ρ_{ABC} and ψ_E have the same spectrum, and since ρ_{BC} and ψ_{AE} have the same spectrum, we obtain

$$H(A|BC)_\rho = H(ABC)_\rho - H(BC)_\rho \tag{7.2.141}$$
$$= H(E)_\psi - H(AE)_\psi \tag{7.2.142}$$
$$= -H(A|E)_\psi \tag{7.2.143}$$
$$\geq -H(A) \tag{7.2.144}$$
$$\geq -\log_2 d_A, \tag{7.2.145}$$

where the first inequality follows from (7.2.140), and the second inequality, as before, from the fact that $H(A)_\rho \leq \log_2 d_A$ for every state ρ. Therefore,

$$H(A|BC)_\rho \geq -H(A)_\rho \geq -\log_2 d_A, \tag{7.2.146}$$

which means that $I(A;B|C)_\rho = H(A|C)_\rho - H(A|BC)_\rho \leq 2H(A)_\rho \leq 2\log_2 d_A$. By applying symmetry under exchange of A and B and the same argument, we conclude that $I(A;B|C)_\rho \leq 2H(B)_\rho \leq 2\log_2 d_B$. Thus, we conclude (7.2.129).

If the system A is classical (as in (7.2.130)), then the state is separable with respect to the bipartite cut $A|BC$. As such, the lower bound in (7.2.146) improves to $H(A|BC)_\rho \geq 0$ (as a consequence of Corollary 7.8), implying that the upper bound improves to $I(A;B|C)_\rho \leq H(A)_\rho \leq \log_2 d_A$ in this case.

3. For $\rho_{ABC} = \sigma_{AB} \otimes \tau_C$, we have

$$I(A;B|C)_\rho = H(A|C)_{\sigma\otimes\tau} + H(B|C)_{\sigma\otimes\tau} - H(AB|C)_{\sigma\otimes\tau}. \tag{7.2.147}$$

Now, we use the fact that

$$H(A|C)_{\sigma\otimes\tau} = H(A)_\sigma, \tag{7.2.148}$$

which follows from (7.2.104). This means that $H(B|C)_{\sigma\otimes\tau} = H(B)_\sigma$ and $H(AB|C)_{\sigma\otimes\tau} = H(AB)_\sigma$, and we find that

$$I(A;B|C)_\rho = H(A)_\sigma + H(B)_\sigma - H(AB)_\sigma = I(A;B)_\sigma. \tag{7.2.149}$$

4. By writing

$$\begin{aligned} I(A_1A_2;B_1B_2|C_1C_2)_{\rho\otimes\tau} = H(A_1A_2|C_1C_2)_{\rho\otimes\tau} + H(B_1B_2|C_1C_2)_{\rho\otimes\tau} \\ - H(A_1A_1B_1B_2|C_1C_2)_{\rho\otimes\tau}, \end{aligned} \tag{7.2.150}$$

and recognizing that each conditional entropy on the right-hand side is evaluated on a product state, we can use (7.2.104) to obtain the desired result. As an example, we evaluate the first term on the right-hand side:

$$\begin{aligned} &H(A_1A_2|C_1C_2)_{\rho\otimes\tau} \\ &\quad = H(A_1A_2C_1C_2)_{\rho\otimes\tau} - H(C_1C_2)_{\rho\otimes\tau} &(7.2.151)\\ &\quad = H(A_1C_1)_\rho + H(A_2C_2)_\tau - H(C_1)_\rho - H(C_2)_\tau &(7.2.152)\\ &\quad = H(A_1|C_1)_\rho + H(A_2|C_2)_\tau. &(7.2.153) \end{aligned}$$

5. By definition, we have that

$$I(A;B|CX)_\sigma = H(A|CX)_\sigma + H(B|CX)_\sigma - H(AB|CX)_\sigma. \tag{7.2.154}$$

Let us consider the first term on the right-hand side. Using the definition of the quantum conditional entropy and using the direct-sum property of the quantum entropy, as stated in (7.2.109), we obtain

$$\begin{aligned} &H(A|CX)_\sigma \\ &= H(ACX)_\sigma - H(CX)_\sigma &(7.2.155)\\ &= H(p) + \sum_{x\in\mathcal{X}} p(x)H(AC)_{\rho^x} - H(p) - \sum_{x\in\mathcal{X}} p(x)H(C)_{\rho^x} &(7.2.156)\\ &= \sum_{x\in\mathcal{X}} p(x)\left(H(AC)_{\rho^x} - H(C)_{\rho^x}\right) &(7.2.157)\\ &= \sum_{x\in\mathcal{X}} p(x)H(A|C)_{\rho^x}. &(7.2.158) \end{aligned}$$

The other terms on the right-hand side of (7.2.154) are evaluated similarly, and we ultimately arrive at

$$I(A;B|CX)_\sigma = \sum_{x\in\mathcal{X}} p(x)\left(H(A|C)_{\rho^x} + H(B|C)_{\rho^x} - H(AB|C)_{\rho^x}\right)$$
$$= \sum_{x\in\mathcal{X}} p(x) I(A;B|C)_{\rho^x}. \tag{7.2.159}$$

6. This follows straightforwardly by using the definition of the quantum conditional mutual information to expand both sides of (7.2.136) to confirm that they are equal to each other.

7. Let us first prove the following inequality,

$$I(A_1A_2;B_1B_2|C)_\rho \geq I(A_1;B_1|C)_\rho \tag{7.2.160}$$

for every state $\rho_{A_1A_2B_1B_2C}$. This inequality implies that discarding parts of the non-conditioning systems never increases the quantum conditional mutual information. Applying the chain rule in (7.2.136), along with strong subadditivity, we find that

$$I(A_1A_2;B_1B_2|C)_\rho = I(A_1A_2;B_1|C)_\rho + I(A_1A_2;B_2|B_1C)_\rho \tag{7.2.161}$$
$$\geq I(A_1A_2;B_1|C)_\rho, \tag{7.2.162}$$

where the last line follows from strong subadditivity, which implies that $I(A_1A_2;B_2|B_1C)_\rho \geq 0$. Applying the chain rule in (7.2.136) and strong subadditivity once again, we conclude that

$$I(A_1A_2;B_1|C)_\rho = I(A_1;B_1|C)_\rho + I(A_2;B_1|A_1C)_\rho \tag{7.2.163}$$
$$\geq I(A_1;B_1|C)_\rho. \tag{7.2.164}$$

So we have $I(A_1A_2;B_1B_2|C)_\rho \geq I(A_1;B_1|C)_\rho$.

Now, let $V_{A\to A'E_1}$ and $W_{B\to B'E_2}$ be isometric extensions of the channels $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$, respectively, with appropriate environment systems E_1 and E_2. Then, by (7.2.160) and the isometric invariance of the quantum entropy, we find that

$$I(A';B'|C)_\omega \leq I(A'E_1;B'E_2|C)_{(V\otimes W)\rho(V\otimes W)^\dagger} \tag{7.2.165}$$
$$= I(A;B|C)_\rho, \tag{7.2.166}$$

as required. ■

Proposition 7.10 Uniform Continuity of Conditional Mutual Information

Let ρ_{ABC} and σ_{ABC} be tripartite quantum states such that

$$\frac{1}{2}\left\|\rho_{ABC}-\sigma_{ABC}\right\|_1 \leq \varepsilon, \tag{7.2.167}$$

for $\varepsilon \in [0,1]$. Then the following bound applies to their conditional mutual informations:

$$\left|I(A;B|C)_\rho - I(A;B|C)_\sigma\right| \leq 2\varepsilon \log_2(\min\{d_A, d_B\}) + 2g_2(\varepsilon), \tag{7.2.168}$$

where $g_2(\varepsilon) := (\varepsilon+1)\log_2(\varepsilon+1) - \varepsilon\log_2\varepsilon$. If the system A is classical (so that both ρ_{ABC} and σ_{ABC} are classical–quantum–quantum states of the form in (7.2.130)), then the following bound holds

$$\left|I(A;B|C)_\rho - I(A;B|C)_\sigma\right| \leq \varepsilon \log_2 d_A + 2g_2(\varepsilon). \tag{7.2.169}$$

REMARK: See Section 2.3 for a definition of uniform continuity.

PROOF: Suppose without loss of generality that $\varepsilon > 0$ (otherwise the statement trivially holds). Let $\omega^\lambda_{ABC} := \lambda\rho_{ABC} + (1-\lambda)\,\sigma_{ABC}$ for $\lambda \in [0,1]$. Then the following inequality holds

$$\lambda I(A;B|C)_\rho + (1-\lambda)\, I(A;B|C)_\sigma \leq I(A;B|C)_{\omega^\lambda} + h_2(\lambda), \tag{7.2.170}$$

because for the classical–quantum state

$$\omega^\lambda_{ABCX} := \lambda\rho_{ABC}\otimes|0\rangle\langle 0|_X + (1-\lambda)\,\sigma_{ABC}\otimes|1\rangle\langle 1|_X, \tag{7.2.171}$$

we have that

$$\begin{aligned}
\lambda I(A;B|C)_\rho + (1-\lambda)\, I(A;B|C)_\sigma &= I(A;B|CX)_{\omega^\lambda} && (7.2.172)\\
&\leq I(AX;B|C)_{\omega^\lambda} && (7.2.173)\\
&= I(A;B|C)_{\omega^\lambda} + I(X;B|CA)_{\omega^\lambda} && (7.2.174)\\
&\leq I(A;B|C)_{\omega^\lambda} + H(X)_{\omega^\lambda} && (7.2.175)\\
&= I(A;B|C)_{\omega^\lambda} + h_2(\lambda). && (7.2.176)
\end{aligned}$$

The first equality follows from (7.2.135). The first inequality follows from the chain rule and strong subadditivity. The second equality follows from the chain rule. The second inequality follows from the local entropy bound in (7.2.131). We also have that

$$I(A;B|C)_{\omega^\lambda} \leq I(AX;B|C)_{\omega^\lambda} \tag{7.2.177}$$
$$= I(X;B|C)_{\omega^\lambda} + I(A;B|CX)_{\omega^\lambda} \tag{7.2.178}$$
$$\leq h_2(\lambda) + \lambda I(A;B|C)_\rho + (1-\lambda)\, I(A;B|C)_\sigma, \tag{7.2.179}$$

which together imply that

$$\left|\lambda I(A;B|C)_\rho + (1-\lambda)\, I(A;B|C)_\sigma - I(A;B|C)_{\omega^\lambda}\right| \leq h_2(\lambda). \tag{7.2.180}$$

Then consider the state

$$\zeta_{ABC} \coloneqq \frac{1}{1+\varepsilon}\left(\rho_{ABC} + [\sigma_{ABC} - \rho_{ABC}]_+\right), \tag{7.2.181}$$

where $[\cdot]_+$ denotes the positive part of an operator, and for this choice, we have that

$$\frac{1}{1+\varepsilon}\rho_{ABC} + \frac{\varepsilon}{1+\varepsilon}\xi^1_{ABC} = \zeta_{ABC} \tag{7.2.182}$$
$$= \frac{1}{1+\varepsilon}\sigma_{ABC} + \frac{\varepsilon}{1+\varepsilon}\xi^2_{ABC}, \tag{7.2.183}$$

analogous to the approach of Thales of Milete, where the states ξ^1_{ABC} and ξ^2_{ABC} are defined as

$$\xi^1_{ABC} \coloneqq \frac{1}{\varepsilon}\left[\sigma_{ABC} - \rho_{ABC}\right]_+, \tag{7.2.184}$$
$$\xi^2_{ABC} \coloneqq \frac{1}{\varepsilon}\left((1+\varepsilon)\zeta_{ABC} - \sigma_{ABC}\right). \tag{7.2.185}$$

Applying (7.2.180) to the convex decompositions above, we find that

$$\frac{1}{1+\varepsilon}\left(I(A;B|C)_\rho - I(A;B|C)_\sigma\right)$$
$$\leq \frac{\varepsilon}{1+\varepsilon}\left(I(A;B|C)_{\xi^2} - I(A;B|C)_{\xi^1}\right) + 2h_2\!\left(\frac{\varepsilon}{1+\varepsilon}\right) \tag{7.2.186}$$
$$\leq \frac{\varepsilon}{1+\varepsilon} I(A;B|C)_{\xi^2} + 2h_2\!\left(\frac{\varepsilon}{1+\varepsilon}\right) \tag{7.2.187}$$
$$\leq \frac{\varepsilon}{1+\varepsilon} 2\log(\min\{d_A, d_B\}) + 2h_2\!\left(\frac{\varepsilon}{1+\varepsilon}\right), \tag{7.2.188}$$

where the last line follows from the dimension bound in (7.2.129). Multiplying through by $1+\varepsilon$ and using the fact that $g_2(\varepsilon) = (1+\varepsilon)\, h_2\!\left(\frac{\varepsilon}{1+\varepsilon}\right)$, we conclude that

$$I(A;B|C)_\rho - I(A;B|C)_\sigma \leq 2\varepsilon \log_2(\min\{d_A, d_B\}) + 2g_2(\varepsilon). \tag{7.2.189}$$

To arrive at the other inequality, we again apply (7.2.180) to the convex decompositions above and find that

$$\frac{1}{1+\varepsilon}\left(I(A;B|C)_\sigma - I(A;B|C)_\rho\right) \\ \leq \frac{\varepsilon}{1+\varepsilon}\left(I(A;B|C)_{\xi^1} - I(A;B|C)_{\xi^2}\right) + 2h_2\!\left(\frac{\varepsilon}{1+\varepsilon}\right). \quad (7.2.190)$$

Then we apply the same reasoning as above to find that

$$I(A;B|C)_\sigma - I(A;B|C)_\rho \leq 2\varepsilon\log_2(\min\{d_A, d_B\}) + 2g_2(\varepsilon). \quad (7.2.191)$$

The inequality in (7.2.169) follows from the same proof, but applying observation that the A system of the states ξ^1_{ABC} and ξ^2_{ABC} in (7.2.184)–(7.2.185) are classical when ρ_{ABC} and σ_{ABC} are classical on A. Here, we also apply the dimension bound in (7.2.130) in (7.2.188) above. ∎

7.2.3 Quantum Mutual Information

The quantum mutual information of a bipartite state ρ_{AB} is a measure of all correlations between the A and B systems. We defined it previously in (7.1.8) and (7.2.96), and we recall its definition here:

$$I(A;B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho. \quad (7.2.192)$$

It can be understood as a special case of the conditional mutual information in which the C system is trivial (i.e., formally, a one-dimensional system that is in tensor product with ρ_{AB} and equal to the number one). As such, all of the properties of the conditional mutual information apply directly to mutual information, and we list them here for convenience:

Corollary 7.11 Non-Negativity of Quantum Mutual Information

For every state ρ_{AB}, the following inequality holds

$$I(A;B)_\rho \geq 0. \quad (7.2.193)$$

We can view non-negativity in (7.2.193) as a consequence of strong subadditivity in (7.2.114). However, the conclusion in (7.2.193) follows more easily as a consequence of non-negativity of quantum relative entropy for quantum states from (7.2.72) because

$$I(A;B)_\rho = D(\rho_{AB}\|\rho_A \otimes \rho_B) \geq 0. \tag{7.2.194}$$

Proposition 7.12 Properties of Quantum Mutual Information

The quantum mutual information has the following properties:

1. *Symmetry*: For every state ρ_{AB}, we have $I(A;B)_\rho = I(B;A)_\rho$.
2. *Local entropy and dimension bounds*: For every state ρ_{AB}, the following bounds hold

$$I(A;B)_\rho \leq 2\min\{H(A)_\rho, H(B)_\rho\} \tag{7.2.195}$$

$$\leq 2\log_2(\min\{d_A, d_B\}). \tag{7.2.196}$$

   Let σ_{XB} be a classical–quantum state of the form

$$\sigma_{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_B^x, \tag{7.2.197}$$

   where $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and $\{\rho_B^x\}_{x\in\mathcal{X}}$ is a set of states. Then the following bounds hold

$$I(X;B)_\sigma \leq H(X)_\sigma \leq \log_2|\mathcal{X}|. \tag{7.2.198}$$

3. *Additivity*: For states $\rho_{A_1B_1}$ and $\tau_{A_2B_2}$, the following equality holds for the product state $\rho_{A_1B_1} \otimes \tau_{A_2B_2}$:

$$I(A_1A_2;B_1B_2)_{\rho\otimes\tau} = I(A_1;B_1)_\rho + I(A_2;B_2)_\tau. \tag{7.2.199}$$

4. *Direct-sum property*: Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $\{\rho_{AB}^x\}_{x\in\mathcal{X}}$ be a set of states. Then, for the classical–quantum state

$$\sigma_{XAB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_{AB}^x \tag{7.2.200}$$

   the following equality holds

$$I(A;B|X)_\sigma = \sum_{x\in\mathcal{X}} p(x) I(A;B)_{\rho^x}. \tag{7.2.201}$$

5. *Data-processing inequality for local channels*: For every state ρ_{AB}, the quantum mutual information is non-increasing under the action of local channels. In other words, for every state ρ_{AB} and all local channels $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$, the following inequality holds

$$I(A;B)_\rho \geq I(A';B')_\omega, \tag{7.2.202}$$

where $\omega_{A'B'} \coloneqq (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{AB})$.

Proposition 7.13 Uniform Continuity of Mutual Information

Let ρ_{AB} and σ_{AB} be bipartite quantum states such that

$$\frac{1}{2}\left\|\rho_{AB} - \sigma_{AB}\right\|_1 \leq \varepsilon, \tag{7.2.203}$$

for $\varepsilon \in [0,1]$. Then the following bound applies to their mutual informations:

$$\left|I(A;B)_\rho - I(A;B)_\sigma\right| \leq 2\varepsilon \log_2(\min\{d_A, d_B\}) + 2g_2(\varepsilon), \tag{7.2.204}$$

where $g_2(\varepsilon) \coloneqq (\varepsilon+1)\log_2(\varepsilon+1) - \varepsilon\log_2\varepsilon$.

Proposition 7.14 Mutual Information of Classical–Quantum States

Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states. Then, for the classical–quantum state

$$\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \tag{7.2.205}$$

the following equalities hold

$$I(X;A)_\rho = H(\overline{\rho}_A) - \sum_{x\in\mathcal{X}} p(x)H(\rho_A^x) \tag{7.2.206}$$

$$= \sum_{x\in\mathcal{X}} p(x) D(\rho_A^x \| \overline{\rho}_A), \tag{7.2.207}$$

where $\overline{\rho}_A \coloneqq \sum_{x\in\mathcal{X}} p(x)\rho_A^x$.

REMARK: The mutual information $I(X;A)_\rho$ of a classical–quantum state ρ_{XA} is often called *Holevo information*, a term we use throughout this book.

PROOF: Consider from (7.2.96) that

$$I(X;A)_\rho = D(\rho_{XA} \| \rho_X \otimes \rho_A) \tag{7.2.208}$$

$$= \text{Tr}[\rho_{XA} \log_2 \rho_{XA}] - \text{Tr}[\rho_{XA} \log_2(\rho_X \otimes \rho_A)]. \tag{7.2.209}$$

Using (7.2.65) in the proof of Proposition 7.3, we find that

$$\log_2 \rho_{XA} = \sum_{x \in \mathcal{X}} |x\rangle\langle x|_X \otimes \log_2 p(x) \mathbb{1}_A + \sum_{x \in \mathcal{X}} |x\rangle\langle x|_X \otimes \log \rho_A^x, \tag{7.2.210}$$

which leads to

$$\text{Tr}[\rho_{XA} \log \rho_{XA}] = \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) + \sum_{x \in \mathcal{X}} p(x) \text{Tr}[\rho_A^x \log_2 \rho_A^x] \tag{7.2.211}$$

$$= -H(X) - \sum_{x \in \mathcal{X}} p(x) H(\rho_A^x). \tag{7.2.212}$$

Then, using $\log_2(\rho_X \otimes \rho_A) = \log_2 \rho_X \otimes \mathbb{1}_A + \mathbb{1}_X \otimes \log_2 \rho_A$, we conclude that

$$\text{Tr}[\rho_{XA} \log_2(\rho_X \otimes \rho_A)] = \text{Tr}[\rho_X \log \rho_X] + \text{Tr}[\rho_A \log_2 \rho_A] \tag{7.2.213}$$

$$= -H(X) - H(\rho_A). \tag{7.2.214}$$

However, $\rho_A = \overline{\rho}_A$, so that

$$I(X;A)_\rho = H(\overline{\rho}_A) - \sum_{x \in \mathcal{X}} p(x) H(\rho_A^x), \tag{7.2.215}$$

which is the statement in (7.2.206).

Now we prove (7.2.207). Starting from (7.2.206), consider that

$$H(\overline{\rho}_A) - \sum_{x \in \mathcal{X}} p(x) H(\rho_A^x)$$

$$= -\text{Tr}[\overline{\rho}_A \log_2 \overline{\rho}_A] + \sum_{x \in \mathcal{X}} p(x) \text{Tr}[\rho_A^x \log_2 \rho_A^x] \tag{7.2.216}$$

$$= -\sum_{x \in \mathcal{X}} p(x) \text{Tr}[\rho_A^x \log_2 \overline{\rho}_A] + \sum_{x \in \mathcal{X}} p(x) \text{Tr}[\rho_A^x \log_2 \rho_A^x] \tag{7.2.217}$$

$$= \sum_{x \in \mathcal{X}} p(x) \text{Tr}[\rho_A^x (\log_2 \rho_A^x - \log_2 \overline{\rho}_A)] \tag{7.2.218}$$

$$= \sum_{x \in \mathcal{X}} p(x) D(\rho_A^x \| \overline{\rho}_A), \tag{7.2.219}$$

which establishes (7.2.207). ■

7.3 Generalized Divergences

The quantum relative entropy is a function on states and positive semi-definite operators that satisfies the data-processing inequality under quantum channels. The data-processing inequality is quite powerful and a unifying concept, in the sense that it allows for establishing many properties of information and distinguishability measures. As such, this motivates the definition of generalized divergence as a function on states and positive semi-definite operators that satisfies the data-processing inequality.

Definition 7.15 Generalized Divergence

For every Hilbert space $\mathcal{H}$, a function $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ is called a *generalized divergence* if it satisfies the data-processing inequality under every channel, i.e., for all channels $\mathcal{N}$, states ρ, and positive semi-definite operators σ,

$$\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.3.1}$$

Many of the strong converse theorems in this book rely heavily on the data-processing inequality, and so we employ the generalized divergence to emphasize this point.

We have already mentioned the quantum relative entropy as an example of a generalized divergence. Other examples discussed later in this chapter, which are relevant in the context of channel capacity theorems, are the Petz–, sandwiched, and geometric Rényi relative entropies.

From the fact that generalized divergences satisfy the data-processing inequality by definition, we immediately obtain two properties of interest.

Proposition 7.16 Basic Properties of the Generalized Divergence

For every generalized divergence $\boldsymbol{D}$, for every state ρ, and every positive semi-definite operator σ:

1. The generalized divergence is isometrically invariant; i.e., for every isometry V,

$$\boldsymbol{D}(\rho\|\sigma) = \boldsymbol{D}(V\rho V^\dagger\|V\sigma V^\dagger). \tag{7.3.2}$$

2. For every state τ,

$$\boldsymbol{D}(\rho\|\sigma) = \boldsymbol{D}(\rho \otimes \tau\|\sigma \otimes \tau). \tag{7.3.3}$$

PROOF:

1. We follow the same approach discussed after (7.2.80). Since the map $\rho \mapsto V\rho V^\dagger$ is a channel, we immediately obtain $\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(V\rho V^\dagger\|V\sigma V^\dagger)$. To prove that $\boldsymbol{D}(\rho\|\sigma) \leq \boldsymbol{D}(V\rho V^\dagger\|V\sigma V^\dagger)$, consider the channel $\mathcal{R}_V$, which was defined in (4.4.13) as

$$\mathcal{R}_V(\omega) = V^\dagger\omega V + \mathrm{Tr}[(\mathbb{1} - VV^\dagger)\omega]\tau \tag{7.3.4}$$

for every positive semi-definite operator ω, where τ is an arbitrary (but fixed) state. Recall from Section 4.4.3 that this is a general way to take the map $\rho \mapsto V^\dagger\rho V$, which is completely positive but not trace preserving, and make it trace preserving and hence a channel. Then, recall that $\mathcal{R}_V(V\rho V^\dagger) = \rho$ and $\mathcal{R}_V(V\sigma V^\dagger) = \sigma$. Therefore, by the data-processing inequality,

$$\boldsymbol{D}(V\rho V^\dagger\|V\sigma V^\dagger) \geq \boldsymbol{D}(\mathcal{R}_V(V\rho V^\dagger)\|\mathcal{R}_V(V\sigma V^\dagger)) = \boldsymbol{D}(\rho\|\sigma), \tag{7.3.5}$$

and so $\boldsymbol{D}(\rho\|\sigma) = \boldsymbol{D}(V\rho V^\dagger\|V\sigma V^\dagger)$.

2. Since taking the tensor product with a fixed state is a channel (recall Definition 4.7), by definition of generalized divergence we obtain $\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(\rho \otimes \tau\|\sigma \otimes \tau)$. On the other hand, the partial trace is also a channel, and so by discarding the second system in the operators $\rho \otimes \tau$ and $\sigma \otimes \tau$, we obtain

$$\boldsymbol{D}(\rho\|\sigma) = \boldsymbol{D}(\mathrm{Tr}_2(\rho \otimes \tau)\|\mathrm{Tr}_2(\sigma \otimes \tau)) \leq \boldsymbol{D}(\rho \otimes \tau\|\sigma \otimes \tau), \tag{7.3.6}$$

which means that $\boldsymbol{D}(\rho\|\sigma) = \boldsymbol{D}(\rho \otimes \tau\|\sigma \otimes \tau)$. ■

Proposition 7.17

Suppose that the generalized divergence obeys the following direct-sum property:

$$\boldsymbol{D}(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho_A^x\|\sigma_A^x), \tag{7.3.7}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0, 1]$ is a probability distribution,

$\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states, $\{\sigma_A^x\}_{x\in\mathcal{X}}$ is a set of positive semi-definite operators, and

$$\rho_{XA} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.3.8}$$

$$\sigma_{XA} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.3.9}$$

Then the generalized divergence is jointly convex; i.e., the following inequality holds

$$\sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho_A^x\|\sigma_A^x) \geq \boldsymbol{D}(\overline{\rho}_A\|\overline{\sigma}_A), \tag{7.3.10}$$

where $\overline{\rho}_A := \sum_{x\in\mathcal{X}} p(x)\rho_A^x$ and $\overline{\sigma}_A := \sum_{x\in\mathcal{X}} p(x)\sigma_A^x$.

Proof: The proof is the same as the proof of Proposition 7.5 with D replaced by $\boldsymbol{D}$. ∎

Now, just as we defined entropic quantities like the entropy, conditional entropy, and mutual information using the quantum relative entropy, we can define their generalized counterparts using the generalized divergence.

Definition 7.18 Generalized Information Measures for States

Let $\boldsymbol{D}$ be a generalized divergence, as given in Definition 7.15.

1. The *generalized quantum entropy* $\boldsymbol{H}(\rho)$, for a state ρ, is defined as

$$\boldsymbol{H}(\rho) := -\boldsymbol{D}(\rho\|\mathbb{1}). \tag{7.3.11}$$

2. The *generalized quantum conditional entropy* $\boldsymbol{H}(A|B)_\rho$, for a bipartite state ρ_{AB}, is defined as

$$\boldsymbol{H}(A|B)_\rho := -\inf_{\sigma_B\in\mathrm{D}(\mathcal{H}_B)} \boldsymbol{D}(\rho_{AB}\|\mathbb{1}_A\otimes\sigma_B), \tag{7.3.12}$$

   and the *generalized coherent information* $\boldsymbol{I}(A\rangle B)_\rho$, for a bipartite state ρ_{AB}, is defined as

$$\boldsymbol{I}(A\rangle B)_\rho := \inf_{\sigma_B\in\mathrm{D}(\mathcal{H}_B)} \boldsymbol{D}(\rho_{AB}\|\mathbb{1}_A\otimes\sigma_B). \tag{7.3.13}$$

3. The *generalized quantum mutual information* $\boldsymbol{I}(A;B)_\rho$, for a bipartite state ρ_{AB}, is defined as

$$\boldsymbol{I}(A;B)_\rho := \inf_{\sigma_B \in \mathrm{D}(\mathcal{H}_B)} \boldsymbol{D}(\rho_{AB} \| \rho_A \otimes \sigma_B). \tag{7.3.14}$$

The data-processing inequality for every generalized divergence translates to the derived generalized information measures for states in the following way.

Proposition 7.19 Data-Processing Inequality for Generalized Information Measures for States

Let $\boldsymbol{D}$ be a generalized divergence.

1. The generalized quantum entropy does not decrease under the action of a unital channel, i.e.,

$$\boldsymbol{H}(\rho) \leq \boldsymbol{H}(\mathcal{N}(\rho)), \tag{7.3.15}$$

for every state ρ and every unital channel $\mathcal{N}$.

2. For every bipartite state ρ_{AB}, the generalized quantum conditional entropy does not decrease under the action of an arbitrary unital channel on A and an arbitrary channel on B, i.e.,

$$\boldsymbol{H}(A|B)_\rho \leq \boldsymbol{H}(A'|B')_{\rho'}, \tag{7.3.16}$$

where $\rho'_{A'B'} := (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{AB})$, with $\mathcal{N}_{A\to A'}$ an arbitrary unital channel and $\mathcal{M}_{B\to B'}$ an arbitrary channel. It follows by definition that

$$\boldsymbol{I}(A\rangle B)_\rho \geq \boldsymbol{I}(A'\rangle B')_{\rho'}. \tag{7.3.17}$$

3. For every bipartite state ρ_{AB}, the generalized quantum mutual information does not increase under the action of arbitrary channels on A and B, i.e.,

$$\boldsymbol{I}(A;B)_\rho \geq \boldsymbol{I}(A';B')_{\rho'}, \tag{7.3.18}$$

where $\rho'_{A'B'} = (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{AB})$, with $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ arbitrary channels.

PROOF:

1. Let ρ be an arbitrary state, and let $\mathcal{N}$ be a unital channel. The unitality of $\mathcal{N}$ means, by definition, that $\mathcal{N}(\mathbb{1}) = \mathbb{1}$. Using this, along with the data-processing inequality for the generalized divergence $\boldsymbol{D}$, we obtain

$$\boldsymbol{H}(\mathcal{N}(\rho)) = -\boldsymbol{D}(\mathcal{N}(\rho)\|\mathbb{1}) \tag{7.3.19}$$
$$= -\boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\mathbb{1})) \tag{7.3.20}$$
$$\geq -\boldsymbol{D}(\rho\|\mathbb{1}) \tag{7.3.21}$$
$$= \boldsymbol{H}(\rho), \tag{7.3.22}$$

as required.

2. Let ρ_{AB} be an arbitrary bipartite state, let $\mathcal{N}_{A\to A'}$ be an arbitrary unital channel, and let $\mathcal{M}_{B\to B'}$ be an arbitrary channel. Also, let σ_B be an arbitrary state. Then, using the data-processing inequality of the generalized divergence $\boldsymbol{D}$ and the unitality of $\mathcal{N}$, we obtain

$$\boldsymbol{D}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \geq \boldsymbol{D}((\mathcal{N}\otimes\mathcal{M})(\rho_{AB})\|(\mathcal{N}\otimes\mathcal{M})(\mathbb{1}_A\otimes\sigma_B)) \tag{7.3.23}$$
$$= \boldsymbol{D}(\rho'_{A'B'}\|\mathcal{N}(\mathbb{1}_A)\otimes\mathcal{M}(\sigma_B)) \tag{7.3.24}$$
$$= \boldsymbol{D}(\rho'_{A'B'}\|\mathbb{1}_{A'}\otimes\mathcal{M}(\sigma_B)) \tag{7.3.25}$$
$$\geq \inf_{\sigma_{B'}} \boldsymbol{D}(\rho'_{A'B'}\|\mathbb{1}_{A'}\otimes\sigma_{B'}) \tag{7.3.26}$$
$$= -\boldsymbol{H}(A'|B')_{\rho'}. \tag{7.3.27}$$

Since the state σ_B is arbitrary, we find that

$$\boldsymbol{H}(A|B)_\rho = -\inf_{\sigma_B}\boldsymbol{D}(\rho_{AB}\|\mathbb{1}_A\otimes\sigma_B) \leq \boldsymbol{H}(A'|B')_{\rho'}, \tag{7.3.28}$$

as required. The data-processing inequality in (7.3.17) for the generalized coherent information follows from the fact that, by definition, $\boldsymbol{I}(A\rangle B)_\rho = -\boldsymbol{H}(A|B)_\rho$.

3. Let ρ_{AB} be an arbitrary bipartite state, and let $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ be arbitrary channels. Also, let σ_B be an arbitrary state. Then, using the data-processing inequality for the generalized divergence $\boldsymbol{D}$, we obtain

$$\boldsymbol{D}(\rho_{AB}\|\rho_A\otimes\sigma_B) \geq \boldsymbol{D}((\mathcal{N}\otimes\mathcal{M})(\rho_{AB})\|(\mathcal{N}\otimes\mathcal{M})(\rho_A\otimes\sigma_B)) \tag{7.3.29}$$
$$= \boldsymbol{D}(\rho'_{A'B'}\|\rho'_{A'}\otimes\mathcal{M}(\sigma_B)) \tag{7.3.30}$$
$$\geq \inf_{\sigma_{B'}}\boldsymbol{D}(\rho'_{A'B'}\|\rho'_{A'}\otimes\sigma_{B'}) \tag{7.3.31}$$

$$= \boldsymbol{I}(A';B')_{\rho'}. \quad (7.3.32)$$

Since the state σ_B is arbitrary, we find that

$$\boldsymbol{I}(A;B)_\rho = \inf_{\sigma_B} \boldsymbol{D}(\rho_{AB}\|\rho_A \otimes \sigma_B) \geq \boldsymbol{I}(A';B')_{\rho'}, \quad (7.3.33)$$

as required. ■

In the above, we have proved most properties of a generalized divergence by employing its defining property only (i.e., by employing the data-processing inequality in (7.3.1)). In further applications, it can be useful to make some very minimal extra assumptions about a generalized divergence. In what follows, we list two of these minimal assumptions. If we ever employ these assumptions in future applications, we indicate this explicitly.

1. A first assumption is that
$$\boldsymbol{D}(1\|c) \geq 0 \quad (7.3.34)$$
for $c \in (0, 1]$. That is, if we plug in a trivial one-dimensional density operator ρ (i.e., the number 1) and a trivial positive semi-definite operator with trace less than or equal to one, then the generalized divergence evaluates to a non-negative real. This assumption is satisfied by all examples of generalized divergences that we employ in this book.

   A consequence of the assumption in (7.3.34) is that
$$\boldsymbol{D}(\rho\|\sigma) \geq 0 \quad (7.3.35)$$
for every density operator ρ and positive semi-definite operator σ satisfying $\mathrm{Tr}[\sigma] \leq 1$. This follows from (7.3.1) and (7.3.34) by applying the trace-out channel.

2. A second minimal assumption is that
$$\boldsymbol{D}(\rho\|\rho) = 0 \quad (7.3.36)$$
for every state ρ. We should clarify that this assumption is quite minimal. The reason is that it is essentially a direct consequence of (7.3.1) up to an inessential additive factor. That is, (7.3.1) implies that there exists a constant c such that
$$\boldsymbol{D}(\rho\|\rho) = c \quad (7.3.37)$$

for every state ρ. To see this, consider that one can get from the state ρ to another state ω by means of a trace and replace channel $\mathrm{Tr}[\cdot]\omega$, so that (7.3.1) implies that

$$\boldsymbol{D}(\rho\|\rho) \geq \boldsymbol{D}(\omega\|\omega). \tag{7.3.38}$$

However, by the same argument, $\boldsymbol{D}(\omega\|\omega) \geq \boldsymbol{D}(\rho\|\rho)$, so that the claim holds. So the assumption in (7.3.36) amounts to a redefinition of the generalized divergence as

$$\boldsymbol{D}'(\rho\|\sigma) := \boldsymbol{D}(\rho\|\sigma) - c. \tag{7.3.39}$$

7.4 Petz–Rényi Relative Entropy

One important example of a generalized divergence is the Petz–Rényi relative entropy.

Definition 7.20 Petz–Rényi Relative Entropy

For all $\alpha \in (0,1) \cup (1,\infty)$, we define the *Petz–Rényi relative quasi-entropy* for every state ρ and positive semi-definite operator σ as

$$Q_\alpha(\rho\|\sigma) := \begin{cases} \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}] & \text{if } \alpha \in (0,1), \text{ or} \\ & \alpha \in (1,\infty) \text{ and } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \tag{7.4.1}$$

The *Petz–Rényi relative entropy* is then defined as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log_2 Q_\alpha(\rho\|\sigma). \tag{7.4.2}$$

By following the recipe in (7.3.11), i.e., setting $\sigma = \mathbb{1}$ and applying a minus sign out in front, the Petz–Rényi relative entropy reduces to the Rényi entropy of a quantum state ρ:

$$H_\alpha(\rho) = -D_\alpha(\rho\|\mathbb{1}) = \frac{1}{1-\alpha} \log_2 \mathrm{Tr}[\rho^\alpha] = \frac{\alpha}{1-\alpha} \log_2 \|\rho\|_\alpha . \tag{7.4.3}$$

If the state ρ is defined on system A, we also employ the notation

$$H_\alpha(A)_\rho \equiv H_\alpha(\rho). \tag{7.4.4}$$

The Rényi entropy is defined for all $\alpha \in (0,1) \cup (1,\infty)$, and one evaluates its value at $\alpha \in \{0,1,\infty\}$ by taking limits:

$$H_0(\rho) \coloneqq \lim_{\alpha\to 0} H_\alpha(\rho) = \log_2 \operatorname{rank}(\rho), \tag{7.4.5}$$

$$H_1(\rho) \coloneqq \lim_{\alpha\to 1} H_\alpha(\rho) = -\operatorname{Tr}[\rho \log_2 \rho] = H(\rho), \tag{7.4.6}$$

$$H_\infty(\rho) \coloneqq \lim_{\alpha\to \infty} H_\alpha(\rho) = -\log \lambda_{\max}(\rho). \tag{7.4.7}$$

Turning back to the Petz–Rényi relative entropy, note that $1-\alpha$ is negative for $\alpha > 1$. In this case, the inverse of σ is taken on its support. An alternative to this convention is to define $D_\alpha(\rho\|\sigma)$ for $\alpha > 1$ for only positive definite σ and for positive semi-definite σ define $D_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0} D_\alpha(\rho\|\sigma + \varepsilon\mathbb{1})$. Both alternatives are equivalent, as we now show.

Proposition 7.21

For every state ρ and positive semi-definite operator σ,

$$D_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}\left[\rho^\alpha (\sigma + \varepsilon\mathbb{1})^{1-\alpha}\right]. \tag{7.4.8}$$

Proof: For $\alpha \in (0,1)$, this is immediate from the fact that the logarithm, trace, and power functions are continuous, so that the limit can be brought inside the trace and inside the power $(\sigma + \varepsilon\mathbb{1})^{1-\alpha}$.

For $\alpha \in (1,\infty)$, since $1-\alpha$ is negative and σ is not necessarily invertible, we first decompose the underlying Hilbert space $\mathcal{H}$ as $\mathcal{H} = \operatorname{supp}(\sigma) \oplus \ker(\sigma)$, just as we did in (7.2.6), in order to write

$$\rho = \begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^\dagger & \rho_{1,1} \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.4.9}$$

Then, writing $\mathbb{1} = \Pi_\sigma + \Pi_\sigma^\perp$, where Π_σ is the projection onto the support of σ and $\Pi_\sigma^\perp$ is the projection onto the orthogonal complement of $\operatorname{supp}(\sigma)$, we find that

$$\sigma + \varepsilon\mathbb{1} = \begin{pmatrix} \sigma + \varepsilon\Pi_\sigma & 0 \\ 0 & \varepsilon\Pi_\sigma^\perp \end{pmatrix}, \tag{7.4.10}$$

which implies that

$$(\sigma + \varepsilon\mathbb{1})^{1-\alpha} = \begin{pmatrix} (\sigma + \varepsilon\Pi_\sigma)^{1-\alpha} & 0 \\ 0 & (\varepsilon\Pi_\sigma^\perp)^{1-\alpha} \end{pmatrix}. \tag{7.4.11}$$

If $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then $\rho = \rho_{0,0}$, $\rho_{0,1} = 0$, and $\rho_{1,1} = 0$, which means that

$$\rho^{\alpha}(\sigma + \varepsilon\mathbb{1})^{1-\alpha} = \begin{pmatrix} \rho^{\alpha}(\sigma + \varepsilon\Pi_{\sigma})^{1-\alpha} & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.4.12}$$

so that

$$\begin{aligned} &\lim_{\varepsilon\to 0^+} \frac{1}{\alpha - 1} \log_2 \text{Tr}\left[\rho^{\alpha}(\sigma + \varepsilon\mathbb{1})^{1-\alpha}\right] \\ &\quad = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha - 1} \log_2 \text{Tr}\left[\rho^{\alpha}(\sigma + \varepsilon\Pi_{\sigma})^{1-\alpha}\right] &(7.4.13)\\ &\quad = \frac{1}{\alpha - 1} \log_2 \text{Tr}\left[\rho^{\alpha}\sigma^{1-\alpha}\right] &(7.4.14)\\ &\quad = D_{\alpha}(\rho\|\sigma), &(7.4.15) \end{aligned}$$

as required.

If $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$, then the blocks $\rho_{0,1}$ and $\rho_{1,1}$ are generally non-zero, and we obtain

$$\begin{aligned} &\rho^{\alpha}(\sigma + \varepsilon\mathbb{1})^{1-\alpha} \\ &\quad = \begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^{\dagger} & \rho_{1,1} \end{pmatrix}^{\alpha} \begin{pmatrix} (\sigma + \varepsilon\Pi_{\sigma})^{1-\alpha} & 0 \\ 0 & (\varepsilon\Pi_{\sigma}^{\perp})^{1-\alpha} \end{pmatrix} &(7.4.16)\\ &\quad = \varepsilon^{1-\alpha}\left[\begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^{\dagger} & \rho_{1,1} \end{pmatrix}^{\alpha} \begin{pmatrix} \varepsilon^{\alpha-1}(\sigma + \varepsilon\Pi_{\sigma})^{1-\alpha} & 0 \\ 0 & (\Pi_{\sigma}^{\perp})^{1-\alpha} \end{pmatrix}\right]. &(7.4.17) \end{aligned}$$

Due to the fact that $\alpha \in (1, \infty)$ it holds that $\lim_{\varepsilon\to 0^+} \varepsilon^{1-\alpha} = +\infty$, and since the limit $\varepsilon \to 0^+$ of the matrix in square brackets in (7.4.17) is finite, we find that

$$\lim_{\varepsilon\to 0} \frac{1}{\alpha - 1} \log_2 \text{Tr}\left[\rho^{\alpha}(\sigma + \varepsilon\mathbb{1})^{1-\alpha}\right] = +\infty = D_{\alpha}(\rho\|\sigma) \tag{7.4.18}$$

for the case $\alpha \in (1, \infty)$ and $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$. ■

The Petz–Rényi relative entropy is a natural generalization of the classical Rényi relative entropy, which is defined for a probability distribution $p : \mathcal{X} \to [0, 1]$ and a positive measure $q : \mathcal{X} \to [0, \infty)$ over the same alphabet $\mathcal{X}$ as

$$D_{\alpha}(p\|q) = \frac{1}{\alpha - 1} \log_2 \sum_{x\in\mathcal{X}} p(x)^{\alpha} q(x)^{1-\alpha} \tag{7.4.19}$$

for $\alpha \in (0,1)$. For $\alpha \in (1,\infty)$, the same expression defines $D_\alpha(p\|q)$ whenever $q(x) = 0$ implies $p(x) = 0$ for all $x \in \mathcal{X}$; otherwise, $D_\alpha(p\|q) = +\infty$.

Recall the quantum Chernoff bound from Theorem 5.4 in Section 5.3.1, which states that optimal error exponent for the task of discriminating between the states ρ and σ is

$$\underline{\xi}(\rho,\sigma) = \overline{\xi}(\rho,\sigma) = C(\rho\|\sigma) := \sup_{\alpha\in(0,1)} -\log_2 \operatorname{Tr}[\rho^\alpha \sigma^{1-\alpha}]. \tag{7.4.20}$$

Using the definition of the Petz–Rényi relative entropy, we find that

$$C(\rho\|\sigma) = \inf_{\alpha\in(0,1)} (1-\alpha) D_\alpha(\rho\|\sigma). \tag{7.4.21}$$

The Petz–Rényi relative entropy thus plays a role in providing the optimal error exponent for the task of discriminating between two states (i.e., symmetric hypothesis testing).

Like the quantum relative entropy, the Petz–Rényi relative entropy is faithful, meaning that, for $\alpha \in (0,1) \cup (1,\infty)$ and states ρ, σ,

$$D_\alpha(\rho\|\sigma) = 0 \qquad \Longleftrightarrow \qquad \rho = \sigma. \tag{7.4.22}$$

We prove this in Proposition 7.36 in the next section, as it requires results from both this section and the next one. Also, like the quantum relative entropy, the Petz–Rényi relative entropy is a generalized divergence for certain values of α, which is shown in Theorem 7.24 below.

Before getting to Theorem 7.24, we first discuss several important properties of the Petz–Rényi relative entropy. Let us note that, if ρ and σ act on a d-dimensional Hilbert space and are invertible, then the Petz–Rényi relative quasi-entropy can be written as

$$Q_\alpha(\rho\|\sigma) = \langle \varphi^\rho | (\rho^{-1} \otimes \sigma^{\mathsf{T}})^{1-\alpha} | \varphi^\rho \rangle, \tag{7.4.23}$$

where $|\varphi^\rho\rangle := (\rho^{\frac{1}{2}} \otimes \mathbb{1})|\Gamma\rangle$ is a purification of ρ and $|\Gamma\rangle = \sum_{i=1}^d |i,i\rangle$. This is due to the transpose trick in (2.2.42) and the identity in (2.2.43). We can extend the applicability of this expression to states ρ and positive semi-definite operators σ that are not invertible by noting that

$$Q_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \lim_{\delta\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}\left[((1-\delta)\rho + \delta\pi)^\alpha (\sigma + \varepsilon\mathbb{1})^{1-\alpha} \right]. \tag{7.4.24}$$

We start by establishing the important fact that the quantum relative entropy is a special case of the Petz–Rényi relative entropy in the limit $\alpha \to 1$.

Proposition 7.22

Let ρ be a state and σ a positive semi-definite operator. Then, in the limit $\alpha \to 1$, the Petz–Rényi relative entropy converges to the quantum relative entropy:

$$\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma). \tag{7.4.25}$$

Proof: Let us first consider the case $\alpha \in (1,\infty)$. If $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, then $D_\alpha(\rho\|\sigma) = +\infty$, so that $\lim_{\alpha\to 1^+} D_\alpha(\rho\|\sigma) = +\infty$, consistent with the definition of the quantum relative entropy in this case (see Definition 7.1). If $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then $D_\alpha(\rho\|\sigma)$ is finite and we can write

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log_2 Q_\alpha(\rho\|\sigma). \tag{7.4.26}$$

Now, let us define the function

$$Q_{\alpha,\beta}(\rho\|\sigma) := \text{Tr}[\rho^\alpha\sigma^{1-\beta}], \tag{7.4.27}$$

so that $Q_\alpha(\rho\|\sigma) = Q_{\alpha,\alpha}(\rho\|\sigma)$. By noting that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ implies $Q_1(\rho\|\sigma) = \text{Tr}[\rho\Pi_\sigma] = 1$ (where Π_σ is the projection onto the support of σ), and since $\log_2 1 = 0$, we can write $D_\alpha(\rho\|\sigma)$ as

$$D_\alpha(\rho\|\sigma) = \frac{\log_2 Q_\alpha(\rho\|\sigma) - \log_2 Q_1(\rho\|\sigma)}{\alpha-1}, \tag{7.4.28}$$

so that

$$\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log_2 Q_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} \tag{7.4.29}$$

$$= \frac{1}{\ln(2)}\frac{\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_\alpha(\rho\|\sigma)\big|_{\alpha=1}}{Q_1(\rho\|\sigma)} \tag{7.4.30}$$

$$= \frac{1}{\ln(2)}\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_\alpha(\rho\|\sigma)\bigg|_{\alpha=1}, \tag{7.4.31}$$

where the first equality follows from the definition of the derivative and the second equality from the derivative of the natural logarithm, along with the chain rule. Using the function $Q_{\alpha,\beta}$ and the chain rule, we write

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} = \frac{\mathrm{d}}{\mathrm{d}\alpha}Q_{\alpha,1}(\rho\|\sigma)\bigg|_{\alpha=1} + \frac{\mathrm{d}}{\mathrm{d}\beta}Q_{1,\beta}(\rho\|\sigma)\bigg|_{\beta=1}. \tag{7.4.32}$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_{\alpha,1}(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathrm{Tr}[\rho^\alpha\Pi_\sigma] = \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathrm{Tr}[\rho^\alpha] = \mathrm{Tr}[\rho^\alpha \ln \rho], \tag{7.4.33}$$

where we used the fact that $\rho^\alpha\Pi_\sigma = \rho^\alpha$ since $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. Therefore,

$$\left.\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_{\alpha,1}(\rho\|\sigma)\right|_{\alpha=1} = \mathrm{Tr}[\rho \ln \rho]. \tag{7.4.34}$$

Similarly,

$$\frac{\mathrm{d}}{\mathrm{d}\beta}Q_{1,\beta}(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Tr}[\rho\sigma^{1-\beta}] = -\mathrm{Tr}[\rho\sigma^{1-\beta}\ln\sigma], \tag{7.4.35}$$

so that

$$\left.\frac{\mathrm{d}}{\mathrm{d}\beta}Q_{1,\beta}(\rho\|\sigma)\right|_{\beta=1} = -\mathrm{Tr}[\rho\Pi_\sigma \ln\sigma] = -\mathrm{Tr}[\rho\ln\sigma], \tag{7.4.36}$$

where the last equality follows from the fact that the support condition $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ holds. So we find that

$$\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = \frac{1}{\ln(2)}\left.\frac{\mathrm{d}}{\mathrm{d}\alpha}Q_\alpha(\rho\|\sigma)\right|_{\alpha=1} = \mathrm{Tr}[\rho\log_2\rho] - \mathrm{Tr}[\rho\log_2\sigma] \tag{7.4.37}$$

when $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, which means that for $\alpha \in (1,\infty)$,

$$\begin{aligned}\lim_{\alpha\to 1^+} D_\alpha(\rho\|\sigma) &= \begin{cases} \mathrm{Tr}[\rho\log_2\rho] - \mathrm{Tr}[\rho\log_2\sigma] & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise} \end{cases} \\ &= D(\rho\|\sigma).\end{aligned} \tag{7.4.38}$$

Let us now consider the case $\alpha \in (0,1)$. If $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, then since the limit in (7.4.37) holds from both sides, we find that

$$\lim_{\alpha\to 1^-} D_\alpha(\rho\|\sigma) = \mathrm{Tr}[\rho\log_2\rho] - \mathrm{Tr}[\rho\log_2\sigma]. \tag{7.4.39}$$

If $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$ (and $\mathrm{Tr}[\rho\sigma] \neq 0$), then observe that we can write D_α as

$$D_\alpha(\rho\|\sigma) = \frac{\log_2 Q_\alpha(\rho\|\sigma) - \log_2 Q_1(\rho\|\sigma)}{\alpha - 1} + \frac{\log_2 Q_1(\rho\|\sigma)}{\alpha - 1}, \tag{7.4.40}$$

so that

$$\lim_{\alpha\to 1^-} D_\alpha(\rho\|\sigma) = \left.\frac{\mathrm{d}}{\mathrm{d}\alpha}\log_2 Q_\alpha(\rho\|\sigma)\right|_{\alpha=1} + \lim_{\alpha\to 1^-}\frac{\log_2 \mathrm{Tr}[\rho\Pi_\sigma]}{\alpha-1}, \tag{7.4.41}$$

where we have used $Q_1(\rho\|\sigma) = \mathrm{Tr}[\rho\Pi_\sigma]$. Now, since $\mathrm{Tr}[\rho\sigma] \neq 0$ and $\mathrm{supp}(\rho) \nsubseteq \mathrm{supp}(\sigma)$, we have that $0 < \mathrm{Tr}[\rho\Pi_\sigma] < 1$, which means that $\log_2 \mathrm{Tr}[\rho\Pi_\sigma] < 0$. Since $\lim_{\alpha\to 1^-}\frac{1}{\alpha-1} = -\infty$, we get that the second term in (7.4.41) is equal to $+\infty$, which means that $\lim_{\alpha\to 1^-} D_\alpha(\rho\|\sigma) = +\infty$. Therefore,

$$\begin{aligned}\lim_{\alpha\to 1^-} & D_\alpha(\rho\|\sigma) \\ &= \begin{cases} \mathrm{Tr}[\rho\log_2\rho] - \mathrm{Tr}[\rho\log_2\sigma] & \text{if } \mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma), \\ +\infty & \text{otherwise} \end{cases} \\ &= D(\rho\|\sigma).\end{aligned} \tag{7.4.42}$$

To conclude, we have that $\lim_{\alpha\to 1^+} D_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1^-} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$, which means that

$$\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma), \tag{7.4.43}$$

as required. ■

Proposition 7.23 Properties of the Petz–Rényi Relative Entropy

For all states ρ, ρ_1, ρ_2 and positive semi-definite operators $\sigma, \sigma_1, \sigma_2$, the Petz–Rényi relative entropy satisfies the following properties.

1. *Isometric invariance*: For all $\alpha \in (0,1)\cup(1,\infty)$ and for all isometries V,

$$D_\alpha(\rho\|\sigma) = D_\alpha(V\rho V^\dagger\|V\sigma V^\dagger). \tag{7.4.44}$$

2. *Monotonicity in α*: For all $\alpha \in (0,1)\cup(1,\infty)$, D_α is monotonically increasing in α, i.e., $\alpha < \beta$ implies $D_\alpha(\rho\|\sigma) \le D_\beta(\rho\|\sigma)$.

3. *Additivity*: For all $\alpha \in (0,1)\cup(1,\infty)$,

$$D_\alpha(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2) = D_\alpha(\rho_1\|\sigma_1) + D_\alpha(\rho_2\|\sigma_2). \tag{7.4.45}$$

4. *Direct-sum property*: Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $q : \mathcal{X} \to [0,\infty)$ be a positive function on $\mathcal{X}$. Let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states

on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$Q_\alpha(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} Q_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.4.46}$$

where

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.4.47}$$

$$\sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.4.48}$$

REMARK: Observe that the direct-sum property analogous to that for the quantum relative entropy (see Proposition 7.3) does not hold for the Petz–Rényi relative entropy for every $\alpha \in (0,1)\cup(1,\infty)$. We can instead only make a statement for the Petz–Rényi relative quasi-entropy.

PROOF:

1. *Proof of isometric invariance*: Let us start by writing $D_\alpha(\rho\|\sigma)$ as in (7.4.8):

$$D_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}\left[\rho^\alpha(\sigma+\varepsilon\mathbb{1})^{1-\alpha}\right]. \tag{7.4.49}$$

Then, using the fact that $(V\rho V^\dagger)^\alpha = V\rho^\alpha V^\dagger$, we find that

$$\begin{aligned} &D_\alpha(V\rho V^\dagger\|V\sigma V^\dagger) \\ &\quad = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}\left[(V\rho V^\dagger)^\alpha(V\sigma V^\dagger+\varepsilon\mathbb{1})^{1-\alpha}\right] &(7.4.50)\\ &\quad = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \operatorname{Tr}\left[V\rho^\alpha V^\dagger(V\sigma V^\dagger+\varepsilon\mathbb{1})^{1-\alpha}\right]. &(7.4.51)\end{aligned}$$

Now, let $\Pi \coloneqq VV^\dagger$ be the projection onto the image of V, and let $\hat{\Pi} \coloneqq \mathbb{1}-\Pi$. Then, we can write

$$V\sigma V^\dagger + \varepsilon\mathbb{1} = V\sigma V^\dagger + \varepsilon\Pi + \varepsilon\hat{\Pi} = V(\sigma+\varepsilon\mathbb{1})V^\dagger + \varepsilon\hat{\Pi}. \tag{7.4.52}$$

Since $V(\sigma+\varepsilon\mathbb{1})V^\dagger$ and $\varepsilon\hat{\Pi}$ are supported on orthogonal subspaces, we obtain

$$(V\sigma V^\dagger+\varepsilon\mathbb{1})^{1-\alpha} = V(\sigma+\varepsilon\mathbb{1})^{1-\alpha}V^\dagger + \varepsilon^{1-\alpha}\hat{\Pi}. \tag{7.4.53}$$

Therefore,

$$\operatorname{Tr}\left[V\rho^\alpha V^\dagger(V\sigma V^\dagger+\varepsilon\mathbb{1})^{1-\alpha}\right]$$
$$=\operatorname{Tr}\left[V\rho^\alpha V^\dagger V(\sigma+\varepsilon\mathbb{1})^{1-\alpha}V^\dagger+\varepsilon^{1-\alpha}V\rho^\alpha V^\dagger\hat{\Pi}\right] \tag{7.4.54}$$
$$=\operatorname{Tr}\left[V\rho^\alpha(\sigma+\varepsilon\mathbb{1})^{1-\alpha}V^\dagger\right] \tag{7.4.55}$$
$$=\operatorname{Tr}\left[\rho^\alpha(\sigma+\varepsilon\mathbb{1})^{1-\alpha}\right], \tag{7.4.56}$$

where the second equality follows from the fact that $V^\dagger\hat{\Pi}V=V^\dagger V-V^\dagger VV^\dagger V=\mathbb{1}-\mathbb{1}=0$, and the last equality from cyclicity of the trace. Therefore,

$$D_\alpha(V\rho V^\dagger\|V\sigma V^\dagger)=\lim_{\varepsilon\to 0^+}\frac{1}{\alpha-1}\log_2\operatorname{Tr}\left[\rho^\alpha(\sigma+\varepsilon\mathbb{1})^{1-\alpha}\right] \tag{7.4.57}$$
$$=D_\alpha(\rho\|\sigma), \tag{7.4.58}$$

as required.

2. *Proof of monotonicity in α*: Using the expression in (7.4.2) for D_α along with the form in (7.4.23) for the quasi-entropy Q_α, let us write $D_\alpha(\rho\|\sigma)$ as

$$D_\alpha(\rho\|\sigma)=\frac{1}{\alpha-1}\frac{\ln\langle\varphi^\rho|X^{1-\alpha}|\varphi^\rho\rangle}{\ln(2)}=-\frac{1}{\gamma}\frac{\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}{\ln(2)}, \tag{7.4.59}$$

where $X=\rho^{-1}\otimes\sigma^{\mathsf{T}}$, $\gamma:=1-\alpha$, and $|\varphi^\rho\rangle=(\rho^{\frac{1}{2}}\otimes\mathbb{1})|\Gamma\rangle$ is a purification of ρ. We first prove the result for ρ invertible, and the proof for non-invertible states ρ follows by (7.4.24). Now, since $\frac{\mathrm{d}}{\mathrm{d}\alpha}=\frac{\mathrm{d}}{\mathrm{d}\gamma}\frac{\mathrm{d}\gamma}{\mathrm{d}\alpha}=-\frac{\mathrm{d}}{\mathrm{d}\gamma}$, we find that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}D_\alpha(\rho\|\sigma)$$
$$=\frac{1}{\ln(2)}\frac{\mathrm{d}}{\mathrm{d}\gamma}\left(\frac{1}{\gamma}\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle\right) \tag{7.4.60}$$
$$=\frac{1}{\ln(2)}\left(-\frac{1}{\gamma^2}\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle+\frac{1}{\gamma}\frac{\langle\varphi^\rho|X^\gamma\ln X|\varphi^\rho\rangle}{\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}\right) \tag{7.4.61}$$
$$=\frac{1}{\ln(2)}\frac{-\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle+\gamma\langle\varphi^\rho|X^\gamma\ln X|\varphi^\rho\rangle}{\gamma^2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle} \tag{7.4.62}$$
$$=\frac{1}{\ln(2)}\frac{-\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle+\langle\varphi^\rho|X^\gamma\ln X^\gamma|\varphi^\rho\rangle}{\gamma^2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}. \tag{7.4.63}$$

Letting $g(x):=x\log_2 x$, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}D_\alpha(\rho\|\sigma)=\frac{\langle\varphi^\rho|g(X^\gamma)|\varphi^\rho\rangle-g(\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle)}{\gamma^2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}. \tag{7.4.64}$$

Then, since g is operator convex, by the operator Jensen inequality in (2.3.23), we conclude that

$$\langle\varphi^\rho|g(X^\gamma)|\varphi^\rho\rangle \geq g(\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle), \tag{7.4.65}$$

which means that $\frac{\mathrm{d}}{\mathrm{d}\alpha}D_\alpha(\rho\|\sigma) \geq 0$. Therefore, $D_\alpha(\rho\|\sigma)$ is monotonically increasing in α, as required.

3. *Proof of additivity*: When all quantities are finite, we have that

$$D_\alpha(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \frac{1}{\alpha - 1}\log_2 \mathrm{Tr}\left[(\rho_1 \otimes \rho_2)^\alpha(\sigma_1 \otimes \sigma_2)^{1-\alpha}\right]. \tag{7.4.66}$$

Using the fact that $(X \otimes Y)^\beta = X^\beta \otimes Y^\beta$ for all positive semi-definite operators X, Y and all $\beta \in \mathbb{R}$, we obtain

$$Q_\alpha(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \mathrm{Tr}\left[(\rho_1 \otimes \rho_2)^\alpha(\sigma_1 \otimes \sigma_2)^{1-\alpha}\right] \tag{7.4.67}$$
$$= \mathrm{Tr}\left[(\rho_1^\alpha \otimes \rho_2^\alpha)(\sigma_1^{1-\alpha} \otimes \sigma_2^{1-\alpha})\right] \tag{7.4.68}$$
$$= \mathrm{Tr}\left[\rho_1^\alpha\sigma_1^{1-\alpha} \otimes \rho_2^\alpha\sigma_2^{1-\alpha}\right] \tag{7.4.69}$$
$$= \mathrm{Tr}[\rho_1^\alpha\sigma_1^{1-\alpha}] \cdot \mathrm{Tr}[\rho_2^\alpha\sigma_2^{1-\alpha}] \tag{7.4.70}$$
$$= Q_\alpha(\rho_1\|\sigma_1) \cdot Q_\alpha(\rho_2\|\sigma_2). \tag{7.4.71}$$

Applying $\frac{1}{\alpha-1}\log_2$ and definitions, additivity follows.

4. *Proof of the direct-sum property*: Define the classical–quantum operators

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \quad \sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.4.72}$$

Then,

$$\rho_{XA}^\alpha = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes (p(x)\rho_X^x)^\alpha \tag{7.4.73}$$
$$\sigma_{XA}^{1-\alpha} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes (q(x)\sigma_A^x)^{1-\alpha}, \tag{7.4.74}$$

so that

$$\rho_{XA}^\alpha\sigma_{XA}^{1-\alpha} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes (p(x)\rho_A^x)^\alpha(q(x)\sigma_A^x)^{1-\alpha} \tag{7.4.75}$$

$$= \sum_{x \in \mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} |x\rangle\langle x|_X \otimes (\rho_A^x)^\alpha (\sigma_A^x)^{1-\alpha}, \tag{7.4.76}$$

and

$$Q_\alpha(\rho_{XA} \| \sigma_{XA}) = \mathrm{Tr}[\rho_{XA}^\alpha \sigma_{XA}^{1-\alpha}] \tag{7.4.77}$$
$$= \sum_{x \in \mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} \mathrm{Tr}[(\rho_A^x)^\alpha (\sigma_A^x)^{1-\alpha}] \tag{7.4.78}$$
$$= \sum_{x \in \mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} Q_\alpha(\rho_A^x \| \sigma_A^x), \tag{7.4.79}$$

as required. ■

We now prove the data-processing inequality for the Petz–Rényi relative entropy for $\alpha \in [0,1) \cup (1,2]$.

Theorem 7.24 Data-Processing Inequality for Petz–Rényi Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then, for all $\alpha \in [0,1) \cup (1,2]$,

$$D_\alpha(\rho \| \sigma) \geq D_\alpha(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)). \tag{7.4.80}$$

Proof: We prove the statement for $\alpha \in (0,1) \cup (1,2]$. The case of $\alpha = 0$ then follows by taking the limit $\alpha \to 0$. From Stinespring's theorem (Theorem 4.3), we know that the action of every channel $\mathcal{N}$ on a linear operator X can be written as

$$\mathcal{N}(X) = \mathrm{Tr}_E[VXV^\dagger], \tag{7.4.81}$$

for some V, where V is an isometry and E is an auxiliary system with dimension $d_E \geq \mathrm{rank}(\Gamma^{\mathcal{N}})$. As stated in (7.4.44), D_α is isometrically invariant. Therefore, it suffices to show the data-processing inequality for D_α under partial trace; i.e., it suffices to show that for every state ρ_{AB} and every positive semi-definite operator σ_{AB},

$$D_\alpha(\rho_{AB} \| \sigma_{AB}) \geq D_\alpha(\rho_A \| \sigma_A), \quad \alpha \in (0,1) \cup (1,2]. \tag{7.4.82}$$

We now proceed to prove this inequality. We prove it for ρ_{AB}, and hence ρ_A, invertible, as well as for σ_{AB} and σ_A invertible. The result follows in the general case of ρ_{AB} and/or ρ_A non-invertible, as well as σ_{AB} and/or σ_A non-invertible, by applying the result to the invertible operators $(1-\delta)\rho_{AB}+\delta\pi_{AB}$ and $\sigma_{AB}+\varepsilon\mathbb{1}_{AB}$, with $\delta,\varepsilon>0$, and taking the limit $\delta\to 0^+$, followed by $\varepsilon\to 0^+$, because

$$D_\alpha(\rho_{AB}\|\sigma_{AB}) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} D_\alpha((1-\delta)\rho_{AB}+\delta\pi_{AB}\|\sigma_{AB}+\varepsilon\mathbb{1}_{AB}), \tag{7.4.83}$$

$$D_\alpha(\rho_A\|\sigma_A) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} D_\alpha((1-\delta)\rho_A+\delta\pi_A\|\sigma_A+d_B\varepsilon\mathbb{1}_A), \tag{7.4.84}$$

which can be verified in a similar manner to the proof of (7.4.8) in Proposition 7.21.

Using the quasi-entropy Q_α, we can equivalently write (7.4.82) as

$$\begin{aligned} Q_\alpha(\rho_{AB}\|\sigma_{AB}) &\geq Q_\alpha(\rho_A\|\sigma_A), \qquad \text{for } \alpha\in(1,2], \\ Q_\alpha(\rho_{AB}\|\sigma_{AB}) &\leq Q_\alpha(\rho_A\|\sigma_A), \qquad \text{for } \alpha\in(0,1). \end{aligned} \tag{7.4.85}$$

The remainder of this proof is thus devoted to establishing (7.4.85).

Consider that

$$Q_\alpha(\rho_{AB}\|\sigma_{AB}) = \langle\varphi^{\rho_{AB}}|f(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})|\varphi^{\rho_{AB}}\rangle, \tag{7.4.86}$$

$$Q_\alpha(\rho_A\|\sigma_A) = \langle\varphi^{\rho_A}|f(\rho_A^{-1}\otimes\sigma_{\hat{A}}^{\mathsf{T}})|\varphi^{\rho_A}\rangle, \tag{7.4.87}$$

where we have set

$$f(x) := x^{1-\alpha} \tag{7.4.88}$$

and

$$|\varphi^{\rho_{AB}}\rangle = (\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{AB\hat{A}\hat{B}} = (\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}}, \tag{7.4.89}$$

$$|\varphi^{\rho_A}\rangle = (\rho_A^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{A\hat{A}}. \tag{7.4.90}$$

Now, let us define the isometry $V_{A\hat{A}\to AB\hat{A}\hat{B}}$ as[1]

$$V_{A\hat{A}\to AB\hat{A}\hat{B}} = \rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}}. \tag{7.4.91}$$

Observe then that

$$\begin{aligned} V_{A\hat{A}\to AB\hat{A}\hat{B}}|\varphi^{\rho_A}\rangle_{A\hat{A}} &= \rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})(\rho_A^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}} &(7.4.92)\\ &= (\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}} &(7.4.93)\end{aligned}$$

[1]Observe that the isometry is related to the isometric extension in (4.6.39) of the Petz recovery channel for the partial trace, as discussed in Section 4.6.1.

$$= |\varphi^{\rho_{AB}}\rangle. \tag{7.4.94}$$

We thus obtain, using the operator Jensen inequality (Theorem 2.16),

$$\begin{aligned} Q_\alpha(\rho_{AB}\|\sigma_{AB}) &= \langle\varphi^{\rho_A}|V^\dagger f(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V|\varphi^{\rho_A}\rangle & (7.4.95)\\ &\geq \langle\varphi^{\rho_A}|f(V^\dagger(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V)|\varphi^{\rho_A}\rangle, \quad \text{for } \alpha\in(1,2], & (7.4.96)\end{aligned}$$

and

$$\begin{aligned} Q_\alpha(\rho_{AB}\|\sigma_{AB}) &= \langle\varphi^{\rho_A}|V^\dagger f(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V|\varphi^{\rho_A}\rangle & (7.4.97)\\ &\leq \langle\varphi^{\rho_A}|f(V^\dagger(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V)|\varphi^{\rho_A}\rangle, \quad \text{for } \alpha\in[0,1). & (7.4.98)\end{aligned}$$

Note that the operator Jensen inequality is applicable because for $\alpha \in (1,2]$ the function f in (7.4.88) is operator convex and for $\alpha\in(0,1)$ it is operator concave.[2] Now, consider that

$$\begin{aligned} &V^\dagger(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V\\ &\quad= \langle\Gamma|_{B\hat{B}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})\rho_{AB}^{\frac{1}{2}}(\rho_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})\rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}} & (7.4.99)\\ &\quad= \langle\Gamma|_{B\hat{B}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})(\rho_{AB}^{0}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}} & (7.4.100)\\ &\quad= \langle\Gamma|_{B\hat{B}}(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})(\mathbb{1}_{AB}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})(\rho_A^{-\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}} & (7.4.101)\\ &\quad= \rho_A^{-1}\otimes\langle\Gamma|_{B\hat{B}}\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}|\Gamma\rangle_{B\hat{B}} & (7.4.102)\\ &\quad= \rho_A^{-1}\otimes\sigma_{\hat{A}}^{\mathsf{T}}, & (7.4.103)\end{aligned}$$

where the last equality follows from the fact that

$$\langle\Gamma|_{B\hat{B}}\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}|\Gamma\rangle_{B\hat{B}} = \mathrm{Tr}_{\hat{B}}[\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}] = \sigma_{\hat{A}}^{\mathsf{T}}, \tag{7.4.104}$$

the last equality due to the fact that the transpose is taken on a product basis for $\mathcal{H}_{\hat{A}}\otimes\mathcal{H}_{\hat{B}}$. Therefore, we find that

$$\begin{aligned} Q_\alpha(\rho_{AB}\|\sigma_{AB}) &\geq Q_\alpha(\rho_A\|\sigma_A), \quad \text{for } \alpha\in(1,2], & (7.4.105)\\ Q_\alpha(\rho_{AB}\|\sigma_{AB}) &\leq Q_\alpha(\rho_A\|\sigma_A), \quad \text{for } \alpha\in(0,1), & (7.4.106)\end{aligned}$$

as required. This establishes the data-processing inequality for D_α under partial trace. Combining this with the isometric invariance of D_α and Stinespring's theorem, we conclude that

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad \alpha\in(0,1)\cup(1,2] \tag{7.4.107}$$

for every state ρ, positive semi-definite operator σ, and channel $\mathcal{N}$. ■

[2] Indeed, the function x^β is operator convex for $\beta\in[-1,0)\cup[1,2]$ and operator concave for $\beta\in(0,1]$, where here $\beta = 1-\alpha$.

By taking the limit $\alpha \to 1$ in the statement of data-processing inequality for D_α, and using Proposition 7.22, we immediately obtain the data-processing inequality for the quantum relative entropy, stated previously as Theorem 7.4.

Corollary 7.25 Data-Processing Inequality for Quantum Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then,

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.4.108}$$

The data-processing inequality for the Petz–Rényi relative entropy can be written using the Petz–Rényi relative quasi-entropy Q_α as

$$\frac{1}{\alpha-1}\log_2 Q_\alpha(\rho\|\sigma) \geq \frac{1}{\alpha-1}\log_2 Q_\alpha(\mathcal{N}(\rho)|\mathcal{N}(\sigma)) \tag{7.4.109}$$

for all $\alpha \in [0,1) \cup (1,2]$. Then, since $\alpha - 1$ is negative for $\alpha \in [0,1)$, we can use the monotonicity of the function $\log_2$ to conclude that

$$Q_\alpha(\rho\|\sigma) \geq Q_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad \alpha \in (1,2], \tag{7.4.110}$$

$$Q_\alpha(\rho\|\sigma) \leq Q_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad \alpha \in [0,1). \tag{7.4.111}$$

With the data-processing inequality for the Petz–Rényi relative entropy in hand, it is now straightforward to prove some of the following additional properties.

Proposition 7.26 Additional Properties of Petz–Rényi Relative Entropy

The Petz–Rényi relative entropy D_α satisfies the following properties for every state ρ and positive semi-definite operator σ for $\alpha \in (0,1) \cup (1,2]$:

1. If $\mathrm{Tr}(\sigma) \leq \mathrm{Tr}(\rho) = 1$, then $D_\alpha(\rho\|\sigma) \geq 0$.
2. *Faithfulness*: If $\mathrm{Tr}[\sigma] \leq 1$, we have that $D_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.
3. If $\rho \leq \sigma$, then $D_\alpha(\rho\|\sigma) \leq 0$.
4. For every positive semi-definite operator σ' such that $\sigma' \geq \sigma$, we have $D_\alpha(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma')$.

PROOF:

1. By the data-processing inequality for D_α with respect to the trace channel Tr, and letting $x = \mathrm{Tr}(\rho) = 1$ and $y = \mathrm{Tr}(\sigma)$, we find that

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(x\|y) = \frac{1}{\alpha-1}\log_2 \mathrm{Tr}[x^\alpha y^{1-\alpha}] \tag{7.4.112}$$
$$= \frac{1}{\alpha-1}\log_2(y^{1-\alpha}) \tag{7.4.113}$$
$$= \frac{1-\alpha}{\alpha-1}\log_2 y \tag{7.4.114}$$
$$= -\log_2 y \tag{7.4.115}$$
$$\geq 0, \tag{7.4.116}$$

where the last line follows from the assumption that $y = \mathrm{Tr}(\sigma) \leq 1$.

2. *Proof of faithfulness*: If $\rho = \sigma$, then the following equalities hold for all $\alpha \in (0,1) \cup (1,2)$:

$$D_\alpha(\rho\|\rho) = \frac{1}{\alpha-1}\log_2 \mathrm{Tr}[\rho^\alpha \rho^{1-\alpha}] \tag{7.4.117}$$
$$= \frac{1}{\alpha-1}\log_2 \mathrm{Tr}(\rho) \tag{7.4.118}$$
$$= 0. \tag{7.4.119}$$

Next, suppose that $\alpha \in (0,1) \cup (1,2)$ and $D_\alpha(\rho\|\sigma) = 0$. From the above, we conclude that $D_\alpha(\mathrm{Tr}(\rho)\|\mathrm{Tr}(\sigma)) = -\log_2 y \geq 0$. From the fact that $\log_2 y = 0$ if and only if $y = 1$, we conclude that $D_\alpha(\rho\|\sigma) = 0$ implies $\mathrm{Tr}(\sigma) = \mathrm{Tr}(\rho) = 1$, so that σ is a density operator. Then, for every measurement channel $\mathcal{M}$,

$$D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq D_\alpha(\rho\|\sigma) = 0. \tag{7.4.120}$$

On the other hand, since $\mathrm{Tr}(\sigma) = \mathrm{Tr}(\rho)$,

$$D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \geq D_\alpha(\mathrm{Tr}(\mathcal{M}(\rho))\|\mathrm{Tr}(\mathcal{M}(\sigma))) \tag{7.4.121}$$
$$= D_\alpha(\mathrm{Tr}(\rho)\|\mathrm{Tr}(\sigma)) \tag{7.4.122}$$
$$= 0, \tag{7.4.123}$$

which means that $D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0$ for all measurement channels $\mathcal{M}$. Now, recall that $\mathcal{M}(\rho)$ and $\mathcal{M}(\sigma)$ are effectively probability distributions determined by the measurement. Since the classical Rényi relative entropy is equal to zero if and only if its two arguments are equal, we can conclude that $\mathcal{M}(\rho) = \mathcal{M}(\sigma)$. Since this is true for every measurement channel, we conclude from Theorem 6.4 and the fact that the trace norm is a norm that $\rho = \sigma$.

So we have that $D_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$, as required.

3. Consider that $\rho \le \sigma$ implies that $\sigma - \rho \ge 0$. Then define the following positive semi-definite operators:

$$\hat{\rho} := |0\rangle\langle 0| \otimes \rho, \tag{7.4.124}$$

$$\hat{\sigma} := |0\rangle\langle 0| \otimes \rho + |1\rangle\langle 1| \otimes (\sigma - \rho)\,. \tag{7.4.125}$$

By exploiting the direct-sum property of Petz–Rényi relative entropy in (7.4.46) and the data-processing inequality, we find that

$$0 = D_\alpha(\rho\|\rho) = D_\alpha(\hat{\rho}\|\hat{\sigma}) \ge D_\alpha(\rho\|\sigma), \tag{7.4.126}$$

where the inequality follows from data processing with respect to partial trace over the classical register.

4. Consider the state $\hat{\rho} := |0\rangle\langle 0| \otimes \rho$ and the operator $\hat{\sigma} := |0\rangle\langle 0| \otimes \sigma + |1\rangle\langle 1| \otimes (\sigma' - \sigma)$, which is positive semi-definite because $\sigma' \ge \sigma$ by assumption.

Then

$$\hat{\rho}^\alpha \hat{\sigma}^{1-\alpha} = |0\rangle\langle 0| \otimes \rho^\alpha \sigma^{1-\alpha}, \tag{7.4.127}$$

which implies that

$$D_\alpha(\hat{\rho}\|\hat{\sigma}) = D_\alpha(\rho\|\sigma). \tag{7.4.128}$$

Then, observing that $\mathrm{Tr}_1[\hat{\sigma}] = \sigma'$, and using the data-processing inequality for D_α with respect to the partial trace channel Tr_1, we conclude that

$$D_\alpha(\rho\|\sigma') = D_\alpha(\mathrm{Tr}_1(\hat{\rho})\|\mathrm{Tr}_1(\hat{\sigma})) \le D_\alpha(\hat{\rho}\|\hat{\sigma}) = D_\alpha(\rho\|\sigma), \tag{7.4.129}$$

as required. ■

Proposition 7.27 Joint Convexity & Concavity of the Petz–Rényi Relative Quasi-Entropy

Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$Q_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x) Q_\alpha(\rho_A^x \| \sigma_A^x), \qquad \text{for } \alpha \in (1,2], \tag{7.4.130}$$

$$Q_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \geq \sum_{x\in\mathcal{X}} p(x) Q_\alpha(\rho_A^x \| \sigma_A^x), \qquad \text{for } \alpha \in [0,1), \tag{7.4.131}$$

Furthermore, the Petz–Rényi relative entropy D_α is jointly convex for $\alpha \in [0,1)$:

$$D_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x) D_\alpha(\rho_A^x \| \sigma_A^x), \quad \alpha \in [0,1). \tag{7.4.132}$$

PROOF: By the direct-sum property of Q_α and applying (7.4.110)–(7.4.111) and Proposition 7.17, we conclude (7.4.130)–(7.4.131).

For $\alpha \in [0,1)$, applying $\log_2$ to both sides of (7.4.131) and multiplying by $\frac{1}{\alpha-1}$, which is negative, we conclude that

$$\begin{aligned}\frac{1}{\alpha-1}\log_2 Q_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \\ \leq \frac{1}{\alpha-1}\log_2\left(\sum_{x\in\mathcal{X}} p(x) Q_\alpha(\rho_A^x\|\sigma_A^x)\right).\end{aligned} \tag{7.4.133}$$

Then, since $-\log_2$ is a convex function, and using the definition of D_α in terms of Q_α, we find that

$$D_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x)\frac{1}{\alpha-1}\log_2 Q_\alpha(\rho_A^x\|\sigma_A^x) \tag{7.4.134}$$

$$= \sum_{x \in \mathcal{X}} p(x) D_\alpha(\rho_A^x \| \sigma_A^x), \tag{7.4.135}$$

as required. ■

7.5 Sandwiched Rényi Relative Entropy

A second example of a generalized divergence is the sandwiched Rényi relative entropy, which we define as follows.

Definition 7.28 Sandwiched Rényi Relative Entropy

For all $\alpha \in (0,1) \cup (1,\infty)$, we define the *sandwiched Rényi relative quasi-entropy* for every state ρ and positive semi-definite operator σ as

$$\widetilde{Q}_\alpha(\rho\|\sigma) := \begin{cases} \operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] & \text{if } \alpha \in (0,1), \text{ or} \\ & \alpha \in (1,\infty), \ \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \tag{7.5.1}$$

The *sandwiched Rényi relative entropy* is then defined as

$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1} \log_2 \widetilde{Q}_\alpha(\rho\|\sigma). \tag{7.5.2}$$

Observe that we can use the definition of the Schatten norm from (2.2.86) to write the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ in the following different ways:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{\alpha}{\alpha-1} \log_2 \left\| \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \tag{7.5.3}$$

$$= \frac{\alpha}{\alpha-1} \log_2 \left\| \rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}} \right\|_\alpha \tag{7.5.4}$$

$$= \frac{2\alpha}{\alpha-1} \log_2 \left\| \rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{2\alpha}} \right\|_{2\alpha}. \tag{7.5.5}$$

The expression in (7.5.4) and Proposition 2.8 then lead us to the following variational characterization of $\widetilde{D}_\alpha$ for $\alpha \in (0,1) \cup (1,\infty)$:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \sup_{\substack{\tau>0,\\ \mathrm{Tr}(\tau)=1}} \widetilde{D}_\alpha(\rho\|\sigma;\tau), \tag{7.5.6}$$

where

$$\widetilde{D}_\alpha(\rho\|\sigma;\tau) := \begin{cases} +\infty & \text{if } \alpha > 1,\ \mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma), \\ \frac{\alpha}{\alpha-1}\log_2 \mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\tau^{\frac{\alpha-1}{\alpha}}\right] & \text{otherwise.} \end{cases} \tag{7.5.7}$$

Also observe that for $\alpha = \frac{1}{2}$, the sandwiched Rényi relative entropy $\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma)$ can be expressed as

$$\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma) = -\log_2 F(\rho,\sigma), \tag{7.5.8}$$

where we recall the definition of the fidelity $F(\rho,\sigma)$ from Definition 6.5.

In the case $\alpha \in (1,\infty)$, since $1-\alpha$ is negative, we take the inverse of σ. In case σ is not invertible, we take the inverse of σ on its support. An alternative to this convention is to define $\widetilde{D}_\alpha(\rho\|\sigma)$ for $\alpha > 1$ using only positive definite σ, and for positive semi-definite σ, define

$$\widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \widetilde{D}_\alpha(\rho\|\sigma + \varepsilon\mathbb{1}). \tag{7.5.9}$$

Both alternatives are equivalent, as we now show (similar to what we did in the proofs of Propositions 7.2 and 7.21).

Proposition 7.29

For every state ρ and positive semi-definite operator σ,

$$\widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1}\log_2 \mathrm{Tr}\left[\left(\rho^{\frac{1}{2}}(\sigma+\varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right)^{\alpha}\right]. \tag{7.5.10}$$

Proof: For $\alpha \in (0,1)$, this is immediate from the fact that the logarithm, trace, and power functions are continuous, so that the limit can be brought inside the trace and inside the power $(\sigma+\varepsilon\mathbb{1})^{\frac{1-\alpha}{2\alpha}}$.

For $\alpha \in (1,\infty)$, since $1-\alpha$ is negative and σ is not necessarily invertible, let us start by decomposing the underlying Hilbert space $\mathcal{H}$ as $\mathcal{H} = \mathrm{supp}(\sigma) \oplus \ker(\sigma)$, as in (7.2.6), so that

$$\rho^{\frac{1}{2}} = \begin{pmatrix} (\rho^{\frac{1}{2}})_{0,0} & (\rho^{\frac{1}{2}})_{0,1} \\ (\rho^{\frac{1}{2}})_{0,1}^{\dagger} & (\rho^{\frac{1}{2}})_{1,1} \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.5.11}$$

Then, writing $\mathbb{1} = \Pi_\sigma + \Pi_\sigma^\perp$, where Π_σ is the projection onto the support of σ and $\Pi_\sigma^\perp$ is the projection onto the orthogonal complement of $\text{supp}(\sigma)$, we find that

$$\sigma + \varepsilon\mathbb{1} = \begin{pmatrix} \sigma + \varepsilon\Pi_\sigma & 0 \\ 0 & \varepsilon\Pi_\sigma^\perp \end{pmatrix}, \tag{7.5.12}$$

which implies that

$$(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}} = \begin{pmatrix} (\sigma + \varepsilon\Pi_\sigma)^{\frac{1-\alpha}{\alpha}} & 0 \\ 0 & (\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}} \end{pmatrix}. \tag{7.5.13}$$

If $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then $(\rho^{\frac{1}{2}})_{1,0} = 0$, $(\rho^{\frac{1}{2}})_{1,1} = 0$, and $(\rho^{\frac{1}{2}})_{0,0} = \rho^{\frac{1}{2}}$, which means that

$$\rho^{\frac{1}{2}}(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}} = \begin{pmatrix} (\rho^{\frac{1}{2}})_{0,0}(\sigma + \varepsilon\Pi_\sigma)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{0,0} & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.5.14}$$

so that

$$\lim_{\varepsilon\to 0^+} \frac{1}{\alpha - 1} \log_2 \text{Tr}\left[\left(\rho^{\frac{1}{2}}(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right)^{\alpha}\right] = \widetilde{D}_\alpha(\rho\|\sigma), \quad \alpha \in (1, \infty), \tag{7.5.15}$$

as required.

If $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$, then $(\rho^{\frac{1}{2}})_{1,1}$ is non-zero. In this case, we use the fact that

$$(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}} = \begin{pmatrix} (\sigma + \varepsilon\Pi_\sigma)^{\frac{1-\alpha}{\alpha}} & 0 \\ 0 & (\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}} \end{pmatrix} \geq \begin{pmatrix} 0 & 0 \\ 0 & (\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}} \end{pmatrix} \tag{7.5.16}$$

to conclude that

$$\begin{aligned} &\rho^{\frac{1}{2}}(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}} \\ &\quad\geq \begin{pmatrix} (\rho^{\frac{1}{2}})_{0,1}(\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{0,1}^{\dagger} & (\rho^{\frac{1}{2}})_{0,1}(\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{1,1} \\ (\rho^{\frac{1}{2}})_{1,1}(\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{0,1}^{\dagger} & (\rho^{\frac{1}{2}})_{1,1}(\varepsilon\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{1,1} \end{pmatrix} \end{aligned} \tag{7.5.17}$$

$$= \varepsilon^{\frac{1-\alpha}{\alpha}} \begin{pmatrix} (\rho^{\frac{1}{2}})_{0,1}(\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{0,1}^\dagger & (\rho^{\frac{1}{2}})_{0,1}(\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{1,1} \\ (\rho^{\frac{1}{2}})_{1,1}(\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{0,1}^\dagger & (\rho^{\frac{1}{2}})_{1,1}(\Pi_\sigma^\perp)^{\frac{1-\alpha}{\alpha}}(\rho^{\frac{1}{2}})_{1,1} \end{pmatrix}. \tag{7.5.18}$$

Now, since $\alpha \in (1,\infty)$, we have that $\lim_{\varepsilon\to 0^+} \varepsilon^{\frac{1-\alpha}{\alpha}} = +\infty$; therefore, by continuity arguments similar to those given above, we conclude that

$$\lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \mathrm{Tr}\left[\left(\rho^{\frac{1}{2}}(\sigma+\varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right)^\alpha\right] \geq +\infty, \tag{7.5.19}$$

for the case $\alpha \in (1,\infty)$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$. This implies that

$$\lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1} \log_2 \mathrm{Tr}\left[\left(\rho^{\frac{1}{2}}(\sigma+\varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right)^\alpha\right] = \widetilde{D}_\alpha(\rho\|\sigma), \tag{7.5.20}$$

for the case $\alpha \in (1,\infty)$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$, as required. ■

The Petz–Rényi and sandwiched Rényi relative entropies are two ways of defining a quantum generalization of the classical Rényi relative entropy in (7.4.19). Indeed, if ρ and σ are both classical, commuting states (i.e., both are diagonal in the same basis), then both $D_\alpha(\rho\|\sigma)$ and $\widetilde{D}_\alpha(\rho\|\sigma)$ reduce to the classical Rényi relative entropy in (7.4.19). In general, there are often many (in fact, typically infinitely many) ways to generalize classical quantities to the quantum (i.e., non-commutative) case such that we recover the original classical quantity in the special case of commuting operators. What distinguishes one generalization from another is the role that they play in characterizing operational tasks in quantum information theory, which is a theme explored throughout this book.

We now establish the important fact that the quantum relative entropy is a special case of the sandwiched Rényi relative entropy in the limit $\alpha \to 1$. The proof proceeds very similarly to the proof of the same property for the Petz–Rényi relative entropy.

Proposition 7.30

Let ρ be a state and σ a positive semi-definite operator. Then, in the limit $\alpha \to 1$, the sandwiched Rényi relative entropy converges to the quantum relative entropy:

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma). \tag{7.5.21}$$

Proof: Let us first consider the case $\alpha \in (1, \infty)$. If $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$, then $\widetilde{D}_\alpha(\rho\|\sigma) = +\infty$ for all $\alpha \in (1, \infty)$, so that $\lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\rho\|\sigma) = +\infty$. If $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, then $\widetilde{D}_\alpha(\rho\|\sigma)$ is finite and using (7.5.4) we write

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log_2 \widetilde{Q}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log_2 \operatorname{Tr}\left[\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha\right]. \tag{7.5.22}$$

Let us define the function

$$\widetilde{Q}_{\alpha,\beta}(\rho\|\sigma) \coloneqq \operatorname{Tr}\left[\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\beta\right], \tag{7.5.23}$$

so that $\widetilde{Q}_\alpha(\rho\|\sigma) = \widetilde{Q}_{\alpha,\alpha}(\rho\|\sigma)$. By noting that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ implies $\widetilde{Q}_1(\rho\|\sigma) = \operatorname{Tr}[\rho\Pi_\sigma] = 1$ (where Π_σ is the projection onto the support of σ), and since $\log_2 1 = 0$, we can write $\widetilde{D}_\alpha(\rho\|\sigma)$ as

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{\log_2 \widetilde{Q}_\alpha(\rho\|\sigma) - \log_2 \widetilde{Q}_1(\rho\|\sigma)}{\alpha - 1}, \tag{7.5.24}$$

so that

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\alpha} \log_2 \widetilde{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} \tag{7.5.25}$$

$$= \frac{1}{\ln(2)} \frac{\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_\alpha(\rho\|\sigma)\Big|_{\alpha=1}}{\widetilde{Q}_1(\rho\|\sigma)} \tag{7.5.26}$$

$$= \frac{1}{\ln(2)} \frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1}, \tag{7.5.27}$$

where the first equality follows from the definition of the derivative and the second equality from the derivative of the natural logarithm, along with the chain rule. Using the function $\widetilde{Q}_{\alpha,\beta}$ and the chain rule, we write

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_{\alpha,1}(\rho\|\sigma)\bigg|_{\alpha=1} + \frac{\mathrm{d}}{\mathrm{d}\beta}\widetilde{Q}_{1,\beta}(\rho\|\sigma)\bigg|_{\beta=1}. \tag{7.5.28}$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_{\alpha,1}(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\operatorname{Tr}\left[\rho\sigma^{\frac{1-\alpha}{\alpha}}\right] = -\frac{1}{\alpha^2}\operatorname{Tr}\left[\rho\sigma^{\frac{1-\alpha}{\alpha}} \ln\sigma\right], \tag{7.5.29}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_{\alpha,1}(\rho\|\sigma)\bigg|_{\alpha=1} = -\mathrm{Tr}[\rho\Pi_\sigma \ln \sigma] = -\mathrm{Tr}[\rho \ln \sigma], \tag{7.5.30}$$

where we used the fact that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ to obtain the last equality. Similarly,

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\widetilde{Q}_{1,\beta}(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Tr}\left[\left(\rho^{\frac{1}{2}}\Pi_\sigma\rho^{\frac{1}{2}}\right)^\beta\right] = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Tr}[\rho^\beta] = \mathrm{Tr}[\rho^\beta \ln \rho], \tag{7.5.31}$$

where we again used the fact that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ in order to conclude that $\rho^{\frac{1}{2}}\Pi_\sigma\rho^{\frac{1}{2}} = \rho$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\widetilde{Q}_{1,\beta}(\rho\|\sigma)\bigg|_{\beta=1} = \mathrm{Tr}[\rho \ln \rho]. \tag{7.5.32}$$

So we find that

$$\lim_{\alpha\to 1}\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\ln(2)}\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} = \mathrm{Tr}[\rho \log_2 \rho] - \mathrm{Tr}[\rho \log_2 \sigma] \tag{7.5.33}$$

when $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. Therefore, for $\alpha \in (1, \infty)$,

$$\begin{aligned}
&\lim_{\alpha\to 1^+}\widetilde{D}_\alpha(\rho\|\sigma) \\
&= \begin{cases} \mathrm{Tr}[\rho \log_2 \rho] - \mathrm{Tr}[\rho \log_2 \sigma] & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise} \end{cases} \\
&= D(\rho\|\sigma).
\end{aligned} \tag{7.5.34}$$

Let us now consider the case $\alpha \in (0, 1)$. If $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, then since the limit in (7.5.33) holds from both sides, we find that

$$\lim_{\alpha\to 1^-}\widetilde{D}_\alpha(\rho\|\sigma) = \mathrm{Tr}[\rho \log_2 \rho] - \mathrm{Tr}[\rho \log_2 \sigma]. \tag{7.5.35}$$

If $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$ (and $\mathrm{Tr}[\rho\sigma] \neq 0$), then observe that we can write $\widetilde{D}_\alpha$ as

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{\log_2 \widetilde{Q}_\alpha(\rho\|\sigma) - \log_2 \widetilde{Q}_1(\rho\|\sigma)}{\alpha - 1} + \frac{\log_2 \widetilde{Q}_1(\rho\|\sigma)}{\alpha - 1}, \tag{7.5.36}$$

so that

$$\lim_{\alpha\to 1^-}\widetilde{D}_\alpha(\rho\|\sigma) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log_2 \widetilde{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} + \lim_{\alpha\to 1^-}\frac{\log_2 \mathrm{Tr}[\rho\Pi_\sigma]}{\alpha - 1}, \tag{7.5.37}$$

where we have used $\widetilde{Q}_1(\rho\|\sigma) = \operatorname{Tr}[\rho\Pi_\sigma]$. Now, since $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$ and $\operatorname{Tr}[\rho\sigma] \neq 0$, we have that $0 < \operatorname{Tr}[\rho\Pi_\sigma] < 1$, which means that $\log_2 \operatorname{Tr}[\rho\Pi_\sigma] < 0$. Since $\lim_{\alpha\to 1^-} \frac{1}{\alpha-1} = -\infty$, we find that the second term in (7.5.37) is equal to $+\infty$, which means that $\lim_{\alpha\to 1^-} \widetilde{D}_\alpha(\rho\|\sigma) = +\infty$. Therefore,

$$\begin{aligned}\lim_{\alpha\to 1^-} &\widetilde{D}_\alpha(\rho\|\sigma)\\ &= \begin{cases} \operatorname{Tr}[\rho\log_2\rho] - \operatorname{Tr}[\rho\log_2\sigma] & \text{if } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty & \text{otherwise} \end{cases} \\ &= D(\rho\|\sigma).\end{aligned} \tag{7.5.38}$$

To conclude, we have that $\lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1^-} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$, which means that (7.5.21) holds. ■

In the following proposition, we state several basic properties of the sandwiched Rényi relative entropy. The proofs of the first four properties are analogous to those of the same properties of the Petz–Rényi relative entropy. The last property in the proposition establishes that the sandwiched Rényi relative entropy is always less than or equal to the Petz–Rényi relative entropy.

Proposition 7.31 Properties of Sandwiched Rényi Relative Entropy

For all states ρ, ρ_1, ρ_2 and positive semi-definite operators $\sigma, \sigma_1, \sigma_2$, the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ satisfies the following properties:

1. *Isometric invariance*: For all $\alpha \in (0,1) \cup (1,\infty)$ and for every isometry V,
$$\widetilde{D}_\alpha(\rho\|\sigma) = \widetilde{D}_\alpha(V\rho V^\dagger\|V\sigma V^\dagger). \tag{7.5.39}$$
2. *Monotonicity in α*: For all $\alpha \in (0,1) \cup (1,\infty)$, $\widetilde{D}_\alpha$ is monotonically increasing in α, i.e., $\alpha < \beta$ implies $\widetilde{D}_\alpha(\rho\|\sigma) \leq \widetilde{D}_\beta(\rho\|\sigma)$.
3. *Additivity*: For all $\alpha \in (0,1) \cup (1,\infty)$,
$$\widetilde{D}_\alpha(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \widetilde{D}_\alpha(\rho_1\|\sigma_1) + \widetilde{D}_\alpha(\rho_2\|\sigma_2). \tag{7.5.40}$$
4. *Direct-sum property*: Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $q : \mathcal{X} \to [0,\infty)$ be a positive function on $\mathcal{X}$. Let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states

on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$\widetilde{Q}_\alpha(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha}\widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.5.41}$$

where

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.5.42}$$

$$\sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.5.43}$$

5. If $\rho_1 \leq \gamma\rho_2$ for some $\gamma \geq 1$, then

$$\widetilde{D}_\alpha(\rho_1\|\sigma) \leq \frac{\alpha}{\alpha-1}\log_2\gamma + \widetilde{D}_\alpha(\rho_2\|\sigma), \quad \alpha > 1. \tag{7.5.44}$$

6. For all $\alpha \in (0,1)\cup(1,\infty)$, the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ is always less than or equal to the Petz–Rényi relative entropy D_α, i.e.,

$$\widetilde{D}_\alpha(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma). \tag{7.5.45}$$

Furthermore, for $\alpha \in (0,1)$, we have

$$\alpha D_\alpha(\rho\|\sigma) + (1-\alpha)(-\log_2\operatorname{Tr}[\sigma]) \leq \widetilde{D}_\alpha(\rho\|\sigma). \tag{7.5.46}$$

Proof:

1. *Proof of isometric invariance*: Let us start by writing $\widetilde{D}_\alpha(\rho\|\sigma)$ using the function $\|\cdot\|_\alpha$ as in (7.5.4):

$$\widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0}\frac{\alpha}{\alpha-1}\log_2\left\|\rho^{\frac{1}{2}}(\sigma+\varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha, \tag{7.5.47}$$

where we have also made use of the fact that for positive semi-definite operators, $\widetilde{D}_\alpha(\rho\|\sigma)$ can be defined as in (7.5.10). Now,

$$\begin{aligned}&\widetilde{D}_\alpha(V\rho V^\dagger\|V\sigma V^\dagger)\\&\quad= \lim_{\varepsilon\to 0}\frac{\alpha}{\alpha-1}\log_2\left\|(V\rho V^\dagger)^{\frac{1}{2}}(V\sigma V^\dagger+\varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}(V\rho V^\dagger)^{\frac{1}{2}}\right\|_\alpha.\end{aligned} \tag{7.5.48}$$

Since $(V\rho V^\dagger)^{\frac{1}{2}} = V\rho^{\frac{1}{2}}V^\dagger$, we find that

$$\begin{aligned}
&\left\|(V\rho V^\dagger)^{\frac{1}{2}}(V\sigma V^\dagger + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}(V\rho V^\dagger)^{\frac{1}{2}}\right\|_\alpha \\
&\quad = \left\|V\rho^{\frac{1}{2}}V^\dagger(V\sigma V^\dagger + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V\rho^{\frac{1}{2}}V^\dagger\right\|_\alpha && (7.5.49)\\
&\quad = \left\|\rho^{\frac{1}{2}}V^\dagger(V\sigma V^\dagger + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V\rho^{\frac{1}{2}}\right\|_\alpha, && (7.5.50)
\end{aligned}$$

where the last equality follows from the isometric invariance of the function $\|\cdot\|_\alpha$. Now, let $\Pi := VV^\dagger$ be the projection onto the image of V, and let $\hat{\Pi} := \mathbb{1} - \Pi$. Then, we write

$$V\sigma V^\dagger + \varepsilon\mathbb{1} = V\sigma V^\dagger + \varepsilon\Pi + \varepsilon\hat{\Pi} = V(\sigma + \varepsilon\mathbb{1})V^\dagger + \varepsilon\hat{\Pi}. \tag{7.5.51}$$

Since $V(\sigma + \varepsilon\mathbb{1})V^\dagger$ and $\varepsilon\hat{\Pi}$ are supported on orthogonal subspaces, we obtain

$$(V\sigma V^\dagger + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}} = V(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V^\dagger + \varepsilon^{\frac{1-\alpha}{\alpha}}\hat{\Pi}. \tag{7.5.52}$$

Continuing from (7.5.50), we thus find that

$$\begin{aligned}
&\left\|\rho^{\frac{1}{2}}V^\dagger(V\sigma V^\dagger + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V\rho^{\frac{1}{2}}\right\|_\alpha \\
&\quad = \left\|\rho^{\frac{1}{2}}V^\dagger\left(V(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V^\dagger + \varepsilon^{\frac{1-\alpha}{\alpha}}\hat{\Pi}\right)V\rho^{\frac{1}{2}}\right\|_\alpha && (7.5.53)\\
&\quad = \left\|\rho^{\frac{1}{2}}V^\dagger V(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}V^\dagger V\rho^{\frac{1}{2}} + \varepsilon^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}V^\dagger\hat{\Pi}V\rho^{\frac{1}{2}}\right\|_\alpha && (7.5.54)\\
&\quad = \left\|\rho^{\frac{1}{2}}(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha, && (7.5.55)
\end{aligned}$$

where the last equality follows from the fact that $V^\dagger\hat{\Pi}V = V^\dagger V - V^\dagger VV^\dagger V = \mathbb{1} - \mathbb{1} = 0$. Therefore,

$$\begin{aligned}
\widetilde{D}_\alpha(V\rho V^\dagger\|V\sigma V^\dagger) &= \lim_{\varepsilon\to 0}\frac{\alpha}{\alpha - 1}\log_2\left\|\rho^{\frac{1}{2}}(\sigma + \varepsilon\mathbb{1})^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha \\
&= \widetilde{D}_\alpha(\rho\|\sigma),
\end{aligned} \tag{7.5.56}$$

as required.

2. *Proof of monotonicity in α*: We make use of the function $\widetilde{D}_\alpha(\rho\|\sigma;\tau)$ defined in (7.5.7), which we can write as

$$\widetilde{D}_\alpha(\rho\|\sigma;\tau) = -\frac{1}{\gamma}\log_2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle = -\frac{1}{\gamma}\frac{\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}{\ln(2)}, \tag{7.5.57}$$

where $X = \tau^{-1} \otimes \sigma^{\mathsf{T}}$, $\gamma := \frac{1-\alpha}{\alpha}$ and $|\varphi^\rho\rangle = (\rho^{\frac{1}{2}} \otimes \mathbb{1})|\Gamma\rangle$ is a purification of ρ. We prove monotonicity of this quantity by taking its derivative with respect to α and showing that it is non-negative. Since $\frac{\mathrm{d}\gamma}{\mathrm{d}\alpha} = -\frac{1}{\alpha^2}$, we can express the derivative with respect to α in terms of the derivative with respect to γ using $\frac{\mathrm{d}}{\mathrm{d}\alpha} = \frac{\mathrm{d}}{\mathrm{d}\gamma}\frac{\mathrm{d}\gamma}{\mathrm{d}\alpha} = -\frac{1}{\alpha^2}\frac{\mathrm{d}}{\mathrm{d}\gamma}$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{D}_\alpha(\rho\|\sigma;\tau) = -\frac{1}{\alpha^2}\frac{\mathrm{d}}{\mathrm{d}\gamma}\left(-\frac{1}{\gamma}\frac{\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}{\ln(2)}\right) = \frac{1}{\alpha^2}\frac{\mathrm{d}f}{\mathrm{d}\gamma}, \tag{7.5.58}$$

where

$$f(\gamma) := \frac{1}{\gamma}\frac{\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}{\ln(2)}. \tag{7.5.59}$$

Now,

$$\frac{\mathrm{d}f}{\mathrm{d}\gamma} = \frac{1}{\ln(2)}\left(-\frac{1}{\gamma^2}\ln\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle + \frac{1}{\gamma}\frac{\langle\varphi^\rho|X^\gamma\ln X|\varphi^\rho\rangle}{\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}\right) \tag{7.5.60}$$

$$= \frac{-\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle\log_2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle + \langle\varphi^\rho|X^\gamma\log_2 X^\gamma|\varphi^\rho\rangle}{\gamma^2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}. \tag{7.5.61}$$

Now, let $g(x) := x\log_2 x$. Then, we can write

$$\frac{\mathrm{d}f}{\mathrm{d}\gamma} = \frac{\langle\varphi^\rho|g(X^\gamma)|\varphi^\rho\rangle - g(\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle)}{\gamma^2\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle}. \tag{7.5.62}$$

Since g is operator convex, the operator Jensen inequality in (2.3.23) implies that

$$\langle\varphi^\rho|g(X^\gamma)|\varphi^\rho\rangle \geq g(\langle\varphi^\rho|X^\gamma|\varphi^\rho\rangle), \tag{7.5.63}$$

which implies that $\frac{\mathrm{d}f}{\mathrm{d}\gamma} \geq 0$. Therefore, $\widetilde{D}_\alpha(\rho\|\sigma;\tau)$ is monotonically increasing in α for all ρ, σ, τ. By (7.5.6), we conclude that $\widetilde{D}_\alpha(\rho\|\sigma)$ is monotonically increasing in α, as required.

3. *Proof of additivity*: When all quantities are finite, we have that

$$\begin{aligned}&\widetilde{D}_\alpha(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2)\\ &\quad= \frac{1}{\alpha-1}\log_2\operatorname{Tr}\left[\left((\sigma_1\otimes\sigma_2)^{\frac{1-\alpha}{2\alpha}}(\rho_1\otimes\rho_2)(\sigma_1\otimes\sigma_2)^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right].\end{aligned} \tag{7.5.64}$$

Using the fact that $(X\otimes Y)^\beta = X^\beta\otimes Y^\beta$ for all positive semi-definite operators X, Y and all $\beta\in\mathbb{R}$, we obtain

$$\widetilde{Q}_\alpha(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2)$$

$$= \mathrm{Tr}\left[\left((\sigma_1 \otimes \sigma_2)^{\frac{1-\alpha}{2\alpha}}(\rho_1 \otimes \rho_2)(\sigma_1 \otimes \sigma_2)^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \tag{7.5.65}$$

$$= \mathrm{Tr}\left[\left(\left(\sigma_1^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_2^{\frac{1-\alpha}{2\alpha}}\right)(\rho_1 \otimes \rho_2)\left(\sigma_1^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_2^{\frac{1-\alpha}{2\alpha}}\right)\right)^{\alpha}\right] \tag{7.5.66}$$

$$= \mathrm{Tr}\left[\left(\sigma_1^{\frac{1-\alpha}{2\alpha}}\rho_1\sigma_1^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_2^{\frac{1-\alpha}{2\alpha}}\rho_2\sigma_2^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \tag{7.5.67}$$

$$= \mathrm{Tr}\left[\left(\sigma_1^{\frac{1-\alpha}{2\alpha}}\rho_1\sigma_1^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha} \otimes \left(\sigma_2^{\frac{1-\alpha}{2\alpha}}\rho_2\sigma_2^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \tag{7.5.68}$$

$$= \mathrm{Tr}\left[\left(\sigma_1^{\frac{1-\alpha}{2\alpha}}\rho_1\sigma_1^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \cdot \mathrm{Tr}\left[\left(\sigma_2^{\frac{1-\alpha}{2\alpha}}\rho_2\sigma_2^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \tag{7.5.69}$$

$$= \widetilde{Q}_\alpha(\rho_1\|\sigma_1) \cdot \widetilde{Q}_\alpha(\rho_2\|\sigma_2). \tag{7.5.70}$$

Applying $\frac{1}{\alpha-1}\log_2$ and definitions, additivity follows.

4. *Proof of the direct-sum property*: Define the classical–quantum state and operator, respectively, as

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \quad \sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.5.71}$$

Then, since

$$\sigma_{XA}^{\frac{1-\alpha}{2\alpha}} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes (q(x)\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}, \tag{7.5.72}$$

we find that

$$\sigma_{XA}^{\frac{1-\alpha}{2\alpha}}\rho_{XA}\sigma_{XA}^{\frac{1-\alpha}{2\alpha}} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes (q(x)\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}(p(x)\rho_A^x)(q(x)\sigma_A^x)^{\frac{1-\alpha}{2\alpha}} \tag{7.5.73}$$

$$= \sum_{x\in\mathcal{X}} p(x)q(x)^{\frac{1-\alpha}{\alpha}}|x\rangle\langle x|_X \otimes (\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}\rho_A^x(\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}, \tag{7.5.74}$$

which means that

$$\left(\sigma_{XA}^{\frac{1-\alpha}{2\alpha}}\rho_{XA}\sigma_{XA}^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha} = \sum_{x\in\mathcal{X}} p(x)^{\alpha}q(x)^{1-\alpha}|x\rangle\langle x|_X \otimes \left((\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}\rho_A^x(\sigma_A^x)^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}. \tag{7.5.75}$$

Taking the trace on both sides of this equation, and using the definition of $\widetilde{Q}_\alpha$, we conclude that

$$\widetilde{Q}_\alpha(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)^{\alpha}q(x)^{1-\alpha}\widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.5.76}$$

as required.

5. From the assumption that $\rho_1 \leq \gamma\rho_2$, we have that

$$\sigma^{\frac{1-\alpha}{2\alpha}}\rho_1\sigma^{\frac{1-\alpha}{2\alpha}} \leq \gamma\sigma^{\frac{1-\alpha}{2\alpha}}\rho_2\sigma^{\frac{1-\alpha}{2\alpha}}. \tag{7.5.77}$$

Then, using (2.2.143), we obtain

$$\operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho_1\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \leq \gamma^{\alpha}\operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho_2\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right]. \tag{7.5.78}$$

The result follows after applying the logarithm and dividing by $\alpha - 1$ on both sides of this inequality.

6. This follows from the Araki–Lieb–Thirring inequalities, which we state here without proof (see the Bibliographic Notes in Section 7.13): for positive semi-definite operators A and B acting on a finite-dimensional Hilbert space, and for $q \geq 0$, the following inequalities hold

(a) $\operatorname{Tr}\left[\left(B^{\frac{1}{2}}AB^{\frac{1}{2}}\right)^{rq}\right] \geq \operatorname{Tr}\left[\left(B^{\frac{r}{2}}A^rB^{\frac{r}{2}}\right)^{q}\right]$ for all $r \in [0,1]$.

(b) $\operatorname{Tr}\left[\left(B^{\frac{1}{2}}AB^{\frac{1}{2}}\right)^{rq}\right] \leq \operatorname{Tr}\left[\left(B^{\frac{r}{2}}A^rB^{\frac{r}{2}}\right)^{q}\right]$ for all $r \geq 1$.

For $\alpha \in (0,1)$, we make use of the first of these inequalities. In particular, we set $q = 1$, $r = \alpha$, $A = \rho$ and $B = \sigma^{\frac{1-\alpha}{\alpha}}$. Then, letting $\gamma := \frac{1-\alpha}{2\alpha}$, we obtain

$$\operatorname{Tr}\left[(\sigma^{\gamma}\rho\sigma^{\gamma})^{\alpha}\right] \geq \operatorname{Tr}[\sigma^{\alpha\gamma}\rho^{\alpha}\sigma^{\alpha\gamma}] = \operatorname{Tr}\left[\sigma^{\frac{1-\alpha}{2}}\rho^{\alpha}\sigma^{\frac{1-\alpha}{2}}\right] = \operatorname{Tr}[\rho^{\alpha}\sigma^{1-\alpha}], \tag{7.5.79}$$

where the last equality holds by cyclicity of the trace. Since the logarithm function is monotonically increasing, this inequality implies that

$$\log_2\operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \geq \log_2\operatorname{Tr}\left[\rho^{\alpha}\sigma^{1-\alpha}\right]. \tag{7.5.80}$$

Finally, since $\alpha - 1 < 0$ for all $\alpha \in (0,1)$, we conclude that

$$\frac{1}{\alpha-1}\log_2\operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \leq \frac{1}{\alpha-1}\log_2\operatorname{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]. \tag{7.5.81}$$

That is, $\widetilde{D}_{\alpha}(\rho\|\sigma) \leq D_{\alpha}(\rho\|\sigma)$, as required.

For $\alpha \in (1,\infty)$, we make use of the second Araki–Lieb–Thirring inequality above. As before, we let $q = 1$, $r = \alpha$, $A = \rho$, and $B^{\frac{1}{2}} = \sigma^{\gamma}$. We find that

$$\operatorname{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \leq \operatorname{Tr}\left[\sigma^{\frac{1-\alpha}{2}}\rho^{\alpha}\sigma^{\frac{1-\alpha}{2}}\right] = \operatorname{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]. \tag{7.5.82}$$

Then, since the logarithm function is a monotonically increasing function and $\alpha - 1 > 0$ for all $\alpha \in (1, \infty)$, we conclude that

$$\frac{1}{\alpha - 1} \log_2 \mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \leq \frac{1}{\alpha - 1} \log_2 \mathrm{Tr}[\rho^{\alpha} \sigma^{1-\alpha}], \tag{7.5.83}$$

i.e., $\widetilde{D}_{\alpha}(\rho\|\sigma) \leq D_{\alpha}(\rho\|\sigma)$, as required.

For the inequality in (7.5.46), with ρ a state and σ a positive semi-definite operator, we use the following "reverse" Araki–Lieb–Thirring inequality, which we state here without proof (see the Bibliographic Notes in Section 7.13):

$$\mathrm{Tr}\left[\left(B^{\frac{1}{2}} A B^{\frac{1}{2}}\right)^{rq}\right] \leq \left(\mathrm{Tr}\left[\left(B^{\frac{r}{2}} A^{r} B^{\frac{r}{2}}\right)^{q}\right]\right)^{r} \left\|A^{\frac{1-r}{2}}\right\|_{a}^{2rq} \left\|B^{\frac{1-r}{2}}\right\|_{b}^{2rq}. \tag{7.5.84}$$

This inequality holds for all positive semi-definite operators A and B, as well as for $q > 0$, $r \in (0, 1]$, $a, b \in (0, \infty]$, and for $\frac{1}{2rq} = \frac{1}{2q} + \frac{1}{a} + \frac{1}{b}$. Taking $q = 1$, $r = \alpha$, $A = \rho$, $B = \sigma^{\frac{1-\alpha}{\alpha}}$, $a = \frac{2}{1-\alpha}$, and $b = \frac{2\alpha}{(1-\alpha)^2}$, we obtain

$$\begin{aligned} &\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \\ &\qquad \leq \left(\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2}} \rho \sigma^{\frac{1-\alpha}{2}}\right)\right]\right)^{\alpha} \left\|\rho^{\frac{1-\alpha}{2}}\right\|_{\frac{2}{1-\alpha}}^{2\alpha} \left\|\sigma^{\frac{(1-\alpha)^2}{2\alpha}}\right\|_{\frac{2\alpha}{(1-\alpha)^2}}^{2\alpha}. \end{aligned} \tag{7.5.85}$$

Now, because ρ is a state,

$$\begin{aligned} \left\|\rho^{\frac{1-\alpha}{2}}\right\|_{\frac{2}{1-\alpha}}^{2\alpha} &= \left(\mathrm{Tr}\left[\left|\rho^{\frac{1-\alpha}{2}}\right|^{\frac{2}{1-\alpha}}\right]\right)^{\alpha(1-\alpha)} && (7.5.86) \\ &= \left(\mathrm{Tr}\left[\left(\rho^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}\right]\right)^{\alpha(1-\alpha)} && (7.5.87) \\ &= (\mathrm{Tr}[\rho])^{\alpha(1-\alpha)} && (7.5.88) \\ &= 1. && (7.5.89) \end{aligned}$$

For σ, we obtain

$$\left\|\sigma^{\frac{(1-\alpha)^2}{2\alpha}}\right\|_{\frac{2\alpha}{(1-\alpha)^2}}^{2\alpha} = (\mathrm{Tr}[\sigma])^{(1-\alpha)^2}. \tag{7.5.90}$$

Therefore,

$$\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \leq \left(\mathrm{Tr}[\rho^{\alpha} \sigma^{1-\alpha}]\right)^{\alpha} \cdot (\mathrm{Tr}[\sigma])^{(1-\alpha)^2}. \tag{7.5.91}$$

Taking the logarithm of both sides and multiplying by $\frac{1}{\alpha-1}$, which is negative for $\alpha \in (0,1)$, we obtain the inequality in (7.5.46). ■

The monotonicity in α of the sandwiched Rényi relative entropy establishes an inequality relating the quantum relative entropy and the fidelity of quantum states ρ and σ, by picking $\alpha = 1$ and $\alpha = 1/2$, respectively, and applying Proposition 7.30:

$$D(\rho\|\sigma) \geq -\log_2 F(\rho,\sigma). \tag{7.5.92}$$

We can modify the lower bound a bit to establish an inequality relating the quantum relative entropy and the trace distance:

Corollary 7.32 Quantum Pinsker Inequality

Let ρ and σ be quantum states. Then the following inequality holds

$$D(\rho\|\sigma) \geq \frac{1}{4\ln 2}\|\rho-\sigma\|_1^2. \tag{7.5.93}$$

REMARK: The constant prefactor can be improved from $\frac{1}{4\ln 2}$ to $\frac{1}{2\ln 2}$, but we do not give a proof here (please consult the Bibliographic Notes in Section 7.13).

PROOF: We can rewrite (7.5.92) as follows:

$$D(\rho\|\sigma) \geq -\frac{1}{\ln 2}\ln F(\rho,\sigma) \tag{7.5.94}$$

$$= -\frac{1}{\ln 2}\ln[1-(1-F(\rho,\sigma))] \tag{7.5.95}$$

$$\geq \frac{1}{\ln 2}(1-F(\rho,\sigma)) \tag{7.5.96}$$

$$\geq \frac{1}{4\ln 2}\|\rho-\sigma\|_1^2. \tag{7.5.97}$$

The second inequality follows from $-\ln(1-x) \geq x$, which holds for $x \in [0,1]$. The final inequality follows from Theorem 6.14. ■

Like the quantum relative entropy and the Petz–Rényi relative entropy, the sandwiched Rényi relative entropy is faithful, meaning that for all states ρ, σ and all $\alpha \in (0,1) \cup (1,\infty)$,

$$\widetilde{D}_\alpha(\rho\|\sigma) = 0 \quad \Longleftrightarrow \quad \rho = \sigma. \tag{7.5.98}$$

We prove this in Proposition 7.36 below.

We now prove the data-processing inequality for the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ for $\alpha \in [1/2, 1) \cup (1, \infty)$. This, along with Proposition 7.30, gives us a different way (apart from using data-processing inequality for the Petz–Rényi relative entropy) to prove the data-processing inequality for the quantum relative entropy.

Theorem 7.33 Data-Processing Inequality for Sandwiched Rényi Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then, for all $\alpha \in [1/2, 1) \cup (1, \infty)$,

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.5.99}$$

Proof: This proof follows steps very similar to those in the proof of the data-processing inequality for the Petz–Rényi relative entropy (Theorem 7.24), with the key difference being that in this case we make use of the fact that the sandwiched Rényi relative entropy can be written as the optimization in (7.5.6).

From Stinespring's theorem (Theorem 4.3), we know that the action of a channel $\mathcal{N}$ on a linear operator X can be written as

$$\mathcal{N}(X) = \mathrm{Tr}_E[VXV^\dagger], \tag{7.5.100}$$

for some V, where V is an isometry and E is an auxiliary system with dimension $d_E \geq \mathrm{rank}(\Gamma^{\mathcal{N}})$. As stated in (7.5.39), $\widetilde{D}_\alpha$ is isometrically invariant. Therefore, it suffices to prove the data-processing inequality for $\widetilde{D}_\alpha$ under partial trace; i.e., it suffices to show that for every state ρ_{AB}, every positive semi-definite operator σ_{AB}, and all $\alpha \in [1/2, 1) \cup (1, \infty)$:

$$\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \geq \widetilde{D}_\alpha(\rho_A\|\sigma_A). \tag{7.5.101}$$

We now proceed to prove this inequality. We prove it for ρ_{AB}, and hence ρ_A, invertible, as well as for σ_{AB} and σ_A invertible. The result follows in the general case of ρ_{AB} and/or ρ_A non-invertible, as well as σ_{AB} and/or σ_A non-invertible, by applying the result to the invertible operators $(1-\delta)\rho_{AB} + \delta\pi_{AB}$ and $\sigma_{AB} + \varepsilon\mathbb{1}_{AB}$, with $\delta, \varepsilon > 0$, and taking the limits $\varepsilon \to 0^+$ and $\delta \to 0^+$, since

$$\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \widetilde{D}_\alpha((1-\delta)\rho_{AB} + \delta\pi_{AB}\|\sigma_{AB} + \varepsilon\mathbb{1}_{AB}), \tag{7.5.102}$$

$$\widetilde{D}_\alpha(\rho_A\|\sigma_A) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+}\widetilde{D}_\alpha((1-\delta)\rho_A + \delta\pi_A\|\sigma_A + d_B\varepsilon\mathbb{1}_A), \tag{7.5.103}$$

which can be verified in a similar manner to the proof of (7.5.10) in Proposition 7.29.

Let us start by defining the quantity $\widetilde{Q}_\alpha(\rho\|\sigma;\tau)$ as

$$\widetilde{Q}_\alpha(\rho\|\sigma;\tau) \coloneqq \langle\varphi^\rho|(\tau^{-1}\otimes\sigma^{\mathsf{T}})^{\frac{1-\alpha}{\alpha}}|\varphi^\rho\rangle, \tag{7.5.104}$$

where τ is a positive definite state and

$$|\varphi^\rho\rangle \coloneqq (\rho^{\frac{1}{2}}\otimes\mathbb{1})|\Gamma\rangle \tag{7.5.105}$$

is a purification of ρ. We note that

$$\widetilde{Q}_\alpha(\rho\|\sigma;\tau) = \operatorname{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}}\tau^{\frac{\alpha-1}{\alpha}}\right] \tag{7.5.106}$$

so that

$$\widetilde{D}_\alpha(\rho\|\sigma;\tau) = \frac{\alpha}{\alpha-1}\log_2\widetilde{Q}_\alpha(\rho\|\sigma;\tau), \tag{7.5.107}$$

where we recall the quantity $\widetilde{D}_\alpha(\rho\|\sigma;\tau)$ defined in (7.5.7). Now, to prove (7.5.101), we show that for every positive definite state ω_A, there exists a positive definite state τ_{AB} such that

$$\begin{aligned}\widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) &\geq \widetilde{Q}_\alpha(\rho_A\|\sigma_A;\omega_A), &&\text{for } \alpha\in(1,\infty),\\ \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) &\leq \widetilde{Q}_\alpha(\rho_A\|\sigma_A;\omega_A), &&\text{for } \alpha\in[1/2,1)\,.\end{aligned} \tag{7.5.108}$$

With these two inequalities, along with (7.5.107) and (7.5.6), the result follows.

Consider that

$$\widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) = \langle\varphi^{\rho_{AB}}|f(\tau_{AB}^{-1}\otimes\sigma_{\hat{A}\hat{B}}^{\mathsf{T}})|\varphi^{\rho_{AB}}\rangle, \tag{7.5.109}$$

$$\widetilde{Q}_\alpha(\rho_A\|\sigma_A;\omega_A) = \langle\varphi^{\rho_A}|f(\omega_A^{-1}\otimes\sigma_{\hat{A}}^{\mathsf{T}})|\varphi^{\rho_A}\rangle, \tag{7.5.110}$$

where we have set

$$f(x) \coloneqq x^{\frac{1-\alpha}{\alpha}} \tag{7.5.111}$$

and

$$|\varphi^{\rho_{AB}}\rangle = (\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{AB\hat{A}\hat{B}} = (\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}}, \tag{7.5.112}$$

$$|\varphi^{\rho_A}\rangle = (\rho_A^{\frac{1}{2}}\otimes\mathbb{1}_{\hat{A}})|\Gamma\rangle_{A\hat{A}}. \tag{7.5.113}$$

Now, let us use the same isometry $V_{A\hat{A}\to AB\hat{A}\hat{B}}$ from (7.4.91) that we used in the proof of data-processing inequality for the Petz–Rényi relative entropy; that is, let

$$V_{A\hat{A}\to AB\hat{A}\hat{B}} \coloneqq \rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}}. \tag{7.5.114}$$

Recall that

$$\begin{aligned} V_{A\hat{A}\to AB\hat{A}\hat{B}}|\varphi^{\rho_A}\rangle_{A\hat{A}} &= \rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}})(\rho_A^{\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}} & (7.5.115)\\ &= (\rho_{AB}^{\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}\hat{B}})|\Gamma\rangle_{A\hat{A}}|\Gamma\rangle_{B\hat{B}} & (7.5.116)\\ &= |\varphi^{\rho_{AB}}\rangle. & (7.5.117)\end{aligned}$$

We thus obtain, for all τ_{AB},

$$\begin{aligned} \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) &= \langle\varphi^{\rho_A}|V^\dagger f(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V|\varphi^{\rho_A}\rangle \\ &\geq \langle\varphi^{\rho_A}|f(V^\dagger(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V)|\varphi^{\rho_A}\rangle \end{aligned} \tag{7.5.118}$$

for $\alpha \in (1,\infty)$ and

$$\begin{aligned} \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) &= \langle\varphi^{\rho_A}|V^\dagger f(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V|\varphi^{\rho_A}\rangle \\ &\leq \langle\varphi^{\rho_A}|f(V^\dagger(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V)|\varphi^{\rho_A}\rangle \end{aligned} \tag{7.5.119}$$

for $\alpha \in [1/2, 1)$, where to obtain the last inequality in each case we used the operator Jensen inequality (Theorem 2.16), which is applicable since for $\alpha \in (1,\infty)$ the function f in (7.5.111) is operator convex and for $\alpha \in [1/2, 1)$ it is operator concave.

Now, recall that to conclude (7.5.101), we should perform an optimization over invertible states τ_{AB} as per the definition in (7.5.6) of $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB})$ in order to obtain $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB})$. Since we only require a lower bound on $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB})$, we can obtain the lower bound in (7.5.101) on $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB})$ by simply picking a particular state τ_{AB} in the optimization in (7.5.6). Let us therefore take

$$\tau_{AB} = \xi_{AB}(\omega_A) \coloneqq \rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}}\omega_A\rho_A^{-\frac{1}{2}} \otimes \mathbb{1}_B)\rho_{AB}^{\frac{1}{2}}, \tag{7.5.120}$$

where ω_A is an arbitrary invertible state. Note that this choice of τ_{AB} is indeed a state because it is the result of applying the Petz recovery channel $\mathcal{P}_{\rho_{AB},\mathrm{Tr}_B}$ defined in (4.6.30) to ω_A. It is also invertible; in particular,

$$\tau_{AB}^{-1} = [\xi_{AB}(\omega_A)]^{-1} = \rho_{AB}^{-\frac{1}{2}}(\rho_A^{\frac{1}{2}}\omega_A^{-1}\rho_A^{\frac{1}{2}} \otimes \mathbb{1}_B)\rho_{AB}^{-\frac{1}{2}}. \tag{7.5.121}$$

With the choice in (7.5.120) for τ_{AB}, we find that

$$
\begin{aligned}
&V^{\dagger}(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})V \\
&\quad = \langle\Gamma|_{B\hat{B}}(\rho_A^{-\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}})\rho_{AB}^{\frac{1}{2}}(\tau_{AB}^{-1} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}})\rho_{AB}^{\frac{1}{2}}(\rho_A^{-\frac{1}{2}} \otimes \mathbb{1}_{\hat{A}})|\Gamma\rangle_{B\hat{B}} && (7.5.122)\\
&\quad = \langle\Gamma|_{B\hat{B}}\left(\rho_A^{-\frac{1}{2}}\rho_{AB}^{\frac{1}{2}}\rho_{AB}^{-\frac{1}{2}}(\rho_A^{\frac{1}{2}}\omega_A^{-1}\rho_A^{\frac{1}{2}} \otimes \mathbb{1}_B)\rho_{AB}^{-\frac{1}{2}}\rho_{AB}^{\frac{1}{2}}\rho_A^{-\frac{1}{2}} \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}}\right)|\Gamma\rangle_{B\hat{B}} && (7.5.123)\\
&\quad = \langle\Gamma|_{B\hat{B}}\omega_A^{-1} \otimes \mathbb{1}_B \otimes \sigma_{\hat{A}\hat{B}}^{\mathsf{T}}|\Gamma\rangle_{B\hat{B}} && (7.5.124)\\
&\quad = \omega_A^{-1} \otimes \langle\Gamma|_{B\hat{B}}\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}|\Gamma\rangle_{B\hat{B}} && (7.5.125)\\
&\quad = \omega_A^{-1} \otimes \sigma_{\hat{A}}^{\mathsf{T}}, && (7.5.126)
\end{aligned}
$$

where we have used the fact that $\langle\Gamma|_{B\hat{B}}\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}|\Gamma\rangle_{B\hat{B}} = \mathrm{Tr}_{\hat{B}}[\sigma_{\hat{A}\hat{B}}^{\mathsf{T}}] = \sigma_{\hat{A}}^{\mathsf{T}}$, the last equality due to the fact that the transpose is taken on a product basis for $\mathcal{H}_{\hat{A}} \otimes \mathcal{H}_{\hat{B}}$.

Therefore, for $\alpha \in (1, \infty)$, taking the logarithm on both sides of (7.5.118) and using the state in (7.5.120), we find that

$$
\begin{aligned}
\log_2 \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB}; \xi_{AB}(\omega_A)) &\geq \log_2\langle\varphi^{\rho_A}|f(\omega_A^{-1} \otimes \sigma_{\hat{A}}^{\mathsf{T}})|\varphi^{\rho_A}\rangle && (7.5.127)\\
&= \log_2 \widetilde{Q}_\alpha(\rho_A\|\sigma_A; \omega_A). && (7.5.128)
\end{aligned}
$$

Multiplying both sides of this inequality by $\frac{\alpha}{\alpha-1}$, we obtain

$$
\begin{aligned}
\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) &= \sup_{\substack{\tau_{AB}>0\\ \mathrm{Tr}[\tau_{AB}]=1}} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}; \tau_{AB}) && (7.5.129)\\
&\geq \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}; \xi_{AB}(\omega_A)) && (7.5.130)\\
&\geq \widetilde{D}_\alpha(\rho_A\|\sigma_A; \omega_A) && (7.5.131)
\end{aligned}
$$

for all invertible states ω_A. Finally, taking the supremum over the set $\{\omega_A : \omega_A > 0,\ \mathrm{Tr}[\omega_A] = 1\}$, we conclude that

$$
\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \geq \widetilde{D}_\alpha(\rho_A\|\sigma_A), \qquad \text{for } \alpha \in (1, \infty). \tag{7.5.132}
$$

For $\alpha \in [1/2, 1)$, taking the logarithm on both sides of (7.5.119) and using the state in (7.5.120), we conclude that

$$
\begin{aligned}
\log_2 \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB}; \xi_{AB}(\omega_A)) &\leq \log_2\langle\varphi^{\rho_A}|f(\omega_A^{-1} \otimes \sigma_{\hat{A}}^{\mathsf{T}})|\varphi^{\rho_A}\rangle && (7.5.133)\\
&= \log_2 \widetilde{Q}_\alpha(\rho_A\|\sigma_A; \omega_A). && (7.5.134)
\end{aligned}
$$

Multiplying both sides of this inequality by $\frac{\alpha}{\alpha-1}$, which is negative in this case, so that

$$\frac{\alpha}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho_{AB}\|\sigma_{AB};\xi_{AB}(\omega_A)) \geq \frac{\alpha}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho_A\|\sigma_A;\omega_A), \quad (7.5.135)$$

we obtain

$$\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) = \sup_{\substack{\tau_{AB}>0,\\ \mathrm{Tr}[\tau_{AB}]=1}} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB};\tau_{AB}) \quad (7.5.136)$$

$$\geq \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB};\xi_{AB}(\omega_A)) \quad (7.5.137)$$

$$\geq \widetilde{D}_\alpha(\rho_A\|\sigma_A;\omega_A) \quad (7.5.138)$$

for all invertible states ω_A. Finally, taking the supremum over the set $\{\omega_A : \omega_A > 0,\ \mathrm{Tr}[\omega_A] = 1\}$, we conclude that

$$\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \geq \widetilde{D}_\alpha(\rho_A\|\sigma_A), \qquad \text{for } \alpha \in [1/2, 1)\,. \quad (7.5.139)$$

Having established the data-processing inequality for $\widetilde{D}_\alpha$ under the partial trace channel for $\alpha \in [1/2, 1) \cup (1, \infty)$, we conclude that

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad (7.5.140)$$

for $\alpha \in [1/2, 1) \cup (1, \infty)$, all states ρ, positive semi-definite operators σ, and all channels $\mathcal{N}$. ■

By taking the limit $\alpha \to 1$ in the statement of the data-processing inequality for $\widetilde{D}_\alpha$, along with Proposition 7.30, we immediately obtain the data-processing inequality for the quantum relative entropy.

Corollary 7.34 Data-Processing Inequality for Quantum Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then,

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \quad (7.5.141)$$

With the data-processing inequality for the sandwiched Rényi relative entropy in hand, it is now straightforward to prove some of the following additional properties.

Proposition 7.35 Additional Properties of Sandwiched Rényi Relative Entropy

The sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ satisfies the following properties for every state ρ and positive semi-definite operator σ for $\alpha \in [1/2, 1) \cup (1, \infty)$.

1. If $\mathrm{Tr}(\sigma) \leq \mathrm{Tr}(\rho) = 1$, then $\widetilde{D}_\alpha(\rho\|\sigma) \geq 0$.
2. *Faithfulness*: If $\mathrm{Tr}[\sigma] \leq 1$, we have that $\widetilde{D}_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.
3. If $\rho \leq \sigma$, then $\widetilde{D}_\alpha(\rho\|\sigma) \leq 0$.
4. For every positive semi-definite operator σ' such that $\sigma' \geq \sigma$, we have $\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(\rho\|\sigma')$.

PROOF:

1. By the data-processing inequality for $\widetilde{D}_\alpha$ with respect to the trace channel Tr, and letting $x = \mathrm{Tr}(\rho) = 1$ and $y = \mathrm{Tr}(\sigma)$, we find that

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(x\|y) = \frac{1}{\alpha - 1} \log_2 \mathrm{Tr}[(y^{\frac{1-\alpha}{2\alpha}} x y^{\frac{1-\alpha}{2\alpha}})^\alpha] \tag{7.5.142}$$

$$= \frac{1}{\alpha - 1} \log_2(y^{1-\alpha}) \tag{7.5.143}$$

$$= \frac{1-\alpha}{\alpha - 1} \log_2 y \tag{7.5.144}$$

$$= -\log_2 y \tag{7.5.145}$$

$$\geq 0, \tag{7.5.146}$$

where the last line follows from the assumption that $y = \mathrm{Tr}(\sigma) \leq 1$.

2. *Proof of faithfulness*: If $\rho = \sigma$, then the following equalities hold for all $\alpha \in [1/2, 1) \cup (1, \infty)$:

$$\widetilde{D}_\alpha(\rho\|\rho) = \frac{1}{\alpha - 1} \log_2 \mathrm{Tr}\left[\left(\rho^{\frac{1-\alpha}{2\alpha}} \rho \rho^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] \tag{7.5.147}$$

$$= \frac{1}{\alpha - 1} \log_2 \mathrm{Tr}\left[\rho^{\frac{1-\alpha}{2}} \rho^\alpha \rho^{\frac{1-\alpha}{2}}\right] \tag{7.5.148}$$

$$= \frac{1}{\alpha - 1} \log_2 \mathrm{Tr}[\rho^{1-\alpha} \rho^\alpha] \tag{7.5.149}$$

$$= \frac{1}{\alpha - 1} \log_2 \operatorname{Tr}(\rho) \tag{7.5.150}$$
$$= 0. \tag{7.5.151}$$

Next, suppose that $\alpha \in [1/2, 1) \cup (1, \infty)$ and $\widetilde{D}_\alpha(\rho\|\sigma) = 0$. From the above, we conclude that $\widetilde{D}_\alpha(\operatorname{Tr}(\rho)\|\operatorname{Tr}(\sigma)) = -\log_2 y \geq 0$. From the fact that $\log_2 y = 0$ if and only if $y = 1$, we conclude that $\widetilde{D}_\alpha(\rho\|\sigma) = 0$ implies $\operatorname{Tr}(\sigma) = \operatorname{Tr}(\rho) = 1$, so that σ is a density operator. Then, for every measurement channel $\mathcal{M}$,

$$\widetilde{D}_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq \widetilde{D}_\alpha(\rho\|\sigma) = 0. \tag{7.5.152}$$

On the other hand, since $\operatorname{Tr}(\sigma) = \operatorname{Tr}(\rho)$,

$$D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \geq \widetilde{D}_\alpha(\operatorname{Tr}(\mathcal{M}(\rho))\|\operatorname{Tr}(\mathcal{M}(\sigma))) \tag{7.5.153}$$
$$= \widetilde{D}_\alpha(\operatorname{Tr}(\rho)\|\operatorname{Tr}(\sigma)) \tag{7.5.154}$$
$$= 0, \tag{7.5.155}$$

which means that $\widetilde{D}_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0$ for all measurement channels $\mathcal{M}$. Now, recall that $\mathcal{M}(\rho)$ and $\mathcal{M}(\sigma)$ are effectively probability distributions determined by the measurement. Since the classical Rényi relative entropy is equal to zero if and only if its two arguments are equal, we can conclude that $\mathcal{M}(\rho) = \mathcal{M}(\sigma)$. Since this is true for every measurement channel, we conclude from Theorem 6.4 and the fact that the trace norm is a norm that $\rho = \sigma$. So we have that $\widetilde{D}_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$, as required.

3. Consider that $\rho \leq \sigma$ implies that $\sigma - \rho \geq 0$. Then define the following positive semi-definite operators:

$$\hat{\rho} := |0\rangle\langle 0| \otimes \rho, \tag{7.5.156}$$
$$\hat{\sigma} := |0\rangle\langle 0| \otimes \rho + |1\rangle\langle 1| \otimes (\sigma - \rho)\,. \tag{7.5.157}$$

By exploiting the direct-sum property of sandwiched Rényi relative entropy (Proposition 7.31) and the data-processing inequality, we find that

$$0 = \widetilde{D}_\alpha(\rho\|\rho) = \widetilde{D}_\alpha(\hat{\rho}\|\hat{\sigma}) \geq \widetilde{D}_\alpha(\rho\|\sigma), \tag{7.5.158}$$

where the inequality follows from data processing with respect to partial trace over the classical register.

4. Consider the state $\hat{\rho} := |0\rangle\langle 0| \otimes \rho$ and the operator $\hat{\sigma} := |0\rangle\langle 0| \otimes \sigma + |1\rangle\langle 1| \otimes (\sigma' - \sigma)$, which is positive semi-definite because $\sigma' \geq \sigma$ by assumption. Then

$$\hat{\sigma}^{\frac{1-\alpha}{2\alpha}} \hat{\rho} \hat{\sigma}^{\frac{1-\alpha}{2\alpha}} = |0\rangle\langle 0| \otimes \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}, \tag{7.5.159}$$

which implies that

$$\widetilde{D}_\alpha(\hat{\rho} \| \hat{\sigma}) = \widetilde{D}_\alpha(\rho \| \sigma). \tag{7.5.160}$$

Then, observing that $\mathrm{Tr}_1[\hat{\sigma}] = \sigma'$, and using the data-processing inequality for $\widetilde{D}_\alpha$ with respect to the partial trace channel Tr_1, we conclude that

$$\widetilde{D}_\alpha(\rho \| \sigma') = \widetilde{D}_\alpha(\mathrm{Tr}_1(\hat{\rho}) \| \mathrm{Tr}_1(\hat{\sigma})) \leq \widetilde{D}_\alpha(\hat{\rho} \| \hat{\sigma}) = \widetilde{D}_\alpha(\rho \| \sigma), \tag{7.5.161}$$

as required. ■

Let us now prove the faithfulness of both the Petz–Rényi and sandwiched Rényi relative entropies for the full range of parameters for which they are defined.

Proposition 7.36 Faithfulness of the Petz–Rényi and Sandwiched Rényi Relative Entropies

For all $\alpha \in (0, 1) \cup (1, \infty)$ and for all states ρ, σ, the Petz–Rényi and sandwiched Rényi relative entropies are faithful, meaning that

$$D_\alpha(\rho \| \sigma) = 0 \text{ if and only if } \rho = \sigma, \tag{7.5.162}$$

$$\widetilde{D}_\alpha(\rho \| \sigma) = 0 \text{ if and only if } \rho = \sigma. \tag{7.5.163}$$

Proof: Note that the equality $\widetilde{D}_\alpha(\rho \| \rho) = 0$ for all $\alpha \in (0, 1) \cup (1, \infty)$ is immediate from the definition (see also (7.5.147)–(7.5.151)). The converse statement has already been established in property 2. of Proposition 7.35 for $\alpha \in [1/2, 1) \cup (1, \infty)$. Before getting to the range $\alpha \in (0, 1/2)$, let us consider the Petz–Rényi relative entropy.

It is immediately clear from the definition that $D_\alpha(\rho \| \rho) = 0$ for all $\alpha \in (0, 1) \cup (1, \infty)$. For $\alpha \in [0, 1) \cup (1, 2]$, the converse follows from the data-processing inequality, which holds for this parameter range as shown in Theorem 7.24, as well as from arguments analogous to those in the proof of property 2. in Proposition 7.35. For $\alpha \in (2, \infty)$, we use the fact that $D_\alpha(\rho \| \sigma) \geq \widetilde{D}_\alpha(\rho \| \sigma)$ for all ρ, σ, as shown in Proposition 7.31. In particular, if $D_\alpha(\rho \| \sigma) = 0$, then $\widetilde{D}_\alpha(\rho \| \sigma) \leq 0$. However, because ρ and σ are states, by property 1. of Proposition 7.35, we have that $\widetilde{D}_\alpha(\rho\|\sigma) \geq 0$, which means that $\widetilde{D}_\alpha(\rho\|\sigma) = 0$. Then, by property 2. of Proposition 7.35, we immediately get that $\rho = \sigma$.

Finally, suppose that $\widetilde{D}_\alpha(\rho\|\sigma) = 0$, where $\alpha \in (0, 1/2)$. Then, using (7.5.46), we have that $\alpha D_\alpha(\rho\|\sigma) \leq 0$. However, because ρ and σ are states, by the data-processing inequality we have that

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(\mathrm{Tr}[\rho]\|\mathrm{Tr}[\sigma]) = \frac{1}{\alpha-1}\log_2\left(\mathrm{Tr}[\rho]^\alpha \mathrm{Tr}[\sigma]^{1-\alpha}\right) = 0. \quad (7.5.164)$$

Therefore, $D_\alpha(\rho\|\sigma) = 0$, which implies that $\rho = \sigma$ by the faithfulness of the Petz–Rényi relative entropy, which we just proved. ■

The data-processing inequality for the sandwiched Rényi relative entropy can be written using the sandwiched Rényi relative quasi-entropy $\widetilde{Q}_\alpha$ as

$$\frac{1}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho\|\sigma) \geq \frac{1}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\mathcal{N}(\rho)|\mathcal{N}(\sigma)). \quad (7.5.165)$$

Then, since $\alpha - 1$ is negative for $\alpha \in [1/2, 1)$, we can use the monotonicity of the function $\log_2$ to obtain

$$\widetilde{Q}_\alpha(\rho\|\sigma) \geq \widetilde{Q}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad \text{for } \alpha \in (1,\infty), \quad (7.5.166)$$

$$\widetilde{Q}_\alpha(\rho\|\sigma) \leq \widetilde{Q}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \quad \text{for } \alpha \in [1/2, 1). \quad (7.5.167)$$

Just as with the Petz–Rényi relative entropy, we can use this to prove the joint convexity of the sandwiched Rényi relative entropy.

Proposition 7.37 Joint Convexity & Concavity of Sandwiched Rényi Relative Quasi-Entropy

Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then, for $\alpha \in (1,\infty)$

$$\widetilde{Q}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x)\widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x), \quad (7.5.168)$$

and for $\alpha \in [1/2, 1)$,

$$\widetilde{Q}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \geq \sum_{x\in\mathcal{X}} p(x)\widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.5.169}$$

Consequently, the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ is jointly convex for $\alpha \in [1/2, 1)$:

$$\widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x)\widetilde{D}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.5.170}$$

Proof: By the direct-sum property of $\widetilde{Q}_\alpha$ and applying (7.5.166)–(7.5.167) and Proposition 7.17, we conclude (7.5.168)–(7.5.169).

For $\alpha \in [1/2, 1)$, taking $\log_2$ of both sides and multiplying by $\frac{1}{\alpha-1}$, which is negative, we find that

$$\frac{1}{\alpha-1}\log_2 \widetilde{Q}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \\ \leq \frac{1}{\alpha-1}\log_2\left(\sum_{x\in\mathcal{X}} p(x)\widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x)\right). \tag{7.5.171}$$

Then, since $-\log_2$ is a convex function, and using the definition of $\widetilde{D}_\alpha$ in terms of $\widetilde{Q}_\alpha$, we conclude that

$$\widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}} p(x)\frac{1}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho_A^x\|\sigma_A^x) \tag{7.5.172}$$

$$= \sum_{x\in\mathcal{X}} p(x)\widetilde{D}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.5.173}$$

as required. ■

Although the sandwiched Rényi relative entropy is not jointly convex for $\alpha \in (1, \infty)$, it is *jointly quasi-convex*, in the sense that

$$\widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \max_{x\in\mathcal{X}} \widetilde{D}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.5.174}$$

for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0, 1]$, set $\{\rho_A^x\}_{x\in\mathcal{X}}$ of states, and set $\{\sigma_A^x\}_{x\in\mathcal{X}}$ of positive semi-definite operators. Indeed, from (7.5.168), we immediately obtain

$$\widetilde{Q}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \leq \max_{x\in\mathcal{X}} \widetilde{Q}_\alpha(\rho_A^x \| \sigma_A^x). \tag{7.5.175}$$

Taking the logarithm and multiplying by $\frac{1}{\alpha-1}$ on both sides of this inequality leads to (7.5.174).

7.6 Geometric Rényi Relative Entropy

In the previous two sections, we considered two examples of generalized divergences, the Petz– and sandwiched Rényi relative entropies. Both of these are quantum generalizations of the classical Rényi relative entropy defined in (7.4.19).

In this section, we consider another generalization of the classical Rényi relative entropy, called the geometric Rényi relative entropy. Unlike the two previous Rényi relative entropies, the geometric Rényi relative entropy does not converge to the quantum relative entropy in the $\alpha \to 1$ limit. Instead, it converges to what is called the Belavkin–Staszewski relative entropy, as shown in Section 7.7. This latter quantity represents a different quantum generalization of the classical relative entropy in (7.2.2). The main use of the geometric Rényi and Belavkin–Staszewski relative entropies is in establishing upper bounds on the rates of feedback-assisted quantum communication protocols, the latter of which is the main focus of Part III of this book.

Definition 7.38 Geometric Rényi Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\alpha \in (0, 1) \cup (1, \infty)$. The *geometric Rényi relative quasi-entropy* is defined as

$$\widehat{Q}_\alpha(\rho\|\sigma) \coloneqq \lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \lim_{\varepsilon\to 0^+} \operatorname{Tr}[G_\alpha(\sigma_\varepsilon, \rho)], \tag{7.6.1}$$

where $\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}$ and

$$G_\alpha(\sigma_\varepsilon, \rho) \coloneqq \sigma_\varepsilon^{\frac{1}{2}}\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha \sigma_\varepsilon^{\frac{1}{2}} \tag{7.6.2}$$

is the *weighted operator geometric mean* of σ_ε and ρ. The *geometric Rényi relative entropy* is then defined as

$$\widehat{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log_2 \widehat{Q}_\alpha(\rho\|\sigma). \tag{7.6.3}$$

Remark: In general, the weighted operator geometric mean of two positive definite operators X and Y is defined as

$$G_\beta(X,Y) := X^{\frac{1}{2}}\left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)^\beta X^{\frac{1}{2}}, \tag{7.6.4}$$

where $\beta \in \mathbb{R}$ is the weight parameter. We recover the standard operator geometric mean for $\beta = \frac{1}{2}$.

An important property of the weighted operator geometric mean is that

$$G_\beta(X,Y) = G_{1-\beta}(Y,X) \tag{7.6.5}$$

for all positive definite X, Y, and all $\beta \in \mathbb{R}$. To see this, observe that

$$G_{1-\beta}(Y,X) = Y^{\frac{1}{2}}\left(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}}\right)^{1-\beta} Y^{\frac{1}{2}} \tag{7.6.6}$$

$$= Y^{\frac{1}{2}}\left(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}}\right)\left(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}}\right)^{-\beta} Y^{\frac{1}{2}} \tag{7.6.7}$$

$$= X^{\frac{1}{2}}X^{\frac{1}{2}}Y^{-\frac{1}{2}}\left(Y^{-\frac{1}{2}}X^{\frac{1}{2}}X^{\frac{1}{2}}Y^{-\frac{1}{2}}\right)^{-\beta} Y^{\frac{1}{2}} \tag{7.6.8}$$

Now we apply Lemma 2.5. Specifically, we set $L = X^{\frac{1}{2}}Y^{-\frac{1}{2}}$ and $f(x) = x^{-\beta}$ therein to conclude that

$$G_{1-\beta}(Y,X) = X^{\frac{1}{2}}\left(X^{\frac{1}{2}}Y^{-\frac{1}{2}}Y^{-\frac{1}{2}}X^{\frac{1}{2}}\right)^{-\beta} X^{\frac{1}{2}}Y^{-\frac{1}{2}}Y^{\frac{1}{2}} \tag{7.6.9}$$

$$= X^{\frac{1}{2}}\left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)^{\beta} X^{\frac{1}{2}} \tag{7.6.10}$$

$$= G_\beta(X,Y). \tag{7.6.11}$$

Definition 7.38 of the geometric Rényi relative entropy involves a limit, which has to do with the possibility that σ might not be invertible (i.e., it might not be positive definite). Recall that the same situation arises for the Petz– and sandwiched Rényi relative entropies, which leads to expressions for them in terms of a limit in Propositions 7.21 and 7.29, respectively. For these two quantities, the limits evaluate to a finite value with an explicit expression under the condition $\alpha \in (0,1)$ and $\mathrm{Tr}[\rho\sigma] \neq 0$, or $\alpha \in (1,\infty)$ and $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. For the geometric Rényi relative entropy, however, there are several cases for which the limit in (7.6.1) is finite and has an explicit expression. The following proposition outlines some of the simpler cases in which σ is positive definite:

Proposition 7.39

Let ρ be a state, and let σ be a positive definite operator. Then,

$$\widehat{Q}_\alpha(\rho\|\sigma) = \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}[G_\alpha(\sigma,\rho)] \tag{7.6.12}$$

for all $\alpha \in (0,1) \cup (1,\infty)$. If ρ is a positive definite state and σ a positive definite operator, then

$$\widehat{Q}_\alpha(\rho\|\sigma) = \operatorname{Tr}\left[\rho\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha}\right] = \operatorname{Tr}[G_{1-\alpha}(\rho,\sigma)] \tag{7.6.13}$$

$$= \operatorname{Tr}\left[\rho\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right]. \tag{7.6.14}$$

for all $\alpha \in (0,1) \cup (1,\infty)$.

Proof. If σ is positive definite, then the support of σ is the entire Hilbert space, and so the limit $\varepsilon \to 0^+$ in (7.6.1) simply evaluates to $\widehat{Q}_\alpha(\rho\|\sigma) = \operatorname{Tr}[G_\alpha(\sigma,\rho)]$ for all $\alpha \in (0,1) \cup (1,\infty)$.

If ρ is also positive definite, then by invoking the equality in (7.6.5), we conclude that $\operatorname{Tr}[G_\alpha(\sigma,\rho)] = \operatorname{Tr}[G_{1-\alpha}(\rho,\sigma)]$ for all $\alpha \in (0,1) \cup (1,\infty)$. Furthermore, since both ρ and σ are positive definite, the following equality holds

$$\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha} = \left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}. \tag{7.6.15}$$

Therefore, the equality in (7.6.14) holds. ■

We now provide explicit expressions for the geometric Rényi relative quasi-entropy $\widehat{Q}_\alpha(\rho\|\sigma)$ that are consistent with the limit-based definition in (7.6.1) whenever ρ and/or σ are not positive definite. The expressions given in (7.6.16) below cover all possible values of $\alpha \in (0,1) \cup (1,\infty)$ and support conditions. Additional expressions are given in (7.6.19).

Proposition 7.40 Explicit Expressions for Geometric Rényi Relative Quasi-Entropy

For every state ρ, positive semi-definite operator σ, and $\alpha \in (0,1) \cup (1,\infty)$, the following equality holds for the geometric Rényi relative quasi-entropy:

$$\widehat{Q}_\alpha(\rho\|\sigma) = \begin{cases} \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha\right] & \begin{array}{c}\text{if } \alpha \in (0,1)\cup(1,\infty) \\ \text{and } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)\end{array} \\ \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\tilde{\rho}\sigma^{-\frac{1}{2}}\right)^\alpha\right] & \begin{array}{c}\text{if } \alpha \in (0,1) \\ \text{and } \operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)\end{array} \\ +\infty & \begin{array}{c}\text{if } \alpha \in (1,\infty) \text{ and} \\ \operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma),\end{array} \end{cases} \tag{7.6.16}$$

where

$$\tilde{\rho} := \rho_{0,0} - \rho_{0,1}\rho_{1,1}^{-1}\rho_{0,1}^\dagger, \quad \rho = \begin{pmatrix} \rho_{0,0} & \rho_{0,1} \\ \rho_{0,1}^\dagger & \rho_{1,1} \end{pmatrix}, \tag{7.6.17}$$

$$\rho_{0,0} := \Pi_\sigma\rho\Pi_\sigma, \quad \rho_{0,1} := \Pi_\sigma\rho\Pi_\sigma^\perp, \quad \rho_{1,1} := \Pi_\sigma^\perp\rho\Pi_\sigma^\perp, \tag{7.6.18}$$

Π_σ is the projection onto the support of σ, $\Pi_\sigma^\perp$ is the projection onto the kernel of σ, and the inverses $\sigma^{-\frac{1}{2}}$ and $\rho_{1,1}^{-1}$ are taken on the supports of σ and $\rho_{1,1}$, respectively. We also have the alternative expressions below for certain cases:

$$\widehat{Q}_\alpha(\rho\|\sigma) = \begin{cases} \operatorname{Tr}\left[\rho\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha}\right] & \begin{array}{c}\text{if } \alpha \in (0,1) \\ \text{and } \operatorname{supp}(\sigma) \subseteq \operatorname{supp}(\rho)\end{array} \\ \operatorname{Tr}\left[\rho\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right] & \begin{array}{c}\text{if } \alpha \in (1,\infty) \\ \text{and } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)\end{array} \end{cases}, \tag{7.6.19}$$

where the inverses $\rho^{-\frac{1}{2}}$ and σ^{-1} are taken on the supports of ρ and σ, respectively.

PROOF: The proof is similar in spirit to the proofs of Propositions 7.21 and 7.29, but it is more complicated than these previous proofs. We provide it in Section 7.6.2. ■

Observe that when $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and $\alpha \in (0,1)$, the expression

$\mathrm{Tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^{\alpha}]$ is actually a special case of $\mathrm{Tr}[\sigma(\sigma^{-1/2}\tilde{\rho}\sigma^{-1/2})^{\alpha}]$, because the operators $\rho_{0,1}$ and $\rho_{1,1}$ are both equal to zero in this case, so that $\Pi_\sigma\rho = \rho\Pi_\sigma = \rho$ and $\tilde{\rho} = \rho_{0,0}$.

The main intuition behind the first expression in (7.6.16) and those in (7.6.19) is as follows. If ρ and σ are positive definite, then the following equalities hold

$$\mathrm{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right] = \mathrm{Tr}\left[\rho\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha}\right] \tag{7.6.20}$$

$$= \mathrm{Tr}\left[\rho\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right], \tag{7.6.21}$$

for all $\alpha \in (0,1) \cup (1,\infty)$, as shown previously in Proposition 7.39.

1. If the support condition $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ holds, then we can think of $\mathrm{supp}(\sigma)$ as being the whole Hilbert space and σ being invertible on the whole space. So then generalized inverses like $\sigma^{-\frac{1}{2}}$ or σ^{-1} are true inverses on $\mathrm{supp}(\sigma)$, and the expression $\mathrm{Tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^{\alpha}]$ is sensible for $\alpha \in (0,1) \cup (1,\infty)$, with the only inverse in the expression being $\sigma^{-\frac{1}{2}}$; this expression results after taking the limit $\varepsilon \to 0^+$.

2. Similarly, the expression $\mathrm{Tr}[\rho(\rho^{1/2}\sigma^{-1}\rho^{1/2})^{\alpha-1}]$ is sensible for $\alpha \in (1,\infty)$, with the only inverse in the expression being σ^{-1}; this latter expression results after taking the limit $\varepsilon \to 0^+$.

3. If the support condition $\mathrm{supp}(\sigma) \subseteq \mathrm{supp}(\rho)$ holds, then we can think of $\mathrm{supp}(\rho)$ as being the whole Hilbert space and ρ being invertible on the whole space. So then the generalized inverse $\rho^{-\frac{1}{2}}$ is a true inverse on $\mathrm{supp}(\rho)$, and the expression $\mathrm{Tr}[\rho(\rho^{-1/2}\sigma\rho^{-1/2})^{1-\alpha}]$ is sensible for $\alpha \in (0,1)$, with the only inverse in the expression being $\rho^{-\frac{1}{2}}$; this expression also results after taking the limit $\varepsilon \to 0^+$.

In order to understand the first expression in (7.6.19) further, observe that the following identities hold for all $\alpha \in (0,1) \cup (1,\infty)$:

$$\widehat{Q}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.6.22}$$

$$= \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \mathrm{Tr}[G_\alpha(\sigma_\varepsilon, \rho_\delta)], \tag{7.6.23}$$

where

$$\rho_\delta \coloneqq (1-\delta)\,\rho + \delta\pi, \tag{7.6.24}$$

and π is the maximally mixed state. This holds because the expression for the geometric Rényi relative quasi-entropy in Definition 7.38 does not involve an inverse of the state ρ.

As it turns out, the order of the limits in (7.6.22) does not matter for $\alpha \in (0,1)$:

Lemma 7.41 Limit Interchange for Geometric Rényi Relative Quasi-Entropy

Let ρ be a state and σ a positive semi-definite operator. For $\alpha \in (0,1)$, the following equality holds

$$\widehat{Q}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.25}$$

$$= \inf_{\varepsilon,\delta>0} \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.26}$$

$$= \lim_{\delta\to 0^+}\lim_{\varepsilon\to 0^+} \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.27}$$

where $\rho_\delta \coloneqq (1-\delta)\,\rho + \delta\pi$, $\delta \in (0,1)$, π is the maximally mixed state, $\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}$, and $\varepsilon > 0$.

Proof: See Section 7.6.1. ■

Now, because both σ_ε and ρ_δ are positive definite for $\varepsilon, \delta > 0$, we can use the property in (7.6.5), along with Lemma 7.41, to obtain the following for $\alpha \in (0,1)$:

$$\widehat{Q}_\alpha(\rho\|\sigma) = \lim_{\delta\to 0^+}\lim_{\varepsilon\to 0^+} \mathrm{Tr}[G_{1-\alpha}(\rho_\delta, \sigma_\varepsilon)] \tag{7.6.28}$$

$$= \lim_{\delta\to 0^+}\lim_{\varepsilon\to 0^+} \mathrm{Tr}\left[\rho_\delta\left(\rho_\delta^{-\frac{1}{2}}\sigma_\varepsilon\rho_\delta^{-\frac{1}{2}}\right)^{1-\alpha}\right] \tag{7.6.29}$$

$$= \lim_{\delta\to 0^+} \mathrm{Tr}\left[\rho_\delta\left(\rho_\delta^{-\frac{1}{2}}\sigma\rho_\delta^{-\frac{1}{2}}\right)^{1-\alpha}\right], \tag{7.6.30}$$

where the last equality holds for the analogous reason that (7.6.22) holds, namely, that the inverse of σ is not involved. We are now in a situation that looks like the expression in (7.6.1), except that the roles of ρ and σ are reversed and α is substituted with $1-\alpha$. Then, in the limit $\delta \to 0^+$, if the support condition

$\operatorname{supp}(\sigma) \subseteq \operatorname{supp}(\rho)$ holds, the expression converges to $\operatorname{Tr}[\rho(\rho^{-1/2}\sigma\rho^{-1/2})^{1-\alpha}]$.

It is worthwhile to consider the special case of $\alpha = 2$. In this case, the geometric Rényi relative quasi-entropy collapses to the Petz–Rényi relative quasi-entropy when $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$:

$$\widehat{Q}_2(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^2\right] \tag{7.6.31}$$

$$= \lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)\right] \tag{7.6.32}$$

$$= \lim_{\varepsilon\to 0^+} \operatorname{Tr}[\rho\sigma_\varepsilon^{-1}\rho] \tag{7.6.33}$$

$$= \lim_{\varepsilon\to 0^+} \operatorname{Tr}[\rho^2\sigma_\varepsilon^{-1}] \tag{7.6.34}$$

$$= Q_2(\rho\|\sigma), \tag{7.6.35}$$

with the last line following from Proposition 7.21. The development above implies that the corresponding Rényi relative entropies are equal:

$$\widehat{D}_2(\rho\|\sigma) = D_2(\rho\|\sigma). \tag{7.6.36}$$

The geometric and sandwiched Rényi relative entropies also converge to the same value in the limit $\alpha \to \infty$, as shown in Section 7.8.

A first property of the geometric Rényi relative entropy that we establish is its relation to the sandwiched Rényi relative entropy.

Proposition 7.42 Ordering of Sandwiched and Geometric Rényi Relative Entropies

Let ρ be a state and σ a positive semi-definite operator. The geometric Rényi relative entropy is not smaller than the sandwiched Rényi relative entropy for all $\alpha \in (0,1) \cup (1,\infty)$:

$$\widetilde{D}_\alpha(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma). \tag{7.6.37}$$

Proof: This is a direct consequence of the Araki–Lieb–Thirring inequality (Lemma 2.15), which we recall here for convenience. For positive semi-definite operators X and Y, $q \geq 0$, and $r \in [0,1]$, the following inequality holds

$$\operatorname{Tr}\left[\left(Y^{\frac{1}{2}}XY^{\frac{1}{2}}\right)^{rq}\right] \geq \operatorname{Tr}\left[\left(Y^{\frac{r}{2}}X^rY^{\frac{r}{2}}\right)^q\right]. \tag{7.6.38}$$

For $r \geq 1$, the following inequality holds

$$\mathrm{Tr}\left[\left(Y^{\frac{1}{2}}XY^{\frac{1}{2}}\right)^{rq}\right] \leq \mathrm{Tr}\left[\left(Y^{\frac{r}{2}}X^{r}Y^{\frac{r}{2}}\right)^{q}\right]. \tag{7.6.39}$$

By employing (7.6.38) with $q = 1$, $r = \alpha \in (0,1)$, $Y = \sigma_\varepsilon^{\frac{1}{\alpha}}$, and $X = \sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}$, and recalling that $\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}$, we find that

$$\widehat{Q}_\alpha(\rho\|\sigma_\varepsilon) = \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.6.40}$$

$$= \mathrm{Tr}\left[\left(\sigma_\varepsilon^{\frac{1}{2\alpha}}\right)^{\alpha}\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^{\alpha}\left(\sigma_\varepsilon^{\frac{1}{2\alpha}}\right)^{\alpha}\right] \tag{7.6.41}$$

$$\leq \mathrm{Tr}\left[\left(\sigma_\varepsilon^{\frac{1}{2\alpha}}\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\sigma_\varepsilon^{\frac{1}{2\alpha}}\right)^{\alpha}\right] \tag{7.6.42}$$

$$= \mathrm{Tr}\left[\left(\sigma_\varepsilon^{\frac{1-\alpha}{2\alpha}}\rho\sigma_\varepsilon^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right] \tag{7.6.43}$$

$$= \widetilde{Q}_\alpha(\rho\|\sigma_\varepsilon), \tag{7.6.44}$$

which implies for $\alpha \in (0,1)$, by using definitions, that

$$\widetilde{D}_\alpha(\rho\|\sigma_\varepsilon) \leq \widehat{D}_\alpha(\rho\|\sigma_\varepsilon). \tag{7.6.45}$$

Now taking the limit as $\varepsilon \to 0^+$, employing Proposition 7.29 and Definition 7.38, we arrive at the inequality in (7.6.37).

Since the Araki–Lieb–Thirring inequality is reversed for $r = \alpha \in (1,\infty)$, we can employ similar reasoning as above, using (7.6.39) and definitions, to arrive at (7.6.37) for $\alpha \in (1,\infty)$. ■

If the state ρ is pure, then the geometric Rényi relative entropy simplifies as follows, such that it is independent of α:

Proposition 7.43 Geometric Rényi Relative Entropy for Pure States

Let $\rho = |\psi\rangle\langle\psi|$ be a pure state and σ a positive semi-definite operator. Then the following equality holds for all $\alpha \in (0,1) \cup (1,\infty)$:

$$\widehat{D}_\alpha(\rho\|\sigma) = \begin{cases} \log_2\langle\psi|\sigma^{-1}|\psi\rangle & \text{if } \mathrm{supp}(|\psi\rangle\langle\psi|) \subseteq \mathrm{supp}(\sigma) \\ +\infty & \text{otherwise} \end{cases}, \tag{7.6.46}$$

where the inverse σ^{-1} is taken on the support of σ. If σ is also a rank-one operator, so that $\sigma = |\phi\rangle\!\langle\phi|$ and $\||\phi\rangle\|_2 > 0$, then the following equality holds for all $\alpha \in (0,1) \cup (1,\infty)$:

$$\widehat{D}_\alpha(\rho\|\sigma) = \begin{cases} -\log_2 \||\phi\rangle\|_2^2 & \text{if } \exists c \in \mathbb{C} \text{ such that } |\psi\rangle = c|\phi\rangle \\ +\infty & \text{otherwise} \end{cases}. \tag{7.6.47}$$

In particular, if $\sigma = |\phi\rangle\!\langle\phi|$ is a state so that $\||\phi\rangle\|_2^2 = 1$, then

$$\widehat{D}_\alpha(\rho\|\sigma) = \begin{cases} 0 & \text{if } |\psi\rangle = |\phi\rangle \\ +\infty & \text{otherwise} \end{cases}. \tag{7.6.48}$$

Proof: Defining $\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}$, consider that

$$\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\!\langle\psi|\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.49}$$

$$= \left(\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2\right)^\alpha \operatorname{Tr}\left[\sigma_\varepsilon\left(\frac{\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\!\langle\psi|\sigma_\varepsilon^{-\frac{1}{2}}}{\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2}\right)^\alpha\right] \tag{7.6.50}$$

$$= \left(\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2\right)^\alpha \operatorname{Tr}\left[\sigma_\varepsilon\frac{\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\!\langle\psi|\sigma_\varepsilon^{-\frac{1}{2}}}{\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2}\right] \tag{7.6.51}$$

$$= \left(\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2\right)^{\alpha-1} \operatorname{Tr}\left[\sigma_\varepsilon\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\!\langle\psi|\sigma_\varepsilon^{-\frac{1}{2}}\right] \tag{7.6.52}$$

$$= \left(\left\|\sigma_\varepsilon^{-\frac{1}{2}}|\psi\rangle\right\|_2^2\right)^{\alpha-1} \operatorname{Tr}[|\psi\rangle\!\langle\psi|] \tag{7.6.53}$$

$$= \left[\langle\psi|\sigma_\varepsilon^{-1}|\psi\rangle\right]^{\alpha-1}. \tag{7.6.54}$$

The third equality follows because $|\varphi\rangle\!\langle\varphi|^\alpha = |\varphi\rangle\!\langle\varphi|$ for all $\alpha \in (0,1)\cup(1,\infty)$ when $\||\varphi\rangle\|_2 = 1$. Applying the chain of equalities above, we find that

$$\frac{1}{\alpha-1}\log_2 \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \frac{1}{\alpha-1}\log_2\left[\langle\psi|\sigma_\varepsilon^{-1}|\psi\rangle\right]^{\alpha-1} \tag{7.6.55}$$

$$= \log_2 \langle \psi | \sigma_\varepsilon^{-1} | \psi \rangle. \tag{7.6.56}$$

Now let a spectral decomposition of σ be given by

$$\sigma = \sum_y \mu_y Q_y, \tag{7.6.57}$$

where μ_y are the non-negative eigenvalues and Q_y are the eigenprojections. In this decomposition, we are including values of μ_y for which $\mu_y = 0$. Then it follows that

$$\sigma_\varepsilon = \sigma + \varepsilon \mathbb{1} = \sum_y \left(\mu_y + \varepsilon\right) Q_y, \tag{7.6.58}$$

and we find that

$$\sigma_\varepsilon^{-1} = \sum_y \left(\mu_y + \varepsilon\right)^{-1} Q_y. \tag{7.6.59}$$

We then conclude that

$$\langle \psi | \sigma_\varepsilon^{-1} | \psi \rangle = \langle \psi | \sum_y \left(\mu_y + \varepsilon\right)^{-1} Q_y | \psi \rangle \tag{7.6.60}$$

$$= \sum_y \left(\mu_y + \varepsilon\right)^{-1} \langle \psi | Q_y | \psi \rangle \tag{7.6.61}$$

$$= \sum_{y : \mu_y \neq 0} \left(\mu_y + \varepsilon\right)^{-1} \langle \psi | Q_y | \psi \rangle + \varepsilon^{-1} \langle \psi | Q_{y_0} | \psi \rangle, \tag{7.6.62}$$

where y_0 is the value of y for which $\mu_y = 0$ (if no such value of y exists, then Q_{y_0} is equal to the zero operator). Thus, if $\langle \psi | Q_{y_0} | \psi \rangle \neq 0$ (equivalent to $|\psi\rangle$ being outside the support of σ), then it follows that

$$\lim_{\varepsilon \to 0^+} \log_2 \langle \psi | \sigma_\varepsilon^{-1} | \psi \rangle = +\infty. \tag{7.6.63}$$

Otherwise the expression converges as claimed.

Now suppose that σ is a rank-one operator, so that $\sigma = |\phi\rangle\langle\phi|$ and $\| |\phi\rangle \|_2 > 0$. By defining

$$|\phi'\rangle := \frac{|\phi\rangle}{\sqrt{\| |\phi\rangle \|_2}}, \qquad N := \| |\phi\rangle \|_2^2, \tag{7.6.64}$$

we find that

$$\sigma_\varepsilon = |\phi\rangle\langle\phi| + \varepsilon \mathbb{1} \tag{7.6.65}$$

$$= N\,|\phi'\rangle\!\langle\phi'| + \varepsilon\,(\mathbb{1} - |\phi'\rangle\!\langle\phi'| + |\phi'\rangle\!\langle\phi'|) \tag{7.6.66}$$
$$= (N+\varepsilon)\,|\phi'\rangle\!\langle\phi'| + \varepsilon\,(\mathbb{1} - |\phi'\rangle\!\langle\phi'|)\,, \tag{7.6.67}$$

so that

$$\sigma_\varepsilon^{-1} = (N+\varepsilon)^{-1}\,|\phi'\rangle\!\langle\phi'| + \varepsilon^{-1}\,(\mathbb{1} - |\phi'\rangle\!\langle\phi'|) \tag{7.6.68}$$
$$= \left((N+\varepsilon)^{-1} - \varepsilon^{-1}\right)|\phi'\rangle\!\langle\phi'| + \varepsilon^{-1}\mathbb{1} \tag{7.6.69}$$

and then

$$\langle\psi|\sigma_\varepsilon^{-1}|\psi\rangle = \langle\psi|\left[\left((N+\varepsilon)^{-1} - \varepsilon^{-1}\right)|\phi'\rangle\!\langle\phi'| + \varepsilon^{-1}\mathbb{1}\right]|\psi\rangle \tag{7.6.70}$$
$$= \left((N+\varepsilon)^{-1} - \varepsilon^{-1}\right)|\langle\psi|\phi'\rangle|^2 + \varepsilon^{-1} \tag{7.6.71}$$
$$= \frac{|\langle\psi|\phi'\rangle|^2}{N+\varepsilon} + \frac{1 - |\langle\psi|\phi'\rangle|^2}{\varepsilon}. \tag{7.6.72}$$

Note that we always have $|\langle\psi|\phi'\rangle|^2 \in [0,1]$ because $|\psi\rangle$ and $|\phi'\rangle$ are unit vectors. In the case that $|\langle\psi|\phi'\rangle|^2 \in [0,1)$, then we find that

$$\lim_{\varepsilon\to 0^+} \log_2\!\left[\langle\psi|\sigma_\varepsilon^{-1}|\psi\rangle\right] = \lim_{\varepsilon\to 0^+} \log_2\!\left[\frac{|\langle\psi|\phi'\rangle|^2}{N+\varepsilon} + \frac{1 - |\langle\psi|\phi'\rangle|^2}{\varepsilon}\right] \tag{7.6.73}$$
$$= +\infty. \tag{7.6.74}$$

Otherwise, if $|\langle\psi|\phi'\rangle|^2 = 1$, then

$$\lim_{\varepsilon\to 0^+} \log_2\!\left[\langle\psi|\sigma_\varepsilon^{-1}|\psi\rangle\right] = \lim_{\varepsilon\to 0^+} \log_2\!\left[\frac{|\langle\psi|\phi'\rangle|^2}{N+\varepsilon} + \frac{1 - |\langle\psi|\phi'\rangle|^2}{\varepsilon}\right] \tag{7.6.75}$$
$$= \lim_{\varepsilon\to 0^+} \log_2\!\left[\frac{1}{N+\varepsilon}\right] \tag{7.6.76}$$
$$= -\log_2 N, \tag{7.6.77}$$

concluding the proof. ■

We note here that, for pure states ρ and σ and as indicated by (7.6.48), the geometric Rényi relative entropy is either equal to zero or $+\infty$, depending on whether $\rho = \sigma$. This behavior of the geometric Rényi relative entropy for pure states ρ and σ is very different from that of the Petz– and sandwiched Rényi relative entropies. The latter quantities always evaluate to a finite value if the pure states are non-orthogonal.

The geometric Rényi relative entropy possesses a number of useful properties, similar to those for the Petz– and sandwiched Rényi relative entropies, which we delineate now.

Proposition 7.44 Properties of Geometric Rényi Relative Entropy

For all states ρ, ρ_1, ρ_2 and positive semi-definite operators σ, σ_1, σ_2, the geometric Rényi relative entropy satisfies the following properties:

1. Isometric invariance: For all $\alpha \in (0,1) \cup (1,\infty)$ and for every isometry V,

$$\widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}_\alpha(V\rho V^\dagger\|V\sigma V^\dagger). \tag{7.6.78}$$

2. Monotonicity in α: For all $\alpha \in (0,1) \cup (1,\infty)$, the geometric Rényi relative entropy $\widehat{D}_\alpha$ is monotonically increasing in α; i.e., $\alpha < \beta$ implies $\widehat{D}_\alpha(\rho\|\sigma) \leq \widehat{D}_\beta(\rho\|\sigma)$.

3. Additivity: For all $\alpha \in (0,1) \cup (1,\infty)$,

$$\widehat{D}_\alpha(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \widehat{D}_\alpha(\rho_1\|\sigma_1) + \widehat{D}_\alpha(\rho_2\|\sigma_2). \tag{7.6.79}$$

4. Direct-sum property: Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $q : \mathcal{X} \to [0,\infty)$ be a positive function on $\mathcal{X}$. Let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then,

$$\widehat{Q}_\alpha(\rho_{XA}\|\sigma_{XA}) = \sum_{x\in\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} \widehat{Q}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.6.80}$$

where

$$\rho_{XA} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.6.81}$$

$$\sigma_{XA} := \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.6.82}$$

PROOF:

1. *Proof of isometric invariance*: Let us start by writing $\widehat{D}_\alpha(\rho\|\sigma)$ as in (7.6.1)–(7.6.3):

$$\widehat{D}_\alpha(\rho\|\sigma) = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1}\log_2 \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.83}$$

where

$$\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}. \tag{7.6.84}$$

Let V be an isometry. Then, defining

$$\omega_\varepsilon \coloneqq V\sigma V^\dagger + \varepsilon\mathbb{1}, \tag{7.6.85}$$

we find that

$$\widehat{D}_\alpha(V\rho V^\dagger\|V\sigma V^\dagger) = \lim_{\varepsilon\to 0^+} \frac{1}{\alpha-1}\log_2 \operatorname{Tr}\left[\omega_\varepsilon\left(\omega_\varepsilon^{-\frac{1}{2}}V\rho V^\dagger\omega_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.86}$$

Now let $\Pi \coloneqq VV^\dagger$ be the projection onto the image of V, so that $\Pi V = V$, and let $\hat{\Pi} \coloneqq \mathbb{1} - \Pi$. Then we can write

$$\omega_\varepsilon = V\sigma V^\dagger + \varepsilon\Pi + \varepsilon\hat{\Pi} = V\sigma_\varepsilon V^\dagger + \varepsilon\hat{\Pi}. \tag{7.6.87}$$

Since $V\sigma_\varepsilon V^\dagger$ and $\varepsilon\hat{\Pi}$ are supported on orthogonal subspaces, we obtain

$$\omega_\varepsilon^{-\frac{1}{2}} = V\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger + \varepsilon^{-\frac{1}{2}}\hat{\Pi}. \tag{7.6.88}$$

Consider then that

$$\begin{aligned}
&\omega_\varepsilon^{-\frac{1}{2}}V\rho V^\dagger\omega_\varepsilon^{-\frac{1}{2}} \\
&= \left(V\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger + \varepsilon^{-\frac{1}{2}}\hat{\Pi}\right)\Pi V\rho V^\dagger\Pi\left(V\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger + \varepsilon^{-\frac{1}{2}}\hat{\Pi}\right) &(7.6.89)\\
&= \left(V\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger\right)\Pi V\rho V^\dagger\Pi\left(V\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger\right) &(7.6.90)\\
&= V\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}V^\dagger, &(7.6.91)
\end{aligned}$$

where the second equality follows because $\hat{\Pi}\Pi = \Pi\hat{\Pi} = 0$. Thus,

$$\left(\omega_\varepsilon^{-\frac{1}{2}}V\rho V^\dagger\omega_\varepsilon^{-\frac{1}{2}}\right)^\alpha = V\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha V^\dagger, \tag{7.6.92}$$

and we find that

$$\operatorname{Tr}\left[\omega_\varepsilon\left(\omega_\varepsilon^{-\frac{1}{2}}V\rho V^\dagger\omega_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]$$

$$= \operatorname{Tr}\left[\left(V\sigma_\varepsilon V^\dagger + \varepsilon\hat{\Pi}\right)V\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha V^\dagger\right] \tag{7.6.93}$$

$$= \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.94}$$

Since the equality

$$\operatorname{Tr}\left[\omega_\varepsilon\left(\omega_\varepsilon^{-\frac{1}{2}}V\rho V^\dagger\omega_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.95}$$

holds for all $\varepsilon > 0$, we conclude the proof of isometric invariance by taking the limit $\varepsilon \to 0^+$.

2. *Proof of monotonicity in α*: We prove this by showing that the derivative is non-negative for all $\alpha > 0$. By applying (7.6.22), we can consider ρ and σ to be positive definite without loss of generality. By applying (7.6.14), consider that

$$\widehat{Q}_\alpha(\rho\|\sigma) = \operatorname{Tr}\left[\rho\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha}\right] \tag{7.6.96}$$

$$= \operatorname{Tr}\left[\rho\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right]. \tag{7.6.97}$$

Now defining $|\varphi^\rho\rangle = (\rho^{\frac{1}{2}} \otimes \mathbb{1})|\Gamma\rangle$ as a purification of ρ, and setting

$$\gamma := \alpha - 1, \tag{7.6.98}$$

$$X := \rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}, \tag{7.6.99}$$

we can write the geometric Rényi relative entropy as

$$\widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{\gamma}\log_2\langle\varphi^\rho|X^\gamma \otimes \mathbb{1}|\varphi^\rho\rangle = \frac{1}{\gamma}\frac{\ln\langle\varphi^\rho|X^\gamma \otimes \mathbb{1}|\varphi^\rho\rangle}{\ln(2)}, \tag{7.6.100}$$

where we made use of (7.6.97). Then $\frac{\mathrm{d}}{\mathrm{d}\alpha} = \frac{\mathrm{d}}{\mathrm{d}\gamma}\frac{\mathrm{d}\gamma}{\mathrm{d}\alpha} = \frac{\mathrm{d}}{\mathrm{d}\gamma}$, and so we find that

$$\ln(2)\frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{D}_\alpha(\rho\|\sigma)$$

$$= \frac{\mathrm{d}}{\mathrm{d}\gamma}\left[\frac{1}{\gamma}\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle\right] \tag{7.6.101}$$

$$= \left[-\frac{1}{\gamma^2}\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle+\frac{1}{\gamma}\frac{\mathrm{d}}{\mathrm{d}\gamma}\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle\right] \tag{7.6.102}$$

$$= \left[-\frac{1}{\gamma^2}\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle+\frac{1}{\gamma}\frac{\langle\varphi^\rho|X^\gamma\ln X\otimes\mathbb{1}|\varphi^\rho\rangle}{\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle}\right] \tag{7.6.103}$$

$$= \left[\frac{-\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle+\gamma\langle\varphi^\rho|X^\gamma\ln X\otimes\mathbb{1}|\varphi^\rho\rangle}{\gamma^2\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle}\right] \tag{7.6.104}$$

$$= \left[\frac{-\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle\ln\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle+\langle\varphi^\rho|X^\gamma\ln X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle}{\gamma^2\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle}\right]. \tag{7.6.105}$$

Letting $g(x) := x\log_2 x$, we write

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{D}_\alpha(\rho\|\sigma) = \frac{\langle\varphi^\rho|g(X^\gamma\otimes\mathbb{1})|\varphi^\rho\rangle - g(\langle\varphi^\rho|(X^\gamma\otimes\mathbb{1})|\varphi^\rho\rangle)}{\gamma^2\langle\varphi^\rho|X^\gamma\otimes\mathbb{1}|\varphi^\rho\rangle}. \tag{7.6.106}$$

Then, since $g(x)$ is operator convex, by the operator Jensen inequality in (2.3.23), we conclude that

$$\langle\varphi^\rho|g(X^\gamma\otimes\mathbb{1})|\varphi^\rho\rangle \geq g(\langle\varphi^\rho|(X^\gamma\otimes\mathbb{1})|\varphi^\rho\rangle), \tag{7.6.107}$$

which means that $\frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{D}_\alpha(\rho\|\sigma)\geq 0$. Therefore, $\widehat{D}_\alpha(\rho\|\sigma)$ is monotonically increasing in α, as required.

3. *Proof of additivity*: The proof of (7.6.79) is found by direct evaluation. Consider that

$$\lim_{\varepsilon_1\to 0^+}\widehat{Q}_\alpha(\rho_1\|\sigma_{1,\varepsilon_1})\cdot\lim_{\varepsilon_2\to 0^+}\widehat{Q}_\alpha(\rho_2\|\sigma_{2,\varepsilon_2}) = \lim_{\varepsilon_1\to 0^+}\lim_{\varepsilon_2\to 0^+}\widehat{Q}_\alpha(\rho_1\|\sigma_{1,\varepsilon_1})\cdot\widehat{Q}_\alpha(\rho_2\|\sigma_{2,\varepsilon_2}), \tag{7.6.108}$$

where $\sigma_{1,\varepsilon_1} := \sigma_1+\varepsilon_1\mathbb{1}$ and $\sigma_{2,\varepsilon_2} := \sigma_2+\varepsilon_2\mathbb{1}$. We then find that

$$\widehat{Q}_\alpha(\rho_1\|\sigma_{1,\varepsilon_1})\cdot\widehat{Q}_\alpha(\rho_2\|\sigma_{2,\varepsilon_2}) = \operatorname{Tr}\left[\sigma_{1,\varepsilon_1}\left(\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\rho_1\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\right)^\alpha\right]\operatorname{Tr}\left[\sigma_{2,\varepsilon_2}\left(\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\rho_2\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.109}$$

$$= \operatorname{Tr}\left[\sigma_{1,\varepsilon_1}\left(\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\rho_1\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\right)^{\alpha}\otimes\sigma_{2,\varepsilon_2}\left(\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\rho_2\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.6.110}$$

$$= \operatorname{Tr}\left[(\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2})\left(\left(\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\rho_1\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\right)^{\alpha}\otimes\left(\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\rho_2\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\right)^{\alpha}\right)\right] \tag{7.6.111}$$

$$= \operatorname{Tr}\left[(\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2})\left(\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\rho_1\sigma_{1,\varepsilon_1}^{-\frac{1}{2}}\otimes\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\rho_2\sigma_{2,\varepsilon_2}^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.6.112}$$

$$= \operatorname{Tr}\left[(\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2})\left((\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2})^{-\frac{1}{2}}(\rho_1\otimes\rho_2)(\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2})^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.6.113}$$

$$= \widehat{Q}_\alpha(\rho_1\otimes\rho_2\|\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2}). \tag{7.6.114}$$

By considering that

$$\lim_{\varepsilon_1\to 0^+}\lim_{\varepsilon_2\to 0^+}\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2} = \lim_{\varepsilon\to 0^+}\sigma_1\otimes\sigma_2+\varepsilon\mathbb{1}\otimes\mathbb{1}, \tag{7.6.115}$$

along with continuity of the underlying functions, we conclude that

$$\lim_{\varepsilon_1\to 0^+}\lim_{\varepsilon_2\to 0^+}\widehat{Q}_\alpha(\rho_1\otimes\rho_2\|\sigma_{1,\varepsilon_1}\otimes\sigma_{2,\varepsilon_2}) = \lim_{\varepsilon\to 0^+}\widehat{Q}_\alpha(\rho_1\otimes\rho_2\|\sigma_1\otimes\sigma_2+\varepsilon\mathbb{1}\otimes\mathbb{1}). \tag{7.6.116}$$

Finally, by applying the continuous function $\frac{1}{\alpha-1}\log_2(\cdot)$ to all sides of the equalities established, we conclude that additivity holds.

4. *Proof of direct-sum property*: Define the classical–quantum state ρ_{XA} and operator σ_{XA}, respectively, as

$$\rho_{XA} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X\otimes\rho_A^x, \qquad \sigma_{XA} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X\otimes\sigma_A^x. \tag{7.6.117}$$

Define

$$\sigma_{XA}^{\varepsilon} \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X\otimes\sigma_A^x+\varepsilon\mathbb{1}_X\otimes\mathbb{1}_A \tag{7.6.118}$$

$$= \sum_{x\in\mathcal{X},q(x)\neq 0} q(x)|x\rangle\langle x|_X\otimes\sigma_A^x+\varepsilon\sum_{x\in\mathcal{X}}|x\rangle\langle x|_X\otimes\mathbb{1}_A \tag{7.6.119}$$

$$= \sum_{x\in\mathcal{X},q(x)\neq 0}|x\rangle\langle x|_X\otimes q(x)\sigma_{A,\varepsilon}^x+\sum_{x\in\mathcal{X},q(x)=0}|x\rangle\langle x|_X\otimes\varepsilon\mathbb{1}_A, \tag{7.6.120}$$

where

$$\sigma^x_{A,\varepsilon} \coloneqq \sigma^x_A + \varepsilon \mathbb{1}_A. \tag{7.6.121}$$

Then we find that

$$\left(\sigma^{\varepsilon}_{XA}\right)^{-\frac{1}{2}} = \sum_{x\in\mathcal{X}, q(x)\neq 0} |x\rangle\!\langle x|_X \otimes \left(q(x)\sigma^x_{A,\varepsilon}\right)^{-\frac{1}{2}} + \sum_{x\in\mathcal{X}, q(x)=0} |x\rangle\!\langle x|_X \otimes \varepsilon^{-\frac{1}{2}} \mathbb{1}_A, \tag{7.6.122}$$

so that (omitting some lines of calculation)

$$\begin{aligned}
&\left(\left(\sigma^{\varepsilon}_{XA}\right)^{-\frac{1}{2}} \rho_{XA} \left(\sigma^{\varepsilon}_{XA}\right)^{-\frac{1}{2}}\right)^{\alpha} \\
&\quad = \sum_{\substack{x\in\mathcal{X}, q(x)\neq 0,\\ p(x)\neq 0}} |x\rangle\!\langle x|_X \otimes \left(\frac{p(x)}{q(x)}\right)^{\alpha} \left(\left(\sigma^x_{A,\varepsilon}\right)^{-\frac{1}{2}} \rho^x_A \left(\sigma^x_{A,\varepsilon}\right)^{-\frac{1}{2}}\right)^{\alpha} \\
&\qquad + \sum_{\substack{x\in\mathcal{X}, q(x)=0,\\ p(x)\neq 0}} |x\rangle\!\langle x|_X \otimes \varepsilon^{-\alpha} (\rho^x_A)^{\alpha}.
\end{aligned} \tag{7.6.123}$$

Defining

$$\omega^x_A \coloneqq \left(\sigma^x_{A,\varepsilon}\right)^{-\frac{1}{2}} \rho^x_A \left(\sigma^x_{A,\varepsilon}\right)^{-\frac{1}{2}}, \tag{7.6.124}$$

it then follows that

$$\begin{aligned}
&\widehat{Q}_{\alpha}(\rho_{XA}\|\sigma^{\varepsilon}_{XA}) \\
&= \operatorname{Tr}\!\left[\sigma^{\varepsilon}_{XA} \left(\left(\sigma^{\varepsilon}_{XA}\right)^{-\frac{1}{2}} \rho_{XA} \left(\sigma^{\varepsilon}_{XA}\right)^{-\frac{1}{2}}\right)^{\alpha}\right] \qquad (7.6.125) \\
&= \operatorname{Tr}\!\left[\left(\sum_{\substack{x\in\mathcal{X},\\ q(x)\neq 0}} q(x)|x\rangle\!\langle x|_X \otimes \sigma^x_{A,\varepsilon}\right)\left(\sum_{\substack{x'\in\mathcal{X},\\ q(x')\neq 0,\\ p(x')\neq 0}} |x'\rangle\!\langle x'|_X \otimes \left(\frac{p(x')}{q(x')}\right)^{\alpha} \left(\omega^{x'}_A\right)^{\alpha}\right)\right] \\
&+ \operatorname{Tr}\!\left[\left(\sum_{\substack{x\in\mathcal{X},\\ q(x)\neq 0}} q(x)|x\rangle\!\langle x|_X \otimes \sigma^x_{A,\varepsilon}\right)\left(\sum_{\substack{x'\in\mathcal{X},\\ q(x')=0,\\ p(x')\neq 0}} |x'\rangle\!\langle x'|_X \otimes \varepsilon^{-\alpha} (\rho^{x'}_A)^{\alpha}\right)\right]
\end{aligned}$$

$$+ \operatorname{Tr}\left[\left(\sum_{\substack{x\in\mathcal{X},\\ q(x)=0}} |x\rangle\langle x|_X \otimes \varepsilon \mathbb{1}_A\right)\left(\sum_{\substack{x'\in\mathcal{X},\\ q(x')\neq 0,\\ p(x')\neq 0}} |x'\rangle\langle x'|_X \otimes \left(\frac{p(x')}{q(x')}\right)^{\alpha}\left(\omega_A^{x'}\right)^{\alpha}\right)\right]$$

$$+ \operatorname{Tr}\left[\left(\sum_{\substack{x\in\mathcal{X},\\ q(x)=0}} |x\rangle\langle x|_X \otimes \varepsilon \mathbb{1}_A\right)\left(\sum_{\substack{x'\in\mathcal{X},q(x')=0,\\ p(x')\neq 0}} |x'\rangle\langle x'|_X \otimes \varepsilon^{-\alpha}(\rho_A^{x'})^{\alpha}\right)\right] \tag{7.6.126}$$

$$= \sum_{x\in\mathcal{X},q(x)\neq 0,p(x)\neq 0} p(x)^{\alpha} q(x)^{1-\alpha} \operatorname{Tr}\left[\sigma_{A,\varepsilon}^{x}\left(\omega_A^{x}\right)^{\alpha}\right]$$

$$+ \sum_{x\in\mathcal{X},q(x)=0,p(x)\neq 0} \varepsilon^{1-\alpha}\operatorname{Tr}[(\rho_A^x)^{\alpha}]. \tag{7.6.127}$$

Now observing that $\operatorname{Tr}\left[\sigma_{A,\varepsilon}^{x}\left(\omega_A^{x}\right)^{\alpha}\right] = \widehat{Q}_{\alpha}(\rho_A^x\|\sigma_{A,\varepsilon}^x)$ and taking the limit $\varepsilon \to 0^+$ in the last line above, we find that

$$\lim_{\varepsilon\to 0^+}\left(\sum_{\substack{x\in\mathcal{X},\\ q(x)\neq 0,\\ p(x)\neq 0}} p(x)^{\alpha} q(x)^{1-\alpha}\widehat{Q}_{\alpha}(\rho_A^x\|\sigma_{A,\varepsilon}^x) + \sum_{\substack{x\in\mathcal{X},\\ q(x)=0,\\ p(x)\neq 0}} \varepsilon^{1-\alpha}\operatorname{Tr}[(\rho_A^x)^{\alpha}]\right)$$

$$= \sum_{x\in\mathcal{X}} p(x)^{\alpha} q(x)^{1-\alpha}\widehat{Q}_{\alpha}(\rho_A^x\|\sigma_A^x) \tag{7.6.128}$$

if $\alpha \in (0,1)$ or if $\alpha \in (1,\infty)$, $\operatorname{supp}(\rho_A^x) \subseteq \operatorname{supp}(\sigma_A^x)$, and there does not exist a value of x for which $p(x) \neq 0$ and $q(x) = 0$. The latter support conditions are precisely the same as $\operatorname{supp}(\rho_{XA}) \subseteq \operatorname{supp}(\sigma_{XA})$. If $\alpha \in (1,\infty)$ and the support conditions do not hold, then the limit evaluates to $+\infty$, consistent with the right-hand side above. This concludes the proof of the direct-sum property. ■

We now establish the data-processing inequality for the geometric Rényi relative entropy for $\alpha \in (0,1) \cup (1,2]$.

Theorem 7.45 Data-Processing Inequality for Geometric Rényi Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then, for all $\alpha \in (0,1) \cup (1,2]$,

$$\widehat{D}_\alpha(\rho\|\sigma) \geq \widehat{D}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.6.129}$$

Proof: From Stinespring's dilation theorem (Theorem 4.3), we know that the action of a quantum channel $\mathcal{N}$ on every linear operator X can be written as

$$\mathcal{N}(X) = \mathrm{Tr}_E[VXV^\dagger], \tag{7.6.130}$$

where V is an isometry and E is an auxiliary system with dimension $d_E \geq \mathrm{rank}(\Gamma^{\mathcal{N}}_{AB})$, with $\Gamma^{\mathcal{N}}_{AB}$ the Choi operator for the channel $\mathcal{N}$. As stated in Proposition 7.44, the geometric Rényi relative entropy $\widehat{D}_\alpha$ is isometrically invariant. Therefore, it suffices to establish the data-processing inequality for $\widehat{D}_\alpha$ under partial trace; i.e., it suffices to show that for every state ρ_{AB}, positive semi-definite operator σ_{AB}, and for all $\alpha \in (0,1) \cup (1,2]$:

$$\widehat{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \geq \widehat{D}_\alpha(\rho_A\|\sigma_A). \tag{7.6.131}$$

We now proceed to prove this inequality. We prove it for ρ_{AB}, and hence ρ_A, invertible, as well as for σ_{AB} and σ_A invertible. The result follows in the general case of ρ_{AB} and/or ρ_A non-invertible, as well as σ_{AB} and/or σ_A non-invertible, by applying the result to the invertible operators $(1-\delta)\,\rho_{AB} + \delta\pi_{AB}$ and $\sigma_{AB} + \varepsilon\mathbb{1}_{AB}$, with $\delta \in (0,1)$ and $\varepsilon > 0$, and taking the limit $\delta \to 0^+$ followed by $\varepsilon \to 0^+$, because

$$\widehat{D}_\alpha(\rho_{AB}\|\sigma_{AB}) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \widehat{D}_\alpha((1-\delta)\,\rho_{AB} + \delta\pi_{AB}\|\sigma_{AB} + \varepsilon\mathbb{1}_{AB}), \tag{7.6.132}$$

$$\widehat{D}_\alpha(\rho_A\|\sigma_A) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \widehat{D}_\alpha((1-\delta)\,\rho_A + \delta\pi_A\|\sigma_A + d_B\varepsilon\mathbb{1}_A), \tag{7.6.133}$$

which follows from (7.6.22) and the fact that the dimensional factor d_B does not affect the limit in the second quantity above.

To establish the data-processing inequality, we make use of the Petz recovery channel for partial trace (see Section 4.6.1.1), as well as the operator Jensen inequality (Theorem 2.16). Recall that the Petz recovery channel $\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}$ for partial trace is defined as

$$\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}(X_A) \equiv \mathcal{P}(X_A) := \sigma_{AB}^{\frac{1}{2}}\left(\sigma_A^{-\frac{1}{2}} X_A \sigma_A^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\sigma_{AB}^{\frac{1}{2}}. \tag{7.6.134}$$

The Petz recovery channel has the following property:

$$\mathcal{P}(\sigma_A) = \sigma_{AB}, \tag{7.6.135}$$

which can be verified by inspection. Since $\mathcal{P}_{\sigma_{AB},\mathrm{Tr}_B}$ is completely positive and trace preserving, it follows that its adjoint

$$\mathcal{P}^\dagger(Y_{AB}) := \sigma_A^{-\frac{1}{2}} \mathrm{Tr}_B[\sigma_{AB}^{\frac{1}{2}} Y_{AB} \sigma_{AB}^{\frac{1}{2}}] \sigma_A^{-\frac{1}{2}}, \tag{7.6.136}$$

is completely positive and unital. Observe that

$$\mathcal{P}^\dagger(\sigma_{AB}^{-\frac{1}{2}} \rho_{AB} \sigma_{AB}^{-\frac{1}{2}}) = \sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}}. \tag{7.6.137}$$

We then find for $\alpha \in (1,2]$ that

$$\widehat{Q}_\alpha(\rho_{AB}\|\sigma_{AB}) = \mathrm{Tr}\left[\sigma_{AB}\left(\sigma_{AB}^{-\frac{1}{2}}\rho_{AB}\sigma_{AB}^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.138}$$

$$= \mathrm{Tr}\left[\mathcal{P}(\sigma_A)\left(\sigma_{AB}^{-\frac{1}{2}}\rho_{AB}\sigma_{AB}^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.139}$$

$$= \mathrm{Tr}\left[\sigma_A \mathcal{P}^\dagger\left(\left(\sigma_{AB}^{-\frac{1}{2}}\rho_{AB}\sigma_{AB}^{-\frac{1}{2}}\right)^\alpha\right)\right] \tag{7.6.140}$$

$$\geq \mathrm{Tr}\left[\sigma_A\left(\mathcal{P}^\dagger\left(\sigma_{AB}^{-\frac{1}{2}}\rho_{AB}\sigma_{AB}^{-\frac{1}{2}}\right)\right)^\alpha\right] \tag{7.6.141}$$

$$= \mathrm{Tr}\left[\sigma_A\left(\sigma_A^{-\frac{1}{2}}\rho_A\sigma_A^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.142}$$

$$= \widehat{Q}_\alpha(\rho_A\|\sigma_A). \tag{7.6.143}$$

The second equality follows from (7.6.135). The sole inequality is a consequence of the operator Jensen inequality and the fact that x^α is operator convex for $\alpha \in (1,2]$. Indeed, for $\mathcal{M}$ a completely positive unital map, item 2. of Theorem 2.16 implies that

$$f(\mathcal{M}(X)) \leq \mathcal{M}(f(X)) \tag{7.6.144}$$

for Hermitian X and an operator convex function f. The second-to-last equality follows from (7.6.137).

Applying the same reasoning as above, but using the fact that x^α is operator concave for $\alpha \in (0,1)$, we find for $\alpha \in (0,1)$ that

$$\widehat{Q}_\alpha(\rho_A\|\sigma_A) \geq \widehat{Q}_\alpha(\rho_{AB}\|\sigma_{AB}). \tag{7.6.145}$$

Putting together the above and employing definitions, we find that the following inequality holds for $\alpha \in (0,1) \cup (1,2]$:

$$\widehat{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \geq \widehat{D}_\alpha(\rho_A\|\sigma_A), \tag{7.6.146}$$

concluding the proof. ■

With the data-processing inequality for the geometric Rényi relative entropy in hand, we can establish some additional properties.

Proposition 7.46 Additional Properties of Geometric Rényi Relative Entropy

The geometric Rényi relative entropy $\widehat{D}_\alpha$ satisfies the following properties for every state ρ and positive semi-definite operator σ for $\alpha \in (0,1) \cup (1,2]$:

1. If $\mathrm{Tr}[\sigma] \leq \mathrm{Tr}[\rho] = 1$, then $\widehat{D}_\alpha(\rho\|\sigma) \geq 0$.
2. Faithfulness: Suppose that $\mathrm{Tr}[\sigma] \leq \mathrm{Tr}[\rho] = 1$ and let $\alpha \in (0,1) \cup (1,\infty)$. Then $\widehat{D}_\alpha(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.
3. If $\rho \leq \sigma$, then $\widehat{D}_\alpha(\rho\|\sigma) \leq 0$.
4. For every positive semi-definite operator σ' such that $\sigma' \geq \sigma$, the following inequality holds $\widehat{D}_\alpha(\rho\|\sigma) \geq \widehat{D}_\alpha(\rho\|\sigma')$.

PROOF:

1. Apply the data-processing inequality with the channel being the full trace-out channel:

$$\widehat{D}_\alpha(\rho\|\sigma) \geq \widehat{D}_\alpha(\mathrm{Tr}[\rho]\|\mathrm{Tr}[\sigma]) \tag{7.6.147}$$

$$= \frac{1}{\alpha-1}\log_2\left[(\mathrm{Tr}[\rho])^\alpha\,(\mathrm{Tr}[\sigma])^{1-\alpha}\right] \tag{7.6.148}$$

$$= -\log_2 \mathrm{Tr}[\sigma] \tag{7.6.149}$$

$$\geq 0. \tag{7.6.150}$$

2. If $\rho = \sigma$, then it follows by direct evaluation that $\widehat{D}_\alpha(\rho\|\sigma) = 0$ for $\alpha \in (0,1) \cup (1,\infty)$.

To see the other implication, suppose first that $(0,1)\cup(1,2]$. Then $\widehat{D}_\alpha(\rho\|\sigma) = 0$ implies that equality is achieved in the two inequalities in item 1. above. So then $\mathrm{Tr}[\sigma] = 1$. Furthermore, we conclude from data processing that $\widehat{D}_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0$ for all measurement channels $\mathcal{M}$. This includes the measurement that achieves the trace distance. By applying the faithfulness of the classical Rényi relative entropy on the distributions that result from the optimal trace-distance measurement, we conclude that $\rho = \sigma$. To get the range outside the data-processing interval of $(0,1)\cup(1,2]$, note that $\widehat{D}_\alpha(\rho\|\sigma) = 0$ for $\alpha > 2$ implies by monotonicity (Property 2 of Proposition 7.44) that $\widehat{D}_\alpha(\rho\|\sigma) = 0$ for $\alpha \leq 2$. Then it follows that $\rho = \sigma$.

3. Consider that $\rho \leq \sigma$ implies that $\sigma - \rho \geq 0$. Then define the following positive semi-definite operators:

$$\hat{\rho} := |0\rangle\langle 0| \otimes \rho, \tag{7.6.151}$$

$$\hat{\sigma} := |0\rangle\langle 0| \otimes \rho + |1\rangle\langle 1| \otimes (\sigma - \rho)\,. \tag{7.6.152}$$

By exploiting the direct-sum property of geometric Rényi relative entropy (Proposition 7.44) and the data-processing inequality (Theorem 7.45), we find that

$$0 = \widehat{D}_\alpha(\rho\|\rho) = \widehat{D}_\alpha(\hat{\rho}\|\hat{\sigma}) \geq \widehat{D}_\alpha(\rho\|\sigma), \tag{7.6.153}$$

where the inequality follows from data processing with respect to partial trace over the classical register.

4. Similar to the above proof, the condition $\sigma' \geq \sigma$ implies that $\sigma' - \sigma \geq 0$. Then define the following positive semi-definite operators:

$$\hat{\rho} := |0\rangle\langle 0| \otimes \rho, \tag{7.6.154}$$

$$\hat{\sigma} := |0\rangle\langle 0| \otimes \sigma + |1\rangle\langle 1| \otimes (\sigma' - \sigma)\,. \tag{7.6.155}$$

By exploiting the direct-sum property of geometric Rényi relative entropy (Proposition 7.44) and the data-processing inequality (Theorem 7.45), we find that

$$\widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}_\alpha(\hat{\rho}\|\hat{\sigma}) \geq \widehat{D}_\alpha(\rho\|\sigma'), \tag{7.6.156}$$

where the inequality follows from data processing with respect to partial trace over the classical register. ■

The data-processing inequality for the geometric Rényi relative entropy can be written using the geometric Rényi relative quasi-entropy $\widehat{Q}_\alpha(\rho\|\sigma)$ as

$$\frac{1}{\alpha-1}\log_2\widehat{Q}_\alpha(\rho\|\sigma) \geq \frac{1}{\alpha-1}\log_2\widehat{Q}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.6.157}$$

Since $\alpha-1$ is negative for $\alpha\in(0,1)$, we can use the monotonicity of the function $\log_2$ to obtain

$$\widehat{Q}_\alpha(\rho\|\sigma) \geq \widehat{Q}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \qquad \text{for } \alpha\in(1,2], \tag{7.6.158}$$

$$\widehat{Q}_\alpha(\rho\|\sigma) \leq \widehat{Q}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \qquad \text{for } \alpha\in(0,1). \tag{7.6.159}$$

We can use this to establish some convexity/concavity statements for the geometric Rényi relative entropy.

Proposition 7.47 Joint Convexity & Concavity of the Geometric Rényi Relative Quasi-Entropy

Let $p:\mathcal{X}\to[0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on system A, and let $\{\sigma_A^x\}_{x\in\mathcal{X}}$ be a set of positive semi-definite operators on A. Then, for $\alpha\in(1,2]$,

$$\widehat{Q}_\alpha\left(\sum_{x\in\mathcal{X}}p(x)\rho_A^x\middle\|\sum_{x\in\mathcal{X}}p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}}p(x)\widehat{Q}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.6.160}$$

and for $\alpha\in(0,1)$,

$$\widehat{Q}_\alpha\left(\sum_{x\in\mathcal{X}}p(x)\rho_A^x\middle\|\sum_{x\in\mathcal{X}}p(x)\sigma_A^x\right) \geq \sum_{x\in\mathcal{X}}p(x)\widehat{Q}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.6.161}$$

Consequently, the geometric Rényi relative entropy $\widehat{D}_\alpha$ is jointly convex for $\alpha\in(0,1)$:

$$\widehat{D}_\alpha\left(\sum_{x\in\mathcal{X}}p(x)\rho_A^x\middle\|\sum_{x\in\mathcal{X}}p(x)\sigma_A^x\right) \leq \sum_{x\in\mathcal{X}}p(x)\widehat{D}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.6.162}$$

PROOF: The first two inequalities follow directly from the direct-sum property of the geometric Rényi relative entropy (Proposition 7.44), the data-processing inequality

(Theorem 7.45), and Proposition 7.17. The last inequality follows from the first by applying the logarithm, scaling by $1/(\alpha-1)$, and taking a maximum. ■

Although the geometric Rényi relative entropy is not jointly convex for $\alpha \in (1,2]$, it is jointly quasi-convex, in the sense that

$$\widehat{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \le \max_{x\in\mathcal{X}} \widehat{D}_\alpha(\rho_A^x\|\sigma_A^x), \tag{7.6.163}$$

for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0,1]$, set $\{\rho_A^x\}_{x\in\mathcal{X}}$ of states, and set $\{\sigma_A^x\}_{x\in\mathcal{X}}$ of positive semi-definite operators. Indeed, from (7.6.160), we immediately obtain

$$\widehat{Q}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_A^x\right) \le \max_{x\in\mathcal{X}} \widehat{Q}_\alpha(\rho_A^x\|\sigma_A^x). \tag{7.6.164}$$

Taking the logarithm and multiplying by $\frac{1}{\alpha-1}$ on both sides of this inequality leads to (7.6.163).

The geometric Rényi relative entropy has another interpretation, which is worthwhile to mention.

Proposition 7.48 Geometric Rényi Relative Entropy from Classical Preparations

Let ρ be a state and σ a positive semi-definite operator satisfying $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. For all $\alpha \in (0,1) \cup (1,2]$, the geometric Rényi relative entropy is equal to the smallest value that the classical Rényi relative entropy can take by minimizing over classical–quantum channels that realize the state ρ and the positive semi-definite operator σ. That is, the following equality holds

$$\widehat{D}_\alpha(\rho\|\sigma) = \inf_{\{p,q,\mathcal{P}\}} \{D_\alpha(p\|q) : \mathcal{P}(\omega(p)) = \rho,\ \mathcal{P}(\omega(q)) = \sigma\}, \tag{7.6.165}$$

where the classical Rényi relative entropy is defined in (7.4.19), the channel $\mathcal{P}$ is a classical–quantum channel, $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$, $q : \mathcal{X} \to [0,\infty)$ is a positive function on $\mathcal{X}$, $\omega(p) := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|$, and $\omega(q) := \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|$.

PROOF: First, suppose that there exists a quantum channel $\mathcal{P}$ such that

$$\mathcal{P}(\omega(p)) = \rho, \qquad \mathcal{P}(\omega(q)) = \sigma. \tag{7.6.166}$$

Then consider the following chain of inequalities:

$$\begin{aligned} D_\alpha(p\|q) &= \widehat{D}_\alpha(\omega(p)\|\omega(q)) && (7.6.167)\\ &\geq \widehat{D}_\alpha(\mathcal{P}(\omega(p))\|\mathcal{P}(\omega(q))) && (7.6.168)\\ &= \widehat{D}_\alpha(\rho\|\sigma). && (7.6.169) \end{aligned}$$

The first equality follows because the geometric Rényi relative entropy reduces to the classical Rényi relative entropy for commuting operators. The inequality is a consequence of the data-processing inequality for the geometric Rényi relative entropy (Theorem 7.45). The final equality follows from the constraint in (7.6.166). Since the inequality holds for arbitrary p, q, and $\mathcal{P}$ satisfying (7.6.166), we conclude that

$$\inf_{\{p,q,\mathcal{P}\}} \{D_\alpha(p\|q) : \mathcal{P}(p) = \rho, \mathcal{P}(q) = \sigma\} \geq \widehat{D}_\alpha(\rho\|\sigma). \tag{7.6.170}$$

The equality in (7.6.165) then follows by demonstrating a specific distribution p, positive function q, and preparation channel $\mathcal{P}$ that saturate the inequality in (7.6.170). The optimal choices of p, q, and $\mathcal{P}$ are given by

$$\begin{aligned} p(x) &\coloneqq \lambda_x q(x), && (7.6.171)\\ q(x) &\coloneqq \operatorname{Tr}[\Pi_x\sigma], && (7.6.172)\\ \mathcal{P}(\cdot) &\coloneqq \sum_x \langle x|(\cdot)|x\rangle \frac{\sigma^{\frac{1}{2}}\Pi_x\sigma^{\frac{1}{2}}}{q(x)}, && (7.6.173) \end{aligned}$$

where the spectral decomposition of the positive semi-definite operator $\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}$ is given by

$$\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} = \sum_x \lambda_x \Pi_x. \tag{7.6.174}$$

The choice of $p(x)$ above is a probability distribution because

$$\sum_x p(x) = \sum_x \lambda_x q(x) = \sum_x \lambda_x \operatorname{Tr}[\Pi_x\sigma] = \operatorname{Tr}[\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\sigma] = \operatorname{Tr}[\Pi_\sigma\rho] = 1. \tag{7.6.175}$$

The preparation channel $\mathcal{P}$ is a classical–quantum channel that measures the input in the basis $\{|x\rangle\}_x$ and prepares the state $\frac{\sigma^{\frac{1}{2}}\Pi_x\sigma^{\frac{1}{2}}}{q(x)}$ if the measurement outcome is x. We find that

$$\mathcal{P}(\omega(p)) = \sum_x \frac{p(x)}{q(x)}\sigma^{\frac{1}{2}}\Pi_x\sigma^{\frac{1}{2}} = \sum_x \frac{\lambda_x q(x)}{q(x)}\sigma^{\frac{1}{2}}\Pi_x\sigma^{\frac{1}{2}} = \sigma^{\frac{1}{2}}\left(\sum_x \lambda_x\Pi_x\right)\sigma^{\frac{1}{2}}$$
$$= \sigma^{\frac{1}{2}}\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)\sigma^{\frac{1}{2}} = \Pi_\sigma\rho\Pi_\sigma = \rho, \tag{7.6.176}$$

and

$$\mathcal{P}(\omega(q)) = \sum_x \frac{q(x)}{q(x)}\sigma^{\frac{1}{2}}\Pi_x\sigma^{\frac{1}{2}} = \sigma^{\frac{1}{2}}\left(\sum_x \Pi_x\right)\sigma^{\frac{1}{2}} = \sigma. \tag{7.6.177}$$

Finally, consider the classical Rényi relative quasi-entropy:

$$\sum_x p(x)^\alpha q(x)^{1-\alpha} = \sum_x (\lambda_x q(x))^\alpha q(x)^{1-\alpha} = \sum_x \lambda_x^\alpha q(x) = \sum_x \lambda_x^\alpha \operatorname{Tr}[\Pi_x\sigma]$$
$$= \operatorname{Tr}\left[\sigma\left(\sum_x \lambda_x^\alpha\Pi_x\right)\right] = \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha\right] = \widehat{Q}_\alpha(\rho\|\sigma), \tag{7.6.178}$$

where the second-to-last equality follows from the spectral decomposition in (7.6.174) and the form of the geometric Rényi relative quasi-entropy from Proposition 7.40. As a consequence of the equality

$$\sum_x p(x)^\alpha q(x)^{1-\alpha} = \widehat{Q}_\alpha(\rho\|\sigma), \tag{7.6.179}$$

and the fact that these choices of p, q, and $\mathcal{P}$ satisfy the constraints $\mathcal{P}(p) = \rho$ and $\mathcal{P}(q) = \sigma$, we conclude that

$$D_\alpha(p\|q) = \widehat{D}_\alpha(\rho\|\sigma). \tag{7.6.180}$$

Combining this equality with (7.6.170), we conclude the equality in (7.6.165). ■

The following proposition establishes the ordering between the sandwiched, Petz–, and geometric Rényi relative entropies for the interval $\alpha \in (0,1) \cup (1,2]$. It follows by applying similar reasoning as in the proof of Proposition 7.48.

Proposition 7.49

Let ρ be a state and σ a positive semi-definite operator. For $\alpha \in (0,1) \cup (1,2]$, the following inequalities hold

$$\widetilde{D}_\alpha(\rho\|\sigma) \le D_\alpha(\rho\|\sigma) \le \widehat{D}_\alpha(\rho\|\sigma), \tag{7.6.181}$$

for the sandwiched ($\widetilde{D}_\alpha$), Petz (D_α), and geometric ($\widehat{D}_\alpha$) Rényi relative entropies.

PROOF: The first inequality was stated as the last property of Proposition 7.31. So we establish the proof of the second inequality here. Suppose that $\mathcal{P}$ is a classical–quantum channel, $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$, and $q : \mathcal{X} \to (0,\infty)$ is a positive function on $\mathcal{X}$ satisfying

$$\mathcal{P}(\omega(p)) = \rho, \qquad \mathcal{P}(\omega(q)) = \sigma, \tag{7.6.182}$$

where

$$\omega(p) \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|, \qquad \omega(q) \coloneqq \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|. \tag{7.6.183}$$

Then consider the following chain of inequalities:

$$D_\alpha(p\|q) = D_\alpha(\omega(p)\|\omega(q)) \tag{7.6.184}$$
$$\ge D_\alpha(\mathcal{P}(\omega(p))\|\mathcal{P}(\omega(q))) \tag{7.6.185}$$
$$= D_\alpha(\rho\|\sigma). \tag{7.6.186}$$

The first equality follows because the Petz–Rényi relative entropy reduces to the classical Rényi relative entropy for commuting operators. The inequality follows from the data-processing inequality for the Petz–Rényi relative entropy for $\alpha \in (0,1) \cup (1,2]$ (Theorem 7.24). The final equality follows from the constraint in (7.6.182). Since the inequality above holds for all p, q, and $\mathcal{P}$ satisfying (7.6.182), we conclude that

$$\inf_{\{p,q,\mathcal{P}\}} \{D_\alpha(p\|q) : \mathcal{P}(p) = \rho, \mathcal{P}(q) = \sigma\} \ge D_\alpha(\rho\|\sigma). \tag{7.6.187}$$

Now applying Proposition 7.48, we conclude the second inequality in (7.6.181). ■

7.6.1 Proof of Proposition 7.41

First consider that

$$(1-\delta)\,\rho'_\delta \le \rho_\delta \le \rho'_\delta, \tag{7.6.188}$$

where

$$\rho'_\delta \coloneqq \rho + \delta\pi. \tag{7.6.189}$$

By operator monotonicity of x^α for $\alpha \in (0,1)$, we conclude that

$$(1-\delta)^\alpha \operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \le \operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.190}$$

$$\le \operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.191}$$

The prefactors in these bounds to the left of the trace expressions are uniform and independent of ε, and so it follows that

$$\lim_{\varepsilon\to0^+}\lim_{\delta\to0^+}\operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \lim_{\varepsilon\to0^+}\lim_{\delta\to0^+}\operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.192}$$

$$\lim_{\delta\to0^+}\lim_{\varepsilon\to0^+}\operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \lim_{\delta\to0^+}\lim_{\varepsilon\to0^+}\operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.193}$$

Again from the operator monotonicity of x^α for $\alpha \in (0,1)$, we conclude for fixed $\varepsilon > 0$ that

$$\delta_1 \le \delta_2 \qquad \Rightarrow \qquad \operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_{\delta_1}\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \le \operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_{\delta_2}\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.194}$$

where $\delta_1 > 0$. By exploiting the identity

$$\operatorname{Tr}\!\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\!\left[\rho'_\delta\left(\left(\rho'_\delta\right)^{-\frac{1}{2}}\sigma_\varepsilon\left(\rho'_\delta\right)^{-\frac{1}{2}}\right)^{1-\alpha}\right] \tag{7.6.195}$$

from Proposition 7.39 and operator monotonicity of $x^{1-\alpha}$ for $\alpha \in (0,1)$, we conclude for fixed $\delta > 0$ that

$$\varepsilon_1 \le \varepsilon_2 \qquad \Rightarrow \qquad \operatorname{Tr}\!\left[\sigma_{\varepsilon_1}\left(\sigma_{\varepsilon_1}^{-\frac{1}{2}}\rho_\delta\sigma_{\varepsilon_1}^{-\frac{1}{2}}\right)^\alpha\right] \le \operatorname{Tr}\!\left[\sigma_{\varepsilon_2}\left(\sigma_{\varepsilon_2}^{-\frac{1}{2}}\rho'_\delta\sigma_{\varepsilon_2}^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.196}$$

where $\varepsilon_1 > 0$. Thus, we find that

$$\lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+}\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \inf_{\varepsilon>0}\inf_{\delta>0}\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.197}$$

$$\lim_{\delta\to 0^+}\lim_{\varepsilon\to 0^+}\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \inf_{\delta>0}\inf_{\varepsilon>0}\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho'_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.198}$$

Since infima can be exchanged, we conclude the statement of the proposition.

7.6.2 Proof of Proposition 7.40

First suppose that $\alpha \in (1,\infty)$ and $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$. Then from Propositions 7.29 and 7.42 and the fact that the sandwiched Rényi relative quasi-entropy $\widetilde{Q}_\alpha(\rho\|\sigma) = +\infty$ in this case, it follows that $\widehat{Q}_\alpha(\rho\|\sigma) = +\infty$, thus establishing the third expression in (7.6.16).

Now suppose that $\alpha \in (0,1)\cup(1,\infty)$ and $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. Let us employ the decomposition of the Hilbert space $\mathcal{H}$ as $\mathcal{H} = \operatorname{supp}(\sigma)\oplus\ker(\sigma)$. Then we can write ρ as

$$\rho = \begin{pmatrix}\rho_{0,0} & \rho_{0,1}\\ \rho_{0,1}^\dagger & \rho_{1,1}\end{pmatrix}, \qquad \sigma = \begin{pmatrix}\sigma & 0\\ 0 & 0\end{pmatrix}. \tag{7.6.199}$$

Writing $\mathbb{1} = \Pi_\sigma + \Pi_\sigma^\perp$, where Π_σ is the projection onto the support of σ and $\Pi_\sigma^\perp$ is the projection onto the orthogonal complement of $\operatorname{supp}(\sigma)$, we find that

$$\sigma_\varepsilon = \begin{pmatrix}\sigma + \varepsilon\Pi_\sigma & 0\\ 0 & \varepsilon\Pi_\sigma^\perp\end{pmatrix}, \tag{7.6.200}$$

which implies that

$$\sigma_\varepsilon^{-\frac{1}{2}} = \begin{pmatrix}(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}} & 0\\ 0 & \varepsilon^{-\frac{1}{2}}\Pi_\sigma^\perp\end{pmatrix}. \tag{7.6.201}$$

The condition $\operatorname{supp}(\rho)\subseteq\operatorname{supp}(\sigma)$ implies that $\rho_{0,1} = 0$ and $\rho_{1,1} = 0$. Then

$$\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}} = \begin{pmatrix}(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}}\rho_{0,0}(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}} & 0\\ 0 & 0\end{pmatrix}, \tag{7.6.202}$$

so that

$$\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]$$

$$= \operatorname{Tr}\left[\begin{pmatrix}\sigma+\varepsilon\Pi_\sigma & 0\\ 0 & \varepsilon\Pi_\sigma^\perp\end{pmatrix}\begin{pmatrix}\left[(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}}\rho_{0,0}(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}}\right]^\alpha & 0\\ 0 & 0\end{pmatrix}\right] \tag{7.6.203}$$

$$= \operatorname{Tr}\left[(\sigma+\varepsilon\Pi_\sigma)\left[(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}}\rho_{0,0}(\sigma+\varepsilon\Pi_\sigma)^{-\frac{1}{2}}\right]^\alpha\right]. \tag{7.6.204}$$

Taking the limit $\varepsilon \to 0^+$ then leads to

$$\lim_{\varepsilon\to 0^+}\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho_{0,0}\sigma^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.205}$$

$$= \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.206}$$

thus establishing the first expression in (7.6.16).

We now establish (7.6.19). For $\alpha \in (1,\infty)$ and $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, the same analysis implies that

$$\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\hat{\sigma}_\varepsilon\left(\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.207}$$

where

$$\hat{\sigma}_\varepsilon := \sigma + \varepsilon\Pi_\sigma. \tag{7.6.208}$$

Since

$$\left(\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right)^\alpha = \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\left(\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right)^{\alpha-1} \tag{7.6.209}$$

for $\alpha > 1$, we have that

$$\operatorname{Tr}\left[\hat{\sigma}_\varepsilon\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\left(\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right)^{\alpha-1}\right]$$

$$= \operatorname{Tr}\left[\hat{\sigma}_\varepsilon^{\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\left(\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right)^{\alpha-1}\right] \tag{7.6.210}$$

$$= \operatorname{Tr}\left[\hat{\sigma}_\varepsilon^{\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\left(\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\right)^{\alpha-1}\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\right] \tag{7.6.211}$$

$$= \operatorname{Tr}\left[\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\hat{\sigma}_\varepsilon^{\frac{1}{2}}\rho_{0,0}^{\frac{1}{2}}\left(\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-1}\rho_{0,0}^{\frac{1}{2}}\right)^{\alpha-1}\right] \tag{7.6.212}$$

$$= \operatorname{Tr}\left[\rho_{0,0}\left(\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-1}\rho_{0,0}^{\frac{1}{2}}\right)^{\alpha-1}\right], \tag{7.6.213}$$

where we applied Lemma 2.5 with $f(x) = x^{\alpha-1}$ and $L = \rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}$. Now taking the limit $\varepsilon \to 0^+$, we conclude that

$$\lim_{\varepsilon\to 0^+} \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \lim_{\varepsilon\to 0^+} \mathrm{Tr}\left[\rho_{0,0}\left(\rho_{0,0}^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-1}\rho_{0,0}^{\frac{1}{2}}\right)^{\alpha-1}\right] \tag{7.6.214}$$

$$= \mathrm{Tr}\left[\rho_{0,0}\left(\rho_{0,0}^{\frac{1}{2}}\sigma^{-1}\rho_{0,0}^{\frac{1}{2}}\right)^{\alpha-1}\right] \tag{7.6.215}$$

$$= \mathrm{Tr}\left[\rho\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right], \tag{7.6.216}$$

for the case $\alpha \in (1, \infty)$ and $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, thus establishing (7.6.19).

For the case that $\alpha \in (0, 1)$ and $\mathrm{supp}(\sigma) \subseteq \mathrm{supp}(\rho)$, we can employ the limit exchange from Lemma 7.41 and a similar argument as in (7.6.199)–(7.6.206), but with respect to the decomposition $\mathcal{H} = \mathrm{supp}(\rho) \oplus \ker(\rho)$, to conclude that

$$\widehat{Q}_\alpha(\rho\|\sigma) = \mathrm{Tr}\left[\rho\left(\rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}\right)^{1-\alpha}\right], \tag{7.6.217}$$

thus establishing the second expression in (7.6.16). This case amounts to the exchange $\rho \leftrightarrow \sigma$ and $\alpha \leftrightarrow 1-\alpha$.

We finally consider the case $\alpha \in (0, 1)$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$, which is the most involved case. Consider that

$$\sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1} = \begin{pmatrix} \hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon\Pi_\sigma^\perp \end{pmatrix}, \tag{7.6.218}$$

where $\hat{\sigma}_\varepsilon \coloneqq \sigma + \varepsilon\Pi_\sigma$. Let us define

$$\rho_\delta \coloneqq (1-\delta)\,\rho + \delta\pi, \tag{7.6.219}$$

with $\delta \in (0, 1)$ and π the maximally mixed state. By invoking Lemma 7.41, we conclude that the following exchange of limits is possible for $\alpha \in (0, 1)$:

$$\lim_{\varepsilon\to 0^+} D_\alpha(\rho\|\sigma_\varepsilon) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} D_\alpha(\rho_\delta\|\sigma_\varepsilon) = \lim_{\delta\to 0^+}\lim_{\varepsilon\to 0^+} D_\alpha(\rho_\delta\|\sigma_\varepsilon). \tag{7.6.220}$$

Now define

$$\rho_{0,0}^\delta \coloneqq \Pi_\sigma\rho_\delta\Pi_\sigma, \quad \rho_{0,1}^\delta \coloneqq \Pi_\sigma\rho_\delta\Pi_\sigma^\perp, \quad \rho_{1,1}^\delta \coloneqq \Pi_\sigma^\perp\rho_\delta\Pi_\sigma^\perp, \tag{7.6.221}$$

so that

$$\rho_\delta = \begin{pmatrix} \rho^\delta_{0,0} & \rho^\delta_{0,1} \\ (\rho^\delta_{0,1})^\dagger & \rho^\delta_{1,1} \end{pmatrix}. \tag{7.6.222}$$

Then

$$D_\alpha(\rho_\delta\|\sigma_\varepsilon) = \frac{1}{\alpha-1}\log_2 \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.223}$$

Consider that

$$\begin{aligned}
\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}} &= \begin{pmatrix} \hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon\Pi^\perp_\sigma \end{pmatrix}^{-\frac{1}{2}} \begin{pmatrix} \rho^\delta_{0,0} & \rho^\delta_{0,1} \\ (\rho^\delta_{0,1})^\dagger & \rho^\delta_{1,1} \end{pmatrix} \begin{pmatrix} \hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon\Pi^\perp_\sigma \end{pmatrix}^{-\frac{1}{2}} &(7.6.224)\\
&= \begin{pmatrix} \hat{\sigma}_\varepsilon^{-\frac{1}{2}} & 0 \\ 0 & \varepsilon^{-\frac{1}{2}}\Pi^\perp_\sigma \end{pmatrix} \begin{pmatrix} \rho^\delta_{0,0} & \rho^\delta_{0,1} \\ (\rho^\delta_{0,1})^\dagger & \rho^\delta_{1,1} \end{pmatrix} \begin{pmatrix} \hat{\sigma}_\varepsilon^{-\frac{1}{2}} & 0 \\ 0 & \varepsilon^{-\frac{1}{2}}\Pi^\perp_\sigma \end{pmatrix} &(7.6.225)\\
&= \begin{pmatrix} \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1}\Pi^\perp_\sigma \\ \varepsilon^{-\frac{1}{2}}\Pi^\perp_\sigma(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-1}\Pi^\perp_\sigma\rho^\delta_{1,1}\Pi^\perp_\sigma \end{pmatrix} &(7.6.226)\\
&= \begin{pmatrix} \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1} \\ \varepsilon^{-\frac{1}{2}}(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-1}\rho^\delta_{1,1} \end{pmatrix}. &(7.6.227)
\end{aligned}$$

So then

$$\begin{aligned}
&\mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \\
&= \mathrm{Tr}\left[\begin{pmatrix} \hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon\Pi^\perp_\sigma \end{pmatrix}\left(\begin{pmatrix} \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1} \\ \varepsilon^{-\frac{1}{2}}(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{-1}\rho^\delta_{1,1} \end{pmatrix}\right)^\alpha\right] &(7.6.228)\\
&= \mathrm{Tr}\left[\begin{pmatrix} \hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon\Pi^\perp_\sigma \end{pmatrix}\left(\varepsilon^{-1}\begin{pmatrix} \varepsilon\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1} \\ \varepsilon^{\frac{1}{2}}(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \rho^\delta_{1,1} \end{pmatrix}\right)^\alpha\right] &(7.6.229)\\
&= \mathrm{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi^\perp_\sigma \end{pmatrix}\begin{pmatrix} \varepsilon\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1} \\ \varepsilon^{\frac{1}{2}}(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \rho^\delta_{1,1} \end{pmatrix}^\alpha\right] &(7.6.230)
\end{aligned}$$

Let us define

$$K(\varepsilon) := \begin{pmatrix} \varepsilon\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,0}\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \varepsilon^{\frac{1}{2}}\hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho^\delta_{0,1} \\ \varepsilon^{\frac{1}{2}}(\rho^\delta_{0,1})^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}} & \rho^\delta_{1,1} \end{pmatrix}, \tag{7.6.231}$$

so that we can write

$$\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\begin{pmatrix}\varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp\end{pmatrix}(K(\varepsilon))^\alpha\right]. \tag{7.6.232}$$

Now let us invoke Lemma 7.50 with the substitutions

$$A \leftrightarrow \rho_{1,1}^\delta, \tag{7.6.233}$$

$$B \leftrightarrow (\rho_{0,1}^\delta)^\dagger\hat{\sigma}_\varepsilon^{-\frac{1}{2}}, \tag{7.6.234}$$

$$C \leftrightarrow \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\rho_{0,0}^\delta\hat{\sigma}_\varepsilon^{-\frac{1}{2}}, \tag{7.6.235}$$

$$\varepsilon \leftrightarrow \varepsilon^{\frac{1}{2}}. \tag{7.6.236}$$

Defining

$$L(\varepsilon) \coloneqq \begin{pmatrix}\varepsilon S(\rho^\delta,\hat{\sigma}_\varepsilon) & 0 \\ 0 & \rho_{1,1}^\delta + \varepsilon R\end{pmatrix}, \tag{7.6.237}$$

$$S(\rho^\delta,\hat{\sigma}_\varepsilon) \coloneqq \hat{\sigma}_\varepsilon^{-\frac{1}{2}}\left(\rho_{0,0}^\delta - \rho_{0,1}^\delta(\rho_{1,1}^\delta)^{-1}(\rho_{0,1}^\delta)^\dagger\right)\hat{\sigma}_\varepsilon^{-\frac{1}{2}}, \tag{7.6.238}$$

$$R \coloneqq \operatorname{Re}[(\rho_{1,1}^\delta)^{-1}(\rho_{0,1}^\delta)^\dagger(\hat{\sigma}_\varepsilon)^{-1}(\rho_{0,1}^\delta)], \tag{7.6.239}$$

we conclude from Lemma 7.50 that

$$\left\|K(\varepsilon) - e^{-i\sqrt{\varepsilon}G}L(\varepsilon)e^{i\sqrt{\varepsilon}G}\right\|_\infty \le o(\varepsilon), \tag{7.6.240}$$

where G in Lemma 7.50 is defined from A and B above. The inequality in (7.6.240) in turn implies the following operator inequalities:

$$e^{-i\sqrt{\varepsilon}G}L(\varepsilon)e^{i\sqrt{\varepsilon}G} - o(\varepsilon)\mathbb{1} \le K(\varepsilon) \le e^{-i\sqrt{\varepsilon}G}L(\varepsilon)e^{i\sqrt{\varepsilon}G} + o(\varepsilon)\mathbb{1}. \tag{7.6.241}$$

Observe that

$$e^{-i\sqrt{\varepsilon}G}L(\varepsilon)e^{i\sqrt{\varepsilon}G} + o(\varepsilon)\mathbb{1} = e^{-i\sqrt{\varepsilon}G}\left[L(\varepsilon) + o(\varepsilon)\mathbb{1}\right]e^{i\sqrt{\varepsilon}G}. \tag{7.6.242}$$

Now invoking these and the operator monotonicity of the function x^α for $\alpha \in (0,1)$, we find that

$$\operatorname{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] \tag{7.6.243}$$

$$= \operatorname{Tr}\left[\begin{pmatrix}\varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp\end{pmatrix}(K(\varepsilon))^\alpha\right] \tag{7.6.244}$$

$$\leq \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix} \left(e^{-i\sqrt{\varepsilon}G}\left[L(\varepsilon) + o(\varepsilon)\mathbb{1}\right] e^{i\sqrt{\varepsilon}G}\right)^\alpha\right] \tag{7.6.245}$$

$$= \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix} e^{-i\sqrt{\varepsilon}G}\left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha e^{i\sqrt{\varepsilon}G}\right]. \tag{7.6.246}$$

Consider that

$$(L(\varepsilon) + o(\varepsilon)\mathbb{1})^\alpha$$

$$= \begin{pmatrix} \varepsilon S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(\varepsilon)\mathbb{1} & 0 \\ 0 & \rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1} \end{pmatrix}^\alpha \tag{7.6.247}$$

$$= \begin{pmatrix} \left(\varepsilon S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha & 0 \\ 0 & \left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha \end{pmatrix} \tag{7.6.248}$$

$$= \begin{pmatrix} \varepsilon^\alpha\left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha & 0 \\ 0 & \left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha \end{pmatrix}. \tag{7.6.249}$$

Now expanding $e^{i\sqrt{\varepsilon}G}$ to first order in order to evaluate (7.6.246) (higher order terms will end up being irrelevant), we find that

$$\operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix} e^{-i\sqrt{\varepsilon}G}\left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha e^{i\sqrt{\varepsilon}G}\right]$$

$$= \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix}\left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha\right]$$
$$+ \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix}\left(-i\sqrt{\varepsilon}G\right)\left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha\right]$$
$$+ \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix}\left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha\left(i\sqrt{\varepsilon}G\right)\right] + o(1) \tag{7.6.250}$$

$$= \operatorname{Tr}\left[\begin{pmatrix} \hat{\sigma}_\varepsilon\left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp\left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha \end{pmatrix}\right]$$
$$- i\sqrt{\varepsilon}\operatorname{Tr}\left[\begin{pmatrix} \left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha\Pi_\sigma^\perp \end{pmatrix} G\right]$$
$$+ i\sqrt{\varepsilon}\operatorname{Tr}\left[\begin{pmatrix} \hat{\sigma}_\varepsilon\left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp\left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha \end{pmatrix} G\right] + o(1) \tag{7.6.251}$$

$$= \operatorname{Tr}\left[\begin{pmatrix} \hat{\sigma}_\varepsilon \left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha \end{pmatrix}\right] + o(1) \tag{7.6.252}$$

$$= \operatorname{Tr}\left[\hat{\sigma}_\varepsilon \left(S(\rho^\delta, \hat{\sigma}_\varepsilon) + o(1)\mathbb{1}\right)^\alpha\right] + \varepsilon^{1-\alpha} \operatorname{Tr}[\Pi_\sigma^\perp \left(\rho_{1,1}^\delta + \varepsilon R + o(\varepsilon)\mathbb{1}\right)^\alpha] + o(1). \tag{7.6.253}$$

By observing the last line, we see that higher order terms for $e^{i\sqrt{\varepsilon}G}$ include prefactors of ε (or higher powers), which vanish in the $\varepsilon \to 0^+$ limit. Now taking the limit $\varepsilon \to 0^+$, we find that

$$\lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix} e^{-i\sqrt{\varepsilon}G} \left(L(\varepsilon) + o(\varepsilon)\mathbb{1}\right)^\alpha e^{i\sqrt{\varepsilon}G}\right]$$
$$= \operatorname{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}} \left(\rho_{0,0}^\delta - \rho_{0,1}^\delta (\rho_{1,1}^\delta)^{-1} (\rho_{0,1}^\delta)^\dagger\right) \sigma^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.254}$$

where the inverses are taken on the support of σ. By proceeding in a similar way, but using the lower bound in (7.6.241), we find the following lower bound on (7.6.243):

$$\operatorname{Tr}\left[\begin{pmatrix} \varepsilon^{-\alpha}\hat{\sigma}_\varepsilon & 0 \\ 0 & \varepsilon^{1-\alpha}\Pi_\sigma^\perp \end{pmatrix} e^{-i\sqrt{\varepsilon}G} \left(L(\varepsilon) - o(\varepsilon)\mathbb{1}\right)^\alpha e^{i\sqrt{\varepsilon}G}\right]. \tag{7.6.255}$$

Then by the same argument above, the lower bound on (7.6.243) after taking the limit $\varepsilon \to 0^+$ is the same as in (7.6.254). So we conclude that

$$\lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon \left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_\delta \sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}} \left(\rho_{0,0}^\delta - \rho_{0,1}^\delta (\rho_{1,1}^\delta)^{-1} (\rho_{0,1}^\delta)^\dagger\right) \sigma^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.6.256}$$

Now consider that

$$\lim_{\delta\to 0^+} \rho_{0,0}^\delta - \rho_{0,1}^\delta (\rho_{1,1}^\delta)^{-1} (\rho_{0,1}^\delta)^\dagger = \rho_{0,0} - \rho_{0,1}\rho_{1,1}^{-1}\rho_{0,1}^\dagger, \tag{7.6.257}$$

where the inverse on the right is taken on the support of $\rho_{1,1}$. This follows because the image of $\rho_{0,1}^\dagger$ is contained in the support of $\rho_{1,1}$. Thus, we take the limit $\delta \to 0^+$, and find that

$$\lim_{\delta\to 0^+} \lim_{\varepsilon\to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon \left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_\delta \sigma_\varepsilon^{-\frac{1}{2}}\right)^\alpha\right] = \operatorname{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}} \left(\rho_{0,0} - \rho_{0,1}\rho_{1,1}^{-1}\rho_{0,1}^\dagger\right) \sigma^{-\frac{1}{2}}\right)^\alpha\right], \tag{7.6.258}$$

where all inverses are taken on the support. This concludes the proof.

Lemma 7.50

Let A be an invertible Hermitian operator, B a linear operator, C a Hermitian operator, and let $\varepsilon > 0$. Then with

$$M(\varepsilon) \coloneqq \begin{bmatrix} A & \varepsilon B \\ \varepsilon B^\dagger & \varepsilon^2 C \end{bmatrix}, \tag{7.6.259}$$

$$D(\varepsilon) \coloneqq \begin{bmatrix} A + \varepsilon^2 \operatorname{Re}[A^{-1}BB^\dagger] & 0 \\ 0 & \varepsilon^2 \left(C - B^\dagger A^{-1} B\right) \end{bmatrix}, \tag{7.6.260}$$

$$G \coloneqq \begin{bmatrix} 0 & -iA^{-1}B \\ iB^\dagger A^{-1} & 0 \end{bmatrix}, \tag{7.6.261}$$

the following inequality holds

$$\left\| M(\varepsilon) - e^{-i\varepsilon G} D(\varepsilon) e^{i\varepsilon G} \right\|_\infty \leq o(\varepsilon^2). \tag{7.6.262}$$

Proof. Observe that G is Hermitian and consider that

$$e^{i\varepsilon G} M(\varepsilon) e^{-i\varepsilon G} = \left(I + i\varepsilon G - \frac{\varepsilon^2}{2} G^2\right) M(\varepsilon) \left(I - i\varepsilon G - \frac{\varepsilon^2}{2} G^2\right) + o(\varepsilon^2).$$

Then we find that

$$\left(I + i\varepsilon G - \frac{\varepsilon^2}{2} G^2\right) M(\varepsilon) \left(I - i\varepsilon G - \frac{\varepsilon^2}{2} G^2\right) = M(\varepsilon) + i\varepsilon \left[G M(\varepsilon) - M(\varepsilon) G\right] \\ + \varepsilon^2 \left[G M(\varepsilon) G - \frac{1}{2} G^2 M(\varepsilon) - \frac{1}{2} M(\varepsilon) G^2\right] + o(\varepsilon^2). \tag{7.6.263}$$

Now observe that

$$G M(\varepsilon) = \begin{bmatrix} 0 & -iA^{-1}B \\ iB^\dagger A^{-1} & 0 \end{bmatrix} \begin{bmatrix} A & \varepsilon B \\ \varepsilon B^\dagger & \varepsilon^2 C \end{bmatrix} \tag{7.6.264}$$

$$= \begin{bmatrix} -i\varepsilon A^{-1} B B^\dagger & -i\varepsilon^2 A^{-1} B C \\ iB^\dagger & i\varepsilon B^\dagger A^{-1} B \end{bmatrix} \tag{7.6.265}$$

$$= \begin{bmatrix} -i\varepsilon A^{-1} B B^\dagger & o(\varepsilon) \\ iB^\dagger & i\varepsilon B^\dagger A^{-1} B \end{bmatrix}, \tag{7.6.266}$$

$$M(\varepsilon) G = [G M(\varepsilon)]^\dagger \tag{7.6.267}$$

$$= \begin{bmatrix} i\varepsilon B B^\dagger A^{-1} & -iB \\ o(\varepsilon) & -i\varepsilon B^\dagger A^{-1} B \end{bmatrix}, \tag{7.6.268}$$

which implies that

$$
\begin{aligned}
& i\varepsilon \left[GM(\varepsilon) - M(\varepsilon)G \right] \\
&= i\varepsilon \left(\begin{bmatrix} -i\varepsilon A^{-1}BB^{\dagger} & o(\varepsilon) \\ iB^{\dagger} & i\varepsilon B^{\dagger}A^{-1}B \end{bmatrix} - \begin{bmatrix} i\varepsilon BB^{\dagger}A^{-1} & -iB \\ o(\varepsilon) & -i\varepsilon B^{\dagger}A^{-1}B \end{bmatrix} \right) && (7.6.269) \\
&= \begin{bmatrix} 2\varepsilon^2 \operatorname{Re}[A^{-1}BB^{\dagger}] & -\varepsilon B + o(\varepsilon^2) \\ -\varepsilon B^{\dagger} + o(\varepsilon^2) & -2\varepsilon^2 B^{\dagger}A^{-1}B \end{bmatrix}. && (7.6.270)
\end{aligned}
$$

Also, observe that

$$
\begin{aligned}
GM(\varepsilon)G &= \begin{bmatrix} o(1) & o(\varepsilon) \\ iB^{\dagger} & o(1) \end{bmatrix} \begin{bmatrix} 0 & -iA^{-1}B \\ iB^{\dagger}A^{-1} & 0 \end{bmatrix} && (7.6.271) \\
&= \begin{bmatrix} o(\varepsilon) & o(1) \\ o(1) & B^{\dagger}A^{-1}B \end{bmatrix}, && (7.6.272) \\
G^2 M(\varepsilon) &= G[GM(\varepsilon)] && (7.6.273) \\
&= \begin{bmatrix} 0 & -iA^{-1}B \\ iB^{\dagger}A^{-1} & 0 \end{bmatrix} \begin{bmatrix} o(1) & o(\varepsilon) \\ iB^{\dagger} & o(1) \end{bmatrix} && (7.6.274) \\
&= \begin{bmatrix} A^{-1}BB^{\dagger} & o(1) \\ o(1) & o(\varepsilon) \end{bmatrix}, && (7.6.275) \\
M(\varepsilon)G^2 &= \left[G^2 M(\varepsilon) \right]^{\dagger} && (7.6.276) \\
&= \begin{bmatrix} BB^{\dagger}A^{-1} & o(1) \\ o(1) & o(\varepsilon) \end{bmatrix}. && (7.6.277)
\end{aligned}
$$

So then we find that

$$
\begin{aligned}
& \varepsilon^2 \left[GM(\varepsilon)G - \frac{1}{2}G^2 M(\varepsilon) - \frac{1}{2}M(\varepsilon)G^2 \right] \\
&= \varepsilon^2 \left(\begin{bmatrix} o(\varepsilon) & o(1) \\ o(1) & B^{\dagger}A^{-1}B \end{bmatrix} - \frac{1}{2} \begin{bmatrix} A^{-1}BB^{\dagger} & o(1) \\ o(1) & o(\varepsilon) \end{bmatrix} - \frac{1}{2} \begin{bmatrix} BB^{\dagger}A^{-1} & o(1) \\ o(1) & o(\varepsilon) \end{bmatrix} \right) && (7.6.278) \\
&= \begin{bmatrix} -\varepsilon^2 \operatorname{Re}[A^{-1}BB^{\dagger}] + o(\varepsilon^3) & o(\varepsilon^2) \\ o(\varepsilon^2) & \varepsilon^2 B^{\dagger}A^{-1}B + o(\varepsilon^3) \end{bmatrix}. && (7.6.279)
\end{aligned}
$$

So then

$$
\begin{aligned}
& \left(I + i\varepsilon G - \frac{\varepsilon^2}{2}G^2 \right) M(\varepsilon) \left(I - i\varepsilon G - \frac{\varepsilon^2}{2}G^2 \right) \\
&= M(\varepsilon) + i\varepsilon \left[GM(\varepsilon) - M(\varepsilon)G \right]
\end{aligned}
$$

$$+\varepsilon^2\left[GM(\varepsilon)G-\frac{1}{2}G^2M(\varepsilon)-\frac{1}{2}M(\varepsilon)G^2\right]+o(\varepsilon^2) \tag{7.6.280}$$

$$=\begin{bmatrix}A & \varepsilon B\\ \varepsilon B^\dagger & \varepsilon^2 C\end{bmatrix}+\begin{bmatrix}2\varepsilon^2\operatorname{Re}[A^{-1}BB^\dagger] & -\varepsilon B+o(\varepsilon^2)\\ -\varepsilon B^\dagger+o(\varepsilon^2) & -2\varepsilon^2B^\dagger A^{-1}B\end{bmatrix}$$

$$+\begin{bmatrix}-\varepsilon^2\operatorname{Re}[A^{-1}BB^\dagger]+o(\varepsilon^3) & o(\varepsilon^2)\\ o(\varepsilon^2) & \varepsilon^2B^\dagger A^{-1}B+o(\varepsilon^3)\end{bmatrix}+o(\varepsilon^2) \tag{7.6.281}$$

$$=\begin{bmatrix}A+\varepsilon^2\operatorname{Re}[A^{-1}BB^\dagger] & 0\\ 0 & \varepsilon^2\left(C-B^\dagger A^{-1}B\right)\end{bmatrix}+o(\varepsilon^2) \tag{7.6.282}$$

$$=D(\varepsilon)+o(\varepsilon^2). \tag{7.6.283}$$

So we conclude that

$$e^{i\varepsilon G}M(\varepsilon)e^{-i\varepsilon G}=D(\varepsilon)+o(\varepsilon^2), \tag{7.6.284}$$

which in turn implies that

$$M(\varepsilon)=e^{-i\varepsilon G}D(\varepsilon)e^{i\varepsilon G}+o(\varepsilon^2), \tag{7.6.285}$$

from which we conclude the claim in (7.6.262). ■

7.7 Belavkin–Staszewski Relative Entropy

A different quantum generalization of the classical relative entropy is given by the Belavkin–Staszewski[3] relative entropy:

Definition 7.51 Belavkin–Staszewski Relative Entropy

The Belavkin–Staszewski relative entropy of a quantum state ρ and a positive semi-definite operator σ is defined as

$$\widehat{D}(\rho\|\sigma):=\begin{cases}\operatorname{Tr}\left[\rho\log_2\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right] & \text{if } \operatorname{supp}(\rho)\subseteq\operatorname{supp}(\sigma)\\ +\infty & \text{otherwise}\end{cases}, \tag{7.7.1}$$

where the inverse σ^{-1} is taken on the support of σ and the logarithm is evaluated on the support of ρ.

[3] The name Staszewski is pronounced Stah·shev·ski, with emphasis on the second syllable.

This quantum generalization of classical relative entropy is not known to possess an information-theoretic meaning. However, it is quite useful for obtaining upper bounds on quantum channel capacities and quantum channel discrimination rates, as considered in Part III of this book.

An important property of the Belavkin–Staszewski relative entropy is that it is the limit of the geometric Rényi relative entropy as $\alpha \to 1$.

Proposition 7.52

Let ρ be a state and σ a positive semi-definite operator. Then, in the limit $\alpha \to 1$, the geometric Rényi relative entropy converges to the Belavkin–Staszewski relative entropy:

$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma). \tag{7.7.2}$$

Proof: Suppose at first that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. Then $\widehat{D}_\alpha(\rho\|\sigma)$ is finite for all $\alpha \in (0,1) \cup (1,\infty)$, and we can write the following explicit formula for the geometric Rényi relative entropy by employing Proposition 7.40:

$$\widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log_2 \widehat{Q}_\alpha(\rho\|\sigma) \tag{7.7.3}$$

$$= \frac{1}{\alpha-1} \log_2 \text{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha\right]. \tag{7.7.4}$$

Our assumption implies that $\text{Tr}[\Pi_\sigma \rho] = 1$, and we find that

$$\widehat{Q}_1(\rho\|\sigma) = \text{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)\right] \tag{7.7.5}$$

$$= \text{Tr}[\Pi_\sigma \rho] \tag{7.7.6}$$

$$= 1. \tag{7.7.7}$$

Since $\log_2 1 = 0$, we can write

$$\widehat{D}_\alpha(\rho\|\sigma) = \frac{\log_2 \widehat{Q}_\alpha(\rho\|\sigma) - \log_2 \widehat{Q}_1(\rho\|\sigma)}{\alpha - 1}, \tag{7.7.8}$$

so that

$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1} \frac{\log_2 \widehat{Q}_\alpha(\rho\|\sigma) - \log_2 \widehat{Q}_1(\rho\|\sigma)}{\alpha - 1} \tag{7.7.9}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\alpha} \log_2 \widehat{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} \tag{7.7.10}$$

$$= \frac{1}{\ln(2)} \frac{\frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{Q}_\alpha(\rho\|\sigma)\Big|_{\alpha=1}}{\widehat{Q}_1(\rho\|\sigma)} \tag{7.7.11}$$

$$= \frac{1}{\ln(2)} \frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1}. \tag{7.7.12}$$

Then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}\widehat{Q}_\alpha(\rho\|\sigma)\bigg|_{\alpha=1} &= \frac{\mathrm{d}}{\mathrm{d}\alpha}\operatorname{Tr}\!\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]\bigg|_{\alpha=1} \\ &= \operatorname{Tr}\!\left[\sigma\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]\bigg|_{\alpha=1}.\end{aligned}$$

For a positive semi-definite operator X with spectral decomposition

$$X = \sum_z \nu_z \Pi_z, \tag{7.7.13}$$

it follows that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}X^\alpha\bigg|_{\alpha=1} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\sum_z \nu_z^\alpha \Pi_z\bigg|_{\alpha=1} \tag{7.7.14}$$

$$= \sum_z \left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\nu_z^\alpha\bigg|_{\alpha=1}\right)\Pi_z \tag{7.7.15}$$

$$= \sum_z \left(\nu_z^\alpha \ln \nu_z^\alpha\Big|_{\alpha=1}\right)\Pi_z \tag{7.7.16}$$

$$= \sum_z (\nu_z \ln \nu_z)\,\Pi_z \tag{7.7.17}$$

$$= X \ln_* X, \tag{7.7.18}$$

where

$$\ln_*(x) := \begin{cases} \ln(x) & x > 0 \\ 0 & x = 0 \end{cases}. \tag{7.7.19}$$

Thus we find that

$$\operatorname{Tr}\!\left[\sigma\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]\bigg|_{\alpha=1}$$

$$= \mathrm{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)\ln_*\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)\right] \tag{7.7.20}$$
$$= \mathrm{Tr}\left[\sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\ln_*\left(\sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}}\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\right)\right] \tag{7.7.21}$$
$$= \mathrm{Tr}\left[\sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}\ln_*\left(\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}}\right)\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\right] \tag{7.7.22}$$
$$= \mathrm{Tr}\left[\rho^{\frac{1}{2}}\Pi_\sigma\rho^{\frac{1}{2}}\ln_*\left(\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}}\right)\right] \tag{7.7.23}$$
$$= \mathrm{Tr}\left[\rho\ln\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right]. \tag{7.7.24}$$

The third equality follows from Lemma 2.5. The final equality follows from the assumption $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ and by applying the interpretation of the logarithm exactly as stated in Definition 7.51. Then we find that

$$\lim_{\alpha\to 1}\widehat{D}_\alpha(\rho\|\sigma) = \mathrm{Tr}\left[\rho\log_2\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right], \tag{7.7.25}$$

for the case in which $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$.

Now suppose that $\alpha \in (1,\infty)$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$. Then $\widehat{D}_\alpha(\rho\|\sigma) = +\infty$, so that $\lim_{\alpha\to 1^+}\widehat{D}_\alpha(\rho\|\sigma) = +\infty$, consistent with the definition of the Belavkin–Staszewski relative entropy in this case (see Definition 7.51).

Suppose that $\alpha \in (0,1)$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$. Employing Proposition 7.42, we have that $\widehat{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(\rho\|\sigma)$ for all $\alpha \in (0,1)$. Since $\lim_{\alpha\to 1^-}\widetilde{D}_\alpha(\rho\|\sigma) = +\infty$ in this case, it follows that $\lim_{\alpha\to 1^-}\widehat{D}_\alpha(\rho\|\sigma) = +\infty$.

Therefore,

$$\lim_{\alpha\to 1^-}\widehat{D}_\alpha(\rho\|\sigma)$$
$$= \begin{cases} \mathrm{Tr}\left[\rho\log_2\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right] & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ +\infty & \text{otherwise} \end{cases} \tag{7.7.26}$$
$$= \widehat{D}(\rho\|\sigma). \tag{7.7.27}$$

To conclude, we have established that $\lim_{\alpha\to 1^+}\widehat{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1^-}\widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma)$, which means that

$$\lim_{\alpha\to 1}\widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma), \tag{7.7.28}$$

as required. ■

The following inequality relates the quantum relative entropy to the Belavkin–Staszewski relative entropy:

Proposition 7.53

Let ρ be a state and σ a positive semi-definite operator. Then the quantum relative entropy is never larger than the Belavkin–Staszewski relative entropy:

$$D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma). \tag{7.7.29}$$

PROOF: If $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, then there is nothing to prove in this case because both

$$D(\rho\|\sigma) = \widehat{D}(\rho\|\sigma) = +\infty, \tag{7.7.30}$$

and so the inequality in (7.7.29) holds trivially in this case. So let us suppose instead that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. From Propositions 7.42 and 7.40, we conclude for all $\alpha \in (0, 1) \cup (1, \infty)$ that

$$\widetilde{D}_\alpha(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma). \tag{7.7.31}$$

From Proposition 7.30, we know that

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma). \tag{7.7.32}$$

While from Proposition 7.52, we know that

$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma). \tag{7.7.33}$$

Thus, applying the limit $\alpha \to 1$ to (7.7.31) and the two equalities above, we conclude (7.7.29). ■

Similar to what was shown in Proposition 7.2, Definition 7.51 is consistent with the following limit:

Proposition 7.54

For every state ρ and positive semi-definite operator σ, the following limit holds

$$\widehat{D}(\rho\|\sigma) = \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} \mathrm{Tr}\left[\rho_\delta \log_2\left(\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-1}\rho_\delta^{\frac{1}{2}}\right)\right], \tag{7.7.34}$$

where $\delta \in (0, 1)$ and

$$\rho_\delta \coloneqq (1-\delta)\,\rho + \delta\pi, \qquad \sigma_\varepsilon \coloneqq \sigma + \varepsilon\mathbb{1}, \tag{7.7.35}$$

with π the maximally mixed state.

Proof: Suppose first that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. We follow an approach similar to that given in the proof of Proposition 7.40. Let us employ the decomposition of the Hilbert space into $\mathrm{supp}(\sigma) \oplus \ker(\sigma)$. Then we can write ρ and σ as in (7.6.199), so that

$$\sigma_\varepsilon^{-1} = \begin{pmatrix} (\sigma + \varepsilon\Pi_\sigma)^{-1} & 0 \\ 0 & \varepsilon^{-1}\Pi_\sigma^\perp \end{pmatrix}, \tag{7.7.36}$$

where we have followed the developments in (7.6.199)–(7.6.201). The condition $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ implies that $\rho_{0,1} = 0$ and $\rho_{1,1} = 0$. It thus follows that $\lim_{\delta\to 0^+}\rho_\delta = \rho_{0,0}$. We then find that

$$\mathrm{Tr}\left[\rho_\delta \log_2\left(\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-1}\rho_\delta^{\frac{1}{2}}\right)\right] = \mathrm{Tr}\left[\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{\frac{1}{2}}\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta^{\frac{1}{2}} \log_2\left(\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-\frac{1}{2}}\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta^{\frac{1}{2}}\right)\right] \tag{7.7.37}$$

$$= \mathrm{Tr}\left[\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{\frac{1}{2}} \log_2\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta^{\frac{1}{2}}\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-\frac{1}{2}}\right)\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta^{\frac{1}{2}}\right] \tag{7.7.38}$$

$$= \mathrm{Tr}\left[\log_2\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)\sigma_\varepsilon\right] \tag{7.7.39}$$

$$= \mathrm{Tr}\left[\sigma_\varepsilon\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)\log_2\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)\right] \tag{7.7.40}$$

$$= \mathrm{Tr}\left[\sigma_\varepsilon\eta\left(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}\right)\right], \tag{7.7.41}$$

where the second equality follows from applying Lemma 2.5 with $f = \log_2$ and $L = \rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-\frac{1}{2}}$. The second-to-last equality follows because $\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}}$ commutes with $\log_2(\sigma_\varepsilon^{-\frac{1}{2}}\rho_\delta\sigma_\varepsilon^{-\frac{1}{2}})$, and by employing cyclicity of trace. In the last line, we made use of the following function:

$$\eta(x) \coloneqq x\log_2 x, \tag{7.7.42}$$

defined for all $x \in [0, \infty)$ with $\eta(0) = 0$. By appealing to the continuity of the function $\eta(x)$ on $x \in [0, \infty)$ and the fact that $\lim_{\delta \to 0^+} \rho_\delta = \rho_{0,0}$, we find that

$$\lim_{\delta \to 0^+} \operatorname{Tr}\left[\sigma_\varepsilon \eta\left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_\delta \sigma_\varepsilon^{-\frac{1}{2}}\right)\right] = \operatorname{Tr}\left[\sigma_\varepsilon \eta\left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0} \sigma_\varepsilon^{-\frac{1}{2}}\right)\right]. \tag{7.7.43}$$

Now recall the function $\log_{2,*}$ defined in (7.7.19). Using it, we can write

$$\begin{aligned}
&\operatorname{Tr}\left[\sigma_\varepsilon \eta\left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0} \sigma_\varepsilon^{-\frac{1}{2}}\right)\right] \\
&= \operatorname{Tr}\left[\sigma_\varepsilon \sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0} \sigma_\varepsilon^{-\frac{1}{2}} \log_{2,*}\left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0} \sigma_\varepsilon^{-\frac{1}{2}}\right)\right] && (7.7.44)\\
&= \operatorname{Tr}\left[\sigma_\varepsilon^{\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}} \sigma_\varepsilon^{-\frac{1}{2}} \log_{2,*}\left(\sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}} \sigma_\varepsilon^{-\frac{1}{2}}\right)\right] && (7.7.45)\\
&= \operatorname{Tr}\left[\sigma_\varepsilon^{\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}} \log_{2,*}\left(\rho_{0,0}^{\frac{1}{2}} \sigma_\varepsilon^{-\frac{1}{2}} \sigma_\varepsilon^{-\frac{1}{2}} \rho_{0,0}^{\frac{1}{2}}\right) \rho_{0,0}^{\frac{1}{2}} \sigma_\varepsilon^{-\frac{1}{2}}\right] && (7.7.46)\\
&= \operatorname{Tr}\left[\rho_{0,0} \log_{2,*}\left(\rho_{0,0}^{\frac{1}{2}} \sigma_\varepsilon^{-1} \rho_{0,0}^{\frac{1}{2}}\right)\right] && (7.7.47)\\
&= \operatorname{Tr}\left[\rho_{0,0} \log_{2,*}\left(\rho_{0,0}^{\frac{1}{2}} (\sigma + \varepsilon \Pi_\sigma)^{-1} \rho_{0,0}^{\frac{1}{2}}\right)\right], && (7.7.48)
\end{aligned}$$

where the last line follows because

$$\begin{aligned}
&\rho_{0,0}^{\frac{1}{2}} (\sigma + \varepsilon \Pi_\sigma)^{-1} \rho_{0,0}^{\frac{1}{2}} \\
&= \begin{pmatrix} \rho_{0,0}^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} (\sigma + \varepsilon \Pi_\sigma)^{-1} & 0 \\ 0 & \varepsilon^{-1} \Pi_\sigma^\perp \end{pmatrix} \begin{pmatrix} \rho_{0,0}^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix} && (7.7.49)\\
&= \begin{pmatrix} \rho_{0,0}^{\frac{1}{2}} (\sigma + \varepsilon \Pi_\sigma)^{-1} \rho_{0,0}^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix}. && (7.7.50)
\end{aligned}$$

Now taking the limit as $\varepsilon \to 0^+$, and appealing to continuity of $\log_{2,*}(x)$ and x^{-1} for $x > 0$, we find that

$$\begin{aligned}
&\lim_{\varepsilon \to 0^+} \operatorname{Tr}\left[\rho_{0,0} \log_{2,*}\left(\rho_{0,0}^{\frac{1}{2}} (\sigma + \varepsilon \Pi_\sigma)^{-1} \rho_{0,0}^{\frac{1}{2}}\right)\right] \\
&= \operatorname{Tr}\left[\rho_{0,0} \log_{2,*}\left(\rho_{0,0}^{\frac{1}{2}} \sigma^{-1} \rho_{0,0}^{\frac{1}{2}}\right)\right] && (7.7.51)
\end{aligned}$$

$$= \text{Tr}\left[\rho \log_2\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right] \tag{7.7.52}$$

where the formula in the last line is interpreted exactly as stated in Definition 7.51. Thus, we conclude that

$$\lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+}\text{Tr}\left[\rho_\delta \log_2\left(\rho_\delta^{\frac{1}{2}}\sigma_\varepsilon^{-1}\rho_\delta^{\frac{1}{2}}\right)\right] = \text{Tr}\left[\rho \log_2\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right]. \tag{7.7.53}$$

Now suppose that $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$. Then applying Proposition 7.53, we find that the following inequality holds for all $\delta \in (0, 1)$ and $\varepsilon > 0$:

$$\widehat{D}(\rho_\delta\|\sigma_\varepsilon) \geq D(\rho_\delta\|\sigma_\varepsilon). \tag{7.7.54}$$

Now taking limits and applying Proposition 7.2, we find that

$$\lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+}\widehat{D}(\rho_\delta\|\sigma_\varepsilon) \geq \lim_{\varepsilon\to 0^+}\lim_{\delta\to 0^+} D(\rho_\delta\|\sigma_\varepsilon) \tag{7.7.55}$$

$$= \lim_{\varepsilon\to 0^+} D(\rho\|\sigma_\varepsilon) \tag{7.7.56}$$

$$= +\infty. \tag{7.7.57}$$

This concludes the proof. ■

By taking the limit $\alpha \to 1$ in the statement of the data-processing inequality for $\widehat{D}_\alpha$, and applying Proposition 7.52, we immediately obtain the data-processing inequality for the Belavkin–Staszewski relative entropy.

Corollary 7.55 Data-Processing Inequality for Belavkin–Staszewski Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then

$$\widehat{D}(\rho\|\sigma) \geq \widehat{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.7.58}$$

Some basic properties of the Belavkin–Staszewski relative entropy are as follows:

Proposition 7.56 Properties of Belavkin–Staszewski Relative Entropy

The Belavkin–Staszewski relative entropy satisfies the following properties for states ρ, ρ_1, ρ_2 and positive semi-definite operators $\sigma, \sigma_1, \sigma_2$.

1. *Isometric invariance*: For every isometry V,

$$\widehat{D}(V\rho V^\dagger \| V\sigma V^\dagger) = \widehat{D}(\rho\|\sigma). \tag{7.7.59}$$

2. (a) If $\mathrm{Tr}[\sigma] \leq 1$, then $\widehat{D}(\rho\|\sigma) \geq 0$.
   (b) *Faithfulness*: Suppose that $\mathrm{Tr}[\sigma] \leq \mathrm{Tr}[\rho] = 1$. Then $\widehat{D}(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.
   (c) If $\rho \leq \sigma$, then $\widehat{D}(\rho\|\sigma) \leq 0$.
   (d) If $\sigma \leq \sigma'$, then $\widehat{D}(\rho\|\sigma) \geq \widehat{D}(\rho\|\sigma')$.

3. *Additivity*:

$$\widehat{D}(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) = \widehat{D}(\rho_1\|\sigma_1) + D(\rho_2\|\sigma_2). \tag{7.7.60}$$

   As a special case, for every $\beta \in (0, \infty)$,

$$\widehat{D}(\rho\|\beta\sigma) = \widehat{D}(\rho\|\sigma) + \log_2\left(\frac{1}{\beta}\right). \tag{7.7.61}$$

4. *Direct-sum property*: Let $p : \mathcal{X} \to [0, 1]$ be a probability distribution over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, and let $q : \mathcal{X} \to [0, \infty)$ be a positive function on $\mathcal{X}$. Let $\{\rho_A^x : x \in \mathcal{X}\}$ be a set of states on a system A, and let $\{\sigma_A^x : x \in \mathcal{X}\}$ be a set of positive semi-definite operators on A. Then,

$$\widehat{D}(\rho_{XA}\|\sigma_{XA}) = \widehat{D}(p\|q) + \sum_{x\in\mathcal{X}} p(x)\widehat{D}(\rho_A^x\|\sigma_A^x). \tag{7.7.62}$$

   where

$$\rho_{XA} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \tag{7.7.63}$$

$$\sigma_{XA} := \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x. \tag{7.7.64}$$

Proof:

1. Isometric invariance is a direct consequence of Propositions 7.44 and 7.52.

2. All of the properties in the second item follow from data processing (Corollary 7.55).

   (a) Applying the trace-out channel, we find that

$$\begin{aligned}\widehat{D}(\rho\|\sigma) &\geq \widehat{D}(\mathrm{Tr}[\rho]\|\,\mathrm{Tr}[\sigma]) &(7.7.65)\\ &= \mathrm{Tr}[\rho]\log_2(\mathrm{Tr}[\rho]/\mathrm{Tr}[\sigma]) &(7.7.66)\\ &= -\log_2\mathrm{Tr}[\sigma] &(7.7.67)\\ &\geq 0. &(7.7.68)\end{aligned}$$

   (b) If $\rho = \sigma$, then it follows by direct evalution that $\widehat{D}(\rho\|\sigma) = 0$. If $\widehat{D}(\rho\|\sigma) = 0$ and $\mathrm{Tr}[\sigma] \leq 1$, then $D(\rho\|\sigma) = 0$ by Proposition 7.53 and we conclude that $\rho = \sigma$ from faithfulness of the quantum relative entropy (Proposition 7.3).

   (c) If $\rho \leq \sigma$, then $\sigma - \rho$ is positive semi-definite, and the following operator is positive semi-definite:

$$\hat{\sigma} := |0\rangle\langle 0| \otimes \rho + |1\rangle\langle 1| \otimes (\sigma - \rho)\,. \qquad (7.7.69)$$

   Defining $\hat{\rho} := |0\rangle\langle 0| \otimes \rho$, we find from the direct-sum property that

$$0 = \widehat{D}(\rho\|\rho) = \widehat{D}(\hat{\rho}\|\hat{\sigma}) \geq \widehat{D}(\rho\|\sigma), \qquad (7.7.70)$$

   where the inequality follows from data processing by tracing out the first classical register of $\hat{\rho}$ and $\hat{\sigma}$.

   (d) If $\sigma \leq \sigma'$, then the operator $\sigma' - \sigma$ is positive semi-definite and so is the following one:

$$\hat{\sigma} := |0\rangle\langle 0| \otimes \sigma + |1\rangle\langle 1| \otimes (\sigma' - \sigma)\,. \qquad (7.7.71)$$

   Defining $\hat{\rho} := |0\rangle\langle 0| \otimes \rho$, we find from the direct-sum property that

$$\widehat{D}(\rho\|\sigma) = \widehat{D}(\hat{\rho}\|\hat{\sigma}) \geq \widehat{D}(\rho\|\sigma'), \qquad (7.7.72)$$

   where the inequality follows from data processing by tracing out the first classical register of $\hat{\rho}$ and $\hat{\sigma}$.

3. Additivity follows by direct evaluation.

4. The direct-sum property follows by direct evaluation. ■

A statement similar to that made by Proposition 7.48 holds for the Belavkin–Staszewski relative entropy:

Proposition 7.57 Belavkin–Staszewski Relative Entropy from Classical Preparations

Let ρ be a state and σ a positive semi-definite operator satisfying $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. The Belavkin–Staszewski relative entropy is equal to the smallest value that the classical relative entropy can take by minimizing over classical–quantum channels that realize the state ρ and the positive semi-definite operator σ. That is, the following equality holds

$$\widehat{D}(\rho\|\sigma) = \inf_{\{p,q,\mathcal{P}\}} \{D(p\|q) : \mathcal{P}(\omega(p)) = \rho,\ \mathcal{P}(\omega(q)) = \sigma\}, \tag{7.7.73}$$

where the classical relative entropy is defined in (7.2.2), the channel $\mathcal{P}$ is a classical–quantum channel, $p : \mathcal{X} \to [0,1]$ is a probability distribution over a finite alphabet $\mathcal{X}$, $q : \mathcal{X} \to [0,\infty)$ is a positive function on $\mathcal{X}$, $\omega(p) := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|$, and $\omega(q) := \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|$.

Proof: The proof is very similar to the proof of Proposition 7.48, and so we use the same notation to provide a brief proof. By following the same reasoning that leads to (7.6.170), it follows that

$$\inf_{\{p,q,\mathcal{P}\}} \{D(p\|q) : \mathcal{P}(p) = \rho, \mathcal{P}(q) = \sigma\} \geq \widehat{D}(\rho\|\sigma). \tag{7.7.74}$$

The optimal choices of p, q, and $\mathcal{P}$ saturating the inequality in (7.7.74) are again given by (7.6.171)–(7.6.173). Consider for those choices that

$$\sum_x p(x)\log_2\!\left(\frac{p(x)}{q(x)}\right) = \sum_x p(x)\log_2(\lambda_x) \tag{7.7.75}$$

$$= \sum_x \lambda_x q(x)\log_2(\lambda_x) \tag{7.7.76}$$

$$= \sum_x \lambda_x \operatorname{Tr}[\Pi_x\sigma]\log_2(\lambda_x) \tag{7.7.77}$$

$$= \mathrm{Tr}\left[\sigma \left(\sum_x \lambda_x \log_2(\lambda_x)\, \Pi_x\right)\right] \tag{7.7.78}$$

$$= \mathrm{Tr}\left[\sigma \left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right) \log_2\!\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)\right] \tag{7.7.79}$$

$$= \mathrm{Tr}\left[\rho \log_2\!\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right], \tag{7.7.80}$$

where the last equality follows from reasoning similar to that used to justify (7.7.20)–(7.7.24). Then by following the reasoning at the end of the proof of Proposition 7.48, we conclude (7.7.73). ■

7.8 Max-Relative Entropy

An important generalized divergence that appears in the context of placing upper bounds on communication rates of feedback-assisted protocols is the max-relative entropy.

Definition 7.58 Max-Relative Entropy

The *max-relative entropy* $D_{\max}(\rho\|\sigma)$ of a state ρ and a positive semi-definite operator σ is defined as

$$D_{\max}(\rho\|\sigma) := \log_2\left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\infty, \tag{7.8.1}$$

if $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$; otherwise, $D_{\max}(\rho\|\sigma) = +\infty$.

The max-relative entropy has the following equivalent representations:

$$D_{\max}(\rho\|\sigma) = \log_2\left\|\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right\|_\infty \tag{7.8.2}$$

$$= 2\log_2\left\|\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\right\|_\infty \tag{7.8.3}$$

$$= \log_2 \inf_{\lambda\geq 0}\{\lambda : \rho \leq \lambda\sigma\} \tag{7.8.4}$$

$$= \inf_{\lambda\in\mathbb{R}}\{\lambda : \rho \leq 2^\lambda\sigma\} \tag{7.8.5}$$

$$= \log_2 \sup_{M\geq 0}\{\mathrm{Tr}[M\rho] : \mathrm{Tr}[M\sigma] \leq 1\}. \tag{7.8.6}$$

The second-to-last equality demonstrates that $D_{\max}(\rho\|\sigma)$ can be calculated using a semi-definite program (SDP) (see Section 2.4). Indeed, the optimization in (7.8.4) can be cast in the standard form in (2.4.4), i.e.,

$$\inf\{\lambda : \rho \leq \lambda\sigma\} = \begin{cases} \text{infimum} & \mathrm{Tr}[BY] \\ \text{subject to} & \Phi^\dagger(Y) \geq A, \\ & Y \geq 0, \end{cases} \tag{7.8.7}$$

with $Y \equiv \lambda$, $A \equiv \rho$, $B \equiv 1$, and $\Phi^\dagger(Y) = Y\sigma$. (Note that taking the trace on both sides of the constraint $\rho \leq \lambda\sigma$ in (7.8.4) results in $\lambda \geq 1/\mathrm{Tr}[\sigma]$, so that $\lambda \geq 0$.) The final equality in (7.8.6) results from calculating the SDP dual to that in (7.8.4).

The max-relative entropy has the following alternative representation, which, when ρ and σ are states, allows for thinking of it as being related to the largest weight that one can place on ρ to realize σ as a probabilistic mixture of ρ and some other state.

Lemma 7.59

The max-relative entropy $D_{\max}(\rho\|\sigma)$ of a state ρ and a positive semi-definite operator σ can be written as follows:

$$D_{\max}(\rho\|\sigma) = \inf_{\lambda\in\mathbb{R},\omega\geq 0} \left\{\lambda : \sigma = 2^{-\lambda}\rho + \left(1 - 2^{-\lambda}\right)\omega,\ \mathrm{Tr}[\omega] = 1\right\}. \tag{7.8.8}$$

Proof: Let $\mu \in \mathbb{R}$ be such that $\rho \leq 2^\mu\sigma$. Then it follows that $2^\mu\sigma - \rho \geq 0$, so that $\omega := \frac{2^\mu\sigma-\rho}{2^\mu-1}$ is a quantum state. Now consider that

$$2^{-\mu}\rho + (1 - 2^{-\mu})\,\omega = 2^{-\mu}\rho + (1 - 2^{-\mu})\,\frac{2^\mu\sigma - \rho}{2^\mu - 1} \tag{7.8.9}$$

$$= 2^{-\mu}\rho + (1 - 2^{-\mu})\,\frac{2^\mu\left(\sigma - 2^{-\mu}\rho\right)}{2^\mu - 1} \tag{7.8.10}$$

$$= 2^{-\mu}\rho + \sigma - 2^{-\mu}\rho \tag{7.8.11}$$

$$= \sigma. \tag{7.8.12}$$

Thus, μ and ω satisfy the constraints for the optimization problem in (7.8.8), and we conclude that

$$\mu \geq \inf_{\lambda\in\mathbb{R},\omega\geq 0} \left\{\lambda : \sigma = 2^{-\lambda}\rho + \left(1 - 2^{-\lambda}\right)\omega,\ \mathrm{Tr}[\omega] = 1\right\}. \tag{7.8.13}$$

By taking an infimum over all μ satisfying $\rho \leq 2^{\mu}\sigma$ and applying (7.8.5), we conclude that

$$D_{\max}(\rho\|\sigma) \geq \inf_{\lambda\in\mathbb{R},\omega\geq 0}\left\{\lambda : \sigma = 2^{-\lambda}\rho + \left(1-2^{-\lambda}\right)\omega,\ \operatorname{Tr}[\omega]=1\right\}. \tag{7.8.14}$$

Now we prove the opposite inequality. Let $\lambda \in \mathbb{R}$ and let ω be an arbitrary state satisfying $\sigma = 2^{-\lambda}\rho + \left(1-2^{-\lambda}\right)\omega$. Then it follows that

$$\sigma = 2^{-\lambda}\rho + \left(1-2^{-\lambda}\right)\omega \geq 2^{-\lambda}\rho, \tag{7.8.15}$$

from which we conclude that $\rho \leq 2^{\lambda}\sigma$. So it follows that $\lambda \geq D_{\max}(\rho\|\sigma)$. Since ω and λ are arbitrary, we conclude that

$$\inf_{\lambda\in\mathbb{R},\omega\geq 0}\left\{\lambda : \sigma = 2^{-\lambda}\rho + \left(1-2^{-\lambda}\right)\omega,\ \operatorname{Tr}[\omega]=1\right\} \geq D_{\max}(\rho\|\sigma), \tag{7.8.16}$$

which completes the proof. ■

Proposition 7.60 Data-Processing Inequality for Max-Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then,

$$D_{\max}(\rho\|\sigma) \geq D_{\max}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.8.17}$$

Remark: This result holds more generally for positive maps that are not necessarily trace preserving.

Proof: To see this, let $\lambda \in \mathbb{R}$ be such that the operator inequality $\rho \leq 2^{\lambda}\sigma$ holds. Then the operator inequality $\mathcal{N}(\rho) \leq 2^{\lambda}\mathcal{N}(\sigma)$ holds because the quantum channel $\mathcal{N}$ is a positive map. Then

$$D_{\max}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) = \inf_{\mu\in\mathbb{R}}\{\mu : \mathcal{N}(\rho) \leq 2^{\mu}\mathcal{N}(\sigma)\} \leq \lambda. \tag{7.8.18}$$

Since the inequality holds for all choices of λ such that $\rho \leq 2^{\lambda}\sigma$ holds, we conclude that

$$D_{\max}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq \inf_{\lambda\in\mathbb{R}}\{\lambda : \rho \leq 2^{\lambda}\sigma\} = D_{\max}(\rho\|\sigma). \tag{7.8.19}$$

This concludes the proof. ■

It turns out that the max-relative entropy is a limiting case of the sandwiched and geometric Rényi relative entropies, as we now show.

Proposition 7.61

The sandwiched and geometric Rényi relative entropies converge to the max-relative entropy in the limit $\alpha \to \infty$:

$$\lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to\infty} \widehat{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma). \tag{7.8.20}$$

Proof: We begin with the case in which $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$. We trivially have $\widetilde{D}_\alpha(\rho\|\sigma) = \widehat{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma) = +\infty$ for all $\alpha > 1$, which implies the equality in (7.8.20) in this case.

In the case that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, we can consider, without loss of generality, that $\operatorname{supp}(\sigma) = \mathcal{H}$, so that $\lambda_{\min}(\sigma) > 0$.

We begin with the sandwiched Rényi relative entropy. Consider that

$$\begin{aligned}\widetilde{D}_\alpha(\rho\|\sigma) &= \frac{1}{\alpha-1}\log_2 \operatorname{Tr}[(\rho^{\frac{1}{2}}\sigma^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{2}})^\alpha] &(7.8.21)\\ &= \frac{1}{\alpha-1}\log_2 \operatorname{Tr}[(\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\sigma^{\frac{1}{\alpha}}\sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}})^\alpha]. &(7.8.22)\end{aligned}$$

By the operator inequalities $[\lambda_{\min}(\sigma)]^{\frac{1}{\alpha}}\mathbb{1} \leq \sigma^{\frac{1}{\alpha}} \leq [\lambda_{\max}(\sigma)]^{\frac{1}{\alpha}}\mathbb{1}$ and the monotonicity $\operatorname{Tr}[X^\alpha] \leq \operatorname{Tr}[Y^\alpha]$ for positive semi-definite X and Y satisfying $X \leq Y$ (see Lemma 2.14), we find for $\alpha > 1$ that

$$\begin{aligned}\lambda_{\min}(\sigma)\operatorname{Tr}[(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}})^\alpha] &\leq \operatorname{Tr}[(\rho^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\sigma^{\frac{1}{\alpha}}\sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}})^\alpha] &(7.8.23)\\ &\leq \lambda_{\max}(\sigma)\operatorname{Tr}[(\rho^{1/2}\sigma^{-1}\rho^{\frac{1}{2}})^\alpha]. &(7.8.24)\end{aligned}$$

Using the fact that

$$\operatorname{Tr}[(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}})^\alpha] = \left\|\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right\|_\alpha^\alpha, \tag{7.8.25}$$

and taking a logarithm followed by multiplication of $\frac{1}{\alpha-1}$, we find that

$$\begin{aligned}\frac{1}{\alpha-1}\log_2\lambda_{\min}(\sigma) + \frac{\alpha}{\alpha-1}\log_2\left\|\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right\|_\alpha &\leq \widetilde{D}_\alpha(\rho\|\sigma)\\ &\leq \frac{1}{\alpha-1}\log_2\lambda_{\max}(\sigma) + \frac{\alpha}{\alpha-1}\log_2\left\|\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right\|_\alpha. \quad (7.8.26)\end{aligned}$$

Now taking the limit $\alpha \to \infty$ and applying the fact that $\lim_{\alpha\to\infty} \|X\|_\alpha = \|X\|_\infty$ (Proposition 2.9), we conclude the equality $\lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$.

We now consider the geometric Rényi relative entropy. Since we have that

$$\lambda_{\min}(\sigma)\mathbb{1} \le \sigma \le \lambda_{\max}(\sigma)\mathbb{1}, \tag{7.8.27}$$

it follows that

$$\lambda_{\min}(\sigma)\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right] \le \operatorname{Tr}\left[\sigma\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right] \tag{7.8.28}$$

$$\le \lambda_{\max}(\sigma)\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]. \tag{7.8.29}$$

Now taking a logarithm, dividing by $\alpha - 1$, and applying definitions, we find that the following inequalities hold for $\alpha > 1$:

$$\frac{1}{\alpha-1}\log_2\lambda_{\min}(\sigma) + \frac{1}{\alpha-1}\log_2\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]$$

$$\le \widehat{D}_\alpha(\rho\|\sigma) \tag{7.8.30}$$

$$\le \frac{1}{\alpha-1}\log_2\lambda_{\max}(\sigma) + \frac{1}{\alpha-1}\log_2\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]. \tag{7.8.31}$$

Rewriting

$$\frac{1}{\alpha-1}\log_2\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right] = \frac{\alpha}{\alpha-1}\log_2\!\left(\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right]\right)^{\frac{1}{\alpha}} \tag{7.8.32}$$

$$= \frac{\alpha}{\alpha-1}\log_2\left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\alpha. \tag{7.8.33}$$

Then by applying $\lim_{\alpha\to\infty}\|X\|_\alpha = \|X\|_\infty$, it follows that

$$\lim_{\alpha\to\infty}\frac{1}{\alpha-1}\log_2\operatorname{Tr}\!\left[\left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^{\alpha}\right] = D_{\max}(\rho\|\sigma). \tag{7.8.34}$$

Combining this limit with the inequalities in (7.8.30) and (7.8.31), we arrive at the equality $\lim_{\alpha\to\infty}\widehat{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$. ■

As a consequence of Proposition 7.61, it is customary to use the notations

$$\widetilde{D}_\infty(\rho\|\sigma) \equiv D_{\max}(\rho\|\sigma) \equiv \widehat{D}_\infty(\rho\|\sigma) \tag{7.8.35}$$

to denote the max-relative entropy. It also follows that the max-relative entropy satisfies the properties of isometric invariance and additivity, as stated in Proposition 7.31. Proposition 7.31 also tells us that the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ is monotonically increasing in α, which means that the max-relative entropy has the largest value among all sandwiched Rényi relative entropies. Due to Proposition 7.44, a similar conclusion holds for the geometric Rényi relative entropies. The max-relative entropy also satisfies all of the properties stated in Proposition 7.35.

The conditional entropy arising from the max-relative entropy (according to the general definition in (7.3.12)) is known as the *conditional min-entropy*:

$$H_{\min}(A|B)_\rho := \widetilde{H}_\infty(A|B)_\rho = -\inf_{\sigma_B} D_{\max}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{7.8.36}$$

for all bipartite states ρ_{AB}, where the optimization is over states σ_B. Since the max-relative entropy has the largest value among all the sandwiched Rényi relative entropies, the quantity in (7.8.36) has the smallest value among all conditional sandwiched Rényi entropies, which is why it is called the conditional min-entropy. Note that the conditional sandwiched Rényi entropy is defined through (7.3.12), with the generalized divergence $\boldsymbol{D}$ therein replaced by the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$, the latter defined in (7.5.2). On the other hand, the quantity

$$H_{\max}(A|B)_\rho := \widetilde{H}_{\frac{1}{2}}(A|B)_\rho = -\inf_{\sigma_B} \widetilde{D}_{\frac{1}{2}}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{7.8.37}$$

is known as the *conditional max-entropy* of the state ρ_{AB}, where the optimization is over states σ_B. This name comes from the fact that $\alpha = \frac{1}{2}$ is the smallest value of α for which the sandwiched Rényi relative entropy is known to satisfy the data-processing inequality (recall Theorem 7.33). Since the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ is monotonically increasing in α, the quantity in (7.8.37) is known as the conditional max-entropy because it has the largest value among all conditional sandwiched Rényi entropies for which the data-processing inequality is known to hold.

REMARK: Let ρ_{XB} be a classical–quantum state of the form

$$\rho_{XB} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_B^x, \tag{7.8.38}$$

where $\mathcal{X}$ is a finite alphabet $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\rho_B^x\}_{x\in\mathcal{X}}$ is a set of states. Using the duality of semi-definite programs (see Section 2.4), as done in (5.3.122), it follows that

$$H_{\min}(X|B)_\rho = -\log_2 p^*_{\text{succ}}(\{(p(x), \rho_B^x)\}_x), \tag{7.8.39}$$

where $p^*_{\text{succ}}(\{(p(x), \rho_B^x)\})$, defined in (5.3.119), is the optimal success probability for multiple state discrimination. The conditional min-entropy of a classical–quantum state thus has an operational interpretation in terms of the optimal success probability for multiple state discrimination.

7.8.1 Smooth Max-Relative Entropy

For the analysis of lower bounds on quantum and private communication rates, we require the smooth max-relative entropy, which is an example of a smooth generalized divergence. A smooth generalized divergence, denoted by $\boldsymbol{D}^{\varepsilon}(\rho\|\sigma)$, is defined by taking a generalized divergence $\boldsymbol{D}(\rho\|\sigma)$, for a given state ρ and positive semi-definite operator σ, and optimizing the quantity $\boldsymbol{D}(\widetilde{\rho}\|\sigma)$ over states $\widetilde{\rho}$ that are within a distance ε from the given state ρ. Specifically, it is defined as follows:

$$\boldsymbol{D}^{\varepsilon}(\rho\|\sigma) := \inf_{\widetilde{\rho}\in\mathcal{B}^{\varepsilon}(\rho)} \boldsymbol{D}(\widetilde{\rho}\|\sigma), \tag{7.8.40}$$

where

$$\mathcal{B}^{\varepsilon}(\rho) := \{\tau : \tau \geq 0,\ \mathrm{Tr}[\tau] = 1,\ P(\rho,\tau) \leq \varepsilon\} \tag{7.8.41}$$

is the set of states τ that are ε-close to ρ in terms of the sine distance (Definition 6.16).

Definition 7.62 Smooth Max-Relative Entropy

Let ρ be a state and σ a positive semi-definite operator. Then, the *ε-smooth max-relative entropy*, for $\varepsilon \in [0, 1)$, is defined as

$$D^{\varepsilon}_{\max}(\rho\|\sigma) := \inf_{\widetilde{\rho}\in\mathcal{B}^{\varepsilon}(\rho)} D_{\max}(\widetilde{\rho}\|\sigma). \tag{7.8.42}$$

Just like the max-relative entropy, the smooth max-relative entropy is a generalized divergence, satisfying the data-processing inequality:

Proposition 7.63 Data-Processing Inequality for Smooth Max-Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. The smooth max-relative entropy obeys the following data-processing inequality for all $\varepsilon \in (0, 1)$:

$$D^{\varepsilon}_{\max}(\rho\|\sigma) \geq D^{\varepsilon}_{\max}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.8.43}$$

PROOF: To see this, let $\widetilde{\rho}$ be an arbitrary state such that

$$P(\widetilde{\rho},\rho) \leq \varepsilon. \tag{7.8.44}$$

Then from the data-processing inequality for the sine distance under positive trace-preserving maps (see (6.2.114)), it follows that

$$P(\mathcal{N}(\widetilde{\rho}), \mathcal{N}(\rho)) \leq \varepsilon. \tag{7.8.45}$$

So it follows that

$$D_{\max}(\widetilde{\rho}\|\sigma) \geq D_{\max}(\mathcal{N}(\widetilde{\rho})\|\mathcal{N}(\sigma)) \tag{7.8.46}$$

$$\geq D^{\varepsilon}_{\max}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.8.47}$$

Since the inequality holds for an arbitrary state $\widetilde{\rho}$ satisfying (7.8.44), we conclude (7.8.43). ■

Remark: The proof given above holds more generally when $\mathcal{N}$ is a positive, trace-preserving map, so that (7.8.43) holds in this more general case.

The smooth max-relative entropy can be related to the sandwiched Rényi relative entropy as follows:

Proposition 7.64 Smooth Max- to Sandwiched Rényi Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, $\alpha \in (1, \infty)$, and $\varepsilon \in (0, 1)$. Then,

$$D^{\varepsilon}_{\max}(\rho\|\sigma) \leq \widetilde{D}_{\alpha}(\rho\|\sigma) + \frac{1}{\alpha - 1}\log_2\left(\frac{1}{\varepsilon^2}\right) + \log_2\left(\frac{1}{1-\varepsilon^2}\right). \tag{7.8.48}$$

Proof: The statement is trivially true if $\rho = \sigma$ or if $\operatorname{supp}(\rho) \not\subseteq \operatorname{supp}(\sigma)$. In the former case, $D^{\varepsilon}_{\max}(\rho\|\sigma) = \widetilde{D}_{\alpha}(\rho\|\sigma) = 0$, and in the latter, $\widetilde{D}_{\alpha}(\rho\|\sigma) = +\infty$.

So going forward, we assume that $\rho \neq \sigma$ and $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. As mentioned in (7.8.6), the SDP dual of $D_{\max}(\tau\|\omega)$ is given by

$$D_{\max}(\tau\|\omega) = \log_2 \sup_{\Lambda \geq 0} \{\operatorname{Tr}[\Lambda\tau] : \operatorname{Tr}[\Lambda\omega] \leq 1\}, \tag{7.8.49}$$

implying that

$$D^{\varepsilon}_{\max}(\rho\|\sigma) = \log_2 \inf_{\widetilde{\rho}: P(\widetilde{\rho}, \rho) \leq \varepsilon} \sup_{\Lambda \geq 0, \operatorname{Tr}[\Lambda\sigma] \leq 1} \operatorname{Tr}[\Lambda\widetilde{\rho}]. \tag{7.8.50}$$

Since the objective function $\mathrm{Tr}[\Lambda\widetilde{\rho}]$ is linear in Λ and $\widetilde{\rho}$, the set $\{\Lambda : \Lambda \geq 0, \mathrm{Tr}[\Lambda\sigma] \leq 1\}$ is compact and concave, and the set

$$\{\widetilde{\rho} : P(\widetilde{\rho}, \rho) \leq \varepsilon,\ \widetilde{\rho} \geq 0,\ \mathrm{Tr}[\widetilde{\rho}] = 1\} \tag{7.8.51}$$

is compact and convex (due to convexity of sine distance), the minimax theorem (Theorem 2.24) applies and we find that

$$D^{\varepsilon}_{\max}(\rho\|\sigma) = \log_2 \sup_{\Lambda\geq 0, \mathrm{Tr}[\Lambda\sigma]\leq 1} \ \inf_{\widetilde{\rho}:P(\widetilde{\rho},\rho)\leq\varepsilon} \mathrm{Tr}[\Lambda\widetilde{\rho}]. \tag{7.8.52}$$

For a fixed operator $\Lambda \geq 0$ with spectral decomposition

$$\Lambda = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|, \tag{7.8.53}$$

let us define the following set, for a choice of $\lambda > 0$ to be specified later:

$$\mathcal{S} \coloneqq \left\{ i : \langle\phi_i|\rho|\phi_i\rangle > 2^{\lambda}\langle\phi_i|\sigma|\phi_i\rangle \right\}. \tag{7.8.54}$$

Let

$$\Pi \coloneqq \sum_{i\in\mathcal{S}} |\phi_i\rangle\langle\phi_i|. \tag{7.8.55}$$

Then from the definition, we find that

$$\mathrm{Tr}[\Pi\rho] > 2^{\lambda}\,\mathrm{Tr}[\Pi\sigma], \tag{7.8.56}$$

which implies that

$$\frac{\mathrm{Tr}[\Pi\rho]}{\mathrm{Tr}[\Pi\sigma]} > 2^{\lambda}. \tag{7.8.57}$$

Now consider from the data-processing inequality under the channel

$$\Delta(\omega) \coloneqq \mathrm{Tr}[\Pi\omega]|0\rangle\langle 0| + \mathrm{Tr}[\hat{\Pi}\omega]|1\rangle\langle 1| \tag{7.8.58}$$

that

$$\widetilde{D}_{\alpha}(\rho\|\sigma) \geq \widetilde{D}_{\alpha}(\Delta(\rho)\|\Delta(\sigma)) \tag{7.8.59}$$

$$= \frac{1}{\alpha-1}\log_2\begin{pmatrix} (\mathrm{Tr}[\Pi\rho])^{\alpha}\,(\mathrm{Tr}[\Pi\sigma])^{1-\alpha} \\ +\left(\mathrm{Tr}[\hat{\Pi}\rho]\right)^{\alpha}\left(\mathrm{Tr}[\hat{\Pi}\sigma]\right)^{1-\alpha}\end{pmatrix} \tag{7.8.60}$$

$$\geq \frac{1}{\alpha-1}\log_2\!\left((\mathrm{Tr}[\Pi\rho])^{\alpha}\,(\mathrm{Tr}[\Pi\sigma])^{1-\alpha}\right) \tag{7.8.61}$$

$$= \frac{1}{\alpha - 1} \log_2 \left(\mathrm{Tr}[\Pi\rho] \left(\frac{\mathrm{Tr}[\Pi\rho]}{\mathrm{Tr}[\Pi\sigma]} \right)^{\alpha - 1} \right) \tag{7.8.62}$$

$$= \frac{1}{\alpha - 1} \log_2(\mathrm{Tr}[\Pi\rho]) + \log_2 \left(\frac{\mathrm{Tr}[\Pi\rho]}{\mathrm{Tr}[\Pi\sigma]} \right) \tag{7.8.63}$$

$$\geq \frac{1}{\alpha - 1} \log_2(\mathrm{Tr}[\Pi\rho]) + \lambda. \tag{7.8.64}$$

Now picking

$$\lambda = \widetilde{D}_\alpha(\rho\|\sigma) + \frac{1}{\alpha - 1} \log_2 \left(\frac{1}{\varepsilon^2} \right), \tag{7.8.65}$$

we conclude that

$$\mathrm{Tr}[\Pi\rho] \leq \varepsilon^2. \tag{7.8.66}$$

Defining $\hat{\Pi} \coloneqq \mathbb{1} - \Pi$, this means that

$$\mathrm{Tr}[\hat{\Pi}\rho] \geq 1 - \varepsilon^2. \tag{7.8.67}$$

Thus, the state

$$\rho' \coloneqq \frac{\hat{\Pi}\rho\hat{\Pi}}{\mathrm{Tr}[\hat{\Pi}\rho]} \tag{7.8.68}$$

is such that

$$F(\rho, \rho') \geq 1 - \varepsilon^2, \tag{7.8.69}$$

by applying Lemma 6.15, and in turn that

$$P(\rho, \rho') \leq \varepsilon. \tag{7.8.70}$$

We also have that

$$\rho' \leq \frac{\hat{\Pi}\rho\hat{\Pi}}{1 - \varepsilon^2}. \tag{7.8.71}$$

Now let Λ be an arbitrary operator satisfying $\Lambda \geq 0$ and $\mathrm{Tr}[\Lambda\sigma] \leq 1$, and let Π be the projection defined in (7.8.55) for this choice of Λ. Then we find that

$$\left(1 - \varepsilon^2\right) \mathrm{Tr}[\Lambda\rho'] \leq \mathrm{Tr}[\Lambda\hat{\Pi}\rho\hat{\Pi}] \tag{7.8.72}$$

$$= \mathrm{Tr}[\hat{\Pi}\Lambda\hat{\Pi}\rho] \tag{7.8.73}$$

$$= \sum_{i \notin \mathcal{S}} \lambda_i \langle\phi_i|\rho|\phi_i\rangle \tag{7.8.74}$$

$$\leq 2^{\lambda} \sum_{i \notin \mathcal{S}} \lambda_i \langle \phi_i | \sigma | \phi_i \rangle \tag{7.8.75}$$

$$\leq 2^{\lambda} \operatorname{Tr}[\Lambda \sigma] \tag{7.8.76}$$

$$\leq 2^{\lambda}. \tag{7.8.77}$$

Thus, we have found the following uniform bound for any operator Λ satisfying $\Lambda \geq 0$ and $\operatorname{Tr}[\Lambda\sigma] \leq 1$, with ρ' the state in (7.8.68) depending on Λ and satisfying $P(\rho, \rho') \leq \varepsilon$:

$$\operatorname{Tr}[\Lambda \rho'] \leq 2^{\lambda + \log_2\left(\frac{1}{1-\varepsilon^2}\right)}. \tag{7.8.78}$$

Then it follows that

$$D_{\max}^{\varepsilon}(\rho\|\sigma) = \log_2 \sup_{\Lambda \geq 0, \operatorname{Tr}[\Lambda\sigma] \leq 1} \inf_{\widetilde{\rho}: P(\widetilde{\rho}, \rho) \leq \varepsilon} \operatorname{Tr}[\Lambda \widetilde{\rho}] \tag{7.8.79}$$

$$\leq \log_2 \sup_{\Lambda \geq 0, \operatorname{Tr}[\Lambda\sigma] \leq 1} \operatorname{Tr}[\Lambda \rho'] \tag{7.8.80}$$

$$\leq \lambda + \log_2\left(\frac{1}{1-\varepsilon^2}\right). \tag{7.8.81}$$

This concludes the proof. ■

A quantity of interest is the *smooth conditional min-entropy*, which is a conditional entropy that we define via the general construction of conditional entropies in (7.3.12). For every bipartite state ρ_{AB}, and every $\varepsilon \in [0, 1)$, we define it as

$$H_{\min}^{\varepsilon}(A|B)_{\rho} := -\inf_{\sigma_B} D_{\max}^{\varepsilon}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B), \tag{7.8.82}$$

where we take the infimum over states σ_B.

Using the definition of the smooth conditional min-entropy in (7.8.82) and applying Proposition 7.64, we conclude that

$$H_{\min}^{\varepsilon}(A|B)_{\rho} \geq \widetilde{H}_{\alpha}(A|B)_{\rho} - \frac{1}{\alpha - 1}\log_2\left(\frac{1}{\varepsilon^2}\right) - \log_2\left(\frac{1}{1-\varepsilon^2}\right), \tag{7.8.83}$$

for all $\alpha > 1$ and $\varepsilon \in (0, 1)$. Note that the conditional sandwiched Rényi entropy $\widetilde{H}_{\alpha}(A|B)_{\rho}$ is defined through (7.3.12), with the generalized divergence $\boldsymbol{D}$ therein replaced by the sandwiched Rényi relative entropy $\widetilde{D}_{\alpha}$, the latter defined in (7.5.2).

7.9 Hypothesis Testing Relative Entropy

We now explore another important generalized divergence, called the hypothesis testing relative entropy. This particular entropy is defined to be the optimal value of an operationally defined problem in the context of quantum hypothesis testing. As such, it is debatable as to whether such a quantity should be given the name "entropy." However, our perspective is that the advantages of doing so far outweigh this semantic point about nomenclature, and so we adopt this perspective here and throughout the book. At the most fundamental level, the hypothesis testing relative entropy obeys the quantum data-processing inequality, and for this reason and others, it is useful for characterizing the optimal limits of various communication protocols.

Definition 7.65 ε-Hypothesis Testing Relative Entropy

Given a state ρ, a positive semi-definite operator σ, and $\varepsilon \in [0,1]$, the *ε-hypothesis testing relative entropy* is defined as

$$D_H^{\varepsilon}(\rho\|\sigma) \coloneqq -\log_2 \beta_{\varepsilon}(\rho\|\sigma), \tag{7.9.1}$$

where

$$\beta_{\varepsilon}(\rho\|\sigma) \coloneqq \inf_{\Lambda} \left\{ \mathrm{Tr}[\Lambda\sigma] : 0 \leq \Lambda \leq \mathbb{1},\ \mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon \right\}. \tag{7.9.2}$$

Observe that $D_H^{\varepsilon}(\rho\|\sigma)$ can be written as

$$D_H^{\varepsilon}(\rho\|\sigma) = \sup_{\Lambda}\{-\log_2 \mathrm{Tr}[\Lambda\sigma] : 0 \leq \Lambda \leq \mathbb{1},\ \mathrm{Tr}[\Lambda\rho] = 1-\varepsilon\}. \tag{7.9.3}$$

That is, the monotonicity of the $\log_2$ function allows us to bring $-\log_2$ inside the minimization in the definition of $\beta_{\varepsilon}(\rho\|\sigma)$, and it suffices to optimize over measurement operators that meet the constraint $\mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon$ with equality. This follows because for every measurement operator Λ such that $\mathrm{Tr}[\Lambda\rho] > 1-\varepsilon$, we can modify it by scaling it by a positive number $\lambda \in [0,1)$ such that $\mathrm{Tr}[(\lambda\Lambda)\rho] = 1-\varepsilon$. The new operator $\lambda\Lambda$ is a legitimate measurement operator and the error probability $\mathrm{Tr}[(\lambda\Lambda)\sigma]$ only decreases under this scaling (i.e., $\mathrm{Tr}[(\lambda\Lambda)\sigma] < \mathrm{Tr}[\Lambda\sigma]$), which allows us to conclude (7.9.3).

The hypothesis testing relative entropy can be computed using a semi-definite program, as indicated in the following proposition:

Proposition 7.66 Hypothesis Testing Relative Entropy as an SDP

For every state ρ, positive semi-definite operator σ, and $\varepsilon \in [0,1]$, the ε-hypothesis testing relative entropy can be expressed as the following SDPs:

$$D_H^\varepsilon(\rho\|\sigma) = -\log_2 \inf_{\Lambda \geq 0} \{\mathrm{Tr}[\Lambda\sigma] : \Lambda \leq \mathbb{1},\ \mathrm{Tr}[\Lambda\rho] \geq 1-\varepsilon\} \tag{7.9.4}$$

$$= -\log_2 \sup_{\mu\geq 0, Z \geq 0} \{\mu(1-\varepsilon) - \mathrm{Tr}[Z] : \mu\rho \leq \sigma + Z\}. \tag{7.9.5}$$

Complementary slackness implies that the following equalities hold for optimal Λ, μ, and Z:

$$\Lambda(\sigma + Z) = \mu\Lambda\rho, \qquad \mathrm{Tr}[\Lambda\rho]\mu = (1-\varepsilon)\mu, \qquad \Lambda Z = Z. \tag{7.9.6}$$

Proof: The primal formulation in (7.9.4) is immediate from Definition 7.65. Indeed, considering the standard form in (2.4.4), we see that we can set

$$B = \sigma, \quad Y = \Lambda, \quad \Phi^\dagger(Y) = \begin{pmatrix} \mathrm{Tr}[\Lambda\rho] & 0 \\ 0 & -\Lambda \end{pmatrix}, \quad A = \begin{pmatrix} 1-\varepsilon & 0 \\ 0 & -\mathbb{1} \end{pmatrix}. \tag{7.9.7}$$

To figure out the dual and having already identified A, B, and $\Phi^\dagger$, we need to determine the map Φ and plug into the standard form in (2.4.3). Letting

$$X := \begin{pmatrix} \mu & 0 \\ 0 & Z \end{pmatrix}, \tag{7.9.8}$$

we find that

$$\mathrm{Tr}[\Phi^\dagger(Y)X] = \mathrm{Tr}\left[\begin{pmatrix} \mathrm{Tr}[\Lambda\rho] & 0 \\ 0 & -\Lambda \end{pmatrix}\begin{pmatrix} \mu & 0 \\ 0 & Z \end{pmatrix}\right] \tag{7.9.9}$$

$$= \mu\mathrm{Tr}[\Lambda\rho] - \mathrm{Tr}[\Lambda Z] \tag{7.9.10}$$

$$= \mathrm{Tr}[\Lambda(\mu\rho - Z)], \tag{7.9.11}$$

which implies that

$$\Phi(X) = \mu\rho - Z. \tag{7.9.12}$$

Now substituting into (2.4.3) and simplifying, we conclude that the right-hand side of (7.9.5) is the dual SDP.

To show that this is equal to $D_H^\varepsilon(\rho\|\sigma)$, we should demonstrate that the primal and dual SDPs satisfy the strong duality property. It is clear that $\Lambda = \mathbb{1}$ is a feasible point for the primal SDP. Furthermore, the choices $\mu = 1$ and $Z = \mathbb{1} + [\sigma - \rho]_+$, where $[\sigma - \rho]_+$ is the positive part of $\sigma - \rho$, are strictly feasible for the dual. Thus, we conclude (7.9.5) by applying Theorem 2.28.

The complementary slackness conditions in (7.9.6) follow directly from Proposition 2.29. ■

Proposition 7.67 Optimal Measurement for Hypothesis Testing Relative Entropy

For every state ρ, positive semi-definite operator σ, and $\varepsilon \in [0,1]$, the ε-hypothesis testing relative entropy $D_H^\varepsilon(\rho\|\sigma)$ is achieved by the following measurement operator:

$$\Lambda(\mu^*, p^*) := \Pi_{\mu^*\rho>\sigma} + p^*\Pi_{\mu^*\rho=\sigma}, \tag{7.9.13}$$

where $\Pi_{\mu^*\rho>\sigma}$ is the projection onto the strictly positive part of $\mu^*\rho - \sigma$, the projection $\Pi_{\mu^*\rho=\sigma}$ projects onto the zero eigenspace of $\mu^*\rho - \sigma$, and $\mu^* \geq 0$ and $p^* \in [0,1]$ are chosen as follows:

$$\mu^* := \sup\left\{\mu : \mathrm{Tr}[\Pi_{\mu\rho>\sigma}\rho] \leq 1 - \varepsilon\right\}, \tag{7.9.14}$$

$$p^* := \frac{1 - \varepsilon - \mathrm{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]}{\mathrm{Tr}[\Pi_{\mu^*\rho=\sigma}\rho]}. \tag{7.9.15}$$

Proof: To find the form of an optimal measurement operator for the hypothesis testing relative entropy, let Λ be a measurement operator satisfying $\mathrm{Tr}[\Lambda\rho] = 1 - \varepsilon$ and let $\mu \geq 0$. Then

$$\mathrm{Tr}[\Lambda\sigma] = \mathrm{Tr}[\Lambda\sigma] + \mu\,(1 - \varepsilon - \mathrm{Tr}[\Lambda\rho]) \tag{7.9.16}$$

$$= -\mu\varepsilon + \mathrm{Tr}[(I - \Lambda)\mu\rho] + \mathrm{Tr}[\Lambda\sigma] \tag{7.9.17}$$

$$\geq -\mu\varepsilon + \frac{1}{2}\left(\mathrm{Tr}[\mu\rho + \sigma] - \|\mu\rho - \sigma\|_1\right) \tag{7.9.18}$$

$$= -\mu\varepsilon + \frac{1}{2}\left(\mu + \mathrm{Tr}[\sigma] - \|\mu\rho - \sigma\|_1\right). \tag{7.9.19}$$

The sole inequality follows as an application of Theorem 5.3, with $B = \sigma$ and $A = \mu\rho$. Observe that the final expression is a universal bound independent of

Λ. To determine an optimal measurement operator, we can look to Theorem 5.3. There, it was established that the following measurement operator is an optimal one for $\inf_{\Lambda:0\leq\Lambda\leq I}\{\text{Tr}[(I-\Lambda)\mu\rho]+\text{Tr}[\Lambda\sigma]\}$:

$$\Lambda(\mu, p) := \Pi_{\mu\rho>\sigma} + p\Pi_{\mu\rho=\sigma}, \tag{7.9.20}$$

where $\Pi_{\mu\rho>\sigma}$ is the projection onto the strictly positive part of $\mu\rho-\sigma$, the projection $\Pi_{\mu\rho=\sigma}$ projects onto the zero eigenspace of $\mu\rho-\sigma$, and $p\in[0,1]$. The measurement operator $\Lambda(\mu,p)$ is called a quantum Neyman–Pearson test. We still need to choose the parameters $\mu\geq 0$ and $p\in[0,1]$. Let us pick μ according to the following optimization:

$$\mu^* := \sup\left\{\mu : \text{Tr}[\Pi_{\mu\rho>\sigma}\rho] \leq 1-\varepsilon\right\}. \tag{7.9.21}$$

If it so happens that μ^* is such that $\text{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]=1-\varepsilon$, then we are done; we can pick $p=0$. However, if μ^* is such that $\text{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]<1-\varepsilon$, then we pick $p^*\in[0,1]$ such that

$$p^* := \frac{1-\varepsilon-\text{Tr}[\Pi_{\mu^*\rho>\sigma}\rho]}{\text{Tr}[\Pi_{\mu^*\rho=\sigma}\rho]}, \tag{7.9.22}$$

with it following that $p^*\in[0,1]$ because

$$\text{Tr}[\Pi_{\mu^*\rho>\sigma}\rho] < 1-\varepsilon \leq \text{Tr}[\Pi_{\mu^*\rho\geq\sigma}\rho]. \tag{7.9.23}$$

With these choices, we then find that

$$\text{Tr}[\Lambda(\mu^*, p^*)\rho] = 1-\varepsilon. \tag{7.9.24}$$

By the analysis in (7.9.16)–(7.9.19), it then follows that

$$\text{Tr}[\Lambda\sigma] \geq \text{Tr}[\Lambda(\mu^*, p^*)\sigma] \tag{7.9.25}$$

for all measurement operators Λ satisfying $0\leq\Lambda\leq I$ and $\text{Tr}[\Lambda\rho]=1-\varepsilon$. ■

Note that the other generalized divergences we have considered so far satisfy $\boldsymbol{D}(\rho\|\rho)=0$ for all states ρ. The ε-hypothesis testing relative entropy, however, does not have this property unless $\varepsilon=0$. In fact, it is clear from the definition, along with (7.9.3), that

$$D_H^{\varepsilon}(\rho\|\rho) = -\log_2(1-\varepsilon) \tag{7.9.26}$$

for all states ρ and $\varepsilon\in[0,1]$.

Like the quantum relative entropy, the Petz–Rényi relative entropy, the sandwiched Rényi relative entropy, and the max-relative entropy, the ε-hypothesis testing relative entropy is also a generalized divergence, meaning that is satisfies the data-processing inequality.

Theorem 7.68 Data-Processing Inequality for Hypothesis Testing Relative Entropy

Let ρ be a state, σ a positive semi-definite operator, and $\mathcal{N}$ a quantum channel. Then, for all $\varepsilon \in [0,1]$,

$$D_H^\varepsilon(\rho\|\sigma) \geq D_H^\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.9.27}$$

Proof: The intuition for this proof is as follows: A measurement operator Λ satisfying the constraints $0 \leq \Lambda \leq \mathbb{1}$ and $\mathrm{Tr}[\Lambda\mathcal{N}(\rho)] \geq 1-\varepsilon$ can be understood as a particular measurement strategy for distinguishing ρ from σ in which we first apply the channel $\mathcal{N}$ and then apply the measurement operator Λ. Then this particular measurement strategy cannot perform better than the optimal measurement strategy for distinguishing ρ from σ.

We start by writing $D_H^\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ as in (7.9.1):

$$D_H^\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) = \sup_{\Lambda}\{-\log_2 \mathrm{Tr}[\Lambda\mathcal{N}(\sigma)] : 0 \leq \Lambda \leq \mathbb{1},\ \mathrm{Tr}[\Lambda\mathcal{N}(\rho)] \geq 1-\varepsilon\}. \tag{7.9.28}$$

Fix Λ such that $0 \leq \Lambda \leq \mathbb{1}$ and $\mathrm{Tr}[\Lambda\mathcal{N}(\rho)] \geq 1-\varepsilon$. By definition of the adjoint, we have that

$$\mathrm{Tr}[\Lambda\mathcal{N}(\sigma)] = \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\sigma], \quad \mathrm{Tr}[\Lambda\mathcal{N}(\rho)] = \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\rho]. \tag{7.9.29}$$

Also, note that $0 \leq \mathcal{N}^\dagger(\Lambda) \leq \mathbb{1}$. The leftmost inequality is due to the fact that $\mathcal{N}^\dagger$ is a positive map because $\mathcal{N}$ is. The rightmost inequality is due to the fact that $\mathcal{N}^\dagger$ is subunital because $\mathcal{N}$ is trace non-increasing. By the positivity of $\mathcal{N}^\dagger$, we obtain

$$\Lambda \leq \mathbb{1} \Rightarrow \mathbb{1}-\Lambda \geq 0 \Rightarrow \mathcal{N}^\dagger(\mathbb{1}-\Lambda) \geq 0 \Rightarrow \mathcal{N}^\dagger(\Lambda) \leq \mathcal{N}^\dagger(\mathbb{1}) \leq \mathbb{1}. \tag{7.9.30}$$

Since Λ is arbitrary, we bound (7.9.28) as

$$D_H^\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq \sup_{\Lambda}\{-\log_2 \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\sigma] : 0 \leq \mathcal{N}^\dagger(\Lambda) \leq \mathbb{1},\ \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\rho] \geq 1-\varepsilon\}. \tag{7.9.31}$$

Now, by enlarging the optimization set from measurement operators $\mathcal{N}^\dagger(\Lambda)$ satisfying $0 \leq \mathcal{N}^\dagger(\Lambda) \leq \mathbb{1}$ and $\mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\rho] \geq 1-\varepsilon$ to all measurement operators, say Λ', satisfying $0 \leq \Lambda' \leq \mathbb{1}$ and $\mathrm{Tr}[\Lambda'\rho] \geq 1-\varepsilon$, we obtain

$$D_H^\varepsilon(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$$

$$\leq \sup_{\Lambda}\{-\log_2 \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\sigma] : 0 \leq \mathcal{N}^\dagger(\Lambda) \leq \mathbb{1},\ \mathrm{Tr}[\mathcal{N}^\dagger(\Lambda)\rho] \geq 1-\varepsilon\} \tag{7.9.32}$$

$$\leq \sup_{\Lambda'}\{-\log_2 \mathrm{Tr}[\Lambda'\sigma] : 0 \leq \Lambda' \leq \mathbb{1},\ \mathrm{Tr}[\Lambda'\rho] \geq 1-\varepsilon\} \tag{7.9.33}$$

$$= D_H^\varepsilon(\rho\|\sigma), \tag{7.9.34}$$

as required. ■

Remark: Inspection of the proof above reveals that it holds more generally for $\mathcal{N}$ a positive, trace-non-increasing map.

Proposition 7.69 Properties of Hypothesis Testing Relative Entropy

The ε-hypothesis testing relative entropy satisfies the following properties for all $\varepsilon \in [0,1]$:

1. If $\varepsilon' \in (\varepsilon, 1]$, then
$$D_H^\varepsilon(\rho\|\sigma) \leq D_H^{\varepsilon'}(\rho\|\sigma). \tag{7.9.35}$$

2. The following limit holds
$$\lim_{\varepsilon\to 0} D_H^\varepsilon(\rho\|\sigma) = D_0(\rho\|\sigma), \tag{7.9.36}$$
where $D_0(\rho\|\sigma) = -\log_2 \mathrm{Tr}[\Pi_\rho\sigma]$ is the Petz–Rényi relative entropy of order zero and Π_ρ is the projection onto the support of ρ.

3. For every state ρ and positive semi-definite operators σ, σ' such that $\sigma' \geq \sigma$, we have that $D_H^\varepsilon(\rho\|\sigma) \geq D_H^\varepsilon(\rho\|\sigma')$.

4. For every state ρ, positive semi-definite operator σ, and $\alpha > 0$, we have that $D_H^\varepsilon(\rho\|\alpha\sigma) = D_H^\varepsilon(\rho\|\sigma) - \log_2 \alpha$.

5. Let $p, q : \mathcal{X} \to [0,1]$ be two probability distributions over a finite alphabet $\mathcal{X}$ with associated $|\mathcal{X}|$-dimensional system X, let $\{\rho_A^x\}_{x\in\mathcal{X}}$ be a set of states on a system A, and let σ_A be a state on system A. Then,

$$D_H^{\varepsilon}\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \middle\| \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x\right) \geq \min_{x\in\mathcal{X}} D_H^{\varepsilon}(\rho_A^x\|\sigma_A^x). \quad (7.9.37)$$

PROOF:

1. Eq. (7.9.35) follows from Definition 7.65: increasing ε increases the set of measurement operators Λ over which we can optimize, and $D_H^{\varepsilon}(\rho\|\sigma)$ does not decrease under such a change.

2. Consider that the following inequality holds for all $\varepsilon \in (0,1)$:

$$D_H^{\varepsilon}(\rho\|\sigma) \geq D_0(\rho\|\sigma), \quad (7.9.38)$$

because the measurement operator Π_ρ (projection onto support of ρ) satisfies $\text{Tr}[\Pi_\rho \rho] \geq 1-\varepsilon$ for all $\varepsilon \in (0,1)$. So we conclude that

$$\liminf_{\varepsilon\to 0} D_H^{\varepsilon}(\rho\|\sigma) \geq D_0(\rho\|\sigma). \quad (7.9.39)$$

Alternatively, suppose that Λ is a measurement operator satisfying $\text{Tr}[\Lambda\rho] = 1-\varepsilon$ (note that when optimizing D_H^{ε}, it suffices to optimize over measurement operators satisfying the constraint $\text{Tr}[\Lambda\rho] \geq 1-\varepsilon$ with equality, as mentioned in (7.9.3)). Then applying the data-processing inequality for $D_\alpha(\rho\|\sigma)$ under the measurement $\{\Lambda, I-\Lambda\}$, which holds for $\alpha \in (0,1)$, we find that

$$D_\alpha(\rho\|\sigma) \geq \frac{1}{\alpha-1}\log_2\left[(1-\varepsilon)^\alpha \,\text{Tr}[\Lambda\sigma]^{1-\alpha} + \varepsilon^\alpha\,(1-\text{Tr}[\Lambda\sigma])^{1-\alpha}\right]. \quad (7.9.40)$$

Since this bound holds for all measurement operators Λ satisfying $\text{Tr}[\Lambda\rho] = 1-\varepsilon$, we conclude the following bound for all $\alpha \in (0,1)$:

$$D_\alpha(\rho\|\sigma) \geq \frac{1}{\alpha-1}\log_2\left[(1-\varepsilon)^\alpha \left(2^{-D_H^{\varepsilon}(\rho\|\sigma)}\right)^{1-\alpha} + \varepsilon^\alpha \left(1-2^{-D_H^{\varepsilon}(\rho\|\sigma)}\right)^{1-\alpha}\right]. \quad (7.9.41)$$

Now taking the limit of the right-hand side as $\varepsilon \to 0$, we find that the following bound holds for all $\alpha \in (0,1)$:

$$D_\alpha(\rho\|\sigma) \geq \limsup_{\varepsilon\to 0} D_H^{\varepsilon}(\rho\|\sigma). \quad (7.9.42)$$

Since the bound holds for all $\alpha \in (0,1)$, we can take the limit on the left-hand side to arrive at

$$\lim_{\alpha\to 0} D_\alpha(\rho\|\sigma) = D_0(\rho\|\sigma) \geq \limsup_{\varepsilon\to 0} D_H^\varepsilon(\rho\|\sigma). \tag{7.9.43}$$

Now putting together (7.9.39) and (7.9.43), we conclude (7.9.36).

3. Fix an operator Λ satisfying $0 \leq \Lambda \leq \mathbb{1}$. The assumption $\sigma' \geq \sigma$ implies that $\text{Tr}[\Lambda\sigma'] \geq \text{Tr}[\Lambda\sigma]$, which in turn implies that

$$-\log_2 \text{Tr}[\Lambda\sigma'] \leq -\log_2 \text{Tr}[\Lambda\sigma]. \tag{7.9.44}$$

Since the operator Λ is arbitrary, we obtain $D_H^\varepsilon(\rho\|\sigma) \geq D_H^\varepsilon(\rho\|\sigma')$, as required.

4. This follows immediately from the fact that

$$-\log_2 \text{Tr}[\Lambda(\alpha\sigma)] = -\log_2(\alpha\text{Tr}[\Lambda\sigma]) = -\log_2 \text{Tr}[\Lambda\sigma] - \log_2\alpha \tag{7.9.45}$$

for all operators Λ satisfying $0 \leq \Lambda \leq \mathbb{1}$.

5. Since $D_H^\varepsilon(\rho\|\sigma) = -\log_2 \beta_\varepsilon(\rho\|\sigma)$, where $\beta_\varepsilon(\rho\|\sigma)$ is defined in (5.3.125), we can equivalently show that

$$\beta_\varepsilon\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \middle\| \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x\right) \leq \max_{x\in\mathcal{X}} \beta_\varepsilon(\rho_A^x\|\sigma_A^x). \tag{7.9.46}$$

Now, in the definition of β_ε on the left-hand side of the inequality above, let us restrict the infimum over all measurement operators to those of the form $\Lambda_{XA} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes M_A^x$ such that $0 \leq M_A^x \leq \mathbb{1}_A$ and $\text{Tr}[M_A^x\rho_A^x] \geq 1-\varepsilon$ for all $x \in \mathcal{X}$. Doing this leads to

$$\beta_\varepsilon\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x \middle\| \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma_A^x\right) \tag{7.9.47}$$

$$\leq \inf_{\{M_A^x\}_{x\in\mathcal{X}}} \left\{\sum_{x\in\mathcal{X}} q(x)\text{Tr}[M_A^x\sigma_A^x] : 0 \leq M_A^x \leq \mathbb{1}_A, \text{Tr}[M_A^x\rho_A^x] \geq 1-\varepsilon\ \forall x \in \mathcal{X}\right\} \tag{7.9.48}$$

$$= \sum_{x \in \mathcal{X}} q(x) \inf_{M_A^x} \left\{ \operatorname{Tr}[M_A^x \sigma_A^x] : 0 \leq M_A^x \leq \mathbb{1}_A, \right.$$

$$\left. \operatorname{Tr}[M_A^x \rho_A^x] \geq 1 - \varepsilon \right\} \tag{7.9.49}$$

$$= \sum_{x \in \mathcal{X}} q(x) \beta_\varepsilon(\rho_A^x \| \sigma_A^x) \tag{7.9.50}$$

$$\leq \max_{x \in \mathcal{X}} \beta_\varepsilon(\rho_A^x \| \sigma_A^x). \tag{7.9.51}$$

The last inequality follows because $\beta_\varepsilon(\rho_A^x \| \sigma_A^x) \leq \max_{x \in \mathcal{X}} \beta_\varepsilon(\rho_A^x \| \sigma_A^x)$ for all $x \in \mathcal{X}$. So we have shown that the inequality (7.9.46) holds, which completes the proof. ■

7.9.1 Connection to Quantum Relative Entropy

We now prove a bound relating the ε-hypothesis testing relative entropy to the quantum relative entropy.

Proposition 7.70 Hypothesis Testing to Quantum Relative Entropy

Fix $\varepsilon \in [0, 1)$. Let ρ be a state and σ a positive semi-definite operator. Then the following bound relates the ε-hypothesis testing relative entropy to the quantum relative entropy:

$$D_H^\varepsilon(\rho \| \sigma) \leq \frac{1}{1 - \varepsilon} \left(D(\rho \| \sigma) + h_2(\varepsilon) + \varepsilon \log_2 \operatorname{Tr}[\sigma] \right), \tag{7.9.52}$$

where $h_2(\varepsilon)$ is the binary entropy defined in (7.1.4).

Proof: To see this, we use the fact that the optimization in the definition of $D_H^\varepsilon(\rho \| \sigma)$ can be restricted as in (7.9.3), i.e.,

$$D_H^\varepsilon(\rho \| \sigma) = \sup_{\Lambda} \{ -\log_2 \operatorname{Tr}[\Lambda \sigma] : 0 \leq \Lambda \leq \mathbb{1}, \ \operatorname{Tr}[\Lambda \rho] = 1 - \varepsilon \}. \tag{7.9.53}$$

For every measurement operator Λ such that $\operatorname{Tr}[\Lambda \rho] = 1 - \varepsilon$, the data-processing inequality for the quantum relative entropy (Theorem 7.4) implies that

$$D(\rho \| \sigma) \geq D(\{1 - \varepsilon, \varepsilon\} \| \{ \operatorname{Tr}[\Lambda \sigma], \operatorname{Tr}[\sigma] - \operatorname{Tr}[\Lambda \sigma] \})$$

$$= (1 - \varepsilon) \log_2 \left(\frac{1 - \varepsilon}{\operatorname{Tr}[\Lambda \sigma]} \right) + \varepsilon \log_2 \left(\frac{\varepsilon}{\operatorname{Tr}[\sigma] - \operatorname{Tr}[\Lambda \sigma]} \right) \tag{7.9.54}$$

$$= (1-\varepsilon)\log_2(1-\varepsilon) - (1-\varepsilon)\log_2(\mathrm{Tr}[\Lambda\sigma]) + \varepsilon\log_2(\varepsilon) + \varepsilon\log_2\left(\frac{1}{\mathrm{Tr}[\sigma]\,(1-\mathrm{Tr}[\Lambda\sigma/\mathrm{Tr}[\sigma]])}\right) \tag{7.9.55}$$

$$= -(1-\varepsilon)\log_2\mathrm{Tr}[\Lambda\sigma] - h_2(\varepsilon) + \varepsilon\log_2\left(\frac{1}{1-\mathrm{Tr}[\Lambda\sigma/\mathrm{Tr}[\sigma]]}\right) - \varepsilon\log_2\mathrm{Tr}[\sigma] \tag{7.9.56}$$

$$\geq -(1-\varepsilon)\log_2\mathrm{Tr}[\Lambda\sigma] - h_2(\varepsilon) - \varepsilon\log_2\mathrm{Tr}[\sigma], \tag{7.9.57}$$

where the inequality holds because $\varepsilon\log_2\left(\frac{1}{1-\mathrm{Tr}[\Lambda\sigma/\mathrm{Tr}[\sigma]]}\right) \geq 0$. Rewriting this gives

$$-\log\mathrm{Tr}[\Lambda\sigma] \leq \frac{1}{1-\varepsilon}\left[D(\rho\|\sigma) + h_2(\varepsilon) + \varepsilon\log_2\mathrm{Tr}[\sigma]\right]. \tag{7.9.58}$$

Since this bound holds for all measurement operators Λ satisfying $\mathrm{Tr}[\Lambda\rho] = 1-\varepsilon$, we conclude (7.9.52). ■

7.9.2 Connections to Quantum Rényi Relative Entropies

We now show a connection between the hypothesis testing relative entropy and the Petz– and sandwiched Rényi relative entropies.

Proposition 7.71 Hypothesis Testing to Sandwiched Rényi Relative Entropy

Let ρ be a state and σ a positive semi-definite operator. Let $\alpha \in (1,\infty)$ and $\varepsilon \in [0,1)$. Then the following inequality holds

$$D_H^{\varepsilon}(\rho\|\sigma) \leq \widetilde{D}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{7.9.59}$$

In particular, in the limit $\alpha \to \infty$,

$$D_H^{\varepsilon}(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma) + \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{7.9.60}$$

Proof: If the support condition $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ does not hold, then the right-hand side of (7.9.59) is equal to $+\infty$, and so the result is trivially true. Thus, in what follows, we suppose that the support condition $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ holds. Let Λ denote a measurement operator such that $\text{Tr}[\Lambda\rho] = 1 - \varepsilon$. Let $q \coloneqq \text{Tr}[\Lambda\sigma]$. By the data-processing inequality for the sandwiched Rényi relative entropy for $\alpha > 1$ (Theorem 7.33), under the measurement channel

$$\omega \mapsto \text{Tr}[\Lambda\omega]|0\rangle\langle 0| + \text{Tr}[(\mathbb{1} - \Lambda)\omega]|1\rangle\langle 1|, \tag{7.9.61}$$

we find that

$$\begin{aligned}
\widetilde{D}_\alpha(\rho\|\sigma) &\geq \widetilde{D}_\alpha(\{1-\varepsilon, \varepsilon\}\|\{q, \text{Tr}[\sigma] - q\}) && (7.9.62)\\
&= \frac{1}{\alpha - 1}\log_2\left[(1-\varepsilon)^\alpha q^{1-\alpha} + \varepsilon^\alpha (\text{Tr}[\sigma] - q)^{1-\alpha}\right] && (7.9.63)\\
&\geq \frac{1}{\alpha - 1}\log_2[(1-\varepsilon)^\alpha q^{1-\alpha}] && (7.9.64)\\
&= \frac{\alpha}{\alpha - 1}\log_2(1-\varepsilon) - \log_2 q. && (7.9.65)
\end{aligned}$$

The statement in (7.9.59) follows by taking the supremum over all Λ such that $\text{Tr}[\Lambda\rho] = 1 - \varepsilon$. Furthermore, (7.9.60) follows because $\lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$ (as shown in Proposition 7.61) and the fact that $\lim_{\alpha\to\infty} \frac{\alpha}{\alpha-1} = 1$. ■

The following proposition establishes an inequality relating hypothesis testing relative entropy and the Petz–Rényi relative entropy, and it represents a counterpart to Proposition 7.71. After giving its proof, we show how Proposition 7.71 and the following proposition lead to a proof of the quantum Stein's lemma.

Proposition 7.72 Hypothesis Testing to Petz–Rényi Relative Entropy

Let ρ be a state, and let σ be a positive semi-definite operator. Let $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1]$. Then the following inequality holds:

$$D_H^\varepsilon(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha - 1}\log_2\left(\frac{1}{\varepsilon}\right). \tag{7.9.66}$$

PROOF: The statement is trivially true if $\rho\sigma = 0$ because both $D_H^\varepsilon(\rho\|\sigma) = +\infty$ and $D_\alpha(\rho\|\sigma) = +\infty$ in this case. So we consider the non-trivial case when this equality does not hold. Recall from Lemma 5.5 that the following inequality holds for positive semi-definite operators A and B and for $\alpha \in (0, 1)$:

$$\inf_{\Lambda: 0\leq\Lambda\leq\mathbb{1}} \text{Tr}[(\mathbb{1} - \Lambda)A] + \text{Tr}[\Lambda B] = \frac{1}{2}\left(\text{Tr}[A + B] - \|A - B\|_1\right) \tag{7.9.67}$$

$$\leq \mathrm{Tr}[A^{\alpha}B^{1-\alpha}], \tag{7.9.68}$$

where the first equality is the result of Theorem 5.3. For $p \in (0, 1)$, pick $A = p\rho$ and $B = (1-p)\,\sigma$. Plugging in to the inequality above, we find that there exists a measurement operator $\Lambda^* = \Lambda(p, \rho, \sigma)$ such that

$$p\mathrm{Tr}[(\mathbb{1} - \Lambda^*)\rho] + (1-p)\mathrm{Tr}[\Lambda^*\sigma] \leq p^{\alpha}(1-p)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]. \tag{7.9.69}$$

This implies that

$$p\mathrm{Tr}[(\mathbb{1} - \Lambda^*)\rho] \leq p^{\alpha}(1-p)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}], \tag{7.9.70}$$

and in turn that

$$\mathrm{Tr}[(\mathbb{1} - \Lambda^*)\rho] \leq \left(\frac{1-p}{p}\right)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]. \tag{7.9.71}$$

For a given $\varepsilon \in (0, 1]$ and $\alpha \in (0, 1)$, we pick $p \in (0, 1)$ such that

$$\left(\frac{1-p}{p}\right)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}] = \varepsilon. \tag{7.9.72}$$

This is possible because we can rewrite the equation above as

$$\begin{aligned}
\varepsilon &= \left(\frac{1-p}{p}\right)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}] \\
&= \left(\frac{1}{p} - 1\right)^{1-\alpha}\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}] && (7.9.73)\\
\Longleftrightarrow \quad \left(\frac{1}{p} - 1\right)^{1-\alpha} &= \frac{\varepsilon}{\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]} && (7.9.74)\\
\Longleftrightarrow \quad \frac{1}{p} &= \left(\frac{\varepsilon}{\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]}\right)^{\frac{1}{1-\alpha}} + 1 && (7.9.75)\\
\Longleftrightarrow \quad p &= \frac{1}{\left(\frac{\varepsilon}{\mathrm{Tr}[\rho^{\alpha}\sigma^{1-\alpha}]}\right)^{\frac{1}{1-\alpha}} + 1} \in (0, 1)\,. && (7.9.76)
\end{aligned}$$

This means that $\Lambda^* = \Lambda(p, \rho, \sigma)$, with p selected as above, is a measurement operator such that

$$\mathrm{Tr}[(\mathbb{1} - \Lambda^*)\rho] \leq \varepsilon. \tag{7.9.77}$$

Now, we use the fact that

$$(1-p)\text{Tr}[\Lambda^*\sigma] \leq p^\alpha(1-p)^{1-\alpha}\text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.78}$$

implies

$$\text{Tr}[\Lambda^*\sigma] \leq \left(\frac{p}{1-p}\right)^\alpha \text{Tr}[\rho^\alpha\sigma^{1-\alpha}]. \tag{7.9.79}$$

Considering that

$$\varepsilon = \left(\frac{1-p}{p}\right)^{1-\alpha} \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] = \left(\frac{p}{1-p}\right)^{\alpha-1} \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.80}$$

implies that

$$\left(\frac{\varepsilon}{\text{Tr}[\rho^\alpha\sigma^{1-\alpha}]}\right)^{\frac{1}{\alpha-1}} = \frac{p}{1-p}, \tag{7.9.81}$$

we get that

$$\text{Tr}[\Lambda^*\sigma] \leq \left(\frac{p}{1-p}\right)^\alpha \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.82}$$

$$= \left(\left(\frac{\varepsilon}{\text{Tr}[\rho^\alpha\sigma^{1-\alpha}]}\right)^{\frac{1}{\alpha-1}}\right)^\alpha \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.83}$$

$$= \varepsilon^{\frac{\alpha}{\alpha-1}} \left(\text{Tr}[\rho^\alpha\sigma^{1-\alpha}]\right)^{\frac{\alpha}{1-\alpha}} \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.84}$$

$$= \varepsilon^{\frac{\alpha}{\alpha-1}} \left(\text{Tr}[\rho^\alpha\sigma^{1-\alpha}]\right)^{\frac{1}{1-\alpha}}. \tag{7.9.85}$$

Then, by taking the negative logarithm and optimizing over all Λ^* satisfying (7.9.77), we find that

$$D_H^\varepsilon(\rho\|\sigma) \geq -\log_2 \text{Tr}[\Lambda^*\sigma] \tag{7.9.86}$$

$$\geq -\log_2\left(\varepsilon^{\frac{\alpha}{\alpha-1}} \left(\text{Tr}[\rho^\alpha\sigma^{1-\alpha}]\right)^{\frac{1}{1-\alpha}}\right) \tag{7.9.87}$$

$$= -\frac{\alpha}{\alpha-1}\log_2(\varepsilon) + \frac{1}{\alpha-1}\log_2 \text{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.9.88}$$

$$= -\frac{\alpha}{\alpha-1}\log_2(\varepsilon) + D_\alpha(\rho\|\sigma). \tag{7.9.89}$$

Rearranging this inequality leads to (7.9.66), as required. ■

7.10 Quantum Stein's Lemma

In this section, we show how Propositions 7.71 and 7.72 lead to a proof of the quantum Stein's lemma, which is one of the most important results in the asymptotic theory of quantum hypothesis testing. Before doing so, let us discuss the task of asymmetric hypothesis testing.

The general setting of asymmetric i.i.d. hypothesis testing is illustated in Figure 7.2. Bob is given n copies of a quantum system, each of which is either in the state ρ or in the state σ, and his task is to determine in which state the systems have been prepared. Bob's strategy consists of performing a joint measurement on all of the systems at once, described by the POVM $\{\Lambda^{(n)}, \mathbb{1}^{\otimes n} - \Lambda^{(n)}\}$, guessing "$\rho$" if the outcome corresponds to $\Lambda^{(n)}$ and guessing "σ" if the outcome corresponds to $\mathbb{1}^{\otimes n} - \Lambda^{(n)}$. In this case, there are two types of errors that can occur.

1. *Type-I Error*: Bob guesses "σ", but the systems are in the state $\rho^{\otimes n}$. The probability of this occurring is $\text{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}]$.

2. *Type-II Error*: Bob guesses "ρ", but the systems are in the state $\sigma^{\otimes n}$. The probability of this occurring is $\text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}]$.

The quantity $\beta_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$, defined from (5.3.125), can be interpreted as the minimum type-II error probability subject to the constraint $1 - \text{Tr}[\Lambda^{(n)}\rho] = \text{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho] \leq \varepsilon$ on the type-I error probability. The *rate* of this protocol, given a type-I error probability constraint of ε, is $\frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$, which is essentially the minimum type-II error probability exponent per copy of the states ρ and σ. By increasing the number n of copies, one might imagine that this normalized minimum type-II error probability exponent can be increased. Also, there is generally a trade-off between the type-I and type-II error probabilities, meaning that both cannot be made arbitrarily small simultaneously. However, by increasing the number n of copies of the states ρ and σ, we might expect that the type-I error probability can be brought down all the way to zero. We thus define the *maximum rate for hypothesis testing* of the states ρ and σ as the largest value of the normalized type-II error probability exponent, as $n \to \infty$, such that the type-I error probability vanishes in this limit, i.e.,

$$\inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} \frac{D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})}{n}. \tag{7.10.1}$$

A tractable expression for this optimal rate is given by the *quantum Stein's lemma*,

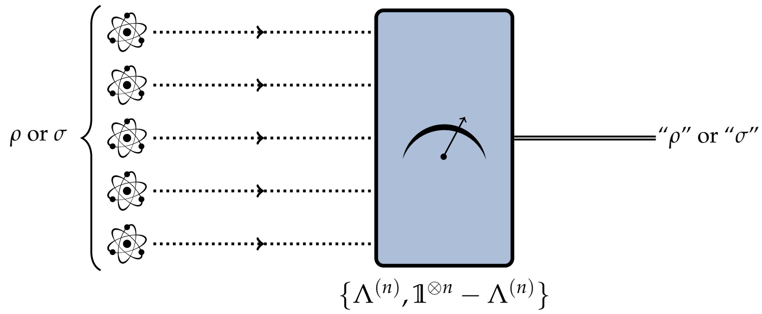


FIGURE 7.2: Schematic depiction of hypothesis testing. Many copies, say n, of an identical quantum system are each prepared either in the state ρ or the state σ. A joint binary measurement, described by the POVM $\{\Lambda^{(n)}, \mathbb{1}^{\otimes n} - \Lambda^{(n)}\}$, is made on the overall state, which is either $\rho^{\otimes n}$ or $\sigma^{\otimes n}$.

which we state and prove in Theorem 7.78 below.

The task of asymmetric hypothesis testing is similar to the task of state discrimination that we considered in Section 5.3.1. While in hypothesis testing we consider two error probabilities, the type-I and type-II error probabilities, in state discrimination we consider only one error probability that is in fact the average of the type-I and type-II error probabilities taken with respect to a prior probability distribution. If $\lambda \in [0, 1]$ is the probability of choosing the state ρ, then the average of the type-I and type-II error probabilities is

$$\lambda \mathrm{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}] + (1-\lambda)\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \\ = p_{\mathrm{err}}(\{\Lambda^{(n)}, \mathbb{1}^{\otimes n} - \Lambda^{(n)}\}, \{\rho^{\otimes n}, \sigma^{\otimes n}\}), \quad (7.10.2)$$

where we recall the definition of the error probability $p_{\mathrm{err}}(\{\Lambda^{(n)}, \mathbb{1}^{\otimes n} - \Lambda^{(n)}\}; \{\rho, \sigma\})$ of state discrimination from (5.3.1). Due to the fact that we combine both the type-I and type-II error probabilities, the task of state discrimination is often referred to as *symmetric hypothesis testing*, while the task of hypothesis testing that we are considering in this section is referred to as *asymmetric hypothesis testing*.

More formally, a *hypothesis testing protocol* is defined by the four elements $(n, \rho, \sigma, \Lambda^{(n)})$, where n is the number of copies of the system, each of which is either in the state ρ or σ, and $0 \leq \Lambda^{(n)} \leq \mathbb{1}^{\otimes n}$ is the operator defining the POVM $\{\Lambda^{(n)}, \mathbb{1}^{\otimes n} - \Lambda^{(n)}\}$ used to decide the state of the system.

Definition 7.73 $(n, \varepsilon_{\mathrm{II}}, \varepsilon_{\mathrm{I}})$ Hypothesis Testing Protocol

Let $(n, \rho, \sigma, \Lambda^{(n)})$ be the elements of a hypothesis testing protocol. The protocol is called an *$(n, \varepsilon_{II}, \varepsilon_I)$ protocol for ρ and σ* if the type-I error probability $\mathrm{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}]$ satisfies

$$\mathrm{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}] \leq \varepsilon_{\mathrm{I}}, \tag{7.10.3}$$

and the type-II error probability $\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}]$ satisfies

$$\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq \varepsilon_{\mathrm{II}}. \tag{7.10.4}$$

Observe that, by definition, if there exists an $(n, \varepsilon_{\mathrm{II}}, \varepsilon_{\mathrm{I}})$ hypothesis testing protocol for ρ and σ, then there exists an $(n, \varepsilon'_{\mathrm{II}}, \varepsilon_{\mathrm{I}})$ hypothesis testing protocol for all $\varepsilon'_{\mathrm{II}} \geq \varepsilon_{\mathrm{II}}$ because every measurement operator $\Lambda^{(n)}$ for which $\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq \varepsilon_{\mathrm{II}}$ clearly satisfies $\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq \varepsilon'_{\mathrm{II}}$. Similarly, if there does not exist an $(n, \varepsilon_{\mathrm{II}}, \varepsilon_{\mathrm{I}})$ hypothesis testing protocol, then there does not exist an $(n, \varepsilon'_{\mathrm{II}}, \varepsilon_{\mathrm{I}})$ hypothesis testing protocol for all $\varepsilon'_{\mathrm{II}} \leq \varepsilon_{\mathrm{II}}$.

We define the *rate* of a hypothesis testing protocol $(n, \rho, \sigma, \Lambda^{(n)})$ as

$$R(n, \rho, \sigma, \Lambda^{(n)}) := -\frac{1}{n}\log_2 \mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}], \tag{7.10.5}$$

The rate is also called the *(normalized) type-II error exponent* because the type-II error probability $\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}]$ can be expressed as $\mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] = 2^{-nR}$. The goal of the hypothesis testing protocol is to find the highest rate such that the type-I error probability tends to zero as n increases.

Definition 7.74 Achievable Rate for Hypothesis Testing

Given quantum states ρ and σ, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for hypothesis testing of ρ and σ* if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol for ρ and σ.

Remark: When we say that there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol for ρ and σ for all $\varepsilon \in (0, 1]$, $\delta > 0$, and "sufficiently large n", we mean that for all $\varepsilon \in (0, 1]$ and $\delta > 0$, there exists a number $N_{\varepsilon,\delta} \in \mathbb{N}$ such that for all $n \geq N_{\varepsilon,\delta}$, there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol for ρ and σ. This convention with the nomenclature "sufficiently large n" is taken throughout the rest of the book.

Note that, by definition, for every achievable rate R there exists a value of n such that the type-I error probability ε becomes arbitrarily close to zero.

Definition 7.75 Optimal Achievable Rate for Hypothesis Testing

Given quantum states ρ and σ, the *optimal achievable rate*, denoted by $E(\rho, \sigma)$, is defined as the supremum of all achievable rates for hypothesis testing of ρ and σ, i.e.,

$$E(\rho, \sigma) \coloneqq \sup\{R : R \text{ is an achievable rate for } \rho, \sigma\}. \tag{7.10.6}$$

Definition 7.76 Strong Converse Rate for Hypothesis Testing

Given quantum states ρ and σ, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for hypothesis testing of ρ and σ* if for all $\varepsilon \in [0, 1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{-n(R+\delta)}, \varepsilon)$ hypothesis testing protocol for ρ and σ.

Note that, by definition, for every strong converse rate R there exists a value of n such that the type-I error probability ε is arbitrarily close to one.

Definition 7.77 Optimal Strong Converse Rate for Hypothesis Testing

Given quantum states ρ and σ, the *optimal strong converse rate*, denoted by $\widetilde{E}(\rho, \sigma)$, is defined as the infimum of all strong converse rates for hypothesis testing of ρ and σ, i.e.,

$$\widetilde{E}(\rho, \sigma) \coloneqq \inf\{R : R \text{ is a strong converse rate for } \rho, \sigma\}. \tag{7.10.7}$$

Note that the following inequality is a direct consequence of definitions:

$$E(\rho, \sigma) \leq \widetilde{E}(\rho, \sigma). \tag{7.10.8}$$

Indeed, suppose for a contradiction that this is not true, i.e., that $E(\rho, \sigma) > \widetilde{E}(\rho, \sigma)$ holds. This means that there exists an achievable rate R such that $\widetilde{E}(\rho, \sigma) < R < E(\rho, \sigma)$, so that for all $\varepsilon \in (0, 1]$, all $\delta > 0$, and all sufficiently large n there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol. On the other hand, since $\widetilde{E}(\rho, \sigma)$ is the optimal strong converse rate, there exists a $\delta > 0$ such that $R - \delta > \widetilde{E}(\rho, \sigma)$ is a strong converse rate. This implies, by definition, that an $(n, 2^{-n(R-\delta+\delta')}, \varepsilon)$ protocol, with $\delta' \in (0, \delta)$, does not exist. However, since R was claimed to be an achievable rate, such a protocol should exist since $R - \delta + \delta' < R$. We have thus reached a contradiction. The inequality $E(\rho, \sigma) > \widetilde{E}(\rho, \sigma)$ therefore cannot be true, which means that $E(\rho, \sigma) \leq \widetilde{E}(\rho, \sigma)$.

We now state and prove the quantum Stein's lemma: both the optimal achievable and strong converse rates are equal to the quantum relative entropy. As alluded to at the beginning of Section 7.2 on the quantum relative entropy, the quantum Stein's lemma gives the quantum relative entropy its most fundamental operational meaning as the optimal rate in asymmetric quantum hypothesis testing.

Theorem 7.78 Quantum Stein's Lemma

For all states ρ and σ, the optimal achievable and strong converse rates are equal to the quantum relative entropy of ρ and σ, i.e.,

$$E(\rho, \sigma) = \widetilde{E}(\rho, \sigma) = D(\rho\|\sigma). \tag{7.10.9}$$

Proof: Note that if $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, then $D(\rho\|\sigma) = +\infty$ by definition. In this singular case, the optimal strong converse rate $\widetilde{E}(\rho, \sigma)$ is undefined, and so we prove that $E(\rho, \sigma) = +\infty$.

Fix $\varepsilon \in (0, 1]$, and let Π_σ be the projection onto the support of σ. Note that since $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, we have that $\text{Tr}[\Pi_\sigma \rho] < 1$. Now, pick n large enough so that $(\text{Tr}[\Pi_\sigma \rho])^n \leq \varepsilon$. Define the measurement operator $\Lambda^{(n)}$ as $\Lambda^{(n)} := \mathbb{1} - (\Pi_\sigma)^{\otimes n}$. Observe that the type-I error probability is

$$\text{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}] = (\text{Tr}[\Pi_\sigma \rho])^n \leq \varepsilon \tag{7.10.10}$$

and the type-II error probability is

$$\text{Tr}[\Lambda^{(n)} \sigma^{\otimes n}] = 1 - (\text{Tr}[\Pi_\sigma \sigma])^n = 0. \tag{7.10.11}$$

Therefore, for all sufficiently large n such that $(\text{Tr}[\Pi_\sigma \rho])^n \leq \varepsilon$ holds, the elements $(n, \rho, \sigma, \Lambda^{(n)})$ constitute an $(n, 0, \varepsilon)$ hypothesis testing protocol for ρ and σ, the rate of which is $+\infty = D(\rho\|\sigma)$. Since ε is arbitrary, we conclude that, for all $\varepsilon \in (0, 1]$ and sufficiently large n, there exists a hypothesis testing protocol for ρ

and σ with rate $R = +\infty$. This implies that $E(\rho, \sigma) = +\infty$ in the singular case of $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$.

For the remainder of the proof, we assume that the support condition $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ holds, so that $D(\rho\|\sigma)$ is finite.

Let us first show that $D(\rho\|\sigma)$ is an achievable rate, which establishes that $E(\rho, \sigma) \geq D(\rho\|\sigma)$. To this end, fix $\varepsilon \in (0, 1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta_1 + \delta_2 = \delta. \tag{7.10.12}$$

Set $\alpha \in (0, 1)$ such that

$$\delta_1 \geq D(\rho\|\sigma) - D_\alpha(\rho\|\sigma), \tag{7.10.13}$$

which is possible because $\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ by Proposition 7.22 and D_α is monotonically increasing in α, as established in Proposition 7.23. Then, with this choice of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(1-\alpha)} \log_2\left(\frac{1}{\varepsilon}\right). \tag{7.10.14}$$

Now, let $0 \leq \Lambda^{(n)} \leq \mathbb{1}^{\otimes n}$ be a measurement operator that achieves the ε-hypothesis testing relative entropy $D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$, which means that

$$\text{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}] = \varepsilon \tag{7.10.15}$$

and

$$-\frac{1}{n}\log_2 \text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] = \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}). \tag{7.10.16}$$

The elements $(n, \rho, \sigma, \Lambda^{(n)})$ thus constitute an $(n, 2^{-n\left(\frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})\right)}, \varepsilon)$ hypothesis testing protocol for ρ and σ. We now apply Proposition 7.72 and the additivity of the Petz–Rényi relative entropy from Proposition 7.23 to find that

$$\frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) \geq \frac{1}{n} D_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{\varepsilon}\right) \tag{7.10.17}$$

$$= D_\alpha(\rho\|\sigma) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{\varepsilon}\right). \tag{7.10.18}$$

Rearranging the right-hand side of this inequality and using (7.10.12)–(7.10.14), we conclude that

$$\frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$$

$$\geq D(\rho\|\sigma) - \left(D(\rho\|\sigma) - D_\alpha(\rho\|\sigma) + \frac{\alpha}{n(1-\alpha)}\log_2\!\left(\frac{1}{\varepsilon}\right)\right) \tag{7.10.19}$$

$$\geq D(\rho\|\sigma) - (\delta_1 + \delta_2) \tag{7.10.20}$$

$$\geq D(\rho\|\sigma) - \delta. \tag{7.10.21}$$

We thus have

$$D(\rho\|\sigma) - \delta \leq \frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}). \tag{7.10.22}$$

The error $2^{-n(D(\rho\|\sigma)-\delta)}$ is then greater than or equal to $2^{-n\left(\frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})\right)}$, which means, by the fact stated in the paragraph immediately after Definition 7.73, that there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol with $R = D(\rho\|\sigma)$ for all sufficiently large n such that (7.10.14) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{-n(R-\delta)}, \varepsilon)$ hypothesis testing protocol with $R = D(\rho\|\sigma)$. Then $D(\rho\|\sigma)$ is an achievable rate, so that

$$E(\rho, \sigma) \geq D(\rho\|\sigma). \tag{7.10.23}$$

Let us now show that the quantum relative entropy $D(\rho\|\sigma)$ is a strong converse rate, which establishes that $\widetilde{E}(\rho, \sigma) \leq D(\rho\|\sigma)$. Fix $\varepsilon \in [0, 1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \tag{7.10.24}$$

Set $\alpha \in (1, \infty)$ such that

$$\delta_1 \geq \widetilde{D}_\alpha(\rho\|\sigma) - D(\rho\|\sigma), \tag{7.10.25}$$

which is possible because $\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ by Proposition 7.30 and $\widetilde{D}_\alpha$ is monotonically increasing in α, as established in Proposition 7.31. With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{7.10.26}$$

Now, consider an arbitrary measurement operator $\Lambda^{(n)}$ such that the hypothesis testing protocol given by $(n, \rho, \sigma, \Lambda^{(n)})$ satisfies $\varepsilon \geq \mathrm{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}]$ and $\varepsilon_{\mathrm{II}} \geq \mathrm{Tr}[\Lambda^{(n)}\sigma^{\otimes n}]$. By definition of the hypothesis testing relative entropy, we have that $-\log_2 \text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$. Applying Proposition 7.71, we thus find that

$$-\frac{1}{n}\log_2 \text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq \frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) \tag{7.10.27}$$

$$\leq \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) + \frac{1}{n}\widetilde{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \tag{7.10.28}$$

$$= \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) + \widetilde{D}_\alpha(\rho\|\sigma), \tag{7.10.29}$$

where the second line follows from the additivity of the sandwiched Rényi relative entropy, as stated in Proposition 7.31. Rearranging the right-hand side of this inequality and using (7.10.24)–(7.10.26), we obtain

$$-\frac{1}{n}\log_2 \text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq D(\rho\|\sigma) + \widetilde{D}_\alpha(\rho\|\sigma) - D(\rho\|\sigma) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{7.10.30}$$

$$\leq D(\rho\|\sigma) + \delta' \tag{7.10.31}$$

$$< D(\rho\|\sigma) + \delta. \tag{7.10.32}$$

We thus have that $\text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] > 2^{-n(D(\rho\|\sigma)+\delta)}$. Since $\Lambda^{(n)}$ is an arbitrary measurement operator satisfying $\varepsilon \geq \text{Tr}[(\mathbb{1}^{\otimes n} - \Lambda^{(n)})\rho^{\otimes n}]$, we see that, for all sufficiently large n such that (7.10.26) holds, an $(n, 2^{-n(D(\rho\|\sigma)+\delta)}, \varepsilon)$ hypothesis testing protocol cannot exist, for if it did we would have $\text{Tr}[\Lambda^{(n)}\sigma^{\otimes n}] \leq 2^{-n(D(\rho\|\sigma)+\delta)}$ for some $\Lambda^{(n)}$. Since ε and δ are arbitrary, we have that for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{-n(D(\rho\|\sigma)+\delta)}, \varepsilon)$ hypothesis testing protocol for ρ and σ, which means that $D(\rho\|\sigma)$ is a strong converse rate, so that

$$\widetilde{E}(\rho,\sigma) \leq D(\rho\|\sigma). \tag{7.10.33}$$

Using (7.10.23), (7.10.33), and (7.10.8), we obtain

$$E(\rho,\sigma) \leq \widetilde{E}(\rho,\sigma) \leq D(\rho\|\sigma) \leq E(\rho,\sigma), \tag{7.10.34}$$

which means that $E(\rho,\sigma) = \widetilde{E}(\rho,\sigma) = D(\rho\|\sigma)$. ■

We can conclude the main result of Theorem 7.78 in a different yet related way. Recall that an alternate definition of the optimal type-II error exponent is given in

(7.10.1), i.e.,

$$E(\rho,\sigma) = \inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} \frac{D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})}{n}. \tag{7.10.35}$$

It is straightforward to show that

$$\inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} \frac{D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})}{n} = D(\rho\|\sigma). \tag{7.10.36}$$

Indeed, using (7.10.18), we find that

$$\inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) \geq D_\alpha(\rho\|\sigma) \tag{7.10.37}$$

for all $\alpha \in (0,1)$, so that taking the supremum over $\alpha \in (0,1)$ on the right-hand side leads to

$$E(\rho,\sigma) \geq D(\rho\|\sigma). \tag{7.10.38}$$

We can also write the optimal strong converse type-II error exponent as

$$\widetilde{E}(\rho,\sigma) = \sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}). \tag{7.10.39}$$

Similarly, using (7.10.29), we conclude that

$$\sup_{\varepsilon\in(0,1)} \limsup_{n\to\infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq \widetilde{D}_\alpha(\rho\|\sigma) \tag{7.10.40}$$

for all $\alpha > 1$, so that taking the infimum over $\alpha \in (1,\infty)$ on the right-hand side leads to

$$\widetilde{E}(\rho,\sigma) \leq D(\rho\|\sigma). \tag{7.10.41}$$

We therefore conclude that $E(\rho,\sigma) = \widetilde{E}(\rho,\sigma) = D(\rho\|\sigma)$. Note that in the arguments above for the lower bound on $E(\rho,\sigma)$ and the upper bound on $\widetilde{E}(\rho,\sigma)$, we did not have to explicitly take the infimum or supremum over $\varepsilon \in (0,1)$, respectively.

7.10.1 Error and Strong Converse Exponents

Given states ρ and σ, as well as the number n of copies of the states, we can change our perspective a bit from that given in the previous section and instead determine bounds on the type-I error probability ε. In particular, we can change our focus a bit, such that we are now interested in how fast the type-I error probability converges to zero if the type-II error exponent is equal to a constant smaller than the quantum relative entropy, and we are also interested in how fast the type-I error probability converges to one if the type-II error exponent is equal to a constant larger than the quantum relative entropy. To assist with this analysis, we establish the following propositions, whose proofs are closely related to the proofs of Propositions 7.71 and 7.72.

Proposition 7.79

Let ρ be a state, and let σ be a positive semi-definite operator. Let $\alpha > 1$ and $R \geq 0$. Then, for Λ a measurement operator satisfying

$$\operatorname{Tr}[\Lambda\sigma] \leq 2^{-R}, \tag{7.10.42}$$

the following bound holds

$$\operatorname{Tr}[(I - \Lambda)\,\rho] \geq 1 - 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(R - \widetilde{D}_\alpha(\rho\|\sigma)\right)}. \tag{7.10.43}$$

REMARK: Note that the second bound is nontrivial only in the case that $R > D(\rho\|\sigma)$ because $\widetilde{D}_\alpha(\rho\|\sigma) > D(\rho\|\sigma)$ for $\alpha > 1$.

PROOF: The proof is similar to the proof of Proposition 7.71. Let $p := \operatorname{Tr}[\Lambda\rho]$ and $q := \operatorname{Tr}[\Lambda\sigma]$. By applying the data-processing inequality for the sandwiched Rényi relative entropy along with the measurement channel from (7.9.61), we conclude that

$$\begin{aligned}
\widetilde{D}_\alpha(\rho\|\sigma) &\geq \widetilde{D}_\alpha(\{p, 1-p\} \,\|\, \{q, \operatorname{Tr}[\sigma] - q\}) && (7.10.44)\\
&= \frac{1}{\alpha-1}\log_2\left[p^\alpha q^{1-\alpha} + (1-p)^\alpha \left(\operatorname{Tr}[\sigma] - q\right)^{1-\alpha}\right] && (7.10.45)\\
&\geq \frac{1}{\alpha-1}\log_2\left[p^\alpha q^{1-\alpha}\right] && (7.10.46)\\
&= \frac{\alpha}{\alpha-1}\log_2 p - \log_2 q && (7.10.47)\\
&\geq \frac{\alpha}{\alpha-1}\log_2 p + R && (7.10.48)\\
&= \frac{\alpha}{\alpha-1}\log_2 \operatorname{Tr}[\Lambda\rho] + R, && (7.10.49)
\end{aligned}$$

which implies that

$$\mathrm{Tr}[\Lambda\rho] \leq 2^{-\left(\frac{\alpha-1}{\alpha}\right)\left(R-\widetilde{D}_\alpha(\rho\|\sigma)\right)} \tag{7.10.50}$$

Rewriting this gives the bound in the statement of the proposition. ■

Proposition 7.80

Let ρ be a state, and let σ be a positive semi-definite operator. Let $\alpha \in (0,1)$ and $R \geq 0$. Then, there exists a measurement operator Λ such that

$$\mathrm{Tr}[\Lambda\sigma] \leq 2^{-R}, \tag{7.10.51}$$

$$\mathrm{Tr}[(I-\Lambda)\,\rho] \leq 2^{-\left(\frac{1-\alpha}{\alpha}\right)(D_\alpha(\rho\|\sigma)-R)}. \tag{7.10.52}$$

Remark: Note that the second bound above is nontrivial only in the case that $R < D(\rho\|\sigma)$ because $D_\alpha(\rho\|\sigma) < D(\rho\|\sigma)$ for $\alpha \in (0,1)$.

Proof: The proof is similar to the proof of Proposition 7.72. Employing the same measurement operator Λ^* therein, we conclude from the same reasoning in that proof that

$$\mathrm{Tr}[(I-\Lambda^*)\,\rho] \leq \left(\frac{1-p}{p}\right)^{1-\alpha} \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}], \tag{7.10.53}$$

$$\mathrm{Tr}[\Lambda^*\sigma] \leq \left(\frac{p}{1-p}\right)^{\alpha} \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}]. \tag{7.10.54}$$

We then pick $p \in (0,1)$ such that the following equation is satisfied

$$2^{-R} = \left(\frac{p}{1-p}\right)^{\alpha} \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{7.10.55}$$

$$= \left(\frac{p}{1-p}\right)^{\alpha} 2^{(\alpha-1)D_\alpha(\rho\|\sigma)} \tag{7.10.56}$$

$$\iff \qquad 2^{-R}2^{(1-\alpha)D_\alpha(\rho\|\sigma)} = \left(\frac{p}{1-p}\right)^{\alpha} \tag{7.10.57}$$

$$\iff \qquad 2^{R/\alpha}2^{(\alpha-1)D_\alpha(\rho\|\sigma)/\alpha} = \left(\frac{1-p}{p}\right). \tag{7.10.58}$$

We see that picking p in such a way is always possible because one more step of the development above leads to the conclusion that

$$p = \frac{1}{1 + 2^{R/\alpha} 2^{(\alpha-1) D_\alpha(\rho\|\sigma)/\alpha}} \in (0,1) . \tag{7.10.59}$$

Substituting into (7.10.53), we find that

$$\operatorname{Tr}[(I - \Lambda^*)\,\rho] \leq \left(2^{R/\alpha} 2^{(\alpha-1)D_\alpha(\rho\|\sigma)/\alpha}\right)^{1-\alpha} 2^{(\alpha-1)D_\alpha(\rho\|\sigma)} \tag{7.10.60}$$

$$= 2^{\left(\frac{1-\alpha}{\alpha}\right)R} 2^{-\frac{(1-\alpha)^2}{\alpha} D_\alpha(\rho\|\sigma)} 2^{(\alpha-1)D_\alpha(\rho\|\sigma)} \tag{7.10.61}$$

$$= 2^{\left(\frac{1-\alpha}{\alpha}\right)R} 2^{\frac{(\alpha-1)}{\alpha} D_\alpha(\rho\|\sigma)} \tag{7.10.62}$$

$$= 2^{-\left(\frac{1-\alpha}{\alpha}\right)(D_\alpha(\rho\|\sigma)-R)}, \tag{7.10.63}$$

concluding the proof. ■

The inequalities from Propositions 7.79 and 7.80 lead to the following bounds on the type-I error probability ε for quantum hypothesis testing, when the type-II error probability has a fixed rate R:

$$1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \widetilde{D}_\alpha(\rho\|\sigma)\right)} \leq \varepsilon \leq 2^{-n\left(\frac{1-\alpha}{\alpha}\right)(D_\alpha(\rho\|\sigma)-R)}. \tag{7.10.64}$$

The left inequality holds for all $\alpha > 1$, while the right inequality holds for all $\alpha \in (0,1)$. Let us now examine the behavior of ε above and below the optimal rate $D(\rho\|\sigma)$.

1. Consider a sequence $\{(n, 2^{-nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of hypothesis testing protocols for ρ and σ such that the rate of each protocol has some arbitrary (but fixed) value $R < D(\rho\|\sigma)$. By Proposition 7.80, the sequence can be chosen such that each element satisfies the right-most inequality in (7.10.64), so that

$$\varepsilon_n \leq 2^{-n\left(\frac{1-\alpha}{\alpha}\right)(D_\alpha(\rho\|\sigma)-R)}. \tag{7.10.65}$$

   Since the Petz–Rényi relative entropy $D_\alpha(\rho\|\sigma)$ is monotonically increasing in α, as established in Proposition 7.23, there exists an $\alpha^* < 1$ such that $D_{\alpha^*}(\rho\|\sigma)$ lies in between $D(\rho\|\sigma)$ and R; i.e., we have that $R < D_{\alpha^*}(\rho\|\sigma)$. We thus obtain

$$\varepsilon_n \leq 2^{-n\left(\frac{1-\alpha^*}{\alpha^*}\right)(D_{\alpha^*}(\rho\|\sigma)-R)}. \tag{7.10.66}$$

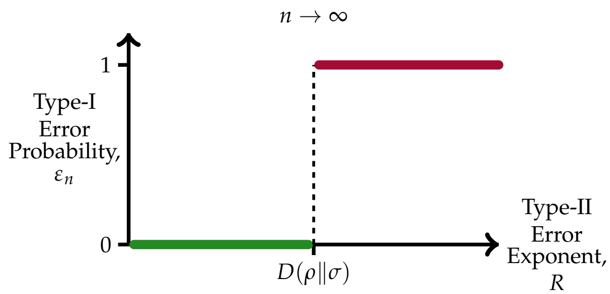


FIGURE 7.3: The type-I error probability ε_n as a function of the rate R, i.e., the type-II error exponent, as the number n of copies of the system approaches infinity for the task of asymmetric hypothesis testing for the states ρ and σ. The optimal rate of $D(\rho\|\sigma)$ for this task, as established by the quantum Stein's lemma in Theorem 7.78, has what is called the *strong converse property*, which means that it is the optimal strong converse rate. Therefore, for every rate above it, the type-I error probability converges to one in the limit of arbitrarily many copies of the system.

Since $R < D_{\alpha^*}(\rho\|\sigma)$, taking the limit $n \to \infty$ on both sides of this inequality gives us $\lim_{n\to\infty} \varepsilon_n \leq 0$. However, $\varepsilon_n \geq 0$ for all n because ε_n is by definition a probability. So we find that

$$\varepsilon_n \to 0 \quad \text{as} \quad n \to \infty \quad \text{if} \quad R < D(\rho\|\sigma). \tag{7.10.67}$$

2. Now consider a sequence $\{(n, 2^{-nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of hypothesis testing protocols for ρ and σ such that the rate of each protocol has some arbitrary (but fixed) value $R > D(\rho\|\sigma)$. In other words, suppose that we would like to perform hypothesis testing at a rate above the optimal rate. Each element of the sequence satisfies the left-most inequality in (7.10.64), so that

$$\varepsilon_n \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\widetilde{D}_\alpha(\rho\|\sigma)\right)}. \tag{7.10.68}$$

Then, recalling from Proposition 7.31 that the sandwiched Rényi relative entropy $\widetilde{D}_\alpha(\rho\|\sigma)$ is monotonically increasing in α, there exists an $\alpha^* > 1$ such that $\widetilde{D}_{\alpha^*}(\rho\|\sigma)$ lies in between $D(\rho\|\sigma)$ and R; i.e., we have that $R > \widetilde{D}_{\alpha^*}(\rho\|\sigma)$. We thus obtain

$$\varepsilon_n \geq 1 - 2^{-n\left(\frac{\alpha^*-1}{\alpha^*}\right)\left(R-\widetilde{D}_{\alpha^*}(\rho\|\sigma)\right)}. \tag{7.10.69}$$

Since $R > \widetilde{D}_{\alpha^*}(\rho\|\sigma)$, taking the limit $n \to \infty$ on both sides of this inequality gives us $\lim_{n\to\infty} \varepsilon_n \geq 1$. However, $\varepsilon_n \leq 1$ for all n because ε_n is by definition a probability. So we conclude that

$$\varepsilon_n \to 1 \quad \text{as} \quad n \to \infty \quad \text{if} \quad R > D(\rho\|\sigma). \tag{7.10.70}$$

So the type-I error probability grows exponentially to one in the limit $n \to \infty$ for every rate above the optimal rate.

The optimal rate $D(\rho\|\sigma)$ is therefore a sharp dividing point below which the type-I error probability ε_n exponentially drops to zero as $n \to \infty$ and above which it exponentially increases to one as $n \to \infty$. This behavior is illustrated in Figure 7.3.

7.11 Information Measures for Quantum Channels

We conclude this chapter with a discussion of information measures for quantum channels.

Given an information measure defined for quantum states, we define a corresponding information measure for channels by sending one share of a bipartite state through a given channel and evaluating the information measure on the corresponding output state. Specifically, for every generalized divergence $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$, which was given in Definition 7.15, we define the *generalized channel divergence* as follows:

Definition 7.81 Generalized Channel Divergence

Given a generalized divergence $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$, a quantum channel $\mathcal{N}_{A\to B}$, and a completely positive map $\mathcal{M}_{A\to B}$, we define the *generalized channel divergence* of $\mathcal{N}$ and $\mathcal{M}$ as

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) \coloneqq \sup_{\rho_{RA}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})), \tag{7.11.1}$$

where the supremum is over all mixed states ρ_{RA} with an arbitrary reference system R.

Proposition 7.82

Let $\boldsymbol{D}$, $\mathcal{N}_{A\to B}$, and $\mathcal{M}_{A\to B}$ be as given in Definition 7.81. It suffices to optimize the generalized channel divergence $\boldsymbol{D}(\mathcal{N}\|\mathcal{M})$ with respect to pure states ψ_{RA}

with the reference system R isomorphic to the input system A:

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) = \sup_{\psi_{RA}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})) \tag{7.11.2}$$

$$= \sup_{\rho_A} \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}). \tag{7.11.3}$$

In the second line, $\Gamma^{\mathcal{N}}_{AB}$ and $\Gamma^{\mathcal{M}}_{AB}$ denote the Choi operators of $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, respectively, and the optimization is with respect to every density operator ρ_A.

Proof: Observe that if we take a purification $|\phi\rangle_{R'RA}$ of the state ρ_{RA} in the optimization in (7.11.1), then we find that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})) \tag{7.11.4}$$

$$= \boldsymbol{D}(\mathcal{N}_{A\to B}(\mathrm{Tr}_{R'}[\phi_{R'RA}])\|\mathcal{M}_{A\to B}(\mathrm{Tr}_{R'}[\phi_{R'RA}])) \tag{7.11.5}$$

$$= \boldsymbol{D}(\mathrm{Tr}_{R'}[\mathcal{N}_{A\to B}(\phi_{R'RA})]\,\|\mathrm{Tr}_{R'}[\mathcal{M}_{A\to B}(\phi_{R'RA})]) \tag{7.11.6}$$

$$\leq \boldsymbol{D}(\mathcal{N}_{A\to B}(\phi_{R'RA})\|\mathcal{M}_{A\to B}(\phi_{R'RA}))\,, \tag{7.11.7}$$

where we have used the data-processing inequality for the generalized divergence in the last line. This means that for every state ρ_{RA}, the generalized divergence in (7.11.4) is never larger than the corresponding generalized divergence evaluated on a purification of ρ_{RA}. This means that it suffices in (7.11.1) to optimize over only pure states. Furthermore, by the Schmidt decomposition theorem (Theorem 2.2), the purifying space $\mathcal{H}_{R'R}$ need not have dimension exceeding that of the dimension of $\mathcal{H}_A$. Therefore, the generalized channel divergence can be written as in (7.11.2).

To see (7.11.3), we first use the fact in (2.2.40), which implies that for every pure state ψ_{RA}, there exists a state ρ_R and a unitary U_R such that

$$|\psi\rangle_{RA} = (U_R\sqrt{\rho_R}\otimes \mathbb{1}_A)|\Gamma\rangle_{RA}. \tag{7.11.8}$$

Thus, it follows that

$$\mathcal{N}_{A\to B}(\psi_{RA}) = \mathcal{N}_{A\to B}((U_R\sqrt{\rho_R}\otimes \mathbb{1}_A)\Gamma_{RA}(\sqrt{\rho_R}U_R^\dagger\otimes \mathbb{1}_A)) \tag{7.11.9}$$

$$= (U_R\sqrt{\rho_R}\otimes \mathbb{1}_B)\mathcal{N}_{A\to B}(\Gamma_{RA})(\sqrt{\rho_R}U_R^\dagger\otimes \mathbb{1}_B) \tag{7.11.10}$$

$$= (U_R\sqrt{\rho_R}\otimes \mathbb{1}_B)\Gamma^{\mathcal{N}}_{RB}(\sqrt{\rho_R}U_R^\dagger\otimes \mathbb{1}_B), \tag{7.11.11}$$

where we employed the Choi representation $\Gamma^{\mathcal{N}}_{RB}$ of the channel $\mathcal{N}$. Similarly, $\mathcal{M}_{A\to B}(\psi_{RA}) = U_R\sqrt{\rho_R}\Gamma^{\mathcal{M}}_{RB}\sqrt{\rho_R}U_R^\dagger$. By employing the unitary invariance of a generalized divergence, we conclude that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})) = \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}). \quad (7.11.12)$$

Then it suffices to optimize over states ρ_A, so that (7.11.3) holds. ■

Proposition 7.83

Let $\boldsymbol{D}$, $\mathcal{N}_{A\to B}$, and $\mathcal{M}_{A\to B}$ be as given in Definition 7.81, and suppose that $\boldsymbol{D}$ obeys the direct-sum property in (7.3.7). Then the function

$$f(\rho_A, \mathcal{M}_{A\to B}) \coloneqq \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}) \quad (7.11.13)$$

is concave in the first argument and convex in the second. If $\mathfrak{M}$ is a convex set of completely positive maps, then

$$\inf_{\mathcal{M}\in\mathfrak{M}}\sup_{\rho_A} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}) \\ = \sup_{\rho_A}\inf_{\mathcal{M}\in\mathfrak{M}} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}). \quad (7.11.14)$$

Equivalently,

$$\inf_{\mathcal{M}\in\mathfrak{M}} \boldsymbol{D}(\mathcal{N}\|\mathcal{M}) = \sup_{\psi_{RA}}\inf_{\mathcal{M}\in\mathfrak{M}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})), \quad (7.11.15)$$

where ψ_{RA} is a pure state with system R isomorphic to system A.

PROOF: To see the concavity, let ψ^0_{RA} and ψ^1_{RA} be pure states with reduced states ψ^0_A and ψ^1_A. Let ψ^λ_{SRA} denote the following pure state:

$$|\psi^\lambda\rangle_{SRA} \coloneqq \sqrt{1-\lambda}|0\rangle_S|\psi^0\rangle_{RA} + \sqrt{\lambda}|1\rangle_S|\psi^1\rangle_{RA}. \quad (7.11.16)$$

Observe that

$$\psi^\lambda_A = (1-\lambda)\,\psi^0_A + \lambda\psi^1_A, \quad (7.11.17)$$

so that the reduced state ψ^λ_A is a convex combination of the reduced states ψ^0_A and ψ^1_A. Define

$$\overline{\psi}^\lambda_{SRA} \coloneqq (1-\lambda)\,|0\rangle\langle 0|_S\otimes\psi^0_{RA} + \lambda|1\rangle\langle 1|_S\otimes\psi^1_{RA}, \quad (7.11.18)$$

which is the state resulting from the action of a completely dephasing qubit channel on system S. Let ϕ^{λ}_{RA} be an arbitrary pure state with reduced state equal to ψ^{λ}_{A}. Then we find that

$$\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\lambda}_{RA})\|\mathcal{M}_{A\to B}(\phi^{\lambda}_{RA})) \\
&\quad = \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{\lambda}_{SRA})\|\mathcal{M}_{A\to B}(\psi^{\lambda}_{SRA})) && (7.11.19)\\
&\quad \geq \boldsymbol{D}(\mathcal{N}_{A\to B}(\overline{\psi}^{\lambda}_{SRA})\|\mathcal{M}_{A\to B}(\overline{\psi}^{\lambda}_{SRA})) && (7.11.20)\\
&\quad = \lambda\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{0}_{RA})\|\mathcal{M}_{A\to B}(\psi^{0}_{RA})) \\
&\qquad\quad + (1-\lambda)\,\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{1}_{RA})\|\mathcal{M}_{A\to B}(\psi^{1}_{RA})). && (7.11.21)
\end{aligned}$$

The first equality follows because every two purifications of the same state are related by an isometric channel acting on the reference system, as well as the isometric invariance of the generalized divergence. The inequality follows from quantum data processing, by the action of a completely dephasing qubit channel on the system S. The final equality follows from the direct-sum property and because the generalized divergence is invariant under tensoring in the same state (Proposition 7.16). Finally, we have the following equalities, by employing the isometric invariance of the generalized divergence (Proposition 7.16), the equality in (7.11.12), and the definition in (7.11.13):

$$\begin{aligned}
\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\lambda}_{RA})\|\mathcal{M}_{A\to B}(\phi^{\lambda}_{RA})) &= f(\psi^{\lambda}_{A},\mathcal{M}_{A\to B}), && (7.11.22)\\
\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{0}_{RA})\|\mathcal{M}_{A\to B}(\psi^{0}_{RA})) &= f(\psi^{0}_{A},\mathcal{M}_{A\to B}), && (7.11.23)\\
\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{1}_{RA})\|\mathcal{M}_{A\to B}(\psi^{1}_{RA})) &= f(\psi^{1}_{A},\mathcal{M}_{A\to B}). && (7.11.24)
\end{aligned}$$

To see the convexity in $\mathcal{M}$, consider that for every $\lambda \in [0,1]$ and completely positive maps $\mathcal{M}_1, \mathcal{M}_2 \in \mathfrak{M}$, the joint convexity of the generalized divergence (Proposition 7.17) gives that

$$\begin{aligned}
&f(\rho_A, \lambda\mathcal{M}_1 + (1-\lambda)\mathcal{M}_2) \\
&\quad = \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\lambda\mathcal{M}_1+(1-\lambda)\mathcal{M}_2}_{AB}\sqrt{\rho_A}) && (7.11.25)\\
&\quad = \boldsymbol{D}(\lambda\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A} + (1-\lambda)\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A} \\
&\qquad\qquad \|\lambda\sqrt{\rho_A}\Gamma^{\mathcal{M}_1}_{AB}\sqrt{\rho_A} + (1-\lambda)\sqrt{\rho_A}\Gamma^{\mathcal{M}_2}_{AB}\sqrt{\rho_A}) && (7.11.26)\\
&\quad \leq \lambda\boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}_1}_{AB}\sqrt{\rho_A}) \\
&\qquad\qquad + (1-\lambda)\boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{M}_2}_{AB}\sqrt{\rho_A}) && (7.11.27)\\
&\quad = \lambda f(\rho_A,\mathcal{M}_1) + (1-\lambda) f(\rho_A,\mathcal{M}_2). && (7.11.28)
\end{aligned}$$

The equality in (7.11.14) follows from what was just shown and Sion's minimax theorem (Theorem 2.24). The equality in (7.11.15) follows from (7.11.12) and Proposition 7.82. ■

The generalized channel divergence takes a simple form if the channel $\mathcal{N}$ and the completely positive map $\mathcal{M}$ both happen to be jointly covariant with respect to a group, as shown in the following proposition.

Proposition 7.84 Generalized Divergence for Jointly Covariant Channels

Let $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ be a generalized divergence and G a finite group with unitary representations $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$. Let $\mathcal{N}_{A\to B}$ be a quantum channel and $\mathcal{M}_{A\to B}$ a completely positive map that are both covariant with respect to G, i.e., (Definition 4.18)

$$\begin{aligned} \mathcal{N}(U_g \rho U_g^\dagger) &= V_g \mathcal{N}(\rho) V_g^\dagger, \\ \mathcal{M}(U_g \rho U_g^\dagger) &= V_g \mathcal{M}(\rho) V_g^\dagger, \end{aligned} \tag{7.11.29}$$

for every $g \in G$ and state ρ. Then, for every state $\psi_{A'A}$, with the dimension of A' equal to the dimension of A,

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A}) \| \mathcal{M}_{A\to B}(\psi_{A'A})) \leq \boldsymbol{D}(\mathcal{N}_{A\to B}(\phi_{RA}^{\overline{\rho}}) \| \mathcal{M}_{A\to B}(\phi_{RA}^{\overline{\rho}})), \tag{7.11.30}$$

where $\rho_A \coloneqq \psi_A = \mathrm{Tr}_{A'}[\psi_{A'A}]$,

$$\overline{\rho}_A \coloneqq \mathcal{T}_G(\rho_A) \coloneqq \frac{1}{|G|} \sum_{g\in G} U_A^g \rho_A U_A^{g\dagger}, \tag{7.11.31}$$

and $\phi_{RA}^{\overline{\rho}}$ is a purification of $\overline{\rho}_A$. Consequently, the generalized channel divergence $\boldsymbol{D}(\mathcal{N}\|\mathcal{M})$ is given by optimizing over pure states $\psi_{A'A}$ such that the reduced state ψ_A is invariant under the channel $\mathcal{T}_G$; i.e.,

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) = \sup_{\psi_{A'A}} \{\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A}) \| \mathcal{M}_{A\to B}(\psi_{A'A})) : \psi_A = \mathcal{T}_G(\psi_A)\}. \tag{7.11.32}$$

In particular, if the representation $\{U_A^g\}_{g\in G}$ is irreducible, then the optimal state in (7.11.32) is the maximally entangled state $\Phi_{A'A}$, so that

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) = \boldsymbol{D}(\rho_{AB}^{\mathcal{N}} \| \rho_{AB}^{\mathcal{M}}), \tag{7.11.33}$$

where $\rho_{AB}^{\mathcal{N}}$ and $\rho_{AB}^{\mathcal{M}}$ are Choi states of $\mathcal{N}$ and $\mathcal{M}$, respectively.

Proof: The inequality

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) \geq \sup_{\psi_{A'A}}\{\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|\mathcal{M}_{A\to B}(\psi_{A'A})) : \psi_A = \mathcal{T}_G(\psi_A)\} \quad (7.11.34)$$

is immediate from the fact that the set $\{\psi_{A'A} : \psi_A = \mathcal{T}_G(\psi_A)\}$ of pure states is a subset of all pure states. The remainder of this proof is devoted to the reverse inequality.

Let $\psi_{A'A}$ be an arbitrary state, let $\rho_A = \psi_A$, and let $\overline{\rho}_A = \mathcal{T}_G(\rho_A)$. Furthermore, let $\phi_{RA}^{\overline{\rho}}$ be a purification of $\overline{\rho}_A$. Let us also consider the following purification of $\overline{\rho}_A$:

$$|\psi^{\overline{\rho}}\rangle_{R'A'A} \coloneqq \frac{1}{\sqrt{|G|}}\sum_{g\in G}|g\rangle_{R'}\otimes(\mathbb{1}_{A'}\otimes U_A^g)|\psi\rangle_{A'A}, \quad (7.11.35)$$

where $\{|g\rangle\}_{g\in G}$ is an orthonormal basis for $\mathcal{H}_{R'}$ indexed by the elements of G. Since all purifications of a state can be mapped to each other by isometries acting on the purifying systems, there exists an isometry $W_{R\to R'A'}$ such that $|\psi^{\overline{\rho}}\rangle_{R'A'A} = W_{R\to R'A'}|\phi^{\overline{\rho}}\rangle_{RA}$. Using this, we find that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{R'A'A}^{\overline{\rho}})\|\mathcal{M}_{A\to B}(\psi_{R'A'A}^{\overline{\rho}})) \quad (7.11.36)$$

$$= \boldsymbol{D}(\mathcal{N}_{A\to B}(\mathcal{W}_{R\to R'A'}(\phi_{RA}^{\overline{\rho}}))\|\mathcal{M}_{A\to B}(\mathcal{W}_{R\to R'A'}(\phi_{RA}^{\overline{\rho}}))) \quad (7.11.37)$$

$$= \boldsymbol{D}(\mathcal{W}_{R\to R'A'}(\mathcal{N}_{A\to B}(\phi_{RA}^{\overline{\rho}}))\|\mathcal{W}_{R\to R'A'}(\mathcal{M}_{A\to B}(\phi_{RA}^{\overline{\rho}}))) \quad (7.11.38)$$

$$= \boldsymbol{D}(\mathcal{N}_{A\to B}(\phi_{RA}^{\overline{\rho}})\|\mathcal{M}_{A\to B}(\phi_{RA}^{\overline{\rho}})), \quad (7.11.39)$$

where the last equality follows from the fact that every generalized divergence is isometrically invariant (recall Proposition 7.16). Now, let us apply the dephasing map $X \mapsto \sum_{g\in G}|g\rangle\langle g|X|g\rangle\langle g|$ to the R' system. Since this map is a channel, by the data-processing inequality for the generalized divergence, we obtain

$$\boldsymbol{D}\left(\mathcal{N}_{A\to B}(\psi_{R'A'A}^{\overline{\rho}})\|\mathcal{M}_{A\to B}(\psi_{R'A'A}^{\overline{\rho}})\right)$$

$$\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_{R'}\otimes(\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{A'A})\right\|$$

$$\left.\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes(\mathcal{M}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{A'A})\right). \quad (7.11.40)$$

Then, because generalized divergences are invariant under unitaries, we can apply the unitary channel given by the unitary $\sum_{g\in G}|g\rangle\!\langle g|\otimes V_B^{g\dagger}$ at the output of $\mathcal{N}$ and $\mathcal{M}$ to obtain

$$\begin{aligned}
&\boldsymbol{D}\left(\mathcal{N}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})\|\mathcal{M}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})\right)\\
&\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes((\mathcal{V}_B^g)^\dagger\circ\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{A'A})\right\|\\
&\qquad\left.\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes((\mathcal{V}_B^g)^\dagger\circ\mathcal{M}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{A'A})\right). \quad (7.11.41)
\end{aligned}$$

Finally, since the group-covariance of $\mathcal{N}$ and $\mathcal{M}$ with respect to the representations $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$ implies that

$$(\mathcal{V}_B^g)^\dagger\circ\mathcal{N}\circ\mathcal{U}_A^g=\mathcal{N},\quad \mathcal{V}_B^{g\dagger}\circ\mathcal{M}\circ\mathcal{U}_A^g=\mathcal{M} \quad (7.11.42)$$

for all $g\in G$, and since from Proposition 7.16 generalized divergences are invariant under tensoring with the same state in both arguments, we obtain

$$\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})\|\mathcal{M}_{A\to B}(\phi^{\overline{\rho}}_{RA})) && (7.11.43)\\
&=\boldsymbol{D}\left(\mathcal{N}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})\|\mathcal{M}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})\right) && (7.11.44)\\
&\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes\mathcal{N}_{A\to B}(\psi_{A'A})\middle\|\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes\mathcal{M}_{A\to B}(\psi_{A'A})\right) && (7.11.45)\\
&=\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|(\mathcal{M}_{A\to B}(\psi_{A'A}))\,, && (7.11.46)
\end{aligned}$$

which is precisely (7.11.30). By definition, the pure state $\phi^{\overline{\rho}}_{RA}$ is such that its reduced state on A is invariant under the channel $\mathcal{T}_G$. Therefore, optimizing over all such pure states, we obtain

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|\mathcal{M}_{A\to B}(\psi_{A'A}))$$

$$\leq \sup_{\phi_{RA}} \{\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi_{RA})\|\mathcal{M}_{A\to B}(\phi_{RA})) : \phi_A = \mathcal{T}_G(\phi_A)\}. \quad (7.11.47)$$

Since this inequality holds for all pure states $\psi_{A'A}$, we obtain

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) = \sup_{\psi_{A'A}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|\mathcal{M}_{A\to B}(\psi_{A'A})) \quad (7.11.48)$$

$$\leq \sup_{\psi_{A'A}} \{\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|\mathcal{M}_{A\to B}(\psi_{A'A})) : \psi_A = \mathcal{T}_G(\psi_A)\}. \quad (7.11.49)$$

Combining this inequality with (7.11.34) gives us (7.11.32).

Finally, if the representation $\{U_A^g\}_{g\in G}$ acting on the input space of the channel $\mathcal{N}$ and the completely positive map $\mathcal{M}$ is irreducible, then for every state $\psi_{A'A}$ such that $\rho_A = \psi_A$, it holds that $\overline{\rho}_A = \mathbb{1}_A/d_A$. Then, since the maximally entangled state is a purification of the maximally mixed state, we let $\phi^{\overline{\rho}}_{RA} = \Phi_{RA}$, which implies by (7.11.39) that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})\|\mathcal{M}_{A\to B}(\psi^{\overline{\rho}}_{R'A'A})) \quad (7.11.50)$$

$$= \boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{M}_{A\to B}(\Phi_{RA})) \quad (7.11.51)$$

$$= \boldsymbol{D}(\rho^{\mathcal{N}}_{RB}\|\rho^{\mathcal{M}}_{RB}). \quad (7.11.52)$$

Then, by (7.11.46), we have that

$$\boldsymbol{D}(\rho^{\mathcal{N}}_{RB}\|\rho^{\mathcal{M}}_{RB}) \geq \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{A'A})\|\mathcal{M}_{A\to B}(\psi_{A'A})) \quad (7.11.53)$$

for all states $\psi_{A'A}$, so that $\boldsymbol{D}(\rho^{\mathcal{N}}_{RB}\|\rho^{\mathcal{M}}_{RB}) \geq \boldsymbol{D}(\mathcal{N}\|\mathcal{M})$. Since the reverse inequality trivially holds, we obtain (7.11.33). ■

We can also define information measures for channels based on information measures for states derived from generalized divergences. Using the generalized coherent information and the generalized mutual information defined in (7.3.13) and (7.3.14), respectively, we make the following definitions.

Definition 7.85 Generalized Information Measures for Quantum Channels

Let $\boldsymbol{D}$ be a generalized divergence, as defined in Definition 7.15, and let $\mathcal{N}_{A\to B}$ be a quantum channel.

1. The *generalized mutual information of* $\mathcal{N}$, denoted by $\boldsymbol{I}(\mathcal{N})$, is defined as

$$\begin{aligned} \boldsymbol{I}(\mathcal{N}) &\coloneqq \sup_{\psi_{RA}} \boldsymbol{I}(R;B)_\omega \\ &= \sup_{\psi_{RA}} \inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B), \end{aligned} \tag{7.11.54}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, ψ_{RA} is a pure state with the dimension of R the same as the dimension of A, and the generalized mutual information $\boldsymbol{I}(A;B)_\rho$ of a bipartite state ρ_{AB} has been defined in (7.3.14).

2. The *generalized coherent information of* $\mathcal{N}$, denoted by $\boldsymbol{I}^c(\mathcal{N})$, is defined as

$$\begin{aligned} \boldsymbol{I}^c(\mathcal{N}) &\coloneqq \sup_{\psi_{RA}} \boldsymbol{I}(R\rangle B)_\omega \\ &= \sup_{\psi_{RA}} \inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA}) \| \mathbb{1}_R \otimes \sigma_B), \end{aligned} \tag{7.11.55}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, ψ_{RA} is a pure state with the dimension of R the same as the dimension of A, and the generalized coherent information $\boldsymbol{I}^c(A\rangle B)_\rho$ of a bipartite state ρ_{AB} has been defined in (7.3.13).

3. The *generalized Holevo information of* $\mathcal{N}$, denoted by $\boldsymbol{\chi}(\mathcal{N})$, is defined as

$$\begin{aligned} \boldsymbol{\chi}(\mathcal{N}) &\coloneqq \sup_{\rho_{XA}} \boldsymbol{I}(X;B)_\omega \\ &= \sup_{\rho_{XA}} \inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{XA}) \| \rho_X \otimes \sigma_B), \end{aligned} \tag{7.11.56}$$

where $\omega_{XA} = \mathcal{N}_{A\to B}(\rho_{XA})$. Here, $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$ is a classical–quantum state, with $\mathcal{X}$ a finite alphabet with corresponding $|\mathcal{X}|$-dimensional system X, $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states, and $p : \mathcal{X} \to [0,1]$ is a probability distribution on $\mathcal{X}$. The supremum is additionally over the classical system X; i.e., no restriction is made on the size of $\mathcal{X}$.

The generalized mutual information $\boldsymbol{I}(\mathcal{N})$ and the generalized coherent information $\boldsymbol{I}^c(\mathcal{N})$ both involve an optimization over pure states. It is straightforward to show that it suffices to optimize over pure states for both quantities. The argument is similar to that in (7.11.4)–(7.11.7) above.

For the generalized mutual information of a covariant channel, we can prove a result that is analogous to Proposition 7.84.

Proposition 7.86 Generalized Mutual Information for Covariant Channels

Let $\mathcal{N}_{A\to B}$ be a G-covariant quantum channel for a finite group G. Then, for all pure states ψ_{RA}, with the dimension of R equal to the dimension of A, we have that

$$\boldsymbol{I}(R;B)_{\omega} \leq \boldsymbol{I}(R;B)_{\overline{\omega}}, \tag{7.11.57}$$

where $\omega_{RB} \coloneqq \mathcal{N}_{A\to B}(\psi_{RA})$, $\overline{\omega}_{RB} \coloneqq \mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})$,

$$\overline{\rho}_A = \frac{1}{|G|}\sum_{g\in G} U^g_A \rho_A U^{g\dagger}_A \eqqcolon \mathcal{T}_G(\rho_A), \tag{7.11.58}$$

$\rho_A = \psi_A = \mathrm{Tr}_R[\psi_{RA}]$, and $\phi^{\overline{\rho}}_{RA}$ is a purification of $\overline{\rho}_A$. Consequently,

$$\boldsymbol{I}(\mathcal{N}) = \sup_{\phi_{RA}}\{\boldsymbol{I}(R;B)_{\omega} : \phi_A = \mathcal{T}_G(\phi_A), \omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\}. \tag{7.11.59}$$

In other words, in order to calculate $\boldsymbol{I}(\mathcal{N})$, it suffices to optimize over pure states ψ_{RA} such that the reduced state ψ_A is invariant under the channel $\mathcal{T}_G$ defined in (7.11.58).

If the representation $\{U^g_A\}_{g\in G}$ is irreducible, then $\boldsymbol{I}(\mathcal{N})$ is equal to the generalized mutual information of the Choi state of the channel, i.e.,

$$\boldsymbol{I}(\mathcal{N}) = \boldsymbol{I}(R;B)_{\rho^{\mathcal{N}}}. \tag{7.11.60}$$

Proof: The proof is similar to the proof of Proposition 7.84 and uses some steps therein.

The inequality

$$\boldsymbol{I}(\mathcal{N}) \geq \sup_{\phi_{RA}}\{\boldsymbol{I}(R;B)_{\omega} : \phi_A = \mathcal{T}_G(\phi_A), \omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\} \tag{7.11.61}$$

holds simply by restricting the optimization in the definition of $\boldsymbol{I}(\mathcal{N})$ to pure states ϕ_{RA} whose reduced states ϕ_A are invariant under the channel $\mathcal{T}_G$. The remainder of the proof is devoted to showing that the reverse inequality holds as well.

Let ψ_{RA} be an arbitrary pure state, let $\rho_A = \psi_A$, and let $\overline{\rho}_A = \mathcal{T}_G(\rho_A)$. Furthermore, let $\phi^{\overline{\rho}}_{RA}$ be a purification of $\overline{\rho}_A$. Let $\mathcal{M}_{A\to B}$ be the replacement channel that traces out the A system and replaces it with a state σ_B. Now following the steps in (7.11.35)–(7.11.41), and incorporating the covariance of the channel $\mathcal{N}$, we conclude that

$$\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{A\to B}(\varphi^{\overline{\rho}}_{RA})\|\mathcal{M}_{A\to B}(\varphi^{\overline{\rho}}_{RA})) \\
&\quad = \boldsymbol{D}(\mathcal{N}_{A\to B}(\varphi^{\overline{\rho}}_{RA})\|\varphi^{\overline{\rho}}_{R}\otimes\sigma_B) && (7.11.62)\\
&\quad \geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_{R'}\otimes\mathcal{N}_{A\to B}(\psi_{RA})\middle\|\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_{R'}\otimes\psi_R\otimes V_B^{g\dagger}\sigma_B V_B^g\right) && (7.11.63)\\
&\quad \geq \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\tau_B), && (7.11.64)
\end{aligned}$$

where the last line follows from the data-processing inequality under the partial-trace channel $\text{Tr}_{R'}$ and we let $\tau_B := \frac{1}{|G|}\sum_{g\in G}V_B^{g\dagger}\sigma_B V_B^g$. By taking the infimum over all states τ_B on the right-hand side of the inequality above, we find that

$$\begin{aligned}
\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})\|\phi^{\overline{\rho}}_{R}\otimes\sigma_B) &\geq \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\tau_B) && (7.11.65)\\
&\geq \inf_{\tau_B}\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\tau_B) && (7.11.66)\\
&= \boldsymbol{I}(R;B)_\omega, && (7.11.67)
\end{aligned}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$. The inequality above holds for all states ψ_{RA} and all states σ_B. Therefore, optimizing over all states σ_B on the left-hand side of the above inequality leads to

$$\inf_{\sigma_B}\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})\|\phi^{\overline{\rho}}_{R}\otimes\sigma_B) = \boldsymbol{I}(R;B)_{\overline{\omega}} \geq \boldsymbol{I}(R;B)_\omega, \tag{7.11.68}$$

where $\overline{\omega}_{RB} = \mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})$. Thus, we conclude (7.11.57).

Next, by construction, the state $\phi^{\overline{\rho}}_{RA}$ is such that its reduced state on A is invariant under the channel $\mathcal{T}_G$. Optimizing over all such states leads to

$$\sup_{\phi_{RA}}\{\boldsymbol{I}(R;B)_\omega : \phi_A = \mathcal{T}_G(\phi_A), \omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\} \geq \boldsymbol{I}(R;B)_\omega. \tag{7.11.69}$$

Since this inequality holds for all pure states ψ_{RA}, we finally obtain

$$\sup_{\phi_{RA}}\{\boldsymbol{I}(R;B)_\omega : \phi_A = \mathcal{T}_G(\phi_A), \omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\}$$
$$\geq \sup_{\psi_{RA}} \boldsymbol{I}(R;B)_\omega = \boldsymbol{I}(\mathcal{N}), \quad (7.11.70)$$

as required.

To prove (7.11.60), note that if $\{U_A^g\}_{g\in G}$ is irreducible, then for every state ψ_{RA}, the state $\rho_A = \psi_A$ satisfies $\overline{\rho}_A = \mathcal{T}_G(\rho_A) = \frac{\mathbb{1}_A}{d_A}$. Then, since the maximally entangled state is a purification of the maximally mixed state, we let $\phi^{\overline{\rho}}_{RA} = \Phi_{RA}$, which implies via (7.11.68) that

$$\inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\pi_R \otimes \sigma_B) = \boldsymbol{I}(R;B)_{\rho^{\mathcal{N}}} \geq \boldsymbol{I}(R;B)_\omega, \quad (7.11.71)$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$. Since the pure state ψ_{RA} is arbitrary, we have that

$$\boldsymbol{I}(R;B)_{\rho^{\mathcal{N}}} \geq \sup_{\psi_{RA}} \boldsymbol{I}(R;B)_\omega = \boldsymbol{I}(\mathcal{N}). \quad (7.11.72)$$

The reverse inequality holds simply by restricting the optimization in the definition of $\boldsymbol{I}(\mathcal{N})$ to the maximally entangled state Φ_{RA}. We thus have (7.11.60), as required. ■

The following proposition is helpful in simplifying the computation of the generalized Holevo information $\boldsymbol{\chi}(\mathcal{N})$ of a quantum channel $\mathcal{N}$:

Proposition 7.87

Let $\mathcal{N}$ be a quantum channel. To compute its generalized Holevo information $\boldsymbol{\chi}(\mathcal{N})$, as defined in (7.11.56), it suffices to optimize over ensembles consisting of pure states. If the underlying generalized divergence is continuous, then no more than d^2 pure states are needed for the optimization, where d is the dimension of the input space of $\mathcal{N}$.

Proof: Let ρ_{XA} be a classical–quantum state of the form

$$\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x, \quad (7.11.73)$$

where $\mathcal{X}$ is a finite alphabet with associated $|\mathcal{X}|$-dimensional system X, $p : \mathcal{X} \to [0,1]$ is a probability distribution on $\mathcal{X}$, and $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states.

First, by simply restricting the optimization in the definition of the Holevo information to ensembles containing pure states only, we obtain

$$\chi(\mathcal{N}) = \sup_{\rho_{XA}} \boldsymbol{I}(X;B)_{\mathcal{N}_{A\to B}(\rho_{XA})} \geq \sup_{\tau_{ZA}} \boldsymbol{I}(Z;B)_{\mathcal{N}_{A\to B}(\tau_{ZA})}, \tag{7.11.74}$$

where τ_{ZA} is a classical–quantum state consisting only of pure states, i.e.,

$$\tau_{ZA} = \sum_{z\in\mathcal{Z}} p(x)|z\rangle\langle z|_Z \otimes |\psi^z\rangle\langle\psi^z|_A. \tag{7.11.75}$$

Now, let each state ρ_A^x in the classical–quantum state ρ_{XA} have a spectral decomposition of the form

$$\rho_A^x = \sum_{y=1}^{r_x} \lambda_y^x |\phi_y^x\rangle\langle\phi_y^x|, \tag{7.11.76}$$

where $r_x = \operatorname{rank}(\rho_A^x)$. So ρ_{XA} can be written as

$$\rho_{XA} = \sum_{x\in\mathcal{X}} \sum_{y=1}^{r_x} p(x)\lambda_y^x |x\rangle\langle x|_X \otimes |\phi_y^x\rangle\langle\phi_y^x|_A. \tag{7.11.77}$$

Now, let us define the state

$$\omega_{XYA} = \sum_{x\in\mathcal{X}} \sum_{y=1}^{r_x} p(x)\lambda_y^x |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes |\phi_y^x\rangle\langle\phi_y^x|_A. \tag{7.11.78}$$

Then, we have that

$$\rho_{XA} = \operatorname{Tr}_Y[\omega_{XYA}]. \tag{7.11.79}$$

Therefore, by the data-processing inequality for the generalized divergence, we find that

$$\begin{aligned}
\boldsymbol{I}(X;B)_{\mathcal{N}(\rho)} &= \inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{XA})\|\rho_X \otimes \sigma_B) && (7.11.80)\\
&= \inf_{\sigma_B} \boldsymbol{D}(\operatorname{Tr}_Y[\mathcal{N}_{A\to B}(\omega_{XYA})]\|\operatorname{Tr}_Y[\omega_{XY}] \otimes \sigma_B) && (7.11.81)\\
&\leq \inf_{\sigma_B} \boldsymbol{D}(\mathcal{N}_{A\to B}(\omega_{XYA})\|\omega_{XY} \otimes \sigma_B) && (7.11.82)\\
&= \boldsymbol{I}(XY;B)_{\mathcal{N}(\omega)}. && (7.11.83)
\end{aligned}$$

Since ω_{XYA} is a classical–quantum state with pure states, it holds that

$$\boldsymbol{I}(XY;B)_{\mathcal{N}(\omega)} \leq \sup_{\tau_{ZA}} \boldsymbol{I}(Z;B)_{\mathcal{N}(\tau)}. \tag{7.11.84}$$

Therefore,

$$\boldsymbol{\chi}(\mathcal{N}) = \sup_{\rho_{XA}} \boldsymbol{I}(X;B)_{\mathcal{N}_{A\to B}(\rho_{XA})} \leq \sup_{\tau_{ZA}} \boldsymbol{I}(Z;B)_{\mathcal{N}_{A\to B}(\tau_{ZA})} \tag{7.11.85}$$

which means that

$$\boldsymbol{\chi}(\mathcal{N}) = \sup_{\tau_{ZA}} \boldsymbol{I}(Z;B)_{\mathcal{N}_{A\to B}(\tau_{ZA})}, \tag{7.11.86}$$

as required.

When the underlying generalized divergence is continuous, the fact that the alphabet $\mathcal{X}$ of the classical–quantum states τ_{ZA} need not exceed d^2 elements is due to the Fenchel–Eggleston–Carathéodory Theorem (Theorem 2.23) and the fact that dimension of the space of density operators on a d-dimensional space is d^2. ■

The information measures for channels on which we primarily focus in this book are those based on the following generalized divergences: the quantum relative entropy, the Petz–Rényi relative entropy, the sandwiched Rényi relative entropy, and the hypothesis testing relative entropy. Specifically, given a channel $\mathcal{N}_{A\to B}$, we are interested in the following mutual information quantities. In each case, ψ_{RA} is a pure state with the dimension of the system R the same as that of A, the state $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and σ_B is a state.

1. *ε-hypothesis testing mutual information* of $\mathcal{N}$, defined for $\varepsilon \in [0,1]$ as

$$I_H^{\varepsilon}(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} I_H^{\varepsilon}(R;B)_{\omega}, \tag{7.11.87}$$

   where

$$I_H^{\varepsilon}(A;B)_{\rho} \coloneqq \inf_{\sigma_B} D_H^{\varepsilon}(\rho_{AB}\|\rho_A \otimes \sigma_B) \tag{7.11.88}$$

   is the *ε-hypothesis testing mutual information* of the bipartite state ρ_{AB}.

2. *Petz–Rényi mutual information* of $\mathcal{N}$:

$$I_{\alpha}(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} I_{\alpha}(R;B)_{\omega} \quad \forall\, \alpha \in [0,1) \cup (1,2], \tag{7.11.89}$$

where

$$I_\alpha(A;B)_\rho := \inf_{\sigma_B} D_\alpha(\rho_{AB}\|\rho_A \otimes \sigma_B) \qquad (7.11.90)$$

is the *Petz–Rényi mutual information* of the bipartite state ρ_{AB}.

3. *Sandwiched Rényi mutual information* of $\mathcal{N}$:

$$\widetilde{I}_\alpha(\mathcal{N}) := \sup_{\psi_{RA}} \widetilde{I}_\alpha(R;B)_\omega \quad \forall\, \alpha \in [1/2, 1) \cup (1, \infty), \qquad (7.11.91)$$

where

$$\widetilde{I}_\alpha(A;B)_\rho := \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB}\|\rho_A \otimes \sigma_B) \qquad (7.11.92)$$

is the *sandwiched Rényi mutual information* of the bipartite state ρ_{AB}.

For each of these quantities, we are also interested in the special case of classical–quantum states. If $\mathcal{X}$ is a finite alphabet with corresponding $|\mathcal{X}|$-dimensional system X, $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states, $p : \mathcal{X} \to [0,1]$ a probability distribution on $\mathcal{X}$, and $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$, then for every channel $\mathcal{N}$ we define the following quantities. In each case, we define $\omega_{XB} := \mathcal{N}_{A\to B}(\rho_{XA})$.

1. *ε-hypothesis testing Holevo information* of $\mathcal{N}$, defined for all $\varepsilon \in [0,1]$ as

$$\chi_H^\varepsilon(\mathcal{N}) := \sup_{\rho_{XA}} I_H^\varepsilon(X;B)_\omega. \qquad (7.11.93)$$

2. *Petz–Rényi Holevo information* of $\mathcal{N}$:

$$\chi_\alpha(\mathcal{N}) := \sup_{\rho_{XA}} I_\alpha(X;B)_\omega \quad \forall\, \alpha \in [0,1) \cup (1,2]. \qquad (7.11.94)$$

3. *Sandwiched Rényi Holevo information* of $\mathcal{N}$:

$$\widetilde{\chi}_\alpha(\mathcal{N}) := \sup_{\rho_{XA}} \widetilde{I}_\alpha(X;B)_\omega \quad \forall\, \alpha \in [1/2, 1) \cup (1, \infty). \qquad (7.11.95)$$

We are also interested in the corresponding coherent information quantities. In each case, ψ_{RA} is a pure state with the dimension of the system R the same as that of A, the state $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and σ_B is a state. The coherent information quantities are defined as follows for every quantum channel $\mathcal{N}_{A\to B}$:

1. *ε-hypothesis testing coherent information* of $\mathcal{N}$, defined for all $\varepsilon \in [0,1]$ as

$$I_H^{c,\varepsilon}(\mathcal{N}) := \sup_{\psi_{RA}} I_H^{\varepsilon}(R\rangle B)_{\omega}, \tag{7.11.96}$$

   where

$$I_H^{\varepsilon}(A\rangle B)_{\rho} := \inf_{\sigma_B} D_H^{\varepsilon}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{7.11.97}$$

   is the *ε-hypothesis testing coherent information* of the bipartite state ρ_{AB}.

2. *Petz–Rényi coherent information* of $\mathcal{N}$:

$$I_{\alpha}^{c}(\mathcal{N}) := \sup_{\psi_{RA}} I_{\alpha}(R\rangle B)_{\omega} \quad \forall\, \alpha \in [0,1) \cup (1,2], \tag{7.11.98}$$

   where

$$I_{\alpha}(A\rangle B)_{\rho} := \inf_{\sigma_B} D_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{7.11.99}$$

   is the *Petz–Rényi coherent information* of the bipartite state ρ_{AB}.

3. *Sandwiched Rényi coherent information* of $\mathcal{N}$:

$$\widetilde{I}_{\alpha}^{c}(\mathcal{N}) := \sup_{\psi_{RA}} \widetilde{I}_{\alpha}(R\rangle B)_{\omega} \quad \forall\, \alpha \in [1/2,1) \cup (1,\infty), \tag{7.11.100}$$

   where

$$\widetilde{I}_{\alpha}(A\rangle B)_{\rho} := \inf_{\sigma_B} \widetilde{D}_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{7.11.101}$$

   is the *sandwiched Rényi coherent information* of the bipartite state ρ_{AB}.

For all of the quantities defined above, we define the corresponding quantities based on the quantum relative entropy by taking the limit $\alpha \to 1$. The key such quantities of interest in this book are the following:

1. *Mutual information of* $\mathcal{N}$, denoted by $I(\mathcal{N})$ and defined as

$$I(\mathcal{N}) := \sup_{\psi_{RA}} I(R;B)_{\omega}, \tag{7.11.102}$$

   where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and we recall from (7.2.96) that the mutual information $I(A;B)_{\rho}$ of a bipartite state ρ_{AB} is given by

$$\begin{aligned} I(A;B)_{\rho} &= H(A)_{\rho} + H(B)_{\rho} - H(AB)_{\rho} && (7.11.103)\\ &= D(\rho_{AB}\|\rho_A \otimes \rho_B) && (7.11.104)\\ &= \inf_{\sigma_B} D(\rho_{AB}\|\rho_A \otimes \sigma_B), && (7.11.105)\end{aligned}$$

   where the optimization in the last line is over states σ_B.

2. *Holevo information of* $\mathcal{N}$, denoted by $\chi(\mathcal{N})$ and defined as

$$\chi(\mathcal{N}) \coloneqq \sup_{\rho_{XA}} I(X;B)_\omega, \tag{7.11.106}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\rho_{XA})$, and the supremum is over all classical-quantum states of the form $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$, with $\mathcal{X}$ a finite alphabet with associated $|\mathcal{X}|$-dimensional system X, $\{\rho_A^x\}_{x\in\mathcal{X}}$ a set of states, and $p : \mathcal{X} \to [0,1]$ a probability distribution on $\mathcal{X}$.

3. *Coherent information of* $\mathcal{N}$, denoted by $I^c(\mathcal{N})$ and defined as

$$I^c(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} I(R\rangle B)_\omega, \tag{7.11.107}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and we recall from (7.2.91) that the coherent information $I(A\rangle B)_\rho$ of a bipartite state ρ_{AB} is given by

$$\begin{aligned} I(A\rangle B)_\rho &= H(B)_\rho - H(AB)_\rho && (7.11.108)\\ &= D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) && (7.11.109)\\ &= \inf_{\sigma_B} D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B), && (7.11.110)\end{aligned}$$

where the optimization in the last line is over states σ_B.

Using (7.11.108), we can write the coherent information of the channel $\mathcal{N}$ as

$$I^c(\mathcal{N}) = \sup_{\psi_{RA}}\{H(B)_\omega - H(RB)_\omega\}. \tag{7.11.111}$$

Given a Stinespring representation of $\mathcal{N}$, so that $\mathcal{N}(\rho) = \mathrm{Tr}_E[V\rho V^\dagger]$, where $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ is an isometry such that $d_E \geq \mathrm{rank}(\Gamma^{\mathcal{N}})$, observe that

$$H(RB)_\omega = H(E)_\tau, \tag{7.11.112}$$

where $\tau_E = \mathcal{N}^c(\psi_A)$. This is due to the fact that the state $\phi_{RBE} \coloneqq V\psi_{RA}V^\dagger$ is pure, meaning that ω_{RB} and $\tau_E = \mathrm{Tr}_{RB}[V\psi_{RA}V^\dagger] = \mathcal{N}^c(\psi_A)$ have the same spectrum. Furthermore, the state on which $H(B)_\omega$ is evaluated is equal to $\mathcal{N}(\psi_A)$. Therefore, we have that

$$I^c(\mathcal{N}) = \sup_{\rho}\{H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho))\}, \tag{7.11.113}$$

where the optimization is over all states ρ.

7.11.1 Simplified Formulas for Rényi Information Measures

In this section, we provide some simplified formulas for the Petz–Rényi information quantities for general bipartite states and for all Rényi information quantities for pure bipartite states.

Proposition 7.88 Quantum Sibson Identities

Let ρ_{AB} be a bipartite state. Then the Petz–Rényi mutual and coherent informations simplify as follows for all $\alpha \in (0,1) \cup (1,\infty)$:

$$I_\alpha(A;B)_\rho = \frac{\alpha}{\alpha-1} \log_2 \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \rho_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right], \tag{7.11.114}$$

$$I_\alpha(A\rangle B)_\rho = \frac{\alpha}{\alpha-1} \log_2 \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\rho_{AB}^\alpha]\right)^{\frac{1}{\alpha}}\right]. \tag{7.11.115}$$

Proof: We show both identities with a unified approach. Let τ_A be a positive semi-definite operator, let σ_B be a state, and let $\omega_B(\alpha)$ denote the following state:

$$\omega_B(\alpha) := \frac{\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}}{\operatorname{Tr}\left[\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]}, \tag{7.11.116}$$

so that

$$\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}} = \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] \cdot \omega_B(\alpha) \tag{7.11.117}$$

We first prove that

$$\begin{aligned} D_\alpha(\rho_{AB}\|\tau_A \otimes \sigma_B) &= D_\alpha(\rho_{AB}\|\tau_A \otimes \omega_B(\alpha)) + D_\alpha(\omega_B(\alpha)\|\sigma_B) && (7.11.118)\\ &\geq D_\alpha(\rho_{AB}\|\tau_A \otimes \omega_B(\alpha)), && (7.11.119) \end{aligned}$$

where the inequality follows because $D_\alpha(\omega_B(\alpha)\|\sigma_B) \geq 0$ for all states. Consider that

$$\begin{aligned} &Q_\alpha(\rho_{AB}\|\tau_A \otimes \sigma_B)\\ &= \operatorname{Tr}[\rho_{AB}^\alpha (\tau_A \otimes \sigma_B)^{1-\alpha}] && (7.11.120)\\ &= \operatorname{Tr}[\rho_{AB}^\alpha(\tau_A^{1-\alpha} \otimes \sigma_B^{1-\alpha})] && (7.11.121) \end{aligned}$$

$$= \mathrm{Tr}_B[\mathrm{Tr}_A[\rho_{AB}^{\alpha}(\tau_A^{1-\alpha} \otimes \sigma_B^{1-\alpha})]] \tag{7.11.122}$$

$$= \mathrm{Tr}[\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\sigma_B^{1-\alpha}] \tag{7.11.123}$$

$$= \mathrm{Tr}\left[\left(\mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] \cdot \omega_B(\alpha)\right)^{\alpha} \sigma_B^{1-\alpha}\right] \tag{7.11.124}$$

$$= \left(\mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]\right)^{\alpha} \cdot \mathrm{Tr}\left[\omega_B(\alpha)^{\alpha}\sigma_B^{1-\alpha}\right]. \tag{7.11.125}$$

Applying the function $(\cdot) \to \frac{1}{\alpha-1}\log_2(\cdot)$ to both sides, we conclude that

$$D_{\alpha}(\rho_{AB}\|\tau_A \otimes \sigma_B) = \frac{\alpha}{\alpha-1}\log_2 \mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] + D_{\alpha}(\omega_B(\alpha)\|\sigma_B). \tag{7.11.126}$$

Now consider that

$$Q_{\alpha}(\rho_{AB}\|\tau_A \otimes \omega_B(\alpha))$$

$$= \mathrm{Tr}[\rho_{AB}^{\alpha}(\tau_A^{1-\alpha} \otimes \omega_B(\alpha)^{1-\alpha})] \tag{7.11.127}$$

$$= \mathrm{Tr}[\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\omega_B(\alpha)^{1-\alpha}] \tag{7.11.128}$$

$$= \mathrm{Tr}\left[\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\left(\frac{(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}])^{\frac{1}{\alpha}}}{\mathrm{Tr}\left[(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}])^{\frac{1}{\alpha}}\right]}\right)^{1-\alpha}\right] \tag{7.11.129}$$

$$= \left(\mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]\right)^{\alpha-1} \mathrm{Tr}\left[\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1-\alpha}{\alpha}}\right] \tag{7.11.130}$$

$$= \left(\mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]\right)^{\alpha-1} \mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.131}$$

$$= \left(\mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]\right)^{\alpha}. \tag{7.11.132}$$

Now applying the function $(\cdot) \to \frac{1}{\alpha-1}\log_2(\cdot)$ to both sides, we conclude that

$$D_{\alpha}(\rho_{AB}\|\tau_A \otimes \omega_B(\alpha)) = \frac{\alpha}{\alpha-1}\log_2 \mathrm{Tr}\left[\left(\mathrm{Tr}_A[\rho_{AB}^{\alpha}\tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]. \tag{7.11.133}$$

So this establishes (7.11.118). We then conclude from (7.11.119) that

$$\inf_{\sigma_B} D_{\alpha}(\rho_{AB}\|\tau_A \otimes \sigma_B) = D_{\alpha}(\rho_{AB}\|\tau_A \otimes \omega_B(\alpha)), \tag{7.11.134}$$

because the lower bound in (7.11.119) is achieved by picking $\sigma_B = \omega_B(\alpha)$. So this establishes that

$$\inf_{\sigma_B} D_\alpha(\rho_{AB}\|\tau_A \otimes \sigma_B) = \frac{\alpha}{\alpha-1}\log_2 \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\rho_{AB}^\alpha \tau_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right]. \tag{7.11.135}$$

We conclude the formula in (7.11.114) by setting $\tau_A = \rho_A$, and we conclude the formula in (7.11.115) by setting $\tau_A = \mathbb{1}_A$. ■

Proposition 7.89 Rényi Information Measures for Pure Bipartite States

Let ψ_{AB} be a pure bipartite state, and let $\alpha \in (0,1)\cup(1,\infty)$. Then the Petz–, sandwiched, and geometric Rényi mutual informations simplify as follows:

$$I_\alpha(A;B)_\psi = 2H_{\frac{2-\alpha}{\alpha}}(A)_\psi, \tag{7.11.136}$$
$$\widetilde{I}_\alpha(A;B)_\psi = 2H_{\frac{1}{2\alpha-1}}(A)_\psi, \tag{7.11.137}$$
$$\widehat{I}_\alpha(A;B)_\psi = 2H_0(A)_\psi, \tag{7.11.138}$$

where the Rényi entropy $H_\alpha(A)$ is defined in (7.4.3). The Petz–, sandwiched, and geometric Rényi coherent informations simplify as follows:

$$I_\alpha(A\rangle B)_\psi = H_{\frac{1}{\alpha}}(A)_\psi, \tag{7.11.139}$$
$$\widetilde{I}_\alpha(A\rangle B)_\psi = H_{\frac{\alpha}{2\alpha-1}}(A)_\psi, \tag{7.11.140}$$
$$\widehat{I}_\alpha(A\rangle B)_\psi = H_{\frac{1}{2}}(A)_\psi. \tag{7.11.141}$$

PROOF: We start by proving (7.11.136). We assume that ψ_{AB} is in its Schmidt form, so that without loss of generality the Hilbert spaces for systems A and B are isomorphic and each have dimension equal to the Schmidt rank of ψ_{AB}. With Γ_{AB} the maximally entangled operator with local bases chosen to match those from the Schmidt decomposition, we have that $\psi_{AB} = \psi_A^{\frac{1}{2}}\Gamma_{AB}\psi_A^{\frac{1}{2}}$, where $\psi_A = \operatorname{Tr}_B[\psi_{AB}]$. We apply the Sibson identity in (7.11.114) to find that

$$2^{\frac{\alpha-1}{\alpha}I_\alpha(A;B)_\psi} = \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\psi_{AB}^\alpha \psi_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] = \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\psi_{AB} \psi_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.142}$$
$$= \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\psi_A^{\frac{1}{2}}\Gamma_{AB}\psi_A^{\frac{1}{2}}\psi_A^{1-\alpha}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.143}$$

$$= \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\Gamma_{AB}\psi_A^{\frac{1}{2}}\psi_A^{1-\alpha}\psi_A^{\frac{1}{2}}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.144}$$

$$= \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\Gamma_{AB}\psi_A^{2-\alpha}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.145}$$

$$= \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\Gamma_{AB}(\psi_B^T)^{2-\alpha}]\right)^{\frac{1}{\alpha}}\right] = \operatorname{Tr}\left[\left((\psi_B^T)^{2-\alpha}\right)^{\frac{1}{\alpha}}\right] \tag{7.11.146}$$

$$= \operatorname{Tr}\left[\psi_B^{\frac{2-\alpha}{\alpha}}\right] = \operatorname{Tr}\left[\psi_A^{\frac{2-\alpha}{\alpha}}\right]. \tag{7.11.147}$$

The fourth equality follows from cyclicity of partial trace and the sixth follows from the transpose trick in (2.2.42). Rearranging the first and last lines gives

$$I_\alpha(A;B)_\psi = \frac{\alpha}{\alpha-1}\log_2 \operatorname{Tr}\left[\psi_A^{\frac{2-\alpha}{\alpha}}\right] \tag{7.11.148}$$

$$= 2\left(\frac{1}{1-\frac{2-\alpha}{\alpha}}\right)\log_2 \operatorname{Tr}\left[\psi_A^{\frac{2-\alpha}{\alpha}}\right] \tag{7.11.149}$$

$$= 2H_{\frac{2-\alpha}{\alpha}}(A)_\psi. \tag{7.11.150}$$

We now prove (7.11.139). It follows from the Sibson identity in (7.11.115):

$$2^{\frac{\alpha-1}{\alpha}I_\alpha(A\rangle B)_\psi} = \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\psi_{AB}^\alpha]\right)^{\frac{1}{\alpha}}\right] = \operatorname{Tr}\left[\left(\operatorname{Tr}_A[\psi_{AB}]\right)^{\frac{1}{\alpha}}\right] \tag{7.11.151}$$

$$= \operatorname{Tr}\left[(\psi_B)^{\frac{1}{\alpha}}\right] = \operatorname{Tr}\left[\psi_A^{\frac{1}{\alpha}}\right]. \tag{7.11.152}$$

Rearranging this gives

$$I_\alpha(A\rangle B)_\psi = \frac{\alpha}{\alpha-1}\log_2 \operatorname{Tr}\left[\psi_A^{\frac{1}{\alpha}}\right] = \frac{1}{1-\frac{1}{\alpha}}\log_2 \operatorname{Tr}\left[\psi_A^{\frac{1}{\alpha}}\right] = H_{\frac{1}{\alpha}}(A)_\psi. \tag{7.11.153}$$

We now prove (7.11.137). Consider, for an arbitrary state σ_B, that

$$\widetilde{Q}_\alpha(\psi_{AB}\|\psi_A\otimes\sigma_B)$$

$$= \operatorname{Tr}\left[\left(\psi_{AB}^{\frac{1}{2}}(\psi_A\otimes\sigma_B)^{\frac{1-\alpha}{\alpha}}\psi_{AB}^{\frac{1}{2}}\right)^\alpha\right] \tag{7.11.154}$$

$$= \operatorname{Tr}\left[\left(|\psi\rangle\langle\psi|_{AB}(\psi_A\otimes\sigma_B)^{\frac{1-\alpha}{\alpha}}|\psi\rangle\langle\psi|_{AB}\right)^\alpha\right] \tag{7.11.155}$$

$$= \left(\langle\psi|_{AB} \left(\psi_A \otimes \sigma_B\right)^{\frac{1-\alpha}{\alpha}} |\psi\rangle_{AB} \right)^{\alpha} \mathrm{Tr}\left[|\psi\rangle\langle\psi|_{AB}^{\alpha}\right] \tag{7.11.156}$$

$$= \left(\langle\psi|_{AB} \left(\psi_A \otimes \sigma_B\right)^{\frac{1-\alpha}{\alpha}} |\psi\rangle_{AB} \right)^{\alpha} \tag{7.11.157}$$

$$= \left(\langle\Gamma|_{AB} \psi_A^{\frac{1}{2}} \left(\psi_A^{\frac{1-\alpha}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{\alpha}} \right) \psi_A^{\frac{1}{2}} |\Gamma\rangle_{AB} \right)^{\alpha} \tag{7.11.158}$$

$$= \left(\langle\Gamma|_{AB} \left(\psi_A^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{\alpha}} \right) |\Gamma\rangle_{AB} \right)^{\alpha} \tag{7.11.159}$$

$$= \left(\langle\Gamma|_{AB} \left(\psi_A^{\frac{1}{\alpha}} \left[\mathrm{T}_A(\sigma_A)\right]^{\frac{1-\alpha}{\alpha}} \otimes \mathbb{1}_B \right) |\Gamma\rangle_{AB} \right)^{\alpha} \tag{7.11.160}$$

$$= \left(\mathrm{Tr}\left[\psi_A^{\frac{1}{\alpha}} \left[\mathrm{T}_A(\sigma_A)\right]^{\frac{1-\alpha}{\alpha}} \right] \right)^{\alpha}. \tag{7.11.161}$$

Now applying the function $(\cdot) \to \frac{1}{\alpha-1}\log_2(\cdot)$ to both sides, we conclude that

$$\widetilde{D}_\alpha(\psi_{AB}\|\psi_A \otimes \sigma_B) = \frac{\alpha}{\alpha-1}\log_2 \mathrm{Tr}\left[\psi_A^{\frac{1}{\alpha}} \left[\mathrm{T}_A(\sigma_A)\right]^{\frac{1-\alpha}{\alpha}} \right], \tag{7.11.162}$$

and applying Proposition 2.8, we conclude that

$$I_\alpha(A;B)_\psi = \inf_{\sigma_B} \widetilde{D}_\alpha(\psi_{AB}\|\psi_A \otimes \sigma_B) \tag{7.11.163}$$

$$= \inf_{\sigma_A} \frac{\alpha}{\alpha-1}\log_2 \mathrm{Tr}\left[\psi_A^{\frac{1}{\alpha}} \left[\mathrm{T}_A(\sigma_A)\right]^{\frac{1-\alpha}{\alpha}} \right] \tag{7.11.164}$$

$$= \frac{\alpha}{\alpha-1}\log_2 \left\| \psi_A^{\frac{1}{\alpha}} \right\|_{\frac{\alpha}{2\alpha-1}} \tag{7.11.165}$$

$$= \frac{\alpha}{\alpha-1}\log_2 \left(\mathrm{Tr}\left[\left(\psi_A^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{2\alpha-1}} \right] \right)^{\frac{2\alpha-1}{\alpha}} \tag{7.11.166}$$

$$= \frac{\alpha}{\alpha-1}\frac{2\alpha-1}{\alpha}\log_2 \mathrm{Tr}\left[\left(\psi_A^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{2\alpha-1}} \right] \tag{7.11.167}$$

$$= \frac{2\alpha-1}{\alpha-1}\log_2 \mathrm{Tr}\left[\psi_A^{\frac{1}{2\alpha-1}} \right] \tag{7.11.168}$$

$$= 2\left(\frac{1}{1-\frac{1}{2\alpha-1}} \right)\log_2 \mathrm{Tr}\left[\psi_A^{\frac{1}{2\alpha-1}} \right] \tag{7.11.169}$$

$$= 2H_{\frac{1}{2\alpha-1}}(A)_\psi. \tag{7.11.170}$$

The third equality follows from Proposition 2.8.

We now prove (7.11.140):

$$\begin{aligned}
&\widetilde{Q}_\alpha(\psi_{AB}\|\mathbb{1}_A\otimes\sigma_B)\\
&=\operatorname{Tr}\left[\left(\psi_{AB}^{\frac{1}{2}}(\mathbb{1}_A\otimes\sigma_B)^{\frac{1-\alpha}{\alpha}}\psi_{AB}^{\frac{1}{2}}\right)^\alpha\right] &(7.11.171)\\
&=\operatorname{Tr}\left[\left(|\psi\rangle\!\langle\psi|_{AB}\left(\mathbb{1}_A\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{AB}\right)^\alpha\right] &(7.11.172)\\
&=\left(\langle\psi|_{AB}\left(\mathbb{1}_A\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\right)|\psi\rangle_{AB}\right)^\alpha\operatorname{Tr}[(|\psi\rangle\!\langle\psi|_{AB})^\alpha] &(7.11.173)\\
&=\left(\langle\psi|_{AB}\left(\mathbb{1}_A\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\right)|\psi\rangle_{AB}\right)^\alpha &(7.11.174)\\
&=\left(\langle\Gamma|_{AB}\left(\psi_A\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\right)|\Gamma\rangle_{AB}\right)^\alpha &(7.11.175)\\
&=\left(\langle\Gamma|_{AB}\left(\psi_A\,[\mathrm{T}_A(\sigma_A)]^{\frac{1-\alpha}{\alpha}}\otimes\mathbb{1}_B\right)|\Gamma\rangle_{AB}\right)^\alpha &(7.11.176)\\
&=\left(\operatorname{Tr}\left[\psi_A\,[\mathrm{T}_A(\sigma_A)]^{\frac{1-\alpha}{\alpha}}\right]\right)^\alpha. &(7.11.177)
\end{aligned}$$

Then consider that

$$\begin{aligned}
\widetilde{I}_\alpha(A\rangle B)_\psi &=\inf_{\sigma_B}\frac{1}{\alpha-1}\log_2\widetilde{Q}_\alpha(\psi_{AB}\|\mathbb{1}_A\otimes\sigma_B) &(7.11.178)\\
&=\inf_{\sigma_A}\frac{1}{\alpha-1}\log_2\left(\operatorname{Tr}\left[\psi_A\,[\mathrm{T}_A(\sigma_A)]^{\frac{1-\alpha}{\alpha}}\right]\right)^\alpha &(7.11.179)\\
&=\frac{\alpha}{\alpha-1}\log_2\|\psi_A\|_{\frac{\alpha}{2\alpha-1}} &(7.11.180)\\
&=\frac{\alpha}{\alpha-1}\log_2\left(\operatorname{Tr}\left[\psi_A^{\frac{\alpha}{2\alpha-1}}\right]\right)^{\frac{2\alpha-1}{\alpha}} &(7.11.181)\\
&=\frac{\alpha}{\alpha-1}\frac{2\alpha-1}{\alpha}\log_2\operatorname{Tr}\left[\psi_A^{\frac{\alpha}{2\alpha-1}}\right] &(7.11.182)\\
&=\frac{2\alpha-1}{\alpha-1}\log_2\operatorname{Tr}\left[\psi_A^{\frac{\alpha}{2\alpha-1}}\right] &(7.11.183)\\
&=\frac{1}{1-\frac{\alpha}{2\alpha-1}}\log_2\operatorname{Tr}\left[\psi_A^{\frac{\alpha}{2\alpha-1}}\right] &(7.11.184)\\
&=H_{\frac{\alpha}{2\alpha-1}}(A)_\psi. &(7.11.185)
\end{aligned}$$

The third equality follows from Proposition 2.8.

We prove (7.11.138). Let σ_B be a state with the same support as ψ_B. Recall the formula in Proposition 7.43 for the geometric Rényi relative entropy when the state ρ is pure. We use this to conclude that

$$\widehat{D}_\alpha(\psi_{AB}\|\psi_A \otimes \sigma_B) = \log_2 \langle\psi|_{AB} (\psi_A \otimes \sigma_B)^{-1} |\psi\rangle_{AB} \tag{7.11.186}$$

$$= \log_2 \langle\psi|_{AB} \left(\psi_A^{-1} \otimes \sigma_B^{-1}\right) |\psi\rangle_{AB} \tag{7.11.187}$$

$$= \log_2 \langle\Gamma|_{AB} \psi_A^{\frac{1}{2}} \left(\psi_A^{-1} \otimes \sigma_B^{-1}\right) \psi_A^{\frac{1}{2}} |\Gamma\rangle_{AB} \tag{7.11.188}$$

$$= \log_2 \langle\Gamma|_{AB} \left(\mathbb{1}_A \otimes \sigma_B^{-1}\right) |\Gamma\rangle_{AB} \tag{7.11.189}$$

$$= \log_2 \mathrm{Tr}\left[\sigma_B^{-1}\right]. \tag{7.11.190}$$

Now consider that the minimum value of $\inf_{\sigma_B} \mathrm{Tr}\left[\sigma_B^{-1}\right]$ occurs when σ_B is the maximally mixed state π_B. This follows from using the Lagrange multiplier method (or alternatively $\inf_{\sigma_B} \mathrm{Tr}\left[\sigma_B^{-1}\right]$ can be evaluated as d_B^2 by applying Proposition 2.8 again, with an implicit identity operator acting on the support of ψ_B). We then conclude that

$$\widehat{I}_\alpha(A;B)_\psi = \inf_{\sigma_B} \log_2 \mathrm{Tr}\left[\sigma_B^{-1}\right] = \log_2 \mathrm{Tr}\left[\pi_B^{-1}\right] \tag{7.11.191}$$

$$= 2\log_2 \mathrm{rank}(\psi_A) = 2H_0(A)_\psi. \tag{7.11.192}$$

We finally prove (7.11.141):

$$\widehat{D}_\alpha(\psi_{AB}\|\mathbb{1}_A \otimes \sigma_B) = \log_2 \langle\psi|_{AB} (\mathbb{1}_A \otimes \sigma_B)^{-1} |\psi\rangle_{AB} \tag{7.11.193}$$

$$= \log_2 \langle\psi|_{AB} \left(\mathbb{1}_A \otimes \sigma_B^{-1}\right) |\psi\rangle_{AB} \tag{7.11.194}$$

$$= \log_2 \langle\Gamma|_{AB} \left(\psi_A \otimes \sigma_B^{-1}\right) |\Gamma\rangle_{AB} \tag{7.11.195}$$

$$= \log_2 \langle\Gamma|_{AB} \left(\mathbb{1}_A \otimes \sigma_B^{-1} \mathrm{T}_B(\psi_B)\right) |\Gamma\rangle_{AB} \tag{7.11.196}$$

$$= \log_2 \mathrm{Tr}\left[\sigma_B^{-1} \mathrm{T}_B(\psi_B)\right]. \tag{7.11.197}$$

Now applying Proposition 2.8, we conclude that

$$\widehat{I}_\alpha(A\rangle B)_\psi = \inf_{\sigma_B} \widehat{D}_\alpha(\psi_{AB}\|\mathbb{1}_A \otimes \sigma_B) = \inf_{\sigma_B} \log_2 \mathrm{Tr}\left[\sigma_B^{-1} \mathrm{T}_B(\psi_B)\right] \tag{7.11.198}$$

$$= \log_2 \|\mathrm{T}_B(\psi_B)\|_{\frac{1}{2}} = \log_2 \|\psi_B\|_{\frac{1}{2}} \tag{7.11.199}$$

$$= \log_2 \|\psi_A\|_{\frac{1}{2}} = H_{\frac{1}{2}}(A)_\psi. \tag{7.11.200}$$

This concludes the proof. ■

7.11.2 Remarks on Defining Channel Quantities from State Quantities

Observe that all of the generalized information measures for quantum channels given in Definition 7.85, as well as all of the channel information measures given above for specific generalized divergences, are defined in a common manner. Specifically, given a function $f : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ for bipartite states, the corresponding function f for quantum channels[4] is defined as

$$f(\mathcal{N}) \coloneqq \sup_{\rho_{RA}} f(R; B)_{\omega}, \tag{7.11.201}$$

where $\mathcal{N}_{A\to B}$ is a quantum channel, ρ_{RA} is a quantum state, with the dimension of R unbounded, and $\omega_{RB} = \mathcal{N}_{A\to B}(\rho_{RA})$. In other words, we define the channel quantity by taking a quantum state ρ_{RA} of a bipartite system consisting of the input system A of the channel and a reference system R (whose dimension is in general unbounded), passing A through the channel, then evaluating the state quantity on the output state $\mathcal{N}_{A\to B}(\rho_{RA})$. We then optimize over all states ρ_{RA}.

A similar principle as in (7.11.201) has been used in Definition 7.81 for the generalized divergence between two quantum channels. In particular, if we have a function $f : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ on two quantum states,[5] then we define a corresponding quantity f for two quantum channels as

$$f(\mathcal{N}, \mathcal{M}) \coloneqq \sup_{\rho_{RA}} f(\mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA})), \tag{7.11.202}$$

where $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ are quantum channels and ρ_{RA} is a quantum state, with the dimension of R unbounded. In other words, we define the channel quantity by evaluating the state quantity on the states $\mathcal{N}_{A\to B}(\rho_{RA})$ and $\mathcal{M}_{A\to B}(\rho_{RA})$ and optimizing over all states ρ_{RA}. We have already seen this principle being used in Chapter 6 to define the diamond distance (Definition 6.18) and fidelity (Definition 6.23) of two quantum channels, in which case the state quantity f is the trace distance or fidelity.

In both (7.11.201) and (7.11.202), properties of the underlying state quantity (namely, the data-processing inequality), as well as the Schmidt decomposition

[4] In a slight abuse of notation, we use the same letter to denote the channel quantity as the state quantity.

[5] We allow for the second argument to be a positive semi-definite operator more generally.

theorem, allow us to restrict the optimizations in (7.11.201) and (7.11.202) to pure states $|\psi\rangle_{RA}$ without loss of generality, with the dimension of R equal to the dimension of A. If the underlying state quantity f in (7.11.201) is invariant with respect to local unitaries, then we can use this simplification to write $f(\mathcal{N})$ as

$$f(\mathcal{N}) = \sup_{\rho_A} f(\rho_A, \mathcal{N}_{A\to B}), \tag{7.11.203}$$

where the optimization is now only over states ρ_A for the input system A for the channel $\mathcal{N}$, and

$$f(\rho_A, \mathcal{N}_{A\to B}) \coloneqq f(A;B)_\omega, \quad \omega_{AB} = \sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}. \tag{7.11.204}$$

This holds due to (2.2.40), which states for every purification $|\psi^\rho\rangle_{AA'}$ of ρ_A there exists an operator Y_A such that $(Y_A \otimes \mathbb{1}_{A'})|\Gamma\rangle_{AA'} = |\psi^\rho\rangle_{AA'}$. Then, by the polar decomposition (Theorem 2.3), and the fact that $Y_A Y_A^\dagger = \rho_A$, it holds that $Y_A = U_A\sqrt{\rho_A}$ for some unitary U_A. Finally, using the definition of the Choi representation $\Gamma^{\mathcal{N}}_{AB}$ and the unitary invariance of f, we obtain (7.11.203). This equivalent formulation of the channel quantity $f(\mathcal{N})$ has been used in (7.11.113) for the coherent information of a channel.

If the underlying state quantity in (7.11.202) is unitarily invariant, then by using the same reasoning as above we can write $f(\mathcal{N},\mathcal{M})$ in a form analogous to (7.11.203):

$$f(\mathcal{N},\mathcal{M}) = \sup_{\rho_A} f(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}, \sqrt{\rho_A}\Gamma^{\mathcal{M}}_{AB}\sqrt{\rho_A}). \tag{7.11.205}$$

7.12 Summary

In this chapter, we studied various entropic quantities, starting with quantum relative entropy. We proved many of its most important properties, and we saw that it acts as a parent quantity for well-known quantities such as von Neumann entropy, quantum conditional entropy, quantum mutual information and conditional mutual information, and coherent information. We then studied the Petz–Rényi, sandwiched Rényi, geometric Rényi, and hypothesis testing relative entropies, and we proved many of their most important properties.

The unifying concept of this chapter is that of generalized divergence. A generalized divergence is a function $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R}$ that satisifes the data-processing inequality: for every state ρ, positive semi-definite operator σ, and quantum channel $\mathcal{N}$,

$$\boldsymbol{D}(\rho\|\sigma) \geq \boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{7.12.1}$$

This inequality holds for all of the quantum relative entropies that we considered in this chapter, and many of their important properties (such as joint convexity) can be derived using it. The data-processing inequality is a core concept in information theory, and it underlies virtually all of the results that we present in this book.

At the end of the chapter, we defined information measures for quantum channels. Given a state information measure (or, more generally, a generalized divergence), we define an information measure for channels in a manner analogous to the way that the diamond norm (a generalized divergence for channels) is defined from the trace norm (a generalized divergence for states): we send one share of a bipartite pure state through the channel, evaluate the state measure, and then optimize over all input states. All of the upper and lower bounds on communication capacities that we present in this book are given in terms of channel information measures defined in this way.

7.13 Bibliographic Notes

The von Neumann quantum entropy was originally defined by von Neumann (1927b). The quantum conditional entropy was considered implicitly by Lieb and Ruskai (1973a,b), proposed as a quantum-information theoretic quantity of interest by Cerf and Adami (1997), and the coherent information thereafter by Schumacher and Nielsen (1996). The modern definition of the quantum mutual information was proposed by Stratonovich (1965). Its non-negativity was proved by Molière and Delbrück (1935); Lanford and Robinson (1968). The quantum conditional mutual information was considered implicitly by Lanford and Robinson (1968); Lieb and Ruskai (1973a,b) and was proposed as a quantum information-theoretic quantity of interest by Cerf and Adami (1997, 1998). Chain rules for quantum conditional entropy and information were employed by Cerf and Adami (1998).

The definition of the quantum relative entropy as presented in Definition 7.1 is due to Umegaki (1962). It took many years after this until the paper by Hiai and Petz (1991) was published, which solidified the operational interpretation of the "Umegaki quantum relative entropy" in terms of quantum hypothesis testing and the quantum Stein's lemma. The strong converse for the quantum Stein's lemma was established by Ogawa and Nagaoka (2005).

The non-negativity of quantum relative entropy follows as a consequence of an inequality by Klein (1931), and its data-processing inequality for quantum channels was established by Lindblad (1975). The data-processing inequality for the quantum relative entropy under partial trace was established by Lieb and Ruskai (1973a), from which its joint convexity follows. The expression in (7.2.100) for the quantum conditional mutual information was given implicitly by Lieb and Ruskai (1973a). Strong subadditivity of quantum entropy (or equivalently, non-negativity of quantum conditional mutual information) was established by Lieb and Ruskai (1973a,b). The data-processing inequality for the quantum conditional mutual information under local channels was established by Christandl and Winter (2004). The uniform continuity bound for quantum conditional mutual information, as presented in Proposition 7.10, was established by Shirokov (2017).

The notion of generalized divergence in the classical case was proposed by Polyanskiy and Verdú (2010), and in the quantum case by Sharma and Warsi (2013). Proposition 7.16 was established by Wilde et al. (2014); Tomamichel et al. (2017). The generalized mutual information and conditional quantum entropy were proposed by Sharma and Warsi (2013).

The Petz–Rényi relative entropy was proposed by Petz (1985, 1986a), wherein its data-processing inequality was established for the case of partial trace. The general definition of Petz–Rényi relative entropy incorporating support conditions and its data-processing inequality for general channels was established by Tomamichel et al. (2009), along with several of its other properties such as monotonicity in α. The expression in (7.4.23) for the Petz–Rényi relative entropy was presented by Tomamichel et al. (2009); Sharma (2012).

The sandwiched Rényi relative entropy was independently proposed by Müller-Lennert et al. (2013) and Wilde et al. (2014), and the alternative expression in (7.5.5) was given by Dupuis and Wilde (2016). The variational expression in (7.5.6), as well as Proposition 7.29, are due to Müller-Lennert et al. (2013). Proposition 7.30 was established by Müller-Lennert et al. (2013); Wilde et al. (2014). The fact that sandwiched Rényi relative entropy is monotone with respect to α is due to Müller-Lennert et al. (2013), with an independent proof for $\alpha > 1$ by Beigi (2013). The inequality in (7.5.45) for $\alpha > 1$ is due to Wilde et al. (2014), as a direct consequence of the inequality by Lieb and Thirring (1976). The inequality in (7.5.45) for $\alpha \in (0, 1)$ is due to Datta and Leditzky (2014), as a direct consequence of the Araki–Lieb–Thirring inequalities by Lieb and Thirring (1976); Araki (1990).

The "reverse" Araki–Lieb–Thirring inequality, which leads to the inequality in (7.5.46), was proved by Iten et al. (2017). The data-processing inequality for the sandwiched Rényi relative entropy was established in a number of papers for various parameter ranges of α: Müller-Lennert et al. (2013); Wilde et al. (2014); Frank and Lieb (2013); Beigi (2013); Mosonyi and Ogawa (2015), being established for the full range $\alpha \in [1/2, \infty]$ by Frank and Lieb (2013). The proof that we presented here is due to Wilde (2018b). Counterexamples to data processing for the sandwiched Rényi relative entropy in the range $\alpha \in (0, 1/2)$ were given by Berta et al. (2017).

The geometric Rényi relative entropy has its roots in work of Petz and Ruskai (1998), and it was further developed by Matsumoto (2013, 2018). See also (Tomamichel, 2015; Hiai and Mosonyi, 2017) for other expositions. It was given the name "geometric Rényi relative entropy" by Fang and Fawzi (2021) because it is a function of the matrix geometric mean of its arguments. See, e.g., Lawson and Lim (2001) for a review of matrix geometric means. Proposition 7.40 was established by Katariya and Wilde (2021), with roots in the earlier work of Matsumoto (2014a,b). In particular, the expression $\mathrm{Tr}[\sigma(\sigma^{-1/2}\tilde{\rho}\sigma^{-1/2})^{\alpha}]$ for $\alpha = 1/2$ and $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$ was identified in Matsumoto (2014a, Section 3) and later generalized to all $\alpha \in (0, 1)$ in Matsumoto (2014b, Section 2). Lemma 7.41 and Proposition 7.43 were also established by Katariya and Wilde (2021), along with monotoniticity in α for all $\alpha \in (0, 1) \cup (1, \infty)$ (in Proposition 7.44). The inequality in Proposition 7.42 was established for the interval $\alpha \in (0, 1) \cup (1, 2]$ in Tomamichel (2015) (by making use of a general result in Matsumoto (2013, 2018)) and for the full interval $\alpha \in (1, \infty)$ in Wang et al. (2019a). Here, we have followed the approach of Wang et al. (2019a) and offered a unified proof in terms of the Araki–Lieb–Thirring inequality Araki (1990); Lieb and Thirring (1976). The first inequality in Proposition 7.49 was established for $\alpha \in (1, 2]$ in Wilde et al. (2014) and for $\alpha \in (0, 1)$ in Datta and Leditzky (2014), by employing the Araki–Lieb–Thirring inequality Araki (1990); Lieb and Thirring (1976). The second inequality was established by Matsumoto (2013, 2018) and reviewed by Tomamichel (2015). Data processing was established by an operator-theoretic approach in Petz and Ruskai (1998) and by an operational method in Matsumoto (2013, 2018). The operator-theoretic approach taken here has its roots in Hiai and Petz (1991, Proposition 2.5) and was reviewed in Hiai and Mosonyi (2017, Corollary 3.31). The interpretation of geometric Rényi relative entropy given in Proposition 7.48 was discovered by Matsumoto (2013, 2018). Lemma 7.50 was presented by Katariya and Wilde (2021) and is based on Zhou and Jiang (2019, Lemma 3).

Belavkin and Staszewski (1982) discovered the quantum generalization of the classical relative entropy given in Section 7.7, now known as the Belavkin–Staszewski relative entropy. Matsumoto (2013, 2018) showed that the Belavkin–Staszewski relative entropy is the limit of the geometric Rényi relative entropy as $\alpha \to 1$. The proof given here was presented by Katariya and Wilde (2021). Hiai and Petz (1991) found the inequality in Proposition 7.53 relating the quantum relative entropy to the Belavkin–Staszewski relative entropy (the proof given here is due to Katariya and Wilde (2021)). Hiai and Petz (1991) established the data-processing inequality for the Belavkin–Staszewski relative entropy, by a method different from that given here. Matsumoto (2013, 2018) found the interpretation of the Belavkin–Staszewski relative entropy given in Proposition 7.57.

The max-relative entropy was proposed by Datta (2009b) as a quantum information-theoretic quantity of interest. Datta also established many basic information processing properties of the max-relative entropy and studied its role in quantum hypothesis testing and entanglement theory Datta (2009b,a). Proposition 7.61 is due to Müller-Lennert et al. (2013); Katariya and Wilde (2021). Proposition 7.64 is due to Wang and Wilde (2019) and Anshu et al. (2019).

The conditional min-entropy was defined by Renner (2005), as well as the smooth conditional min-entropy. The operational interpretations of the conditional min- and max-entropies were examined by Koenig et al. (2009).

The hypothesis testing relative entropy was studied implicitly by a variety of authors for a long time in the context of quantum hypothesis testing: Hiai and Petz (1991); Ogawa and Nagaoka (2005); Hayashi and Nagaoka (2003); Nagaoka (2006); Hayashi (2007). It was proposed as a quantum information-theoretic quantity of interest by Buscemi and Datta (2010a) (in the context of operator smoothing of a one-shot entropic quantity), and given the name hypothesis testing relative entropy by Wang and Renner (2012), wherein its connection to classical communication was explored (see also Hayashi and Nagaoka (2003)). An alternate proof of the data-processing inequality for quantum relative entropy in terms of hypothesis testing was given by Bjelakovic and Siegmund-Schultze (2003). Various properties of the hypothesis testing relative entropy were established by Dupuis et al. (2013) and Datta et al. (2016), including the fact that it can be written as an SDP (see also Wang and Wilde (2019) in this context). Eq. (7.9.36) was established by Wang and Wilde (2019). Proposition 7.67 was considered by Helstrom (1976) and discussed more recently by Vazquez-Vilar (2016). A special case of Proposition 7.70 was established by Wang and Renner (2012);

Matthews and Wehner (2014). Proposition 7.71 was established by Cooney et al. (2016). Proposition 7.72 is essentially due to Hayashi (2007), with a refinement by Audenaert et al. (2012) and a later rediscovery of it, formulated in a different way, by Qi et al. (2018b). Proposition 7.80 is essentially due to Hayashi (2007).

The generalized channel divergence of Definition 7.81 was proposed by Leditzky et al. (2018), and Proposition 7.84 was established as well by Leditzky et al. (2018). The various generalized channel information measures can be found in the papers of Wilde et al. (2014); Gupta and Wilde (2015), and the related channel information measures based on hypothesis testing, Petz–Rényi, and sandwiched Rényi relative entropy are from Koenig and Wehner (2009); Sharma and Warsi (2013); Wilde et al. (2014); Gupta and Wilde (2015); Datta et al. (2016). The channel mutual information was defined by Adami and Cerf (1997), the channel Holevo information by Schumacher and Westmoreland (1997) (based on the Holevo quantity for ensembles Holevo (1973)), and the channel coherent information by Lloyd (1997). These papers together thus developed a general concept of promoting a measure of correlations in a quantum state to a measure of a channel's ability to create the same correlations, by optimizing the state measure with respect to a (subset of) all of the states that can be generated by means of the channel.

The review by Ruskai (2002) is helpful not only for understanding entropy inequalities in quantum information, but also for understanding the history of developments with respect to quantum entropy and information. The book of Tomamichel (2015) provides an exposition of Rényi relative entropies and their properties (see also Leditzky (2016)). The book of Wilde (2017a) provides an overview of entropies in the von Neumann family, their properties, and the derived channel information measures.

Chapter 8

Information Measures for Quantum Channels

Chapter 9

Entanglement Measures

In the previous chapter, we laid the foundation for analyzing quantum communication protocols by defining entropic quantities, such as the Petz– and sandwiched Rényi relative entropies, as well as information measures for quantum states and channels derived from these relative entropies. We now use these information measures to define entanglement measures for quantum states and channels. Given quantum systems A and B, an entanglement measure is a function $E : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ that quantifies the amount of entanglement present in a state ρ_{AB} of these systems. The notion of "quantifying entanglement" is explained in Section 9.1 below, with the defining requirement of an entanglement measure being that it does not increase under channels realized by local operations and classical communication (Definition 4.22). We can think of this requirement of "LOCC monotonicity" as a restricted form of the data-processing inequality, but now applied to a single bipartite state rather than to a pair of states. The data-processing inequality indicates that the distinguishability of two states does not increase under the action of the same quantum channel (Definition 7.15), whereas LOCC monotonicity indicates that entanglement does not increase under the action of an LOCC channel on a bipartite state.

Given an entanglement measure E for states, the corresponding entanglement measure for channels is defined using the general principle in Section 7.11.2, which is to optimize the state measure with respect to all bipartite states that can be shared between the sender and receiver of the channel by making use of the channel. We develop entanglement measures for channels in Chapter 10, and these naturally quantify how much entanglement can be generated by a channel connecting a sender to a receiver.

Entanglement measures feature prominently in the analysis of optimal rates for distillation and communication protocols. In particular, entanglement measures for states arise as upper bounds on their distillable entanglement and secret key, which we examine in Chapters 13 and 15, respectively. Entanglement measures for quantum channels arise as upper bounds on the rates of quantum and private communication over a quantum channel, which we consider in Chapters 14 and 16, respectively, as well as for their feedback-assisted counterparts that we consider in Part III of this book.

Being a uniquely quantum-mechanical property, it is perhaps not surprising that entanglement features prominently in quantum communication protocols, both in the encoding and decoding of messages, as well as in the analysis of their optimal rates. In fact, any state that is not entangled (i.e., separable) is useless for entanglement and secret key distillation, and similarly, entanglement breaking channels are useless for quantum and private communication. The distinguishability of a given state from the set of separable states, which we show in this chapter is an entanglement measure for states, can thus give an indication of how much entanglement or secret key can be distilled from it. Similarly, for communication tasks over quantum channels, the distinguishability of a given quantum channel from the set of entanglement breaking channels, which we show in this chapter is an entanglement measure for channels, can be used to determine how good the channel is for quantum or private communication.

The rest of this chapter proceeds as follows. We start in Section 9.1 by formally defining what it means for a function E to be an entanglement measure for bipartite states and by providing examples of entanglement measures. In Section 9.2, we consider entanglement measures that quantify the distinguishability of a given state ρ_{AB} from the set $\text{SEP}(A\!:\!B)$ of separable states, with the measure given by some generalized divergence $\boldsymbol{D}$ (see Definition 7.15). We also consider in Section 9.3 a class of entanglement measures based on the distinguishability of a given state from the larger set $\text{PPT}' \supset \text{SEP}$. In Section 9.4, we consider a different kind of entanglement measure for states called squashed entanglement.

9.1 Definition and Basic Properties

Recall from Definition 3.5 that a bipartite state ρ_{AB} is called entangled if it is not separable, meaning that it *cannot* be written in the following form

$$\sum_{x\in\mathcal{X}} p(x)\tau_A^x \otimes \omega_B^x, \tag{9.1.1}$$

for some finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0,1]$, and sets $\{\tau_A^x\}_{x\in\mathcal{X}}$ and $\{\omega_B^x\}_{x\in\mathcal{X}}$ of states. Determining whether a given quantum state ρ_{AB} is entangled is a fundamental problem in quantum information theory. In Section 3.2.1, in the discussion after Definition 3.5, we listed the following criteria for the entanglement of pure and mixed states:

- *Schmidt rank criterion*: A pure bipartite state ψ_{AB} is entangled if and only if its Schmidt rank is strictly greater than one.

- *PPT criterion*: If a bipartite mixed state ρ_{AB} has negative partial transpose (i.e., the partial transpose $\rho_{AB}^{\mathsf{T}_B}$ has at least one negative eigenvalue), then it is entangled. If both systems A and B are qubit systems or if one of the systems is a qubit and the other a qutrit, then ρ_{AB} is entangled if and only if $\rho_{AB}^{\mathsf{T}_B}$ has negative partial transpose.

In the case of mixed states, there is generally not a simple necessary and sufficient criterion to determine whether a given bipartite state is entangled, and in fact it is known that it is computationally difficult, in a precise sense, to decide if a state is entangled (please consult the Bibliographic Notes in Section 9.6).

In addition to determining whether or not a given quantum state is entangled, we are interested in quantifying the amount of entanglement present in a quantum state. Doing so allows us to compare quantum states based on the amount of entanglement present in them. An *entanglement measure* is a function $E : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ from the set of density operators acting on the Hilbert space of a bipartite system to the set of real numbers, and it quantifies the entanglement of a quantum state $\rho_{AB} \in \mathrm{D}(\mathcal{H}_{AB})$. (The formal definition of an entanglement measure is given in Definition 9.1.) To indicate the partitioning of the subsystems explicitly, we often write $E(A;B)_\rho$ instead of $E(\rho_{AB})$.

How exactly do we quantify entanglement? Suppose that we have a bipartite state ρ_{AB} and we would like to quantify the entanglement between the systems

A and B. One fundamental observation is that the entanglement of ρ_{AB} cannot increase under the action of a local operations and classical communication (LOCC) channel (recall Definition 4.22). This is intuitive because entanglement is a non-local property of a bipartite state, and so local operations alone do not increase it. Similarly, classical communication should only affect the *classical* correlations between the two systems A and B, and not the quantum correlations, i.e., the entanglement.

Given the reasoning above, the defining property of an entanglement measure $E : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ for the quantum systems A and B is that it does not increase under the action of an LOCC channel:

Definition 9.1 Entanglement Measure

We say that $E(A;B)_\rho$ is an entanglement measure for a bipartite state ρ_{AB} if the following inequality holds for every bipartite state ρ_{AB} and every LOCC channel $\mathcal{L}_{AB\to A'B'}$ that acts on ρ_{AB}:

$$E(A;B)_\rho \geq E(A';B')_\omega, \tag{9.1.2}$$

where $\omega_{A'B'} := \mathcal{L}_{AB\to A'B'}(\rho_{AB})$.

This property of LOCC monotonicity is the most operational property of an entanglement measure, in the sense that it relates directly to the distillation or communication tasks that we consider in this book, which involve LOCC between the two parties sharing a state ρ_{AB} or between sender and receiver at the terminals of a quantum channel, respectively. It is also analogous to the data-processing inequality for generalized divergences.

The defining requirement of LOCC monotonicity implies that an entanglement measure E takes on its minimum value on the set of separable states. To see this, recall from the discussion after Definition 3.5 that a separable state can be prepared by LOCC. Thus, starting from an arbitrary state ρ_{AB}, Alice and Bob can trace out their local systems A and B and perform LOCC to prepare a separable state $\sigma_{A'B'}$. The serial concatenation of these two actions is itself an LOCC channel. Thus, it follows from the definition above that

$$E(A;B)_\rho \geq E(A';B')_\sigma \tag{9.1.3}$$

for every separable state $\sigma_{A'B'}$. Now, given another separable state $\sigma'_{A''B''}$, it is possible to transform between $\sigma_{A'B'}$ and $\sigma'_{A''B''}$ using LOCC, meaning that $E(A';B')_\sigma \geq E(A'';B'')_{\sigma'}$ and $E(A';B')_\sigma \leq E(A'';B'')_{\sigma'}$. Therefore,

$$E(A';B')_\sigma = E(A'';B'')_{\sigma'}, \tag{9.1.4}$$

for all separable states σ_{AB} and $\sigma'_{A'B'}$. As a consequence, an entanglement measure E takes on its minimum value and is equal to a constant $c \in \mathbb{R}$ for all separable states. It is often convenient and simpler if an entanglement measure E is equal to zero for all separable states. If this is not the case, then we can simplify redefine the entanglement measure as $E'(A;B)_\rho = E(A;B)_\rho - c$. By this reasoning and adjustment (if needed), every entanglement measure (as per Definition 9.1) satisfies the following two properties of non-negativity on all states and vanishing on separable states:

1. *Non-negativity*: $E(\rho_{AB}) \geq 0$ for every state ρ_{AB}.

2. *Vanishing for separable states*: $E(\sigma_{AB}) = 0$ for every separable state σ_{AB}.

Other properties that are desirable for an entanglement measure E are as follows:

1. *Faithfulness*: $E(\sigma_{AB}) = 0$ if and only if σ_{AB} is separable, so that $E(\rho_{AB}) > 0$ if and only if ρ_{AB} is entangled.

2. *Invariance under classical communication*: For every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0,1]$ on $\mathcal{X}$, and set $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ of states, define the following classical–quantum state:

$$\rho_{XAB} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}. \tag{9.1.5}$$

Then the entanglement measure E satisfies invariance under classical communication if

$$E(XA;B)_\rho = E(A;BX)_\rho = \sum_{x\in\mathcal{X}} p(x)E(A;B)_{\rho^x}. \tag{9.1.6}$$

This property has the interpretation of invariance under classical communication because the equality $E(XA;B)_\rho = E(A;BX)_\rho$ indicates that the classical value x in register X can be communicated classically to Bob and discarded locally, and the entanglement measure does not change under this action. Furthermore, the value of the entanglement is simply the expected entanglement, where the expectation is calculated with respect to the probability distribution $p(x)$.

This property is also known as the “flags” property in the research literature.

3. *Convexity*: For every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0, 1]$ on $\mathcal{X}$, and set $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ of states,

$$\sum_{x\in\mathcal{X}} p(x)E(\rho^x_{AB}) \geq E\left(\sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}\right). \tag{9.1.7}$$

Convexity is an intuitive property of entanglement, and for entanglement measures that are invariant under classical communication (obeying (9.1.6)), it captures the idea that entanglement should not increase on average if classical information about the identity of a state is lost. (In fact, convexity is an immediate consequence of LOCC monotonicity and invariance under classical communication.)

4. *Additivity*: The entanglement of a tensor-product state $\rho_{A_1A_2B_1B_2} = \tau_{A_1B_1} \otimes \omega_{A_2B_2}$ is the sum of the entanglement of the individual states in the tensor product:

$$E(A_1A_2; B_1B_2)_{\tau\otimes\omega} = E(A_1; B_1)_\tau + E(A_2; B_2)_\omega. \tag{9.1.8}$$

If instead we have only that

$$E(A_1A_2; B_1B_2)_{\tau\otimes\omega} \leq E(A_1; B_1)_\tau + E(A_2; B_2)_\omega \tag{9.1.9}$$

for all states $\tau_{A_1B_1}$ and $\omega_{A_2B_2}$, then the entanglement measure E is *subadditive*.

5. *Selective LOCC monotonicity*: A property stronger than LOCC monotonicity is that E is non-increasing on average under an *LOCC instrument*. In more detail, let ρ_{AB} be a bipartite state, and let $\{\mathcal{L}^x_{AB\to A'B'}\}_{x\in\mathcal{X}}$ be a collection of maps, such that $\mathcal{L}^{\leftrightarrow}_{AB\to A'B'}$ is an LOCC channel of the form:

$$\mathcal{L}^{\leftrightarrow}_{AB\to A'B'} = \sum_{x\in\mathcal{X}} \mathcal{L}^x_{AB\to A'B'}, \tag{9.1.10}$$

for some finite alphabet $\mathcal{X}$ and where each map $\mathcal{L}^x_{AB\to A'B'}$ is completely positive such that the sum map $\mathcal{L}^{\leftrightarrow}_{AB\to A'B'}$ is trace preserving (i.e., a quantum channel). Furthermore, each map $\mathcal{L}^x_{AB\to A'B'}$ can be written in the form of (4.6.52), as follows:

$$\mathcal{L}^x_{AB\to A'B'} = \sum_{y\in\mathcal{Y}} \mathcal{E}^{x,y}_{A\to A'} \otimes \mathcal{F}^{x,y}_{B\to B'}, \tag{9.1.11}$$

where $\{\mathcal{E}^{x,y}_{A\to A'}\}_{x\in\mathcal{X}}$ and $\{\mathcal{F}^{x,y}_{B\to B'}\}_{x\in\mathcal{X}}$ are sets of completely positive maps. Set

$$p(x) := \mathrm{Tr}[\mathcal{L}^x_{AB\to A'B'}(\rho_{AB})], \tag{9.1.12}$$

and for $x \in \mathcal{X}$ such that $p(x) \neq 0$, set

$$\omega_{AB}^{x} := \frac{1}{p(x)} \mathcal{L}_{AB \to A'B'}^{x}(\rho_{AB}). \tag{9.1.13}$$

If the classical value of x is not discarded, then the given state ρ_{AB} is transformed to the ensemble $\{(p(x), \omega_{AB}^{x})\}_{x \in \mathcal{X}}$ via LOCC.

The entanglement measure E satisfies *selective LOCC monotonicity* if

$$E(\rho_{AB}) \geq \sum_{x \in \mathcal{X} : p(x) \neq 0} p(x) E(\omega_{AB}^{x}), \tag{9.1.14}$$

for every ensemble $\{(p(x), \omega_{AB}^{x})\}_{x \in \mathcal{X}}$ that arises from ρ_{AB} via LOCC as specified above. Selective LOCC monotonicity indicates that entanglement does not increase on average under the action of LOCC. Many entanglement measures satisfy this stronger property.

Observe that selective LOCC monotonicity in (9.1.14) implies LOCC monotonicity in (9.1.2), simply because the alphabet $\mathcal{X}$ in (9.1.10) can consist of only one letter.

The entanglement measures that we consider in this chapter satisfy many of the properties listed above.

Given that we would like to quantify entanglement, it makes sense to ask what the basic *unit* of entanglement should be. We take as our unit of entanglement the two-qubit maximally entangled Bell state $|\Phi\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$, and we thus say that the state $|\Phi\rangle$ represents "one ebit." A maximally entangled state of Schmidt rank d is then referred to as having "$\log_2 d$ ebits." All of the entanglement measures that we consider in this chapter are equal to one for a two-qubit maximally entangled state, which is another justification for using it as the unit of entanglement.[1] Similarly, for a maximally entangled state of Schmidt rank d, all of the entanglement measures that we consider in this chapter are equal to $\log_2 d$.

To close out this introductory section, we prove a lemma that helps to reduce the difficulty in determining whether a given function is an entanglement measure.

[1]This "normalization" condition is sometimes taken to be a requirement for an entanglement measure.

Lemma 9.2

Let $E : \mathcal{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ be a function that, for every bipartite state ρ_{AB}, is

1. invariant under classical communication, as defined in (9.1.6), and

2. obeys data processing under local channels, in the sense that

$$E(A;B)_\rho \geq E(A';B')_\omega, \tag{9.1.15}$$

for all channels $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$, where

$$\omega_{A'B'} := (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{AB}). \tag{9.1.16}$$

Then E is convex, as defined in (9.1.7), and a selective LOCC monotone, as defined in (9.1.14).

Proof: We first prove convexity. Let

$$\rho_{XAB} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}, \tag{9.1.17}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ is a set of states. Then

$$\sum_{x\in\mathcal{X}} p(x)E(A;B)_{\rho^x} = E(XA;B)_\rho \tag{9.1.18}$$

$$\geq E(A;B)_\rho, \tag{9.1.19}$$

where the entanglement $E(A;B)$ in the last line is evaluated with respect to the reduced state $\rho_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}$. The equality follows from the assumption of invariance under classical communication, as defined in (9.1.6), and the inequality follows because the partial trace channel Tr_X is a local channel that discards the classical system X.

To establish selective LOCC monotonicity, consider that an LOCC channel $\mathcal{L}^{\leftrightarrow}_{AB\to A'B'}$ of the form in (9.1.10), by Definition 4.22, can be built up as a serial concatenation of one-way LOCC channels. Furthermore, each one-way LOCC channel can keep some classical information and discard some of it. It is helpful conceptually to think of the retained classical information as part of the variable x in (9.1.10) and the discarded classical information as part of the variable y in

(9.1.11). In more detail, each such one-way LOCC channel (from Alice to Bob in the case discussed below) has the following form:

$$\sum_{k,\ell} \mathcal{E}^{k,\ell}_{A\to A'} \otimes \mathcal{F}^{k,\ell}_{B\to B'}, \tag{9.1.20}$$

where $\{\mathcal{E}^{k,\ell}_{A}\}_{k,\ell}$ is a collection of completely positive maps such that the sum map $\sum_{k,\ell} \mathcal{E}^{k,\ell}_{A}$ is trace preserving, and $\{\mathcal{F}^{k,\ell}_{B}\}_{k,\ell}$ is a collection of quantum channels. For now and for simplicity, let use the superindex $m \coloneqq (k,\ell)$. We should think of the classical information k as that which is being *kept* and that in ℓ as that which is being *lost* or *let go*. The one-way LOCC channel in (9.1.20) can be implemented using the following steps:

1. Alice applies the following local quantum channel:

$$\tau_{AB} \mapsto \sum_{m} \mathcal{E}^{m}_{A\to A'}(\tau_{AB}) \otimes |m\rangle\langle m|_{M_A}. \tag{9.1.21}$$

2. Alice employs a classical communication channel

$$(\cdot)_{M_A} \mapsto \sum_{m} |m\rangle_{M_B}\langle m|_{M_A}(\cdot)_{M_A}|m\rangle_{M_A}\langle m|_{M_B} \tag{9.1.22}$$

to communicate the value in M_A to Bob:

$$\sum_{m} \mathcal{E}^{m}_{A\to A'}(\tau_{AB}) \otimes |m\rangle\langle m|_{M_A} \mapsto \sum_{m} \mathcal{E}^{m}_{A\to A'}(\tau_{AB}) \otimes |m\rangle\langle m|_{M_B}. \tag{9.1.23}$$

3. Bob performs the local channel

$$(\cdot)_{BM_B} \mapsto \sum_{m} \mathcal{F}^{m}_{B\to B'}(\cdot) \otimes |m\rangle\langle m|_{M_B}(\cdot)_{M_B}|m\rangle\langle m|_{M_B}, \tag{9.1.24}$$

which can be understood as "looking in the classical register M_B" to determine the value m and performing the local quantum channel $\mathcal{F}^{m}_{B\to B'}$ based on the value m found. Under this local channel, the global state becomes as follows:

$$\sum_{m} \mathcal{E}^{m}_{A\to A'}(\tau_{AB}) \otimes |m\rangle\langle m|_{M_B} \mapsto \sum_{m} (\mathcal{E}^{m}_{A\to A'} \otimes \mathcal{F}^{m}_{B\to B'})(\tau_{AB}) \otimes |m\rangle\langle m|_{M_B}. \tag{9.1.25}$$

4. Bob then discards the ℓ part of the classical information m, as follows:

$$\sum_m (\mathcal{E}^m_{A\to A'} \otimes \mathcal{F}^m_{B\to B'})(\tau_{AB}) \otimes |m\rangle\!\langle m|_{M_B}$$
$$= \sum_{k,\ell} (\mathcal{E}^{k,\ell}_{A\to A'} \otimes \mathcal{F}^{k,\ell}_{B\to B'})(\tau_{AB}) \otimes |k,\ell\rangle\!\langle k,\ell|_{K_B L_B} \tag{9.1.26}$$
$$\mapsto \sum_{k,\ell} (\mathcal{E}^{k,\ell}_{A\to A'} \otimes \mathcal{F}^{k,\ell}_{B\to B'})(\tau_{AB}) \otimes |k\rangle\!\langle k|_{K_B} \tag{9.1.27}$$
$$= \sum_k p(k)\omega^k_{A'B'} \otimes |k\rangle\!\langle k|_{K_B}, \tag{9.1.28}$$

where

$$p(k) := \mathrm{Tr}\left[\sum_\ell (\mathcal{E}^{k,\ell}_{A\to A'} \otimes \mathcal{F}^{k,\ell}_{B\to B'})(\tau_{AB})\right], \tag{9.1.29}$$
$$\omega^k_{A'B'} := \frac{1}{p(k)} \sum_\ell (\mathcal{E}^{k,\ell}_{A\to A'} \otimes \mathcal{F}^{k,\ell}_{B\to B'})(\tau_{AB}). \tag{9.1.30}$$

Bob could, if desired, finally discard the classical register K_B to implement the one-way LOCC channel in (9.1.20). However, it is helpful to hold on to it for our analysis below.

Now we analyze how the entanglement changes under each of these steps, omitting the state subscripts at each step except for the first and last lines, because those not shown are clear from the context:

$$E(A;B)_\tau \geq E(A'M_A;B) \tag{9.1.31}$$
$$= E(A';BM_B) \tag{9.1.32}$$
$$\geq E(A';B'M_B) \tag{9.1.33}$$
$$= E(A';B'K_B L_B) \tag{9.1.34}$$
$$\geq E(A';B'K_B) \tag{9.1.35}$$
$$= \sum_k p(k) E(A';B')_{\omega^k}. \tag{9.1.36}$$

The first inequality follows from data processing under the local channel in (9.1.21). The first equality follows from the assumption of invariance of classical communication, i.e., invariance under the action of the classical channel in (9.1.22). The second inequality follows from data processing under the local channel in

(9.1.25). The second equality is trivial, following because $M_B = (K_B, L_B)$ by definition. The third inequality follows again from data processing under the local channel in (9.1.27). The final equality follows again from invariance under classical communication.

Thus, we have shown selective one-way LOCC monotonicity (from Alice to Bob) in the following sense:

$$E(A;B)_\tau \geq \sum_k p(k) E(A';B')_{\omega^k}, \tag{9.1.37}$$

where the ensemble $\{(p(k), \omega^k_{A'B'})\}_k$ arises from the state τ_{AB} by means of one-way LOCC from Alice to Bob. By the same argument, but flipping the role of Alice and Bob, it follows that selective one-way LOCC monotonicity from Bob to Alice holds for the function E. Since every LOCC channel is built up as a serial concatenation of one-way LOCC channels and since we have proven that selective monotonicity holds for the function E for each of them, it follows that E obeys selective LOCC monotonicity. ■

9.1.1 Examples

Let us now consider some examples of entanglement measures. The first two entanglement measures that we consider are related to the Schmidt rank criterion and the PPT criterion, respectively, stated in Section 3.2.1 and reiterated at the beginning of this chapter. They are known as the entanglement of formation and the log-negativity, respectively, and are some of the simplest and earliest entanglement measures defined. They are also conceptually linked to more complex entanglement measures like squashed entanglement and the Rains relative entropy, the latter of which are the best known upper bounds on distillable entanglement (studied in Chapter 13).

9.1.1.1 Entanglement of Formation

Given a pure bipartite state $\psi_{AB} = |\psi\rangle\langle\psi|_{AB}$, there exists a Schmidt decomposition of $|\psi\rangle_{AB}$ such that

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k} |e_k\rangle_A \otimes |f_k\rangle_B, \tag{9.1.38}$$

where r is the Schmidt rank, $\lambda_k > 0$ are the Schmidt coefficients, and $\{|e_k\rangle_A\}_{k=1}^r$, $\{|f_k\rangle_B\}_{k=1}^r$ are orthonormal sets. Observe that the reduced states $\psi_A := \mathrm{Tr}_B[\psi_{AB}]$ and $\psi_B := \mathrm{Tr}_A[\psi_{AB}]$ have the same non-zero eigenvalues, which means that their entropies are equal, i.e., $H(\psi_A) = H(\psi_B)$. Furthermore, $H(\psi_A) = 0$ if and only if $r = 1$, and $r = 1$ if and only if ψ_{AB} is separable, by the Schmidt rank criterion. Therefore, the entropy of the reduced state of a pure bipartite state provides us with a signature of entanglement for pure bipartite states:

$$\psi_{AB} \text{ entangled} \qquad \Longleftrightarrow \qquad H(\psi_A) > 0. \tag{9.1.39}$$

We let

$$E_F(\psi_{AB}) := H(\mathrm{Tr}_B[\psi_{AB}]) = -\sum_{k=1}^{r} \lambda_k \log_2 \lambda_k \tag{9.1.40}$$

for every pure state ψ_{AB}.

The function E_F is an entanglement measure, as proven in Proposition 9.3 below. When evaluated on pure bipartites as above, it is known as the *entropy of entanglement* or *entanglement entropy*, and it is often simply denoted by $E(\psi_{AB})$ in the research literature. By (9.1.39), it is also a faithful entanglement measure on pure states, i.e., $E_F(\psi_{AB}) = 0$ if and only if ψ_{AB} is a separable state. Recall from Section 3.2.3 that a maximally entangled state is defined by having $\lambda_k = \frac{1}{r}$ for all $1 \leq k \leq r$. For such states, $E_F(\psi_{AB}) = \log_2 r$, which justifies calling them maximally entangled because $\log_2 r$ is the largest value of the quantum entropy for states supported on an r-dimensional space.

The definition in (9.1.40) for the entanglement measure E_F, so far, has been defined only for pure states. To extend the definition to mixed states, we use the fact that a mixed state ρ_{AB} can be decomposed into a convex combination of pure states as follows:

$$\rho_{AB} = \sum_{x \in \mathcal{X}} p(x) \psi_{AB}^x, \tag{9.1.41}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0, 1]$ is a probability distribution, and $\{\psi_{AB}^x\}_{x \in \mathcal{X}}$ is a set of pure states. We can then measure the entanglement of ρ_{AB} by taking the expected entanglement of the pure states involved in the decomposition of ρ_{AB}, i.e., by $\sum_{x \in \mathcal{X}} p(x) H(\psi_A^x)$, where $\psi_A^x := \mathrm{Tr}_B[\psi_{AB}^x]$. However, this strategy can lead to different values for the entanglement of ρ_{AB}, depending on the chosen decomposition, because the decomposition of mixed states into a convex combination of pure states is generally not unique. We can address this issue by minimizing over all possible decompositions, leading to the following definition:

Definition 9.3 Entanglement of Formation

The entanglement of formation of a bipartite state ρ_{AB} is defined as

$$E_F(\rho_{AB}) := \inf_{\{(p(x),\psi^x_{AB})\}_{x\in\mathcal{X}}} \left\{ \sum_{x\in\mathcal{X}} p(x) H(\psi^x_A) : \rho_{AB} = \sum_{x\in\mathcal{X}} p(x)\psi^x_{AB} \right\}, \tag{9.1.42}$$

where $\rho_{AB} = \sum_{x\in\mathcal{X}} p(x)\psi^x_{AB}$ is a pure-state decomposition of ρ_{AB}.

It suffices to take $|\mathcal{X}| \leq \dim(\mathcal{H}_{AB})^2$ in the optimization in (9.1.42), and furthermore, the infimum is achieved by at least one pure-state decomposition. The fact that the alphabet $\mathcal{X}$ need not exceed $\dim(\mathcal{H}_{AB})^2$ elements is due to the entropy being a continuous function and the Fenchel–Eggleston–Carathéodory Theorem (Theorem 2.23) and the fact that dimension of the space of density operators on a $\dim(\mathcal{H}_{AB})$-dimensional space is $\dim(\mathcal{H}_{AB})^2$. The fact that the infimum is achieved follows because the optimization is with respect to a compact space and the function being optimized is continuous.

For every pure bipartite state ψ_{AB}, the equality in (9.1.40) holds because it is not possible to decompose a pure state ψ_{AB} as a mixture of other pure states different from ψ_{AB}. As mentioned above, the entanglement of formation is also known as the *entropy of entanglement* for the special case of a pure bipartite state, due to the fact that it is equal to the entropy of the reduced state.

It is a direct consequence of Definition 9.3 and the non-negativity of quantum entropy in (7.2.107) that the entanglement of formation is non-negative, i.e.,

$$E_F(\rho_{AB}) \geq 0 \tag{9.1.43}$$

for every bipartite state ρ_{AB}.

Not only is the entanglement of formation non-negative, but it is also faithful; and we can make an even more refined statement about approximate faithfulness (when ρ_{AB} is close to separable or when $E_F(\rho_{AB})$ is close to zero).

Before establishing this statement in Proposition 9.5 below, we first prove a uniform continuity bound for the entanglement of formation. *Uniform continuity* is a desirable property of an entanglement measure: if two bipartite states ρ_{AB} and σ_{AB} are not very distinguishable from each other (according to some distinguishability measure), then the difference of their entanglement should be small. (See Section 2.3

for a definition of uniform continuity.)

Proposition 9.4 Uniform Continuity of Entanglement of Formation

Let ρ_{AB} and σ_{AB} be states satisfying

$$\frac{1}{2}\left\|\rho_{AB}-\sigma_{AB}\right\|_1 \leq \varepsilon, \tag{9.1.44}$$

where $\varepsilon \in [0,1]$. Then the entanglement of formations of ρ_{AB} and σ_{AB} satisfy

$$|E_F(\rho_{AB}) - E_F(\sigma_{AB})| \leq \delta \log_2 \min\{d_A, d_B\} + g_2(\delta), \tag{9.1.45}$$

where $\delta := \sqrt{\varepsilon(2-\varepsilon)}$ and $g_2(x) := (x+1)\log_2(x+1) - x\log_2 x$.

Proof: By applying Theorem 6.14 to (9.1.44), we find that

$$\sqrt{F}(\rho_{AB}, \sigma_{AB}) \geq 1-\varepsilon, \tag{9.1.46}$$

which implies that

$$\sqrt{1-F(\rho_{AB},\sigma_{AB})} \leq \sqrt{1-(1-\varepsilon)^2} = \sqrt{\varepsilon(2-\varepsilon)}. \tag{9.1.47}$$

Let

$$|\psi^\rho\rangle_{RAB} := \sum_x \sqrt{p(x)}|x\rangle_R|\psi^x\rangle_{AB}. \tag{9.1.48}$$

be a purification of ρ_{AB}, with $\rho_{AB} = \sum_x p(x)\psi^x_{AB}$ a pure-state decomposition of ρ_{AB}. By applying Uhlmann's theorem (Theorem 6.8), there exists a purification ψ^σ_{RAB} of σ_{AB} such that $F(\psi^\rho_{RAB}, \psi^\sigma_{RAB}) = F(\rho_{AB}, \sigma_{AB})$. By combining this observation with (9.1.47), and the fact that the sine distance of two pure states is equal to the normalized trace distance (see (6.1.1)), we conclude that

$$\frac{1}{2}\left\|\psi^\rho_{RAB} - \psi^\sigma_{RAB}\right\|_1 \leq \sqrt{\varepsilon(2-\varepsilon)}. \tag{9.1.49}$$

We now apply the measurement channel $\mathcal{M}_{R\to X}(\cdot) := \sum_x |x\rangle_X\langle x|_R(\cdot)|x\rangle_R\langle x|_X$ to the R systems, as well as the data-processing inequality for the trace distance, to conclude that

$$\frac{1}{2}\left\|\mathcal{M}_{R\to X}(\psi^\rho_{RAB}) - \mathcal{M}_{R\to X}(\psi^\sigma_{RAB})\right\|_1 \leq \sqrt{\varepsilon(2-\varepsilon)}. \tag{9.1.50}$$

Consider that

$$\rho_{XAB} := \mathcal{M}_{R\to X}(\psi^{\rho}_{RAB}) = \sum_x p(x)|x\rangle\langle x|_X \otimes \psi^x_{AB}, \tag{9.1.51}$$

and there exists a probability distribution $q(x)$ and a set $\{\varphi^x_{AB}\}_x$, satisfying

$$\sigma_{AB} = \sum_x q(x)\varphi^x_{AB}, \tag{9.1.52}$$

such that

$$\sigma_{XAB} := \mathcal{M}_{R\to X}(\psi^{\sigma}_{RAB}) = \sum_x q(x)|x\rangle\langle x|_X \otimes \varphi^x_{AB}. \tag{9.1.53}$$

Now applying the uniform continuity of conditional mutual information (Proposition 7.10), we conclude that

$$\frac{1}{2}I(A;B|X)_\sigma \leq \frac{1}{2}I(A;B|X)_\rho + \delta\log_2\min\{d_A,d_B\} + g_2(\delta), \tag{9.1.54}$$

with $\delta = \sqrt{\varepsilon(2-\varepsilon)}$. Since the states of systems AB are pure when conditioned on the classical system X, for both ρ_{XAB} and σ_{XAB}, consider that

$$\frac{1}{2}I(A;B|X)_\sigma = H(A|X)_\sigma, \qquad \frac{1}{2}I(A;B|X)_\rho = H(A|X)_\rho. \tag{9.1.55}$$

So we conclude that

$$E_F(\sigma_{AB}) \leq H(A|X)_\sigma \leq H(A|X)_\rho + \delta\log_2\min\{d_A,d_B\} + g_2(\delta), \tag{9.1.56}$$

where the first inequality follows from Definition 9.3 and (9.1.52). Since the pure-state decomposition of ρ_{AB} is arbitrary, we conclude that

$$E_F(\sigma_{AB}) \leq E_F(\rho_{AB}) + \delta\log_2\min\{d_A,d_B\} + g_2(\delta). \tag{9.1.57}$$

Running the argument again, but starting from an arbitrary pure-state decomposition of σ_{AB}, we conclude the inequality

$$E_F(\rho_{AB}) \leq E_F(\sigma_{AB}) + \delta\log_2\min\{d_A,d_B\} + g_2(\delta), \tag{9.1.58}$$

which, together with (9.1.57), implies (9.1.45). ■

Proposition 9.5 Faithfulness of Entanglement of Formation

The entanglement of formation is faithful, so that $E_F(\rho_{AB}) = 0$ if and only if the state ρ_{AB} is separable. More quantitatively, for a state ρ_{AB} and $\varepsilon \in [0, 1]$, if

$$\inf_{\sigma_{AB} \in \mathrm{SEP}(A:B)} \frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_1 \leq \varepsilon, \tag{9.1.59}$$

then

$$E_F(\rho_{AB}) \leq \delta \log_2 \min\{d_A, d_B\} + g_2(\delta), \tag{9.1.60}$$

where $\delta := \sqrt{\varepsilon(2-\varepsilon)}$. Conversely, for $\varepsilon \geq 0$, if

$$E_F(\rho_{AB}) \leq \varepsilon, \tag{9.1.61}$$

then

$$\inf_{\sigma_{AB} \in \mathrm{SEP}(A:B)} \frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_1 \leq \sqrt{\varepsilon \ln 2}. \tag{9.1.62}$$

Proof: We begin by proving the first statement. Suppose that the state σ_{AB} is separable. Then by the remark after Definition 3.5, there exists a pure-state decomposition of σ_{AB} as

$$\sigma_{AB} = \sum_x p(x) \phi_A^x \otimes \varphi_B^x. \tag{9.1.63}$$

For this decomposition, we have that $\sum_x p(x) H(\phi_A^x) = 0$ because the quantum entropy is equal to zero for a pure state. This implies by definition that $E_F(\sigma_{AB}) = 0$. The statement in (9.1.59)–(9.1.60) then follows by combining this observation with (9.1.44)–(9.1.45), as well as the fact that the function on the right-hand side of (9.1.60) is monotone in ε.

To see the second statement, let

$$\rho_{AB} = \sum_x p(x) \psi_{AB}^x \tag{9.1.64}$$

be an arbitrary pure-state decomposition of ρ_{AB}. By applying the observation in (9.1.55), we find that

$$\sum_x p(x) H(\psi_A^x) = \frac{1}{2} \sum_x p(x) I(A;B)_{\psi^x} \tag{9.1.65}$$

$$= \frac{1}{2} \sum_x p(x) D(\psi^x_{AB} \| \psi^x_A \otimes \psi^x_B) \tag{9.1.66}$$

$$\geq \frac{1}{4 \ln 2} \sum_x p(x) \left\| \psi^x_{AB} - \psi^x_A \otimes \psi^x_B \right\|_1^2 \tag{9.1.67}$$

$$\geq \frac{1}{4 \ln 2} \left\| \sum_x p(x) \psi^x_{AB} - \sum_x p(x) \psi^x_A \otimes \psi^x_B \right\|_1^2 \tag{9.1.68}$$

$$= \frac{1}{4 \ln 2} \left\| \rho_{AB} - \sum_x p(x) \psi^x_A \otimes \psi^x_B \right\|_1^2 \tag{9.1.69}$$

$$\geq \frac{1}{\ln 2} \left(\inf_{\sigma_{AB} \in \text{SEP}(A:B)} \frac{1}{2} \left\| \rho_{AB} - \sigma_{AB} \right\|_1 \right)^2 . \tag{9.1.70}$$

The second equality follows from rewriting the mutual information in terms of relative entropy (see (7.2.96)). The first inequality follows from the quantum Pinsker inequality (Corollary 7.32 and the remark thereafter). The second inequality follows from convexity of the square function and the trace norm. Since the inequality holds for an arbitrary pure-state decomposition of ρ_{AB}, we conclude that

$$E_F(\rho_{AB}) \geq \frac{1}{\ln 2} \left(\inf_{\sigma_{AB} \in \text{SEP}(A:B)} \frac{1}{2} \left\| \rho_{AB} - \sigma_{AB} \right\|_1 \right)^2 . \tag{9.1.71}$$

From this and (9.1.61), we conclude (9.1.62).

Finally, to conclude exact faithfulness from approximate faithfulness, we argue that the infimum in (9.1.59) and (9.1.62) is achieved (i.e., can be replaced with a minimum). This follows because the trace distance is continuous in σ_{AB} and the set of separable states is compact. ■

As the following proposition indicates, the entanglement of formation is an entanglement measure according to Definition 9.1. The proof detailed below is also a good opportunity to apply Lemma 9.2 to a simple example.

Proposition 9.6

The entanglement of formation is convex, so that (9.1.7) holds with E set to E_F, and it is a selective LOCC monotone, so that (9.1.14) holds with E set to E_F.

PROOF: By Lemma 9.2, we only need to show that the entanglement of formation does not increase under the action of a local channel and is invariant under classical communication. We begin with the first one. Let ρ_{AB} be a bipartite state, and let $\mathcal{N}_{A\to A'}$ be a local quantum channel. Let $\{(p(x), \psi^x_{AB})\}_x$ be a pure-state decomposition of ρ_{AB}, i.e., satisfying $\sum_x p(x)\psi^x_{AB} = \rho_{AB}$. Let $\mathcal{N}_{A\to A'}$ have the Kraus representation $\{N^y_{A\to A'}\}_y$. Then

$$\omega_{A'B} \coloneqq \mathcal{N}_{A\to A'}(\rho_{AB}) = \sum_x p(x)\mathcal{N}_{A\to A'}(\psi^x_{AB}), \tag{9.1.72}$$

and

$$\mathcal{N}_{A\to A'}(\psi^x_{AB}) = \sum_y N^y_{A\to A'}\psi^x_{AB}(N^y_{A\to A'})^\dagger = \sum_y p(y|x)\varphi^{x,y}_{A'B}, \tag{9.1.73}$$

where

$$p(y|x) \coloneqq \mathrm{Tr}[N^y_{A\to A'}\psi^x_{AB}(N^y_{A\to A'})^\dagger], \tag{9.1.74}$$

$$\varphi^{x,y}_{A'B} \coloneqq \frac{1}{p(y|x)} N^y_{A\to A'}\psi^x_{AB}(N^y_{A\to A'})^\dagger. \tag{9.1.75}$$

Thus, $\{(p(x)p(y|x), \varphi^{x,y}_{A'B})\}_{x,y}$ is a pure-state decomposition of $\omega_{A'B}$. Also, observe that

$$\psi^x_B = \mathrm{Tr}_A[\psi^x_{AB}] \tag{9.1.76}$$

$$= \mathrm{Tr}_{A'}[\mathcal{N}_{A\to A'}(\psi^x_{AB})] \tag{9.1.77}$$

$$= \sum_y p(y|x)\,\mathrm{Tr}_{A'}[\varphi^{x,y}_{A'B}] \tag{9.1.78}$$

$$= \sum_y p(y|x)\varphi^{x,y}_B. \tag{9.1.79}$$

Then we have that

$$\sum_x p(x)H(\psi^x_B) \geq \sum_{x,y} p(x)p(y|x)H(\varphi^{x,y}_B) \tag{9.1.80}$$

$$\geq E_F(A';B)_\omega, \tag{9.1.81}$$

where the first inequality follows from the concavity of entropy (see (7.2.106)) and the second from the definition of entanglement of formation. Since the inequality holds for all pure-state decompositions of ρ_{AB}, we conclude the desired inequality:

$$E_F(A;B)_\rho \geq E_F(A';B)_\omega. \tag{9.1.82}$$

By flipping the role of Alice and Bob in the analysis above, we conclude that the entanglement of formation does not increase under the action of a local channel on Bob's system.

Now we prove that E_F is invariant under classical communication. Let ρ_{XAB} be the classical–quantum state defined in (9.1.5). A pure-state decomposition of ρ_{XAB} has the form $\{(p(x)p(y|x), |x\rangle\langle x|_X \otimes \psi_{AB}^{x,y})\}_{x,y}$, where

$$\rho_{AB}^x = \sum_y p(y|x)\psi_{AB}^{x,y}. \tag{9.1.83}$$

This ensemble serves as a decomposition of ρ_{XAB} for $E_F(XA;B)_\rho$. Then

$$\sum_{x,y} p(x)p(y|x)H(\psi_B^{x,y}) \geq \sum_x p(x)E_F(A;B)_{\rho^x}. \tag{9.1.84}$$

Since the inequality holds for all pure-state decompositions of ρ_{XAB}, we conclude that

$$E_F(XA;B)_\rho \geq \sum_x p(x)E_F(A;B)_{\rho^x}. \tag{9.1.85}$$

Now let $\{(p(y|x), \psi_{AB}^{x,y})\}_y$ be a pure-state decomposition of ρ_{AB}^x. Then we find that

$$\sum_{x,y} p(x)p(y|x)H(\psi_B^{x,y}) \geq E_F(XA;B)_\rho \tag{9.1.86}$$

because the ensemble $\{(p(x)p(y|x), |x\rangle\langle x|_X \otimes \psi_{AB}^{x,y})\}_{x,y}$ is a particular pure-state decomposition of ρ_{XAB}. Since the inequality holds for all pure-state decompositions of ρ_{AB}^x, we conclude that

$$\sum_x p(x)E_F(A;B)_{\rho^x} \geq E_F(XA;B)_\rho. \tag{9.1.87}$$

Putting together (9.1.85) and (9.1.87), we conclude that

$$\sum_x p(x)E_F(A;B)_{\rho^x} = E_F(XA;B)_\rho. \tag{9.1.88}$$

By the same argument, but exchanging the roles of Alice and Bob, we conclude that

$$\sum_x p(x)E_F(A;B)_{\rho^x} = E_F(A;BX)_\rho. \tag{9.1.89}$$

This concludes the proof. ■

Proposition 9.7 Subadditivity of Entanglement of Formation

The entanglement of formation is subadditive; i.e., (9.1.9) holds with E set to E_F.

Proof: Let $\sum_x p(x)\psi^x_{A_1B_1}$ and $\sum_y q(y)\phi^y_{A_2B_2}$ be respective pure-state decompositions of $\tau_{A_1B_1}$ and $\omega_{A_2B_2}$. Then $\sum_{x,y} p(x)q(y)\psi^x_{A_1B_1} \otimes \phi^y_{A_2B_2}$ is a pure-state decomposition of $\tau_{A_1B_1} \otimes \omega_{A_2B_2}$. It follows that

$$E_F(A_1A_2; B_1B_2)_{\tau\otimes\omega} \leq \sum_{x,y} p(x)q(y)H(\psi^x_{A_1} \otimes \phi^y_{A_2}) \tag{9.1.90}$$

$$= \sum_{x,y} p(x)q(y)[H(\psi^x_{A_1}) + H(\phi^y_{A_2})] \tag{9.1.91}$$

$$= \sum_{x} p(x)H(\psi^x_{A_1}) + \sum_{y} q(y)H(\phi^y_{A_2}). \tag{9.1.92}$$

Since the inequality holds for arbitrary pure-state decompositions of $\tau_{A_1B_1}$ and $\omega_{A_2B_2}$, we conclude that subadditivity holds. ■

The opposite inequality, superadditivity of entanglement of formation, is known not to hold in general. Thus, the entanglement of formation is non-additive in general. The proof of this statement is highly nontrivial (please consult the Bibliographic Notes in Section 9.6).

The entanglement of formation is connected to an information-theoretic task: $E_F(\rho_{AB})$ is an achievable rate for the task of preparing the state ρ_{AB} from many copies of the two-qubit maximally entangled state $|\Phi\rangle_{AB}$ when allowing LOCC for free (please consult the Bibliographic Notes in Section 9.6).

For two-qubit states ρ_{AB}, the entanglement of formation has the following analytic expression:

$$E_F(\rho_{AB}) = h_2\left(\frac{1 + \sqrt{1 - C(\rho_{AB})^2}}{2}\right) \quad \text{(two-qubit states).} \tag{9.1.93}$$

(please consult the Bibliographic Notes in Section 9.6.) Here,

$$C(\rho_{AB}) = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}, \tag{9.1.94}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the eigenvalues of $\sqrt{\sqrt{\rho_{AB}}\,\widetilde{\rho}_{AB}\sqrt{\rho_{AB}}}$ in decreasing order. The operator $\widetilde{\rho}_{AB}$ is defined as $\widetilde{\rho}_{AB} \coloneqq (Y \otimes Y)\overline{\rho}_{AB}(Y \otimes Y)$, with Y being the Pauli-Y operator (see (4.5.25)) and $\overline{\rho}_{AB}$ being the complex conjugate of ρ_{AB} in the standard basis.

9.1.1.2 Negativity and Logarithmic Negativity

Similar to how we motivated the entanglement of formation from the Schmidt rank criterion for pure states, we can also motivate an entanglement measure from the PPT criterion. The PPT criterion states that if the partial transpose $\mathrm{T}_B(\rho_{AB})$ of a given state ρ_{AB} has a negative eigenvalue, then ρ_{AB} is entangled.[2] We use this fact to define the *negativity* of ρ_{AB} as

$$N(\rho_{AB}) \coloneqq \frac{\|\mathrm{T}_B(\rho_{AB})\|_1 - 1}{2}, \tag{9.1.95}$$

and the *logarithmic negativity* (often written simply as *log-negativity*) of ρ_{AB} as

$$E_N(\rho_{AB}) \coloneqq \log_2 \|\mathrm{T}_B(\rho_{AB})\|_1 . \tag{9.1.96}$$

Both the negativity and the log-negativity quantify the extent to which the partial transpose $\mathrm{T}_B(\rho_{AB})$ has negative eigenvalues. In particular, suppose that $\mathrm{T}_B(\rho_{AB})$ has the following Jordan–Hahn decomposition:

$$\mathrm{T}_B(\rho_{AB}) = P - N, \tag{9.1.97}$$

where P and N are the positive and negative parts of $\mathrm{T}_B(\rho_{AB})$, satisfying $P, N \geq 0$ and $PN = 0$, and we have used (2.2.68) and (2.2.69). By definition of the trace norm,

$$\|\mathrm{T}_B(\rho_{AB})\|_1 = \mathrm{Tr}[P + N]. \tag{9.1.98}$$

On the other hand, observe that $\mathrm{Tr}[\mathrm{T}_B(\rho_{AB})] = \mathrm{Tr}[\rho_{AB}] = 1$, so that

$$1 = \mathrm{Tr}[\mathrm{T}_B(\rho_{AB})] = \mathrm{Tr}[P - N]. \tag{9.1.99}$$

Therefore,

$$N(\rho_{AB}) = \frac{\|\mathrm{T}_B(\rho_{AB})\|_1 - 1}{2} = \frac{\|\mathrm{T}_B(\rho_{AB})\|_1 - \mathrm{Tr}[\mathrm{T}_B(\rho_{AB})]}{2} = \mathrm{Tr}[N]. \tag{9.1.100}$$

[2] Note that it does not matter in which basis the transpose is defined.

So, according to (2.2.69), the negativity is the sum of the absolute values of the negative eigenvalues of $\rho_{AB}^{\mathrm{T}_B}$.

By utilizing Hölder duality and semi-definite programming duality, it is possible to write $\|\mathrm{T}_B(\rho_{AB})\|_1$ as the following primal and dual semi-definite programs:

$$\|\mathrm{T}_B(\rho_{AB})\|_1 = \sup_{R_{AB}} \{\mathrm{Tr}[R_{AB}\rho_{AB}] : -\mathbb{1}_{AB} \leq \mathrm{T}_B(R_{AB}) \leq \mathbb{1}_{AB}\}, \tag{9.1.101}$$

$$= \inf_{K_{AB},L_{AB}\geq 0} \{\mathrm{Tr}[K_{AB} + L_{AB}] : \mathrm{T}_B(K_{AB} - L_{AB}) = \rho_{AB}\}. \tag{9.1.102}$$

where the optimization in the first line is with respect to Hermitian R_{AB}. We give a proof of (9.1.102)–(9.1.101) in Appendix 9.A.

Proposition 9.8

The log-negativity is non-negative for all bipartite states, and it is faithful on the set of PPT states (i.e., it is equal to zero if and only if a state is PPT).

Proof: To see the first statement, we note that the choice $R_{AB} = \mathbb{1}_{AB}$ is feasible for the primal SDP in (9.1.101), so that $\|\mathrm{T}_B(\rho_{AB})\|_1 \geq 1$, and hence $E_N(\rho_{AB}) \geq 0$, for every bipartite state ρ_{AB}.

Suppose that ρ_{AB} is a PPT state. Then $\|\mathrm{T}_B(\rho_{AB})\|_1 = \mathrm{Tr}[\mathrm{T}_B(\rho_{AB})] = \mathrm{Tr}[\rho_{AB}] = 1$ due to the assumption that $\mathrm{T}_B(\rho_{AB}) \geq 0$, implying that $E_N(\rho_{AB}) = 0$ for every PPT state.

Finally, suppose that $E_N(\rho_{AB}) = 0$. Then $\|\mathrm{T}_B(\rho_{AB})\|_1 = 1$, and employing the notation of (9.1.98)–(9.1.99), we conclude that $1 = \mathrm{Tr}[P + N] = \mathrm{Tr}[P - N]$, which implies that $\mathrm{Tr}[N] = 0$. Since $N \geq 0$, this implies that $N = 0$. Thus, $\mathrm{T}_B(\rho_{AB})$ has no negative part and ρ_{AB} is thus a PPT state. ■

Definition 9.9 Selective PPT Monotonicity

As a generalization of selective LOCC monotonicity defined in (9.1.14), we say that a function $E : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ is a selective PPT monotone if it satisfies

$$E(\rho_{AB}) \geq \sum_{x\in\mathcal{X}:p(x)>0} p(x)E(\omega_{A'B'}^x), \tag{9.1.103}$$

for every bipartite state ρ_{AB} and C-PPT-P instrument $\{\mathcal{P}^x_{AB\to A'B'}\}_{x\in\mathcal{X}}$, with

$$p(x) := \mathrm{Tr}[\mathcal{P}^x_{AB\to A'B'}(\rho_{AB})], \tag{9.1.104}$$

$$\omega^x_{A'B'} := \frac{1}{p(x)}\mathcal{P}^x_{AB\to A'B'}(\rho_{AB}). \tag{9.1.105}$$

A C-PPT-P instrument is such that every map $\mathcal{P}^x_{AB\to A'B'}$ is completely positive and $\mathrm{T}_{B'}\circ\mathcal{P}^x_{AB\to A'B'}\circ\mathrm{T}_B$ is completely positive, and the sum map $\sum_{x\in\mathcal{X}}\mathcal{P}^x_{AB\to A'B'}$ is trace preserving.
It follows that E is a PPT monotone if it is a selective PPT monotone, because the former is a special case of the latter in which the alphabet $\mathcal{X}$ has only one letter.

An LOCC instrument in (9.1.10) is a C-PPT-P instrument because every map $\mathcal{L}^x_{AB\to A'B'}$ in an LOCC instrument satisfies the requirements for a C-PPT-P instrument.

The negativity and the log-negativity are entanglement measures, as shown in Proposition 9.10 below. Interestingly, the method of proof does not involve making use of Lemma 9.2 because it is impossible to do so for the log-negativity, as the latter is not convex. In any case, we prove a stronger result than selective LOCC monotonicity for the log-negativity: we prove that it is a selective PPT monotone.

Proposition 9.10

The negativity and the log-negativity are selective PPT monotones, satisfying (9.1.103). The negativity is convex, satisfying (9.1.7), but the log-negativity is not.

Proof: Define ρ_{AB} and $\{(p(x),\omega^x_{A'B'})\}_{x\in\mathcal{X}}$ as in (9.1.104)–(9.1.105), and let $\{\mathcal{P}^x_{AB\to A'B'}\}_{x\in\mathcal{X}}$ be a C-PPT-P instrument. Let K_{AB} and L_{AB} be arbitrary positive semi-definite operators satisfying

$$\mathrm{T}_B(K_{AB}-L_{AB}) = \rho_{AB}. \tag{9.1.106}$$

Then we find that

$$p(x)\omega^x_{A'B'} = \mathcal{P}^x_{AB\to A'B'}(\rho_{AB}) \tag{9.1.107}$$

$$= \mathcal{P}^x_{AB\to A'B'}(\mathrm{T}_B(K_{AB}-L_{AB})) \tag{9.1.108}$$

$$= \mathrm{T}_{B'}(\mathrm{T}_{B'} \circ \mathcal{P}^x_{AB\to A'B'} \circ \mathrm{T}_B)(K_{AB} - L_{AB})). \tag{9.1.109}$$

Let us define

$$K^x_{A'B'} := \frac{1}{p(x)}(\mathrm{T}_{B'} \circ \mathcal{P}^x_{AB\to A'B'} \circ \mathrm{T}_B)(K_{AB}), \tag{9.1.110}$$

$$L^x_{A'B'} := \frac{1}{p(x)}(\mathrm{T}_{B'} \circ \mathcal{P}^x_{AB\to A'B'} \circ \mathrm{T}_B)(L_{AB}), \tag{9.1.111}$$

so that

$$\omega^x_{A'B'} = \mathrm{T}_{B'}(K^x_{A'B'} - L^x_{A'B'}). \tag{9.1.112}$$

Furthermore, since $\mathrm{T}_{B'} \circ \mathcal{P}^x_{AB\to A'B'} \circ \mathrm{T}_B$ is completely positive, it follows that $K^x_{A'B'}, L^x_{A'B'} \geq 0$. Thus, $K^x_{A'B'}$ and $L^x_{A'B'}$ are feasible for the SDP in (9.1.102) for $\left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1$, and we conclude that

$$\mathrm{Tr}[K^x_{A'B'} + L^x_{A'B'}] \geq \left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1 . \tag{9.1.113}$$

Then consider that

$$\mathrm{Tr}[K_{AB} + L_{AB}]$$

$$= \mathrm{Tr}[\mathrm{T}_B(K_{AB} + L_{AB})] \tag{9.1.114}$$

$$= \sum_{x\in\mathcal{X}:p(x)>0} \mathrm{Tr}[\mathcal{P}^x_{AB\to A'B'}(\mathrm{T}_B(K_{AB} + L_{AB}))] \tag{9.1.115}$$

$$= \sum_{x\in\mathcal{X}:p(x)>0} \mathrm{Tr}[(\mathrm{T}_{B'} \circ \mathcal{P}^x_{AB\to A'B'} \circ \mathrm{T}_B)(K_{AB} + L_{AB})] \tag{9.1.116}$$

$$= \sum_{x\in\mathcal{X}:p(x)>0} p(x)\,\mathrm{Tr}[K^x_{A'B'} + L^x_{A'B'}] \tag{9.1.117}$$

$$\geq \sum_{x\in\mathcal{X}:p(x)>0} p(x) \left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1 . \tag{9.1.118}$$

The first and third equalities hold because the trace is invariant under a partial transpose. The second equality follows because the sum map $\sum_x \mathcal{P}^x_{AB\to A'B'}$ is trace preserving. The fourth equality follows from the definitions in (9.1.110)–(9.1.111). The inequality follows from (9.1.113). Since the inequality holds for all $K_{AB}, L_{AB} \geq 0$ satisfying (9.1.106), we conclude that

$$\left\|\mathrm{T}_B(\rho_{AB})\right\|_1 \geq \sum_{x\in\mathcal{X}:p(x)>0} p(x) \left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1 . \tag{9.1.119}$$

By applying the definition in (9.1.95), we conclude that the negativity is a selective PPT monotone. Now considering (9.1.119) and taking the logarithm, and using its monotonicity and concavity, we conclude that the log-negativity is a selective PPT monotone:

$$E_N(\rho_{AB}) = \log_2 \|\mathrm{T}_B(\rho_{AB})\|_1 \tag{9.1.120}$$
$$\geq \log_2\left[\sum_{x\in\mathcal{X}:p(x)>0} p(x) \left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1\right] \tag{9.1.121}$$
$$\geq \sum_{x\in\mathcal{X}:p(x)>0} p(x) \log_2 \left\|\mathrm{T}_B(\omega^x_{A'B'})\right\|_1 \tag{9.1.122}$$
$$= \sum_{x\in\mathcal{X}:p(x)>0} p(x) E_N(\omega^x_{A'B'}). \tag{9.1.123}$$

That the negativity is convex follows directly from the definition, convexity of the trace norm, and linearity of the partial transpose.

The lack of convexity of log-negativity follows from direct evaluation for the states $\Phi_{AB} := \frac{1}{2}\sum_{i,j\in\{0,1\}} |ii\rangle\langle jj|_{AB}$, $\sigma_{AB} := \frac{1}{2}\sum_{i\in\{0,1\}} |ii\rangle\langle ii|_{AB}$, and $\overline{\rho}_{AB} := \frac{1}{2}(\Phi_{AB} + \sigma_{AB})$, for which we have

$$E_N(\Phi_{AB}) = 1, \qquad E_N(\sigma_{AB}) = 0, \qquad E_N(\overline{\rho}_{AB}) = \log_2 \frac{3}{2}, \tag{9.1.124}$$

so that

$$E_N(\overline{\rho}_{AB}) > \frac{1}{2}\left(E_N(\Phi_{AB}) + E_N(\sigma_{AB})\right). \tag{9.1.125}$$

This concludes the proof. ■

Proposition 9.11 Additivity of Log-Negativity

The logarithmic negativity is additive; i.e., (9.1.8) holds with E set to E_N.

Proof: For every two states $\tau_{A_1B_1}$ and $\omega_{A_2B_2}$, consider that

$$E_N(A_1A_2; B_1B_2)_{\tau\otimes\omega} = \log_2 \left\|\mathrm{T}_{B_1B_2}(\tau_{A_1B_1} \otimes \omega_{A_2B_2})\right\|_1 \tag{9.1.126}$$
$$= \log_2 \left\|\mathrm{T}_{B_1}(\tau_{A_1B_1}) \otimes \mathrm{T}_{B_2}(\omega_{A_2B_2})\right\|_1 \tag{9.1.127}$$
$$= \log_2\left(\left\|\mathrm{T}_{B_1}(\tau_{A_1B_1})\right\|_1 \cdot \left\|\mathrm{T}_{B_2}(\omega_{A_2B_2})\right\|_1\right) \tag{9.1.128}$$

$$= \log_2 \left\| \mathrm{T}_{B_1}(\tau_{A_1B_1}) \right\|_1 + \log_2 \left\| \mathrm{T}_{B_2}(\omega_{A_2B_2}) \right\|_1 \tag{9.1.129}$$
$$= E_N(A_1;B_1)_\tau + E_N(A_2;B_2)_\omega. \tag{9.1.130}$$

This concludes the proof. ■

Proposition 9.12 Log-Negativity of Pure Bipartite States

For a pure bipartite state ψ_{AB}, the log-negativity is equal to the Rényi entropy of order $\frac{1}{2}$ of the reduced state ψ_A:

$$E_N(\psi_{AB}) = H_{\frac{1}{2}}(\psi_A). \tag{9.1.131}$$

PROOF: For every pure state $\psi_{AB} = |\psi\rangle\langle\psi|_{AB}$ such that

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k} |e_k\rangle_A \otimes |f_k\rangle_B \tag{9.1.132}$$

is a Schmidt decomposition, we have that

$$|\psi\rangle\langle\psi|_{AB}^{\mathsf{T}_B} = \sum_{k,k'=1}^{r} \sqrt{\lambda_k \lambda_{k'}} |e_k\rangle\langle e_{k'}|_A \otimes |f_{k'}\rangle\langle f_k|_B, \tag{9.1.133}$$

where we have taken the partial transpose with respect to the orthonormal set $\{|f_k\rangle_B\}_{k=1}^r$. Observe that

$$|\psi\rangle\langle\psi|_{AB}^{\mathsf{T}_B} = F_{AB} \left(\sum_{k'=1}^{r} \sqrt{\lambda_{k'}} |e_{k'}\rangle\langle e_{k'}|_A \otimes \sum_{k=1}^{r} \sqrt{\lambda_k} |f_k\rangle\langle f_k|_B \right), \tag{9.1.134}$$

where $F_{AB} = \sum_{k,k'=1}^{r} |e_{k'}\rangle\langle e_k|_A \otimes |f_k\rangle\langle f_{k'}|_B$ is a unitary swap operator. Thus, by unitary invariance of the trace norm, we obtain

$$E_N(\psi_{AB}) = \log_2 \left(\sum_{k=1}^{r} \sqrt{\lambda_k} \right)^2 = 2 \log_2 \left(\sum_{k=1}^{r} \sqrt{\lambda_k} \right). \tag{9.1.135}$$

By comparing with (7.4.3), we conclude the statement of the proposition. ■

For a maximally entangled state ($\lambda_k = \frac{1}{r}$ for all $1 \leq k \leq r$), we find that $E_N(\psi_{AB}) = \log_2 r$, exactly as with the entanglement of formation.

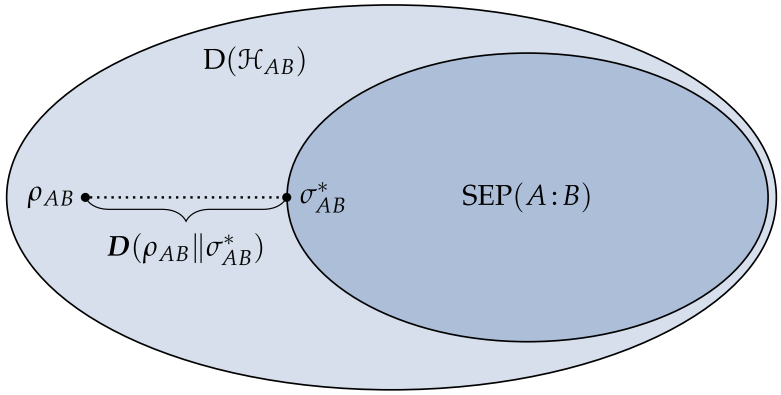


FIGURE 9.1: A simple way to measure the entanglement of a bipartite state ρ_{AB} is to calculate its divergence with the set of separable states. If we use a generalized divergence $\boldsymbol{D}$ as our measure, then the measure of the entanglement in ρ_{AB} is given by the smallest value of $\boldsymbol{D}(\rho_{AB}\|\sigma_{AB})$, where $\sigma_{AB} \in \text{SEP}(A : B)$ is a separable state, i.e., by the quantity $\boldsymbol{D}(\rho_{AB}\|\sigma^*_{AB}) = \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB})$.

9.1.1.3 Divergence-Based Measures

The two entanglement measures considered above are based on specific mathematical properties of entanglement. However, using the fact that entangled states are, by definition, not separable, we can construct a broad class of entanglement measures by finding the divergence of a given state ρ_{AB} with the set of separable states. This idea is illustrated in Figure 9.1. We primarily consider such *divergence-based* entanglement measures in this book (in the research literature, these are also called "distance-based" entanglement measures, even though divergences that are not distances, such as relative entropy, are used in this approach).

As an example of a divergence-based entanglement measure, let us consider a concrete divergence, the normalized trace distance, which we defined in Section 6.1 as $\frac{1}{2}\|\rho - \sigma\|_1$ for every two states ρ and σ. Mathematically, the distance of a point to a set is defined by finding the element of that set that is closest to the given point. With this idea, we define the *trace distance of entanglement* of a state ρ_{AB} as the normalized trace distance from ρ_{AB} to the closest state $\sigma_{AB} \in \text{SEP}(A\!:\!B)$:

$$E_T(A;B)_\rho := \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \frac{1}{2}\|\rho_{AB} - \sigma_{AB}\|_1 . \tag{9.1.136}$$

Note that the infimum is indeed achieved, because SEP($A\!:\!B$) is a compact set and the trace norm is continuous in σ_{AB}, so that there always exists a closest separable state to the given state ρ_{AB}. Recall that we implicitly introduced the trace distance of entanglement in Proposition 9.5, when considering approximate faithfulness of the entanglement of formation.

The quantity E_T is indeed an entanglement measure. To see this, we use the data-processing inequality for the trace distance (Theorem 6.3), and the fact that separable states are preserved under LOCC channels (which follows immediately from the definition of LOCC channels). Then, for every state ρ_{AB}, every LOCC channel $\mathcal{L}_{AB\to A'B'}$, and letting $\omega_{A'B'} = \mathcal{L}_{AB\to A'B'}(\rho_{AB})$, we obtain

$$E_T(A;B)_\rho = \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \frac{1}{2}\left\|\rho_{AB} - \sigma_{AB}\right\|_1 \tag{9.1.137}$$

$$\geq \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \frac{1}{2}\left\|\mathcal{L}_{AB\to A'B'}(\rho_{AB}) - \mathcal{L}_{AB\to A'B'}(\sigma_{AB})\right\|_1 \tag{9.1.138}$$

$$\geq \inf_{\tau_{A'B'}\in\mathrm{SEP}(A':B')} \frac{1}{2}\left\|\mathcal{L}_{AB\to A'B'}(\rho_{AB}) - \tau_{A'B'}\right\|_1 \tag{9.1.139}$$

$$= E_T(A';B')_\omega. \tag{9.1.140}$$

Although the simple proof above makes it clear that the trace distance of entanglement is an LOCC monotone, it is known that the trace distance of entanglement is not a selective LOCC monotone, as defined in (9.1.14) (please consult the Bibliographic Notes in Section 9.6).

The trace distance of entanglement is also faithful, which is due to the fact that the trace distance is a metric in the mathematical sense: $\frac{1}{2}\left\|\rho_{AB} - \sigma_{AB}\right\|_1 \geq 0$ for all states ρ_{AB}, σ_{AB}, and $\frac{1}{2}\left\|\rho_{AB} - \sigma_{AB}\right\|_1 = 0$ if and only if $\rho_{AB} = \sigma_{AB}$.

Beyond the trace distance, we can take any distinguishability measure and define an entanglement measure analogous to the one in (9.1.136). That is, we can take any *generalized divergence* $\boldsymbol{D}$ as our divergence. Recall from Definition 7.15 that a generalized divergence is a function $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R}\cup\{+\infty\}$ that obeys the data-processing inequality. We then define the *generalized divergence of entanglement of* ρ_{AB} as follows:

$$\boldsymbol{E}(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}). \tag{9.1.141}$$

See Figure 9.1 for a visual depiction of the idea behind this quantity. If the generalized divergence $\boldsymbol{D}$ is continuous in its second argument, then the infimum in (9.1.141) is achieved. We study this entanglement measure in much more detail in Section 9.2. By the data-processing inequality for the generalized divergence, as well as the fact that separable states are preserved under LOCC channels, it follows that $\boldsymbol{E}(A;B)_\rho$ is an entanglement measure. We prove this and other properties of the generalized divergence of entanglement in Proposition 9.16. As was the case for the trace distance of entanglement, it does not necessarily follow that the generalized divergence of entanglement is a selective LOCC monotone, as defined in (9.1.14), but it does hold for some important cases.

Although the generalized divergence of entanglement of a bipartite state is conceptually simple, it is in general difficult to optimize over the set of separable states because it does not have a simple characterization (except in low dimensions). This means that the generalized divergence of entanglement is difficult to compute in most cases.

To obtain an entanglement measure that is simpler to compute, one idea is to relax the optimization in (9.1.141) from the set of separable states to some other set of states that contains the set of separable states. It is ideal if this other set is easier to characterize than the set of separable states. As a first step, let us recall the PPT criterion from Section 3.2.9, which states that if a bipartite state is separable, then it is PPT, meaning that it has positive partial transpose (recall Definition 3.17). This fact immediately leads to the containment $\text{SEP}(A\!:\!B) \subsetneq \text{PPT}(A\!:\!B)$. (As stated in Section 3.2.9, if both A and B are qubits, or if one of them is a qubit and the other a qutrit, then $\text{PPT}(A\!:\!B) = \text{SEP}(A\!:\!B)$.) We can thus define the *PPT generalized divergence* of a bipartite state ρ_{AB} as

$$\boldsymbol{E}_{\text{PPT}}(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\text{PPT}(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}). \tag{9.1.142}$$

If the generalized divergence $\boldsymbol{D}$ is continuous in its second argument, then the infimum is achieved. Also, note that $\boldsymbol{E}_{\text{PPT}}(A;B)_\rho = \boldsymbol{E}(A;B)_\rho$ when both A and B are qubits or when one of them is a qubit and the other a qutrit.

Like the generalized divergence of entanglement, the PPT divergence is an entanglement measure. In fact, the PPT divergence is monotone under C-PPT-P channels, as defined in Definition 4.27. This is due to the data-processing inequality for the generalized divergence and the fact that the set of PPT states is closed under C-PPT-P channels (see Proposition 4.28). Since the set of LOCC channels is contained in the set of C-PPT-P channels (Propositions 4.24 and 4.29), it follows that the PPT divergence is an entanglement measure.

Unlike the generalized divergence of entanglement, the PPT divergence is not a faithful entanglement measure. It is true that $\boldsymbol{E}_{\text{PPT}}(A;B)_\rho = 0$ for all separable

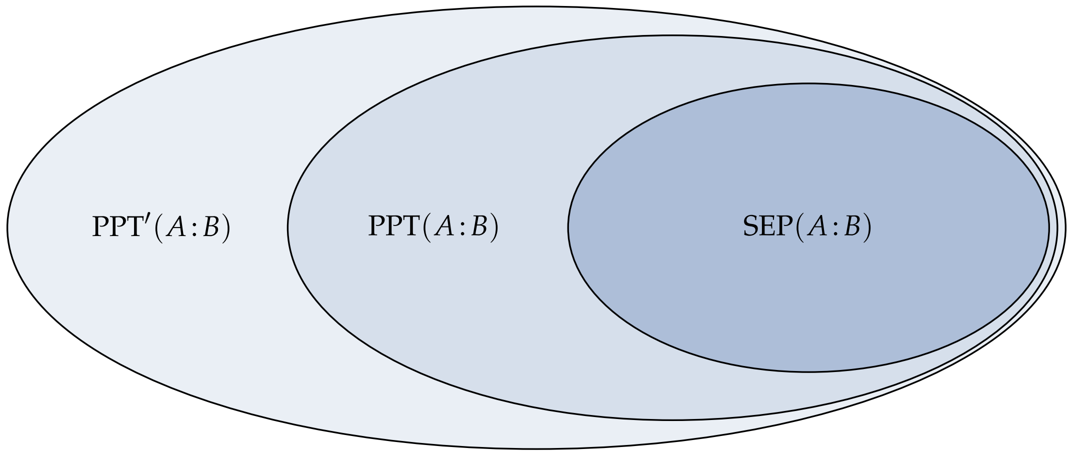


FIGURE 9.2: The set SEP$(A:B)$ of separable states acting on the Hilbert space $\mathcal{H}_{AB}$ is contained in the set PPT$(A:B)$ of positive partial transpose (PPT) states, which in turn is contained in the set PPT$'(A:B)$ of operators defined in (9.1.144). The sets PPT and PPT$'$ are relaxations of the set of separable states that can be easily characterized in terms of semi-definite constraints.

states ρ_{AB} due to the containment

$$\mathrm{SEP}(A:B) \subseteq \mathrm{PPT}(A:B). \tag{9.1.143}$$

However, the converse statement is not true because the infimum in (9.1.142), if achieved, need not be achieved by a separable state. In other words, there exist PPT entangled states ρ_{AB} for which $\boldsymbol{E}_{\mathrm{PPT}}(A;B)_\rho = 0$.

It turns out to be useful to relax the set of PPT states further:

Definition 9.13 PPT'

Let PPT$'(A:B)$ denote the following convex set of positive semi-definite operators:

$$\mathrm{PPT}'(A:B) \coloneqq \{\sigma_{AB} : \sigma_{AB} \geq 0, \ \|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1\}. \tag{9.1.144}$$

Convexity of the set PPT$'(A:B)$ follows from convexity of the trace norm. Furthermore, the set PPT$'(A:B)$ contains the set of PPT states because every PPT state σ_{AB} satisfies $\|\mathrm{T}_B(\sigma_{AB})\|_1 = 1$. Furthermore, every operator $\sigma_{AB} \in \mathrm{PPT}'(A:B)$ is subnormalized, satisfying $\mathrm{Tr}[\sigma_{AB}] \leq 1$, which follows because

$$\mathrm{Tr}[\sigma_{AB}] = \mathrm{Tr}[\mathrm{T}_B(\sigma_{AB})] \leq \|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1. \tag{9.1.145}$$

The set $\mathrm{PPT}'(A:B)$ can be written equivalently as

$$\mathrm{PPT}'(A:B) \coloneqq \{\sigma_{AB} : \sigma_{AB} \geq 0,\ E_N(\sigma_{AB}) \leq 0\}, \tag{9.1.146}$$

by inspecting the formula for log-negativity in (9.1.96). By comparing with (3.2.116), we clearly have the containment

$$\mathrm{SEP}(A:B) \subseteq \mathrm{PPT}(A:B) \subseteq \mathrm{PPT}'(A:B). \tag{9.1.147}$$

See Figure 9.2 for a visual depiction of this containment.

We define the *generalized Rains divergence* of a bipartite state ρ_{AB} as

$$\boldsymbol{R}(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}). \tag{9.1.148}$$

If the underlying generalized divergence $\boldsymbol{D}$ is continuous in its second argument, then the infimum is achieved in (9.1.148). Like the entanglement measures $\boldsymbol{E}(A;B)_\rho$ and $\boldsymbol{E}_{\mathrm{PPT}}(A;B)_\rho$, the generalized Rains divergence is an entanglement measure. In fact, it is monotone under C-PPT-P channels, which follows from the data-processing inequality for the generalized divergence and because the set PPT' is preserved under C-PPT-P channels (a consequence of Lemma 9.14 below). Since the set of LOCC channels is contained in the set of C-PPT-P channels (Propositions 4.24 and 4.29), it follows that the generalized Rains divergence is an entanglement measure.

As with the PPT divergence, the generalized Rains divergence is not a faithful entanglement measure. It is true that $\boldsymbol{R}(A;B)_\rho = 0$ for all separable states ρ_{AB}, due to the containment $\mathrm{SEP}(A:B) \subseteq \mathrm{PPT}'(A:B)$. However, the converse statement is not true because the infimum in (9.1.148) need not be achieved by a separable state.

Depending on the form of the generalized divergence $\boldsymbol{D}$, the relaxation from the set SEP to the set PPT' leads to an entanglement measure that can be computed efficiently via semi-definite programming (Section 2.4). We investigate one such example of an entanglement measure in Section 9.3.1. Also, due to the containments in (9.1.147), we have that

$$\boldsymbol{E}(A;B)_\rho \geq \boldsymbol{E}_{\mathrm{PPT}}(A;B)_\rho \geq \boldsymbol{R}(A;B)_\rho, \tag{9.1.149}$$

for every bipartite state ρ_{AB}. Thus, as we show later in the book, the relaxation from SEP to PPT' via the generalized Rains divergence not only allows for the possibility of efficiently computable entanglement measures, but due to the inequality in

(9.1.149), it also allows for the possibility of obtaining a tighter upper bound on communication rates in certain scenarios. We investigate the properties of the generalized Rains divergence in detail in Section 9.3.

Before proceeding, let us state some properties of the set PPT′.

Lemma 9.14 Properties of the Set PPT′

The set $\text{PPT}'(A\!:\!B)$ defined in (9.1.144) has the following properties:

1. It is closed under completely PPT-preserving channels (recall Definition 4.27). In more detail, let $\mathcal{P}_{AB\to A'B'}$ be a completely PPT-preserving channel. Then, for every state $\rho_{AB} \in \text{PPT}'(A\!:\!B)$, we have that $\mathcal{P}_{AB\to A'B'}(\rho_{AB}) \in \text{PPT}'(A'\!:\!B')$.

2. It is closed under LOCC channels (recall Definition 4.22). In more detail, let $\mathcal{L}_{AB\to A'B'}$ be an LOCC channel. Then, for every state $\rho_{AB} \in \text{PPT}'(A\!:\!B)$, it holds that $\mathcal{L}_{AB\to A'B'}(\rho_{AB}) \in \text{PPT}'(A'\!:\!B')$.

Remark: We emphasize that not all operators in the set PPT′ are quantum states, meaning that not all operators $\sigma_{AB} \in \text{PPT}'(A\!:\!B)$ satisfy $\text{Tr}[\sigma_{AB}] = 1$.

Proof:

1. Let $\sigma_{A'B'} = \mathcal{P}_{AB\to A'B'}(\rho_{AB})$. Since $\mathcal{P}_{AB\to A'B'}$ is a channel, and ρ_{AB} is a state, we have that $\sigma_{A'B'} \geq 0$. Then,

$$\text{T}_{B'}(\sigma_{A'B'}) = (\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'})(\rho_{AB}) \tag{9.1.150}$$
$$= (\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B \circ \text{T}_B)(\rho_{AB}) \tag{9.1.151}$$
$$= (\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B)(\text{T}_B(\rho_{AB})). \tag{9.1.152}$$

Now, consider that the induced trace norm of $\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B$ satisfies $\|\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B\|_1 = 1$, which follows from (2.2.169)–(2.2.170) and the fact that $\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B$ is a channel by definition of $\mathcal{P}_{AB\to A'B'}$. Furthermore, we have that $\|\text{T}_B(\rho_{AB})\|_1 \leq 1$ because $\rho_{AB} \in \text{PPT}'(A:B)$. Putting these observations together, we find that

$$\|\text{T}_{B'}(\sigma_{A'B'})\|_1 = \|(\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B)(\text{T}_B(\rho_{AB}))\|_1 \tag{9.1.153}$$
$$\leq \|\text{T}_{B'} \circ \mathcal{P}_{AB\to A'B'} \circ \text{T}_B\|_1 \, \|\text{T}_B(\rho_{AB})\|_1 \tag{9.1.154}$$
$$\leq 1. \tag{9.1.155}$$

Therefore, $\sigma_{A'B'} \in \mathrm{PPT}'(A\!:\!B)$.

2. Since every LOCC channel is a completely PPT-preserving channel (Propositions 4.24 and 4.29), the result immediately follows from the proof above. ■

9.1.1.4 Squashed Entanglement

In the previous example, we considered the generalized divergence of entanglement, which is simply the generalized divergence of a given bipartite state with the set of separable states. If we restrict ourselves to *product states* only, i.e., states of the form $\tau_A \otimes \sigma_B$, and we consider the quantum relative entropy, then we obtain

$$\inf_{\tau_A,\sigma_B} D(\rho_{AB}\|\tau_A \otimes \sigma_B) = I(A;B)_\rho, \tag{9.1.156}$$

where we recall the expression in (7.2.99) for the mutual information $I(A;B)_\rho$ of the state ρ_{AB}. Thus, the mutual information is the minimal relative entropy between the state of interest and the set of product states. We can thus view the mutual information as a measure of the correlations contained in the bipartite state ρ_{AB}. However, the mutual information quantifies both classical and quantum correlations because it detects any correlation whatsoever. Thus, it cannot be the case that the mutual information is an entanglement measure, because it is strictly positive even for some separable states that have only classical correlations, and so it can in general increase under LOCC channels.

Regardless, we can still use the mutual information in a meaningful way to quantify entanglement. For example, suppose that σ_{AB} is a separable state, so that

$$\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho_A^x \otimes \tau_B^x, \tag{9.1.157}$$

for a finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0,1]$, and sets $\{\rho_A^x\}_{x\in\mathcal{X}}$, $\{\tau_B^x\}_{x\in\mathcal{X}}$ of states. Let us form the following extension of σ_{AB} to a state ω_{ABX}, with X a classical register:

$$\omega_{ABX} = \sum_{x\in\mathcal{X}} p(x)\rho_A^x \otimes \tau_B^x \otimes |x\rangle\!\langle x|_X. \tag{9.1.158}$$

This is indeed an extension because $\mathrm{Tr}_X[\omega_{ABX}] = \sigma_{AB}$. Let us now consider the *conditional* mutual information $I(A;B|X)_\omega$ of this extension (recall the definition of the quantum conditional mutual information in (7.1.11)). Since ω_{ABX} is a classical–quantum state, it follows that

$$I(A;B|X)_\omega = \sum_{x\in\mathcal{X}} p(x) I(A;B)_{\rho^x\otimes\tau^x} = 0. \tag{9.1.159}$$

Therefore, while the mutual information of a separable state can in general be non-zero, the conditional mutual information is always zero. Intuitively, this is due to the fact that the classical system acts as a "probe," which, when measured, reveals a value $x \in \mathcal{X}$ for the classical system X. Conditioned on this value, the joint state $\rho_A^x \otimes \tau_B^x$ is product.

Thus, for every separable state ρ_{AB}, there exists a classical extension ω_{ABX} such that the conditional mutual information $I(A;B|X)_\omega$ is equal to zero. Using this idea, we could propose a potential measure of entanglement as follows:

$$\frac{1}{2}\inf_{\omega_{ABX}} \{I(A;B|X)_\omega : \mathrm{Tr}_X[\omega_{ABX}] = \rho_{AB}\}, \tag{9.1.160}$$

where the optimization is with respect to extensions of ρ_{AB} having a classical extension system X of arbitrary (finite) dimension $|\mathcal{X}|$, so that

$$\omega_{ABX} = \sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \otimes |x\rangle\langle x|_X \tag{9.1.161}$$

for some set $\{\rho_{AB}^x\}_{x\in\mathcal{X}}$ of states and a probability distribution $p(x)$ satisfying $\rho_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x$. The normalization factor of $\frac{1}{2}$ is there for reasons that become apparent later. If we require that every state ρ_{AB}^x in the extension ω_{ABX} should be pure, then the measure in (9.1.160) reduces to the entanglement of formation (this was actually used in (9.1.55) in the proof of Proposition 9.4).

The quantity proposed in (9.1.160) is non-negative for every state ρ_{AB}, due to the non-negativity of mutual information and the fact that conditional mutual information with a classical conditioning system is equal to a convex combination of mutual informations. It is already clear that the quantity proposed in (9.1.160) is equal to zero for every separable state—if a state is separable, then the optimization in (9.1.160) finds the separable decomposition and the value of the quantity is zero, as discussed just after (9.1.159). The converse is also true, which follows from the same proof given for (9.1.61)–(9.1.62). It is actually also possible to show that the quantity in (9.1.160) is an entanglement measure. However, we do not make further use of this quantity in this book, because there is an entanglement measure more suitable for our purposes, as introduced below.

Instead of taking a classical extension of the separable state σ_{AB} in (9.1.157), as we did in (9.1.158), we can take a "quantum extension," in the sense that we could define an extension ω_{ABE} in which the system E is some finite-dimensional quantum system. Optimizing over all such extensions, we obtain the *squashed entanglement*:

$$E_{\text{sq}}(A;B)_\rho := \frac{1}{2} \inf_{\omega_{ABE}} \{I(A;B|E)_\omega : \text{Tr}_E[\omega_{ABE}] = \rho_{AB}\}, \tag{9.1.162}$$

which can only be smaller than the quantity proposed in (9.1.160). Note that we optimize with respect to extensions ω_{ABE} for which the extension system E can have *arbitrary* finite dimension. It is not known whether the optimization can be restricted to extension systems of a certain fixed dimension. In general, therefore, it is not known whether the infimum in (9.1.162) can be replaced by a minimum.

The squashed entanglement is indeed an entanglement measure, as we show in Section 9.4. It is also a faithful entanglement measure. That it vanishes for separable states follows from the arguments presented above. For a proof of the converse direction, please consult the Bibliographic Notes in Section 9.6. We establish other important properties of the squashed entanglement in Section 9.4.

9.2 Generalized Divergence of Entanglement

In this section, we investigate properties of the generalized divergence of entanglement, which is a general construction of an entanglement measure. Let us recall from (9.1.141) above that the generalized divergence of entanglement of a bipartite state is the generalized divergence between that state and the set of separable states.

Definition 9.15 Generalized Divergence of Entanglement

Let $\boldsymbol{D}$ be a generalized divergence (see Definition 7.15). For every bipartite state ρ_{AB}, we define the *generalized divergence of entanglement of* ρ_{AB} as

$$\boldsymbol{E}(A;B)_\rho := \inf_{\sigma_{AB} \in \text{SEP}(A:B)} \boldsymbol{D}(\rho_{AB} \| \sigma_{AB}). \tag{9.2.1}$$

If the underlying generalized divergence $\boldsymbol{D}$ is continuous in its second argument, then the infimum can be replaced by a minimum.

We are particularly interested throughout the rest of this book in the following generalized divergences of entanglement for every state ρ_{AB}:

1. The *relative entropy of entanglement of* ρ_{AB},

$$E_R(A;B)_\rho := \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}), \tag{9.2.2}$$

   where $D(\rho_{AB}\|\sigma_{AB})$ is the quantum relative entropy of ρ_{AB} and σ_{AB} (Definition 7.1).

2. The ε*-hypothesis testing relative entropy of entanglement of* ρ_{AB},

$$E_R^\varepsilon(A;B)_\rho := \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D_H^\varepsilon(\rho_{AB}\|\sigma_{AB}), \tag{9.2.3}$$

   where $D_H^\varepsilon(\rho_{AB}\|\sigma_{AB})$ is the ε-hypothesis testing relative entropy of ρ_{AB} and σ_{AB} (Definition 7.65).

3. The *sandwiched Rényi relative entropy of entanglement of* ρ_{AB},

$$\widetilde{E}_\alpha(A;B)_\rho := \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}), \tag{9.2.4}$$

   where $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB})$, $\alpha\in[1/2,1)\cup(1,\infty)$, is the sandwiched Rényi relative entropy of ρ_{AB} and σ_{AB} (Definition 7.28). Note that $\widetilde{E}_\alpha(A;B)_\rho$ is monotonically increasing in α for all ρ_{AB} (see Proposition 7.31). This fact, along with the fact that $\lim_{\alpha\to1}\widetilde{D}_\alpha = D$ (see Proposition 7.30), leads to

$$\lim_{\alpha\to1}\widetilde{E}_\alpha(A;B)_\rho = E_R(A;B)_\rho \tag{9.2.5}$$

   for every state ρ_{AB}. See Appendix 10.A for details of the proof.

4. The *max-relative entropy of entanglement of* ρ_{AB},

$$E_{\max}(A;B)_\rho := \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}), \tag{9.2.6}$$

   where $D_{\max}(\rho_{AB}\|\sigma_{AB})$ is the max-relative entropy of ρ_{AB} and σ_{AB} (Definition 7.58). Using the fact that $\lim_{\alpha\to\infty}\widetilde{D}_\alpha = D_{\max}$ (see Proposition 7.61), we find that

$$E_{\max}(A;B)_\rho = \lim_{\alpha\to\infty}\widetilde{E}_\alpha(A;B)_\rho \tag{9.2.7}$$

for every state ρ_{AB}. See Appendix 10.A for details of the proof. As a consequence of this fact, and the fact that $\widetilde{E}_\alpha(A;B)_\rho$ is monotonically increasing in α for all ρ_{AB}, we have that

$$E_{\max}(A;B)_\rho \geq \widetilde{E}_\alpha(A;B)_\rho \tag{9.2.8}$$

for all $\alpha \in (1,\infty)$ and every state ρ_{AB}.

In addition to the quantities above, we also can define the Petz– and geometric Rényi relative entropies of entanglement in a similar way, but based on Definitions 7.20 and 7.38, respectively, for the range of α for which data processing holds. These are denoted by $E_\alpha(A;B)_\rho$ and $\widehat{E}_\alpha(A;B)_\rho$, respectively.

Proposition 9.16 Properties of Generalized Divergence of Entanglement

Let $\boldsymbol{D}$ be a generalized divergence that is continuous in its second argument, and consider the generalized divergence of entanglement $\boldsymbol{E}(A;B)_\rho$ of a state ρ_{AB}, as defined in (9.2.1).

1. *Separable monotonicity*: For every separable channel $\mathcal{S}_{AB\to A'B'}$, the generalized divergence of entanglement is monotone non-increasing:

$$\boldsymbol{E}(A;B)_\rho \geq \boldsymbol{E}(A';B')_\omega, \tag{9.2.9}$$

where $\omega_{A'B'} = \mathcal{S}_{AB\to A'B'}(\rho_{AB})$. Since every LOCC channel is a separable channel, the generalized divergence of entanglement is also monotone non-increasing under LOCC channels. It is therefore an entanglement measure as per Definition 9.1.

2. *Faithfulness*: If $\boldsymbol{D}$ satisfies $\boldsymbol{D}(\rho\|\sigma) \geq 0$ and $\boldsymbol{D}(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$ (for all states ρ and σ), then $\boldsymbol{E}(A;B)_\rho = 0$ if and only if $\rho_{AB} \in \mathrm{SEP}(A\!:\!B)$. The generalized divergence of entanglement is then a faithful entanglement measure.

3. *Subadditivity*: If $\boldsymbol{D}$ is additive for product positive semi-definite operators, i.e., $\boldsymbol{D}(\rho\otimes\omega\|\sigma\otimes\tau) = \boldsymbol{D}(\rho\|\sigma) + \boldsymbol{D}(\omega\|\tau)$, then for every two quantum states $\rho_{A_1B_1}$ and $\omega_{A_2B_2}$, the generalized divergence of entanglement is sub-additive:

$$\boldsymbol{E}(A_1A_2;B_1B_2)_{\rho\otimes\omega} \leq \boldsymbol{E}(A_1;B_1)_\rho + \boldsymbol{E}(A_2;B_2)_\omega. \tag{9.2.10}$$

4. *Convexity*: If $\boldsymbol{D}$ is jointly convex, meaning that

$$\boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}\middle\|\sum_{x\in\mathcal{X}} p(x)\sigma^x_{AB}\right) \leq \sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho^x_{AB}\|\sigma^x_{AB}), \qquad (9.2.11)$$

for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0, 1]$, and sets $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$, $\{\sigma^x_{AB}\}_{x\in\mathcal{X}}$ of states, then the generalized divergence of entanglement is convex:

$$\boldsymbol{E}(A; B)_{\overline{\rho}} \leq \sum_{x\in\mathcal{X}} p(x)\boldsymbol{E}(A; B)_{\rho^x}, \qquad (9.2.12)$$

where $\overline{\rho}_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}$.

Properties 1., 2., 3. are satisfied when the generalized divergence is the quantum relative entropy, the Petz–, sandwiched, and geometric Rényi relative entropy, and the max-relative entropy. Property 4. is satisfied when the generalized divergence is the quantum relative entropy and the Petz–, sandwiched, and geometric Rényi relative entropies for the range of $\alpha < 1$ for which data processing holds.

Proof:

1. For $\omega_{A'B'} = \mathcal{S}_{AB\to A'B'}(\rho_{AB})$, we have by definition,

$$\boldsymbol{E}(A'; B')_\omega = \inf_{\tau_{A'B'}\in\text{SEP}(A':B')} \boldsymbol{D}(\omega_{A'B'}\|\tau_{A'B'}) \qquad (9.2.13)$$

$$= \inf_{\tau_{A'B'}\in\text{SEP}(A':B')} \boldsymbol{D}(\mathcal{S}_{AB\to A'B'}(\rho_{AB})\|\tau_{A'B'}). \qquad (9.2.14)$$

Now, recall that every separable channel $\mathcal{S}_{AB\to A'B'}$ takes $\sigma_{AB} \in \text{SEP}(A : B)$ to a state in $\text{SEP}(A' : B')$, as shown already in (4.6.64)–(4.6.65). Therefore, restricting the optimization in (9.2.14) leads to

$$\boldsymbol{E}(A'; B')_\omega \leq \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \boldsymbol{D}(\mathcal{S}_{AB\to A'B'}(\rho_{AB})\|\mathcal{S}_{AB\to A'B'}(\sigma_{AB})) \qquad (9.2.15)$$

$$\leq \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}) \qquad (9.2.16)$$

$$= \boldsymbol{E}(A; B)_\rho, \qquad (9.2.17)$$

as required, where we used the data-processing inequality for the generalized divergence to obtain the second inequality.

2. We have

$$\boldsymbol{E}(A;B)_\rho = \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}). \tag{9.2.18}$$

If $\rho_{AB} \in \mathrm{SEP}(A:B)$, then the state ρ_{AB} itself achieves the minimum in (9.2.18) because $\boldsymbol{D}(\rho_{AB}\|\rho_{AB}) = 0$. We thus have $\boldsymbol{E}(A;B)_\rho = 0$. On the other hand, if $\boldsymbol{E}(A;B)_\rho = 0$, then there exists a separable state σ^*_{AB} such that $\boldsymbol{D}(\rho_{AB}\|\sigma^*_{AB}) = 0$, which by assumption implies that $\rho_{AB} = \sigma^*_{AB}$, i.e., that ρ_{AB} is separable.

3. By definition, the optimization in the definition of $\boldsymbol{E}(A_1A_2;B_1B_2)_{\rho\otimes\omega}$ is over the set $\mathrm{SEP}(A_1A_2 : B_1B_2)$. It is straightforward to see that this set contains states of the form $\xi_{A_1B_1}\otimes\tau_{A_2B_2}$, where $\xi_{A_1B_1} \in \mathrm{SEP}(A_1:B_1)$ and $\tau_{A_2B_2} \in \mathrm{SEP}(A_2:B_2)$. By restricting the optimization to such states, and by using additivity of the generalized divergence $\boldsymbol{D}$, we obtain

$$\boldsymbol{E}(A_1A_2;B_1B_2)_{\rho\otimes\omega} \le \boldsymbol{D}(\rho_{A_1B_1}\otimes\omega_{A_2B_2}\|\xi_{A_1B_1}\otimes\tau_{A_2B_2}) \tag{9.2.19}$$

$$= \boldsymbol{D}(\rho_{A_1B_1}\|\xi_{A_1B_1}) + \boldsymbol{D}(\omega_{A_2B_2}\|\tau_{A_2B_2}). \tag{9.2.20}$$

Since $\xi_{A_1B_1} \in \mathrm{SEP}(A_1:B_1)$ and $\tau_{A_2B_2} \in \mathrm{SEP}(A_2:B_2)$ are arbitrary, the inequality in (9.2.10) follows.

4. We have

$$\boldsymbol{E}(A;B)_{\overline{\rho}} = \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}\middle\|\sigma_{AB}\right). \tag{9.2.21}$$

Let us restrict the optimization over all separable states to an optimization over sets $\{\sigma^x_{AB}\}_{x\in\mathcal{X}}$ of separable states indexed by the alphabet $\mathcal{X}$. Then, $\sum_{x\in\mathcal{X}} p(x)\sigma^x_{AB}$ is a separable state because the set of separable states is convex. Therefore, using the joint convexity of $\boldsymbol{D}$, we obtain

$$\boldsymbol{E}(A;B)_{\overline{\rho}} \le \inf_{\{\sigma^x_{AB}\}_x\subset\mathrm{SEP}(A:B)} \boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}\middle\|\sum_{x\in\mathcal{X}} p(x)\sigma^x_{AB}\right) \tag{9.2.22}$$

$$\le \inf_{\{\sigma^x_{AB}\}_x\subset\mathrm{SEP}(A:B)} \sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho^x_{AB}\|\sigma^x_{AB}) \tag{9.2.23}$$

$$\le \sum_{x\in\mathcal{X}} p(x) \inf_{\sigma^x_{AB}\in\mathrm{SEP}(A:B)} \boldsymbol{D}(\rho^x_{AB}\|\sigma^x_{AB}) \tag{9.2.24}$$

$$= \sum_{x\in\mathcal{X}} p(x)\boldsymbol{E}(A;B)_{\rho^x}, \tag{9.2.25}$$

as required. ■

We now delve a bit more into particular examples of the generalized divergence of entanglement, which are based on the relative entropy and the Petz–, sandwiched, and geometric Rényi relative entropies.

Proposition 9.17

The relative entropy of entanglement is invariant under classical communication; i.e., (9.1.6) holds with E set to E_R.

PROOF: Let ρ_{XAB} be a classical–quantum state of the form in (9.1.5). Let $\{\sigma^x_{AB}\}_{x\in\mathcal{X}}$ be an arbitrary set of separable states, and set

$$\sigma_{XAB} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \sigma^x_{AB}. \tag{9.2.26}$$

Consider that

$$\begin{aligned} E_R(XA;B)_\rho &= \inf_{\sigma_{XAB}\in\mathrm{SEP}(XA:B)} D(\rho_{XAB}\|\sigma_{XAB}) && (9.2.27)\\ &\leq D(\rho_{XAB}\|\sigma_{XAB}) && (9.2.28)\\ &= \sum_{x\in\mathcal{X}} p(x) D(\rho^x_{AB}\|\sigma^x_{AB}), && (9.2.29)\end{aligned}$$

where the last equality follows from the direct-sum property of relative entropy in (7.2.27). Since the inequality holds for every set $\{\sigma^x_{AB}\}_{x\in\mathcal{X}}$ of separable states, we conclude that

$$E_R(XA;B)_\rho \leq \sum_{x\in\mathcal{X}} p(x) E_R(A;B)_{\rho^x}. \tag{9.2.30}$$

Now suppose that σ_{XAB} is an arbitrary separable state of the systems $XA|B$ (here we assume that the system X is not necessarily classical). After performing the completely dephasing channel $\Delta_X(\cdot) \coloneqq \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X(\cdot)|x\rangle\langle x|_X$ on the X system of σ_{XAB}, the resulting state is a classical–quantum state of the following form:

$$\Delta_X(\sigma_{XAB}) = \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma^x_{AB}, \tag{9.2.31}$$

where $q(x)$ is a probability distribution and $\{\sigma^x_{AB}\}_{x\in\mathcal{X}}$ is a set of separable states. Then consider that

$$D(\rho_{XAB}\|\sigma_{XAB}) \geq D(\Delta_X(\rho_{XAB})\|\Delta_X(\sigma_{XAB})) \tag{9.2.32}$$

$$= \sum_{x\in\mathcal{X}} p(x) D(\rho^x_{AB} \| \sigma^x_{AB}) + D(p\|q) \tag{9.2.33}$$

$$\geq \sum_{x\in\mathcal{X}} p(x) D(\rho^x_{AB} \| \sigma^x_{AB}) \tag{9.2.34}$$

$$\geq \sum_{x\in\mathcal{X}} p(x) E_R(A;B)_{\rho^x}. \tag{9.2.35}$$

The first inequality follows from the data-processing inequality for relative entropy (Theorem 7.4). The equality follows from the direct-sum property of relative entropy in (7.2.27). The second inequality follows from the non-negativity of the classical relative entropy $D(p\|q)$. The final inequality follows from the definition of the relative entropy of entanglement and the fact that σ^x_{AB} is separable. Since the chain of inequalities holds for every separable state σ_{XAB}, we conclude that

$$E_R(XA;B)_\rho \geq \sum_{x\in\mathcal{X}} p(x) E_R(A;B)_{\rho^x}. \tag{9.2.36}$$

Putting together (9.2.30) and (9.2.36), and noting that the same argument applies when exchanging the roles of Alice and Bob, we conclude the statement of the proposition. ■

As an immediate corollary of Proposition 9.17, Lemma 9.2, and Property 1. of Proposition 9.16, we conclude that the relative entropy of entanglement is a selective LOCC monotone. However, we can conclude something stronger, which is what we prove in Proposition 9.19 below after defining selective separable monotonicity.

Definition 9.18 Selective Separable Monotonicity

As a generalization of selective LOCC monotonicity defined in (9.1.14) and in the spirit of the selective PPT monotonicity from Definition 9.9, we say that a function $E : \mathrm{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ is a selective separable monotone if it satisfies

$$E(\rho_{AB}) \geq \sum_{x\in\mathcal{X}:p(x)>0} p(x) E(\omega^x_{A'B'}), \tag{9.2.37}$$

for every bipartite state ρ_{AB} and separable instrument $\{\mathcal{S}^x_{AB\to A'B'}\}_{x\in\mathcal{X}}$, with

$$p(x) := \mathrm{Tr}[\mathcal{S}^x_{AB\to A'B'}(\rho_{AB})], \tag{9.2.38}$$

$$\omega^x_{A'B'} := \frac{1}{p(x)} \mathcal{S}^x_{AB\to A'B'}(\rho_{AB}). \tag{9.2.39}$$

A separable instrument is such that every map $\mathcal{S}^x_{AB\to A'B'}$ is completely positive and separable (with Kraus operators of the form in (4.6.63)), and the sum map $\sum_{x\in\mathcal{X}} \mathcal{S}^x_{AB\to A'B'}$ is trace preserving.
It follows that E is a separable monotone if it is a selective separable monotone, because the former is a special case of the latter in which the alphabet $\mathcal{X}$ has only one letter.

Proposition 9.19 Selective Separable Monotonicity of Relative Entropies of Entanglement

The relative entropy of entanglement is a selective separable monotone; i.e., (9.2.37) holds with E set to E_R. The Petz–, sandwiched, and geometric Rényi relative entropies of entanglement are selective separable monotones for the range $\alpha > 1$ for which data processing holds.

PROOF: Let us begin with the relative entropy of entanglement. Let ρ_{AB} be an arbitrary bipartite state, let $\{\mathcal{S}^x_{AB\to A'B'}\}_{x\in\mathcal{X}}$ be a separable instrument, and let $\mathcal{S}_{AB\to XA'B'}$ denote the following quantum channel:

$$\mathcal{S}_{AB\to XA'B'}(\omega_{AB}) := \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \mathcal{S}^x_{AB\to A'B'}(\omega_{AB}). \tag{9.2.40}$$

Let $\tau_{XA'B'} := \mathcal{S}_{AB\to XA'B'}(\rho_{AB})$, and note that

$$\tau_{XA'B'} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \tau^x_{A'B'}, \tag{9.2.41}$$

for some probability distribution $p(x)$ and set $\{\tau^x_{A'B'}\}_{x\in\mathcal{X}}$ of states. Then consider that

$$E_R(A;B)_\rho \geq E_R(XA';B')_\tau \tag{9.2.42}$$

$$= \sum_{x\in\mathcal{X}} p(x)E_R(A';B')_{\tau^x}. \tag{9.2.43}$$

The inequality follows because E_R is monotone under separable channels (Property 1. of Proposition 9.16), and the equality follows from Proposition 9.17.

Let us now consider proving the statement for the Petz–Rényi relative entropy of entanglement for $\alpha \in (1,2]$. Consider the same channel $\mathcal{S}_{AB\to XA'B'}$ and state

$\tau_{XA'B'}$ defined above. Let σ_{AB} be an arbitrary separable state, and consider that

$$\mathcal{S}_{AB\to XA'B'}(\sigma_{AB}) = \sum_{x\in\mathcal{X}} q(x)|x\rangle\langle x|_X \otimes \sigma^x_{A'B'}, \tag{9.2.44}$$

for some probability distribution $q(x)$ and set $\{\sigma^x_{A'B'}\}_{x\in\mathcal{X}}$ of separable states. Then we find that

$$Q_\alpha(\rho_{AB}\|\sigma_{AB}) \geq Q_\alpha(\mathcal{S}_{AB\to XA'B'}(\rho_{AB})\|\mathcal{S}_{AB\to XA'B'}(\sigma_{AB})) \tag{9.2.45}$$
$$= \sum_{x\in\mathcal{X}:p(x)>0} p(x)^\alpha q(x)^{1-\alpha} Q_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'}) \tag{9.2.46}$$
$$= \sum_{x\in\mathcal{X}:p(x)>0} p(x) \left(\frac{p(x)}{q(x)}\right)^{\alpha-1} Q_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'}). \tag{9.2.47}$$

The first inequality follows from the data-processing inequality for the Petz–Rényi relative quasi-entropy (Theorem 7.24), and the first equality follows from its direct-sum property (see (7.4.46)). Now applying the monotonicity and concavity of the function $(\cdot) \to \frac{1}{\alpha-1}\log_2(\cdot)$ for $\alpha \in (1, 2]$, we find that

$$D_\alpha(\rho_{AB}\|\sigma_{AB})$$
$$= \frac{1}{\alpha-1}\log_2 Q_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{9.2.48}$$
$$\geq \frac{1}{\alpha-1}\log_2\left[\sum_{x\in\mathcal{X}:p(x)>0} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} Q_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'})\right] \tag{9.2.49}$$
$$\geq \sum_{x\in\mathcal{X}:p(x)>0} p(x)\frac{1}{\alpha-1}\log_2\left[\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} Q_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'})\right] \tag{9.2.50}$$
$$= \sum_{x\in\mathcal{X}:p(x)>0} p(x)\left[\log_2\left(\frac{p(x)}{q(x)}\right) + D_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'})\right] \tag{9.2.51}$$
$$= D(p\|q) + \sum_{x\in\mathcal{X}:p(x)>0} p(x) D_\alpha(\tau^x_{A'B'}\|\sigma^x_{A'B'}) \tag{9.2.52}$$
$$\geq \sum_{x\in\mathcal{X}:p(x)>0} p(x) E_\alpha(A';B')_{\tau^x}. \tag{9.2.53}$$

The final two equalities follow by direct evaluation and applying definitions. The final inequality follows because $D(p\|q) \geq 0$ for probability distributions p and

q, and it also follows from the definition of the Petz–Rényi relative entropy of entanglement and the fact that the state $\sigma^x_{A'B'}$ is separable. Since the inequality holds for every separable state σ_{AB}, we conclude the desired inequality:

$$E_\alpha(A;B)_\rho \geq \sum_{x\in\mathcal{X}:p(x)>0} p(x)E_\alpha(A';B')_{\tau^x}. \tag{9.2.54}$$

By applying the same method of proof for the sandwiched and geometric Rényi relative entropies for the range of $\alpha > 1$ for which data processing holds, along with their data processing and direct-sum properties, we conclude the same inequality for the sandwiched and geometric Rényi relative entropies of entanglement. ■

The following additional facts are known specifically about the relative entropy of entanglement and the Petz–, sandwiched, and geometric Rényi relative entropies of entanglement.

Proposition 9.20

1. For every bipartite state ρ_{AB},

$$E_R(A;B)_\rho \geq \max\{I(A\rangle B)_\rho, I(B\rangle A)_\rho\}, \tag{9.2.55}$$
$$E_\alpha(A;B)_\rho \geq \max\{I_\alpha(A\rangle B)_\rho, I_\alpha(B\rangle A)_\rho\}, \tag{9.2.56}$$
$$\widetilde{E}_\alpha(A;B)_\rho \geq \max\{\widetilde{I}_\alpha(A\rangle B)_\rho, \widetilde{I}_\alpha(B\rangle A)_\rho\} \tag{9.2.57}$$
$$\widehat{E}_\alpha(A;B)_\rho \geq \max\{\widehat{I}_\alpha(A\rangle B)_\rho, \widehat{I}_\alpha(B\rangle A)_\rho\}, \tag{9.2.58}$$

   where the last three inequalities hold for the range of α for which data processing holds.

2. For every pure bipartite state ψ_{AB},

$$E_R(A;B)_\psi = H(A)_\psi, \tag{9.2.59}$$
$$E_\alpha(A;B)_\psi = H_{\frac{1}{\alpha}}(A)_\psi, \tag{9.2.60}$$
$$\widetilde{E}_\alpha(A;B)_\psi = H_{\frac{\alpha}{2\alpha-1}}(A)_\psi, \tag{9.2.61}$$
$$\widehat{E}_\alpha(A;B)_\psi = H_{\frac{1}{2}}(A)_\psi, \tag{9.2.62}$$

   where the last three equalities hold for the range of α for which data processing holds.

REMARK: Observe that for pure states, the relative entropy of entanglement is equal to the entanglement of formation (see (9.1.40)).

PROOF:

1. Let σ_{AB} be an arbitrary separable state, which can be written as

$$\sigma_{AB} = \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x, \tag{9.2.63}$$

where $\mathcal{X}$ is some finite alphabet $p : \mathcal{X} \to [0, 1]$ is a probability distribution, and $\{\omega_A^x\}_{x\in\mathcal{X}}$, $\{\tau_B^x\}_{x\in\mathcal{X}}$ are sets of states. Since every ω_A^x is a state (in particular, since all of its eigenvalues are bounded from above by one), the inequality $\omega_A^x \leq \mathbb{1}_A$ holds, which implies that $\mathbb{1}_A \otimes \tau_B^x \geq \omega_A^x \otimes \tau_B^x$ for all $x \in \mathcal{X}$, which thus implies that

$$\mathbb{1}_A \otimes \sigma_B = \sum_{x\in\mathcal{X}} p(x)\mathbb{1}_A \otimes \tau_B^x \geq \sum_{x\in\mathcal{X}} p(x)\omega_A^x \otimes \tau_B^x = \sigma_{AB}. \tag{9.2.64}$$

Therefore, using property 2.(d) in Proposition 7.3, we have that

$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{9.2.65}$$
$$\geq \inf_{\sigma_B} D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{9.2.66}$$
$$= I(A\rangle B)_\rho \tag{9.2.67}$$

for all separable states, where the optimization is with respect to every state σ_B on the right-hand side, and where we have used the expression in (7.2.92) for coherent information. We thus have

$$E_R(A; B)_\rho = \inf_{\sigma_{AB}\in\text{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}) \geq I(A\rangle B)_\rho. \tag{9.2.68}$$

By the same argument, but flipping the roles of Alice and Bob, we conclude that

$$E_R(A; B)_\rho \geq I(B\rangle A)_\rho. \tag{9.2.69}$$

Combining this inequality with the one in (9.2.68) leads to the desired result.

The same proof, but using (9.2.64) and Property 4. of Propositions 7.26, 7.35, and 7.46, leads to the inequalities in (9.2.56)–(9.2.58).

2. Let
$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}|e_k\rangle_A \otimes |f_k\rangle_B \tag{9.2.70}$$
be the Schmidt decomposition of $|\psi\rangle_{AB}$, where r is the Schmidt rank, $\lambda_k > 0$ for all $1 \leq k \leq r$, and $\{|e_k\rangle_A\}_{k=1}^{r}$, $\{|f_k\rangle_B\}_{k=1}^{r}$ are orthonormal sets of vectors. Since the entropy $H(AB)_\psi$ vanishes for all pure states, we immediately have
$$I(A\rangle B)_\psi = H(B)_\psi = H(A)_\psi = I(B\rangle A)_\psi, \tag{9.2.71}$$
where the equality $H(B)_\psi = H(A)_\psi$ follows from the Schmidt decomposition in (9.2.70), which tells us that the reduced states ψ_A and ψ_B have the same non-zero eigevalues. Based on the fact that $E_R(A;B)_\psi \geq I(A\rangle B)_\psi$, which we just proved above, we thus have the lower bound
$$E_R(A;B)_\psi \geq H(A)_\psi. \tag{9.2.72}$$
The same reasoning, but using the lower bounds in (9.2.56)–(9.2.58), as well as (7.11.139)–(7.11.141), leads to the inequalities
$$E_\alpha(A;B)_\psi \geq H_{\frac{1}{\alpha}}(A)_\psi, \tag{9.2.73}$$
$$\widetilde{E}_\alpha(A;B)_\psi \geq H_{\frac{\alpha}{2\alpha-1}}(A)_\psi, \tag{9.2.74}$$
$$\widehat{E}_\alpha(A;B)_\psi \geq H_{\frac{1}{2}}(A)_\psi. \tag{9.2.75}$$
For the reverse inequality, let
$$\Pi \coloneqq \sum_{k=1}^{r} |e_k\rangle\langle e_k|_A \otimes |f_k\rangle\langle f_k|_B \tag{9.2.76}$$
be the projection onto the r^2-dimension subspace of $\mathcal{H}_{AB}$ on which ψ_{AB} is supported. Also, define a channel $\mathcal{N}$ as
$$\mathcal{N}(X_{AB}) \coloneqq \Pi X_{AB} \Pi + (\mathbb{1}_{AB} - \Pi) X_{AB} (\mathbb{1}_{AB} - \Pi). \tag{9.2.77}$$
Note that, by definition, $\mathcal{N}(\psi_{AB}) = \psi_{AB}$. Also, let σ_B be a state, and, for $p(k) \coloneqq \langle f_k|\sigma_B|f_k\rangle_B$, set
$$\overline{\sigma}_{AB} \coloneqq \sum_{k=1}^{r} p(k)|e_k\rangle\langle e_k|_A \otimes |f_k\rangle\langle f_k|_B, \tag{9.2.78}$$

which is a separable state. It is straightforward to show that

$$\Pi(\mathbb{1}_A \otimes \sigma_B)\Pi = \overline{\sigma}_{AB}. \tag{9.2.79}$$

Also, we have that

$$\begin{aligned} &D(\mathcal{N}(\psi_{AB})\|\mathcal{N}(\mathbb{1}_A \otimes \sigma_B)) \\ &\qquad = D(\psi_{AB}\|\overline{\sigma}_{AB} + (\mathbb{1}_{AB} - \Pi)(\mathbb{1}_A \otimes \sigma_B)(\mathbb{1}_{AB} - \Pi)) && (9.2.80) \\ &\qquad = D(\psi_{AB}\|\overline{\sigma}_{AB}), && (9.2.81) \end{aligned}$$

because the operator $(\mathbb{1}_{AB} - \Pi)(\mathbb{1}_A \otimes \sigma_B)(\mathbb{1}_{AB} - \Pi)$ is supported on the space orthogonal to the support of ψ_{AB}. Therefore, by the data-processing inequality for quantum relative entropy, we obtain

$$\begin{aligned} E_R(A;B)_\psi &= \inf_{\sigma_{AB} \in \mathrm{SEP}(A:B)} D(\psi_{AB}\|\sigma_{AB}) && (9.2.82) \\ &\leq D(\psi_{AB}\|\overline{\sigma}_{AB}) && (9.2.83) \\ &= D(\mathcal{N}(\psi_{AB})\|\mathcal{N}(\mathbb{1}_A \otimes \sigma_B)) && (9.2.84) \\ &\leq D(\psi_{AB}\|\mathbb{1}_A \otimes \sigma_B). && (9.2.85) \end{aligned}$$

Since the inequality holds for every state σ_B, we conclude that

$$E_R(A;B)_\psi \leq \inf_{\sigma_B} D(\psi_{AB}\|\mathbb{1}_A \otimes \sigma_B) = I(A\rangle B)_\psi = H(A)_\psi. \tag{9.2.86}$$

Combining this with (9.2.72) gives us $E_R(A;B)_\psi = H(A)_\psi$, as required.

Applying the same method of proof, but using the properties of the Petz–, sandwiched, and geometric Rényi relative entropies, as well as (7.11.139)–(7.11.141), we conclude the inequalities

$$\begin{aligned} E_\alpha(A;B)_\psi &\leq H_{\frac{1}{\alpha}}(A)_\psi, && (9.2.87) \\ \widetilde{E}_\alpha(A;B)_\psi &\leq H_{\frac{\alpha}{2\alpha-1}}(A)_\psi, && (9.2.88) \\ \widehat{E}_\alpha(A;B)_\psi &\leq H_{\frac{1}{2}}(A)_\psi, && (9.2.89) \end{aligned}$$

which, combined with (9.2.73)–(9.2.75), leads to (9.2.60)–(9.2.62). ■

9.2.1 Cone Program Formulations

Computing a generalized divergence of entanglement involves an optimization over the set of separable states, which is known to be hard (please consult the

Bibliographic Notes in Section 9.6). The optimization is made more complicated by the fact that most of the generalized divergences we consider in this book are non-linear functions of the input state (for example, sandwiched Rényi relative entropy). However, both the max-relative entropy and hypothesis testing relative entropy can be formulated as semi-definite programs (SDPs). Indeed, recall from (7.8.4) that

$$D_{\max}(\rho\|\sigma) = \log_2 \inf\{\lambda : \rho \leq \lambda\sigma\}, \tag{9.2.90}$$

and recall from (5.3.125) that

$$D_H^{\varepsilon}(\rho\|\sigma) = -\log_2 \inf\{\operatorname{Tr}[\Lambda\sigma] : 0 \leq \Lambda \leq \mathbb{1},\ \operatorname{Tr}[\Lambda\rho] \geq 1-\varepsilon\}. \tag{9.2.91}$$

As discussed earlier, both of these generalized divergences can be cast into the SDP standard forms in Definition 2.26, and thus their corresponding generalized divergence of entanglement can be formulated as a *cone program*. A cone program is an optimization problem over a convex cone[3] with a convex objective function. An SDP is a special case of a cone program in which the convex cone is the set of positive semi-definite operators.

The convex cone of interest here is the set $\widehat{\text{SEP}}(A\!:\!B)$ of all separable operators, which we define as follows: $X_{AB} \in \widehat{\text{SEP}}(A\!:\!B)$ if there exists a positive integer ℓ and positive semi-definite operators $\{P_A^x\}_{x=1}^{\ell}$ and $\{Q_B^x\}_{x=1}^{\ell}$ such that

$$X_{AB} = \sum_{x=1}^{\ell} P_A^x \otimes Q_B^x. \tag{9.2.92}$$

In what follows, we sometimes employ the shorthands SEP and $\widehat{\text{SEP}}$ when the bipartition is clear from the context.

We now show that the max-relative entropy of entanglement can be written as a cone program.

Proposition 9.21 Cone Program for Max-Relative Entropy of Entanglement

Let ρ_{AB} be a bipartite state. Then,

$$E_{\max}(A;B)_\rho = \log_2 G_{\max}(A;B)_\rho, \tag{9.2.93}$$

[3]A subset C of a vector space is called a cone if $\alpha x \in C$ for every $x \in C$ and $\alpha > 0$. A convex cone is one for for which $\alpha x + \beta y \in C$ for all $\alpha, \beta > 0$ and $x, y \in C$.

where

$$G_{\max}(A;B)_\rho := \inf\left\{\mathrm{Tr}[X_{AB}] : \rho_{AB} \leq X_{AB}, X_{AB} \in \widehat{\mathrm{SEP}}\right\}. \tag{9.2.94}$$

PROOF: Employing the expression in (9.2.90), we find that

$$E_{\max}(A;B)_\rho \quad \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB})$$

$$= \log_2 \inf\{\mu : \rho_{AB} \leq \mu\sigma_{AB},\ \sigma_{AB} \in \mathrm{SEP}\} \tag{9.2.95}$$

$$= \log_2 \inf\left\{\mathrm{Tr}[X_{AB}] : \rho_{AB} \leq X_{AB},\ X_{AB} \in \widehat{\mathrm{SEP}}\right\}, \tag{9.2.96}$$

as required, where in the last step we made the change of variable $\mu\sigma_{AB} \equiv X_{AB}$. Since $\sigma_{AB} \in \mathrm{SEP}(A\!:\!B)$ and $\mu \geq 0$, we have that $X_{AB} \in \widehat{\mathrm{SEP}}(A\!:\!B)$. ■

Next, we show that the hypothesis testing relative entropy of entanglement can be written as a cone program.

Proposition 9.22 Cone Program for Hypothesis Testing Relative Entropy of Entanglement

Let ρ_{AB} be a bipartite state. Then, for all $\varepsilon \in [0,1]$,

$$E_R^\varepsilon(A;B)_\rho = -\log_2 \sup\{\mu(1-\varepsilon) - \mathrm{Tr}[Z_{AB}] : \mu \geq 0, Z_{AB} \geq 0, \\ \sigma_{AB} \in \widehat{\mathrm{SEP}},\ \mu\rho_{AB} \leq \sigma_{AB} + Z_{AB},\ \mathrm{Tr}[\sigma_{AB}] = 1\}. \tag{9.2.97}$$

PROOF: This follows from the definition in (9.2.3) and the dual formulation of the hypothesis testing relative entropy stated in (7.9.5). ■

Recall that SEP = PPT in the case of qubit-qubit and qubit-qutrit states, which means that the optimizations in (9.2.94) and (9.2.97) are SDPs when ρ_{AB} is either a two-qubit state or a qubit-qutrit state.

9.3 Generalized Rains Divergence

As explained in Section 9.1.1, the generalized divergence of entanglement is in general complicated to compute because the set of separable states does not admit a simple characterization, making optimization over separable states difficult. By relaxing the set of separable states to the set PPT′ defined in (9.1.144), which has a simple characterization in terms of semi-definite constraints, we defined the generalized Rains divergence in (9.1.148). Recall that

$$\mathrm{PPT}'(A\!:\!B) = \{\sigma_{AB} : \sigma_{AB} \geq 0,\ \ \|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1\}. \tag{9.3.1}$$

In this section, we investigate properties of the generalized Rains divergence. We start by recalling its definition.

Definition 9.23 Generalized Rains Divergence of a Bipartite State

Let $\boldsymbol{D}$ be a generalized divergence (see Definition 7.15). For every bipartite state ρ_{AB}, we define the *generalized Rains divergence of* ρ_{AB} as

$$\boldsymbol{R}(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}). \tag{9.3.2}$$

If $\boldsymbol{D}$ is continuous in its second argument, then the infimum can be replaced by a minimum.

Since SEP ⊆ PPT′, optimizing over states in PPT′ can never lead to a value that is greater than the value obtained by optimizing over separable states. Therefore, as stated in (9.1.149),

$$\boldsymbol{R}(A;B)_\rho \leq \boldsymbol{E}(A;B)_\rho \tag{9.3.3}$$

for every state ρ_{AB}.

We are particularly interested throughout the rest of this book in the following generalized Rains divergences for every state ρ_{AB}:

1. The *Rains relative entropy of* ρ_{AB},

$$R(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D(\rho_{AB}\|\sigma_{AB}), \tag{9.3.4}$$

where $D(\rho_{AB}\|\sigma_{AB})$ is the quantum relative entropy of ρ_{AB} and σ_{AB} (Definition 7.1). For two-qubit states, it is known that the infimum in (9.3.4) is achieved by a separable state, which means that $E_R(A;B)_\rho = R(A;B)_\rho$ for two-qubit states ρ_{AB} (please consult the Bibliographic Notes in Section 9.6).

2. The *ε-hypothesis testing Rains relative entropy of* ρ_{AB},

$$R_H^\varepsilon(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D_H^\varepsilon(\rho_{AB}\|\sigma_{AB}), \tag{9.3.5}$$

where $D_H^\varepsilon(\rho_{AB}\|\sigma_{AB})$ is the ε-hypothesis testing relative entropy of ρ_{AB} and σ_{AB} (Definition 7.65).

3. The *sandwiched Rényi Rains relative entropy of* ρ_{AB},

$$\widetilde{R}_\alpha(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}), \tag{9.3.6}$$

where $\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB})$, $\alpha \in [1/2, 1) \cup (1,\infty)$, is the sandwiched Rényi relative entropy of ρ_{AB} and σ_{AB} (Definition 7.28). Note that $\widetilde{R}_\alpha(A;B)_\rho$ is monotonically increasing in α for all ρ_{AB} (see Proposition 7.31). This fact, along with the fact that $\lim_{\alpha\to1}\widetilde{D}_\alpha = D$ (see Proposition 7.30), leads to

$$\lim_{\alpha\to1}\widetilde{R}_\alpha(A;B)_\rho = R(A;B)_\rho \tag{9.3.7}$$

for every state ρ_{AB}. See Appendix 10.A for details of the proof.

4. The *max-Rains relative entropy of* ρ_{AB},

$$R_{\max}(A;B)_\rho \coloneqq \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}), \tag{9.3.8}$$

where $D_{\max}(\rho_{AB}\|\sigma_{AB})$ is the max-relative entropy of ρ_{AB} and σ_{AB} (Definition 7.58). Using the fact that $\lim_{\alpha\to\infty}\widetilde{D}_\alpha = D_{\max}$ (see Proposition 7.61), we find that

$$R_{\max}(A;B)_\rho = \lim_{\alpha\to\infty}\widetilde{R}_\alpha(A;B)_\rho \tag{9.3.9}$$

for every state ρ_{AB}. See Appendix 10.A for details of the proof. Due to this fact, along with the fact that $\widetilde{R}_\alpha(A;B)_\rho$ is monotonically increasing in α for all ρ_{AB}, we have that

$$R_{\max}(A;B)_\rho \geq \widetilde{R}_\alpha(A;B)_\rho \tag{9.3.10}$$

for all $\alpha \in (1,\infty)$ and every state ρ_{AB}.

The following inequalities relate the log-negativity, as defined in (9.1.96), to the Rains relative entropy, the sandwiched Rényi Rains relative entropy, and the max-Rains relative entropy:

Proposition 9.24 Log-Negativity to Rains Relative Entropies

For a bipartite state ρ_{AB}, the following inequalities hold

$$R(A;B)_\rho \leq R_{\max}(A;B)_\rho \leq E_N(A;B)_\rho. \tag{9.3.11}$$

Furthermore, for all $\alpha, \beta \in [1/2, 1) \cup (1, \infty)$ such that $\alpha < \beta$, we have that

$$\widetilde{R}_\alpha(A;B)_\rho \leq \widetilde{R}_\beta(A;B)_\rho. \tag{9.3.12}$$

Proof: The inequality $R(A;B)_\rho \leq R_{\max}(A;B)_\rho$ and the inequality in (9.3.12) are a direct consequence of the monotonicity in α of the sandwiched Rényi relative entropy (Proposition 7.31), as well as (9.3.7) and (9.3.9).

To see the inequality $R_{\max}(A;B)_\rho \leq E_N(A;B)_\rho$, consider picking

$$\sigma_{AB} = \frac{\rho_{AB}}{\|\mathrm{T}_B(\rho_{AB})\|_1} \tag{9.3.13}$$

in (9.3.8). For this choice, we have that $\sigma_{AB} \geq 0$ and $\|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1$, so that $\sigma_{AB} \in \mathrm{PPT}'(A\!:\!B)$. Thus

$$\begin{aligned} R_{\max}(A;B)_\rho &\leq D_{\max}(\rho_{AB}\|\sigma_{AB}) && (9.3.14)\\ &= D_{\max}(\rho_{AB}\|\rho_{AB}) + \log_2 \|\mathrm{T}_B(\rho_{AB})\|_1 && (9.3.15)\\ &= E_N(A;B)_\rho. && (9.3.16) \end{aligned}$$

The first equality follows by direct evaluation using Definition 7.58. The last equality follows because $D_{\max}(\rho_{AB}\|\rho_{AB}) = 0$ and from the definition in (9.1.96). ■

Proposition 9.25 Properties of Generalized Rains Divergence

Let $\boldsymbol{D}$ be a generalized divergence, and consider the generalized Rains divergence $\boldsymbol{R}(A;B)_\rho$ of a state ρ_{AB} as defined in (9.3.2).

1. *PPT monotonicity*: For every completely PPT-preserving channel $\mathcal{P}_{AB\to A'B'}$,

$$\boldsymbol{R}(A;B)_\rho \geq \boldsymbol{R}(A';B')_\omega, \tag{9.3.17}$$

where $\omega_{A'B'} = \mathcal{P}_{AB\to A'B'}(\rho_{AB})$. In other words, the generalized Rains divergence is monotonically non-increasing under completely PPT-preserving channels. Since every LOCC channel is a completely PPT-preserving channel (Propositions 4.24 and 4.29), the generalized Rains divergence is monotone non-increasing under LOCC channels, and thus it is an entanglement measure as per Definition 9.1.

2. *Subadditivity*: If $\boldsymbol{D}$ is additive for product positive semi-definite operators, i.e., $\boldsymbol{D}(\rho \otimes \omega \| \sigma \otimes \tau) = \boldsymbol{D}(\rho\|\sigma) + \boldsymbol{D}(\omega\|\tau)$, then for every two quantum states $\rho_{A_1B_1}$ and $\omega_{A_2B_2}$ the generalized Rains relative entropy is subadditive:

$$\boldsymbol{R}(A_1A_2; B_1B_2)_{\rho\otimes\omega} \leq \boldsymbol{R}(A_1; B_1)_\rho + \boldsymbol{R}(A_2; B_2)_\omega. \tag{9.3.18}$$

3. *Convexity*: If $\boldsymbol{D}$ is jointly convex, meaning that

$$\boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x\right) \leq \sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho_{AB}^x\|\sigma_{AB}^x) \tag{9.3.19}$$

for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0, 1]$, set $\{\rho_{AB}^x\}_{x\in\mathcal{X}}$ of states, set $\{\sigma_{AB}^x\}_{x\in\mathcal{X}}$ of positive semi-definite operators, then the generalized Rains divergence is convex:

$$\boldsymbol{R}(A; B)_{\overline{\rho}} \leq \sum_{x\in\mathcal{X}} p(x)\boldsymbol{R}(A; B)_{\rho^x}, \tag{9.3.20}$$

where $\overline{\rho}_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x$.

Properties 1. and 2. are satisfied when the generalized divergence is the quantum relative entropy, the Petz–, sandwiched, and geometric Rényi relative entropies, and the max-relative entropy. Property 3. is satisfied when the generalized divergence is the quantum relative entropy and the Petz–, sandwiched, and geometric Rényi relative entropies for the range of $\alpha < 1$ for which data processing holds.

REMARK: Note that the generalized Rains divergence is generally not a faithful entanglement measure. Although $\boldsymbol{R}(A; B)_\rho = 0$ for all separable states ρ_{AB} due to the containment $\text{SEP}(A : B) \subseteq \text{PPT}'(A : B)$, the converse statement is not generally true because the infimum in the definition of $\boldsymbol{R}(A; B)_\rho$ is not generally achieved by a separable state.

PROOF:

1. For $\omega_{A'B'} = \mathcal{P}_{AB\to A'B'}(\rho_{AB})$, we have by definition,

$$\boldsymbol{R}(A';B')_\omega = \inf_{\tau_{A'B'}\in \mathrm{PPT}'(A':B')} \boldsymbol{D}(\omega_{A'B'}\|\tau_{A'B'}) \tag{9.3.21}$$

$$= \inf_{\tau_{A'B'}\in \mathrm{PPT}'(A':B')} \boldsymbol{D}(\mathcal{P}_{AB\to A'B'}(\rho_{AB})\|\tau_{A'B'}). \tag{9.3.22}$$

Now, recall from Lemma 9.14 that the set PPT′ is closed under completely PPT-preserving channels. Based on this, it follows that the output operators of the completely PPT-preserving channel $\mathcal{P}_{AB\to A'B'}$ are in the set $\mathrm{PPT}'(A'\!:\!B')$. In other words, we have

$$\{\mathcal{P}_{AB\to A'B'}(\sigma_{AB}) : \sigma_{AB}\in \mathrm{PPT}'(A\!:\!B)\} \subseteq \mathrm{PPT}'(A'\!:\!B'). \tag{9.3.23}$$

Therefore, restricting the optimization in (9.3.22) to this set leads to

$$\boldsymbol{R}(A';B')_\omega \le \inf_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} \boldsymbol{D}(\mathcal{P}_{AB\to A'B'}(\rho_{AB})\|\mathcal{P}_{AB\to A'B'}(\sigma_{AB})) \tag{9.3.24}$$

$$\le \inf_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} \boldsymbol{D}(\rho_{AB}\|\sigma_{AB}) \tag{9.3.25}$$

$$= \boldsymbol{R}(A;B)_\rho, \tag{9.3.26}$$

as required, where to obtain the second inequality we used the data-processing inequality for the generalized divergence.

2. By definition, the optimization in the definition of $\boldsymbol{R}(A_1A_2;B_1B_2)_{\rho\otimes\omega}$ is over the set

$$\mathrm{PPT}'(A_1A_2\!:\!B_1B_2)$$
$$= \{\sigma_{A_1A_2B_1B_1} : \sigma_{A_1A_2B_1B_2} \ge 0, \left\|\mathrm{T}_{B_1B_2}(\sigma_{A_1A_2B_1B_2})\right\|_1 \le 1\}, \tag{9.3.27}$$

which contains operators of the form $\sigma_{A_1A_2B_1B_2} = \xi_{A_1B_1}\otimes\tau_{A_2B_2}$, where $\xi_{A_1B_1}\in \mathrm{PPT}'(A_1\!:\!B_1)$ and $\tau_{A_2B_2}\in\mathrm{PPT}'(A_2\!:\!B_2)$. By restricting the optimization to such tensor product operators, and by using the additivity of the generalized divergence $\boldsymbol{D}$, we obtain

$$\boldsymbol{R}(A_1A_2;B_1B_2)_{\rho\otimes\omega} \le \boldsymbol{D}(\rho_{A_1B_1}\otimes\omega_{A_2B_2}\|\xi_{A_1B_1}\otimes\tau_{A_2B_2}) \tag{9.3.28}$$

$$= \boldsymbol{D}(\rho_{A_1B_1}\|\xi_{A_1B_1}) + \boldsymbol{D}(\omega_{A_2B_2}\|\tau_{A_2B_2}). \tag{9.3.29}$$

Since $\xi_{A_1B_1}\in\mathrm{PPT}'(A_1:B_1)$ and $\tau_{A_2B_2}\in\mathrm{PPT}'(A_2:B_2)$ are arbitrary, the inequality in (9.3.18) follows.

3. We have

$$\boldsymbol{R}(A;B)_{\overline{\rho}} = \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\| \sigma_{AB}\right). \tag{9.3.30}$$

Let us restrict the optimization over all PPT′ operators to an optimization over sets $\{\sigma_{AB}^x\}_{x\in\mathcal{X}}$ of PPT′ operators indexed by the alphabet $\mathcal{X}$. Then, because $\mathrm{PPT}'(A:B)$ is a convex set, we have that $\sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x \in \mathrm{PPT}'(A:B)$. Therefore, using the joint convexity of $\boldsymbol{D}$, we obtain

$$\boldsymbol{R}(A;B)_{\overline{\rho}} \leq \inf_{\{\sigma_{AB}^x\}_x\subset\mathrm{PPT}'(A:B)} \boldsymbol{D}\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\| \sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x\right) \tag{9.3.31}$$

$$\leq \inf_{\{\sigma_{AB}^x\}_x\subset\mathrm{PPT}'(A:B)} \sum_{x\in\mathcal{X}} p(x)\boldsymbol{D}(\rho_{AB}^x\|\sigma_{AB}^x) \tag{9.3.32}$$

$$\leq \sum_{x\in\mathcal{X}} p(x) \inf_{\sigma_{AB}^x\in\mathrm{PPT}'(A:B)} \boldsymbol{D}(\rho_{AB}^x\|\sigma_{AB}^x) \tag{9.3.33}$$

$$= \sum_{x\in\mathcal{X}} p(x)\boldsymbol{R}(A;B)_{\rho^x}, \tag{9.3.34}$$

as required. ■

By Proposition 9.25, the sandwiched Rényi Rains relative entropy $\widetilde{R}_\alpha$ is convex for $\alpha \in [1/2, 1)$. Although convexity does not hold for $\alpha > 1$, we do have quasi-convexity, which we now prove.

Proposition 9.26 Quasi-Convexity of Rényi Rains Relative Entropy

Let $p : \mathcal{X} \to [0,1]$ be a probability distribution over a finite alphabet $\mathcal{X}$, and let $\{\rho_{AB}^x\}_{x\in\mathcal{X}}$ be a set of states. Then, for all $\alpha \in (1,\infty)$,

$$\widetilde{R}_\alpha(A;B)_{\overline{\rho}} \leq \max_{x\in\mathcal{X}} \widetilde{R}_\alpha(A;B)_{\rho^x}, \tag{9.3.35}$$

where $\overline{\rho}_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x$.

Proof: We have

$$\widetilde{R}_\alpha(A;B)_{\overline{\rho}} = \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\| \sigma_{AB}\right). \tag{9.3.36}$$

Let us restrict the optimization over all PPT′ operators to an optimization over sets $\{\sigma_{AB}^x\}_{x\in\mathcal{X}}$ of PPT′ operators indexed by the alphabet $\mathcal{X}$. Then, because PPT′$(A\!:\!B)$ is a convex set, we have that $\sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x \in \mathrm{PPT}'(A:B)$. Let us also recall from (7.5.174) that the sandwiched Rényi relative entropy is jointly quasi-convex, meaning that

$$\widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\Vert \sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x\right) \leq \max_{x\in\mathcal{X}} \widetilde{D}_\alpha(\rho_{AB}^x\Vert\sigma_{AB}^x), \tag{9.3.37}$$

We thus obtain

$$\widetilde{R}_\alpha(A;B)_\rho \leq \inf_{\{\sigma_{AB}^x\}_x\subset\mathrm{PPT}'(A:B)} \widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)\rho_{AB}^x \middle\Vert \sum_{x\in\mathcal{X}} p(x)\sigma_{AB}^x\right) \tag{9.3.38}$$

$$\leq \inf_{\{\sigma_{AB}^x\}_x\subset\mathrm{PPT}'(A:B)} \max_{x\in\mathcal{X}} \widetilde{D}_\alpha(\rho_{AB}^x\Vert\sigma_{AB}^x) \tag{9.3.39}$$

$$\leq \max_{x\in\mathcal{X}} \inf_{\sigma_{AB}^x\in\mathrm{PPT}'(A:B)} \widetilde{D}_\alpha(\rho_{AB}^x\Vert\sigma_{AB}^x) \tag{9.3.40}$$

$$= \max_{x\in\mathcal{X}} \widetilde{R}_\alpha(A;B)_{\rho^x}, \tag{9.3.41}$$

as required. ■

9.3.1 Semi-Definite Program Formulations

One of the advantages of using the generalized Rains divergence as an entanglement measure is that the set PPT′ involved in its definition has a simple characterization in terms of semi-definite constraints. Indeed, let us recall that

$$\mathrm{PPT}'(A\!:\!B) = \{\sigma_{AB} : \sigma_{AB} \geq 0,\ \ \|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1\}. \tag{9.3.42}$$

Using the expression in (9.1.102) for $\|\mathrm{T}_B(\sigma_{AB})\|_1$, this set can equivalently be written as follows:

$$\mathrm{PPT}'(A\!:\!B) = \{\sigma_{AB} : \sigma_{AB} \geq 0,\ \exists K_{AB}, L_{AB} \geq 0 \text{ such that } \mathrm{Tr}[K_{AB}+L_{AB}] \leq 1,\ \mathrm{T}_B(K_{AB}-L_{AB}) = \sigma_{AB}\}, \tag{9.3.43}$$

which can be further simplified to

$$\mathrm{PPT}'(A\!:\!B) = \{\mathrm{T}_B(K_{AB} - L_{AB}) : \mathrm{T}_B(K_{AB} - L_{AB}) \geq 0,$$
$$K_{AB}, L_{AB} \geq 0,\ \mathrm{Tr}[K_{AB} + L_{AB}] \leq 1\}, \quad (9.3.44)$$

In this section, we show how these characterizations of the set PPT′ allow us to compute both the max-Rains relative entropy and the hypothesis testing Rains relative entropy via semi-definite programs (Section 2.4).

We first consider the max-Rains relative entropy, which we recall is defined as

$$R_{\max}(A;B)_\rho = \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} D_{\max}(\rho_{AB} \| \sigma_{AB}). \quad (9.3.45)$$

Let us also recall from (7.8.4) that $D_{\max}(\rho_{AB}\|\sigma_{AB})$ can be written as follows:

$$D_{\max}(\rho_{AB}\|\sigma_{AB}) = \log_2 \inf\{\lambda : \rho_{AB} \leq \lambda \sigma_{AB}\}. \quad (9.3.46)$$

As shown in the discussion after (7.8.4), the optimization in the equation above is a semi-definite program (SDP). This, along with the definition in (9.3.42) of the set $\mathrm{PPT}'(A\!:\!B)$, leads to the following SDP formulation for $R_{\max}$.

Proposition 9.27 SDPs for the Max-Rains Relative Entropy

Let ρ_{AB} be a bipartite state. Then the max-Rains relative entropy can be written as

$$R_{\max}(A;B)_\rho = \log_2 W_{\max}(A;B)_\rho, \quad (9.3.47)$$

where

$$W_{\max}(A;B)_\rho$$
$$= \inf_{K_{AB}, L_{AB} \geq 0} \{\mathrm{Tr}[K_{AB} + L_{AB}] : \mathrm{T}_B[K_{AB} - L_{AB}] \geq \rho_{AB}\}, \quad (9.3.48)$$
$$= \sup_{Y_{AB} \geq 0} \{\mathrm{Tr}[Y_{AB}\rho_{AB}] : \|\mathrm{T}_B[Y_{AB}]\|_\infty \leq 1\}. \quad (9.3.49)$$

Proof: First, we establish the equality for $W(A;B)_\rho$ in (9.3.48). Due to the fact that the infimum over PPT′ operators in the definition of $R_{\max}$ can be achieved, we have that

$$R_{\max}(A;B)_\rho = \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}) \quad (9.3.50)$$
$$= \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} \log_2 \inf\{\lambda : \rho_{AB} \leq \lambda\sigma_{AB}\} \quad (9.3.51)$$

$$= \log_2 \inf\{\lambda : \rho_{AB} \leq \lambda\sigma_{AB},\ \sigma_{AB} \geq 0,\ \ \|\mathrm{T}_B[\sigma_{AB}]\|_1 \leq 1\}. \tag{9.3.52}$$

The constraint $\rho_{AB} \leq \lambda\sigma_{AB}$ implies (by taking the trace on both sides of the inequality) that $\lambda \geq 1$. This in turn implies that $\lambda\sigma_{AB} \geq 0$ and that $\|\mathrm{T}_B[\lambda\sigma_{AB}]\|_1 \leq \lambda$. So we have

$$R_{\max}(A;B)_\rho = \log_2 \inf\{\lambda : \rho_{AB} \leq \lambda\sigma_{AB},\ \lambda\sigma_{AB} \geq 0,\ \ \|\mathrm{T}_B[\lambda\sigma_{AB}]\|_1 \leq \lambda\} \tag{9.3.53}$$

Let us now make the change of variable $S_{AB} \equiv \lambda\sigma_{AB}$. We then have

$$R_{\max}(A;B)_\rho = \log_2 \inf\{\lambda : \rho_{AB} \leq S_{AB},\ S_{AB} \geq 0,\ \ \|\mathrm{T}_B[S_{AB}]\|_1 \leq \lambda\} \tag{9.3.54}$$

$$= \log_2 \inf\{\|\mathrm{T}_B[S_{AB}]\|_1 : \rho_{AB} \leq S_{AB}\}, \tag{9.3.55}$$

where to obtain the last line we eliminated the constraint $S_{AB} \geq 0$ (because it is implied by $\rho_{AB} \leq S_{AB}$) and we used the fact that $\mu \geq \|\mathrm{T}_B[S_{AB}]\|_1$, meaning that the smallest value of λ is $\|\mathrm{T}_B[S_{AB}]\|_1$.

Now, for an arbitrary operator S_{AB} satisfying $\rho_{AB} \leq S_{AB}$, recall from (2.2.69) that the Jordan–Hahn decomposition of $\mathrm{T}_B[S_{AB}]$ is given by $\mathrm{T}_B(S_{AB}) = K_{AB} - L_{AB}$, with $K_{AB}, L_{AB} \geq 0$ and $K_{AB}L_{AB} = 0$. We then find that

$$\|\mathrm{T}_B[S_{AB}]\|_1 = \mathrm{Tr}[K_{AB} + L_{AB}], \tag{9.3.56}$$

$$S_{AB} = \mathrm{T}_B[K_{AB} - L_{AB}]. \tag{9.3.57}$$

The following inequality thus holds:

$$\inf\{\mathrm{Tr}[K_{AB} + L_{AB}] : \rho_{AB} \leq \mathrm{T}_B[K_{AB} - L_{AB}], K_{AB}, L_{AB} \geq 0\} \\ \leq \inf\{\|\mathrm{T}_B(S_{AB})\|_1 : \rho_{AB} \leq S_{AB}\}. \tag{9.3.58}$$

To see the opposite inequality, let K_{AB} and L_{AB} be arbitrary operators such that $K_{AB}, L_{AB} \geq 0$ and $\rho_{AB} \leq \mathrm{T}_B[K_{AB} - L_{AB}]$. Then, setting $S_{AB} = \mathrm{T}_B[K_{AB} - L_{AB}]$, we find that $\rho_{AB} \leq S_{AB}$ and

$$\|\mathrm{T}_B(S_{AB})\|_1 = \|K_{AB} - L_{AB}\|_1 \tag{9.3.59}$$

$$\leq \|K_{AB}\|_1 + \|L_{AB}\|_1 \tag{9.3.60}$$

$$= \mathrm{Tr}[K_{AB} + L_{AB}], \tag{9.3.61}$$

where we used the triangle inequality. This implies that

$$\inf\{\|\mathrm{T}_B(S_{AB})\|_1 : \rho_{AB} \leq S_{AB}\} \leq \inf\{\mathrm{Tr}[K_{AB} + L_{AB}] : \rho_{AB} \leq \mathrm{T}_B(K_{AB} - L_{AB}), K_{AB}, L_{AB} \geq 0\}. \quad (9.3.62)$$

Putting together (9.3.58) and (9.3.62), we conclude that $W_{\max}(A;B)_\rho$ is given by the equality in (9.3.48).

To see the equality in (9.3.49), we employ semi-definite programming duality (see Section 2.4). We first put (9.3.48) into standard form (see Definition 2.26) as follows:

$$\inf_{X \geq 0}\{\mathrm{Tr}[CX] : \Phi(X) \geq D\}, \quad (9.3.63)$$

with

$$X = \begin{pmatrix} K_{AB} & 0 \\ 0 & L_{AB} \end{pmatrix}, \quad C = \begin{pmatrix} \mathbb{1}_{AB} & 0 \\ 0 & \mathbb{1}_{AB} \end{pmatrix}, \quad (9.3.64)$$

$$\Phi(X) = \mathrm{T}_B[K_{AB} - L_{AB}], \quad D = \rho_{AB}. \quad (9.3.65)$$

The dual program is then given by

$$\sup_{Y \geq 0}\{\mathrm{Tr}[DY] : \Phi^\dagger(Y) \leq C\}. \quad (9.3.66)$$

In order to determine the adjoint $\Phi^\dagger$ of Φ, consider that

$$\begin{aligned} \mathrm{Tr}[\Phi(X)Y] &= \mathrm{Tr}[\mathrm{T}_B[K_{AB} - L_{AB}]Y_{AB}] && (9.3.67) \\ &= \mathrm{Tr}[(K_{AB} - L_{AB})\mathrm{T}_B[Y_{AB}]] && (9.3.68) \\ &= \mathrm{Tr}\left[\begin{pmatrix} K_{AB} & 0 \\ 0 & L_{AB} \end{pmatrix}\begin{pmatrix} \mathrm{T}_B[Y_{AB}] & 0 \\ 0 & -\mathrm{T}_B[Y_{AB}] \end{pmatrix}\right]. && (9.3.69) \end{aligned}$$

Therefore,

$$\Phi^\dagger(Y) = \begin{pmatrix} \mathrm{T}_B[Y_{AB}] & 0 \\ 0 & -\mathrm{T}_B[Y_{AB}] \end{pmatrix} \quad (9.3.70)$$

so that $\Phi^\dagger(Y) \leq C$ is equivalent to

$$\begin{pmatrix} \mathrm{T}_B[Y_{AB}] & 0 \\ 0 & -\mathrm{T}_B[Y_{AB}] \end{pmatrix} \leq \begin{pmatrix} \mathbb{1}_{AB} & 0 \\ 0 & \mathbb{1}_{AB} \end{pmatrix}. \quad (9.3.71)$$

This is equivalent to the condition $-\mathbb{1}_{AB} \leq \mathrm{T}_B[Y_{AB}] \leq \mathbb{1}_{AB}$, which is equivalent to $\|\mathrm{T}_B[Y_{AB}]\|_\infty \leq 1$ (see (2.4.26)). We thus conclude that (9.3.66) is equal to (9.3.49).

To arrive at the equality in (9.3.48)–(9.3.49), we need to verify that strong duality holds, and so we check the conditions of Theorem 2.28. For the dual program in (9.3.49), pick $Y_{AB} = \frac{1}{2}\mathbb{1}_{AB}$. Then the constraints in (9.3.48) are strict, because $Y_{AB} > 0$ and $\|\mathrm{T}_B(Y_{AB})\|_\infty < 1$ for this choice. Furthermore, for the primal in (9.3.48), pick K_{AB} and L_{AB} equal to the positive and negative parts of $\mathrm{T}_B(\rho_{AB})$, respectively. These are positive semi-definite by definition and satisfy $\rho_{AB} = \mathrm{T}_B(K_{AB} - L_{AB})$, so that they are feasible for the primal. ■

It is worthwhile to observe the close connection of the SDP in (9.3.48), related to the max-Rains relative entropy, to that in (9.1.102), related to the log-negativity. In fact, the SDP in (9.3.48) is a relaxation of that in (9.1.102), and this leads to an alternate proof of the rightmost inequality from (9.3.11):

$$R_{\max}(A;B)_\rho \leq E_N(A;B)_\rho, \tag{9.3.72}$$

holding for every bipartite state ρ_{AB}.

Furthermore, the form of the SDP in (9.3.48), along with the same proof given for Proposition 9.10, except with (9.1.106) and (9.1.112) replaced with inequality constraints, leads to the following conclusion:

Proposition 9.28 Max-Rains Relative Entropy is a Selective PPT Monotone

The max-Rains relative entropy is a selective PPT monotone; i.e., (9.1.103) holds with E set to $R_{\max}$.

Due to the additivity of $D_{\max}$, Proposition 9.25 implies that $R_{\max}$ is subadditive:

$$R_{\max}(A_1A_2;B_1B_2)_{\rho\otimes\omega} \leq R_{\max}(A_1;B_1)_\rho + R_{\max}(A_2;B_2)_\omega, \tag{9.3.73}$$

where the inequality holds for all states $\rho_{A_1B_1}$ and $\omega_{A_2B_2}$. Using the dual formulation in (9.3.48) for $R_{\max}$, we find that the reverse inequality also holds, implying that $R_{\max}$ is an additive entanglement measure.

Proposition 9.29 Additivity of Max-Rains Relative Entropy

Let $\rho_{A_1B_1}$ and $\omega_{A_2B_2}$ be quantum states. Then,

$$R_{\max}(A_1A_2;B_1B_2)_{\rho\otimes\omega} = R_{\max}(A_1;B_1)_\rho + R_{\max}(A_2;B_2)_\omega. \tag{9.3.74}$$

PROOF: We prove the inequality reverse to the one in (9.3.73). To this end, we employ the dual formulation of $R_{\max}$ in (9.3.49). Let $Y_{A_1B_1}$ and $S_{A_2B_2}$ be arbitrary operators satisfying

$$\left\|\mathrm{T}_{B_1}[Y_{A_1B_1}]\right\|_\infty \leq 1, \quad Y_{A_1B_1} \geq 0, \tag{9.3.75}$$

$$\left\|\mathrm{T}_{B_2}[S_{A_2B_2}]\right\|_\infty \leq 1, \quad S_{A_2B_2} \geq 0. \tag{9.3.76}$$

Then it follows from multiplicativity of the Schatten ∞-norm under tensor products (see (2.2.95)) that

$$\left\|\mathrm{T}_{B_1B_2}[Y_{A_1B_1} \otimes S_{A_2B_2}]\right\|_\infty = \left\|\mathrm{T}_{B_1}[Y_{A_1B_1}] \otimes \mathrm{T}_{B_2}[S_{A_2B_2}]\right\|_\infty \tag{9.3.77}$$

$$= \left\|\mathrm{T}_{B_1}[Y_{A_1B_1}]\right\|_\infty \left\|\mathrm{T}_{B_2}[S_{A_2B_2}]\right\|_\infty \tag{9.3.78}$$

$$\leq 1. \tag{9.3.79}$$

Furthermore, we have that $Y_{A_1B_1} \otimes S_{A_2B_2} \geq 0$. So it follows that

$$\log_2 \mathrm{Tr}[Y_{A_1B_1}\rho_{A_1B_1}] + \log_2 \mathrm{Tr}[S_{A_2B_2}\omega_{A_2B_2}]$$

$$= \log_2\left(\mathrm{Tr}[Y_{A_1B_1}\rho_{A_1B_1}]\mathrm{Tr}[S_{A_2B_2}\omega_{A_2B_2}]\right) \tag{9.3.80}$$

$$= \log_2\left(\mathrm{Tr}[\left(Y_{A_1B_1} \otimes S_{A_2B_2}\right)\left(\rho_{A_1B_1} \otimes \omega_{A_2B_2}\right)]\right) \tag{9.3.81}$$

$$\leq R_{\max}(A_1A_2; B_1B_2)_{\rho\otimes\omega}. \tag{9.3.82}$$

The inequality follows because $Y_{A_1B_1} \otimes S_{A_2B_2}$ is a particular operator satisfying the constraints in (9.3.48) for $R_{\max}(A_1A_2; B_1B_2)_{\rho\otimes\omega}$. Since the inequality holds for all $Y_{A_1B_1}$ and $S_{A_2B_2}$ satisfying (9.3.75)–(9.3.76), the superadditivity inequality

$$R_{\max}(A_1A_2; B_1B_2)_{\rho\otimes\omega} \geq R_{\max}(A_1; B_1)_\rho + R_{\max}(A_2; B_2)_\omega \tag{9.3.83}$$

follows. ■

We now consider the hypothesis testing Rains relative entropy, which is defined as

$$R_H^\varepsilon(A; B)_\rho = \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D_H^\varepsilon(\rho_{AB}\|\sigma_{AB}), \tag{9.3.84}$$

for $\varepsilon \in [0, 1]$. Recall the primal and dual formulations of D_H^ε from Proposition 7.66. Using the dual formulation, we obtain the following:

Proposition 9.30 SDP for Hypothesis Testing Rains Relative Entropy

Let ρ_{AB} be a bipartite state. Then, the hypothesis testing Rains relative entropy can be written as

$$R_H^\varepsilon(A;B)_\rho = -\log_2 W_H^\varepsilon(A;B)_\rho, \tag{9.3.85}$$

for all $\varepsilon \in [0,1]$, where

$$\begin{aligned} &W_H^\varepsilon(A;B)_\rho \\ &:= \sup_{\substack{\mu\geq 0, Z_{AB},\\ K_{AB},L_{AB}\geq 0}} \{\mu(1-\varepsilon) - \text{Tr}[Z_{AB}] : \mu\rho_{AB} \leq \text{T}_B(K_{AB}-L_{AB}) + Z_{AB}, \\ &\qquad\qquad \text{T}_B(K_{AB}-L_{AB}) \geq 0,\ \text{Tr}[K_{AB}+L_{AB}] \leq 1\} \qquad (9.3.86) \\ &= \inf_{M_{AB},N_{AB}\geq 0} \{\|\text{T}_B(M_{AB}+N_{AB})\|_\infty : \text{Tr}[M_{AB}\rho_{AB}] \geq 1-\varepsilon,\ M_{AB} \leq \mathbb{1}_{AB}\}. \qquad (9.3.87) \end{aligned}$$

Proof: Using the dual SDP formulation of the hypothesis testing relative entropy from Proposition 7.66, we obtain

$$W_H^\varepsilon(A;B)_\rho = \sup_{\mu\geq 0, Z_{AB}\geq 0, \sigma_{AB}\in\text{PPT}'(A:B)} \{\mu(1-\varepsilon) - \text{Tr}[Z_{AB}] : \mu\rho_{AB} \leq \sigma_{AB} + Z_{AB}\} \tag{9.3.88}$$

Now, combining with the characterization of $\text{PPT}'(A:B)$ from (9.3.44), we conclude (9.3.86).

The SDP for $W_H^\varepsilon(A;B)_\rho$ in (9.3.86) can be written in the standard form of (2.4.3) as

$$\sup_{X\geq 0} \{\text{Tr}[AX] : \Phi(X) \leq B\}, \tag{9.3.89}$$

where

$$X = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & Z_{AB} & 0 & 0 \\ 0 & 0 & K_{AB} & 0 \\ 0 & 0 & 0 & L_{AB} \end{pmatrix}, \quad A = \begin{pmatrix} 1-\varepsilon & 0 & 0 & 0 \\ 0 & -\mathbb{1}_{AB} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{9.3.90}$$

$$\Phi(X) = \begin{pmatrix} \mu\rho_{AB} - \text{T}_B(K_{AB}-L_{AB}) - Z_{AB} & 0 & 0 \\ 0 & -\text{T}_B(K_{AB}-L_{AB}) & 0 \\ 0 & 0 & \text{Tr}[K_{AB}+L_{AB}] \end{pmatrix}, \tag{9.3.91}$$

$$B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9.3.92}$$

Setting

$$Y = \begin{pmatrix} M_{AB} & 0 & 0 \\ 0 & N_{AB} & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \tag{9.3.93}$$

we compute the dual map $\Phi^\dagger$ as follows:

$$\begin{aligned} &\operatorname{Tr}[Y\Phi(X)] \\ &= \operatorname{Tr}[M_{AB}\,(\mu\rho_{AB} - \mathrm{T}_B(K_{AB} - L_{AB}) - Z_{AB})] - \operatorname{Tr}[N_{AB}\mathrm{T}_B(K_{AB} - L_{AB})] \\ &\qquad + \lambda \operatorname{Tr}[K_{AB} + L_{AB}] \qquad (9.3.94) \\ &= \mu \operatorname{Tr}[M_{AB}\rho_{AB}] - \operatorname{Tr}[M_{AB}Z_{AB}] + \operatorname{Tr}[(\lambda \mathbb{1}_{AB} - \mathrm{T}_B(M_{AB} + N_{AB}))\, K_{AB}] \\ &\qquad + \operatorname{Tr}[(\lambda \mathbb{1}_{AB} + \mathrm{T}_B(M_{AB} + N_B))\, L_{AB}], \qquad (9.3.95) \end{aligned}$$

which implies that

$$\Phi^\dagger(Y) = \begin{pmatrix} \operatorname{Tr}[M_{AB}\rho_{AB}] & 0 & 0 & 0 \\ 0 & -M_{AB} & 0 & 0 \\ 0 & 0 & \lambda\mathbb{1}_{AB} - \mathrm{T}_B(M_{AB} + N_{AB}) & 0 \\ 0 & 0 & 0 & \lambda\mathbb{1}_{AB} + \mathrm{T}_B(M_{AB} + N_B) \end{pmatrix}. \tag{9.3.96}$$

Then using the standard form of the dual program in (2.4.4), i.e.,

$$\inf_{Y \geq 0} \left\{ \operatorname{Tr}[BY] : \Phi^\dagger(Y) \geq A \right\}, \tag{9.3.97}$$

we find that the dual SDP is given by

$$\inf_{\lambda, M_{AB}, N_{AB} \geq 0} \{\lambda : \operatorname{Tr}[M_{AB}\rho_{AB}] \geq 1 - \varepsilon,\ M_{AB} \leq \mathbb{1}_{AB}, \\ \lambda\mathbb{1}_{AB} \pm \mathrm{T}_B(M_{AB} + N_{AB}) \geq 0\}. \tag{9.3.98}$$

This can alternatively be written as

$$\inf_{M_{AB}, N_{AB} \geq 0} \{\|\mathrm{T}_B(M_{AB} + N_{AB})\|_\infty : \operatorname{Tr}[M_{AB}\rho_{AB}] \geq 1 - \varepsilon,\ M_{AB} \leq \mathbb{1}_{AB}\}. \tag{9.3.99}$$

Finally, we verify that strong duality holds, by applying Theorem 2.28. A strictly feasible choice for the primal variables is $\mu = 1$, $Z_{AB} = \mathbb{1}_{AB}$, $K_{AB} = \pi_{AB}/2$, $L_{AB} = \pi_{AB}/3$. A feasible choice for the dual variables is $M_{AB} = \mathbb{1}_{AB}$ and $N_{AB} = \mathbb{1}_{AB}$. ■

9.4 Squashed Entanglement

In this section, we investigate the properties of the squashed entanglement, which we introduced as an entanglement measure in Section 9.1.1. We first recall the definition of squashed entanglement from (9.1.162).

Definition 9.31 Squashed Entanglement

Let ρ_{AB} be a bipartite state. Then, the squashed entanglement is defined as

$$E_{\text{sq}}(A;B)_\rho := \frac{1}{2} \inf_{\omega_{ABE}} \{ I(A;B|E)_\omega : \text{Tr}_E[\omega_{ABE}] = \rho_{AB} \}, \tag{9.4.1}$$

where the quantum conditional mutual information $I(A;B|E)_\omega$ is defined in (7.1.11) as

$$I(A;B|E)_\omega = H(A|E)_\omega + H(B|E)_\omega - H(AB|E)_\omega \tag{9.4.2}$$
$$= H(AE)_\omega - H(E)_\omega + H(BE)_\omega - H(ABE)_\omega \tag{9.4.3}$$
$$= H(B|E)_\omega - H(B|AE)_\omega \tag{9.4.4}$$

The optimization in the definition of squashed entanglement is with respect to all extensions ω_{ABE} of ρ_{AB}, with the extension system E having arbitrarily large, yet finite dimension, which means that the infimum cannot in general be replaced by a minimum. The fact that the extension system can have arbitrarily large, yet finite dimension also means that computing the squashed entanglement is in general difficult; however, we can always place an upper bound on it by calculating the quantum conditional mutual information of a specific extension.

To understand why the quantity in (9.4.1) is called squashed entanglement, it is helpful to think in terms of a cryptographic scenario. This cryptographic perspective turns out to be useful in the context of secret key agreement, which we consider in Chapters 15 and 20. Consider three parties, the protagonists Alice and Bob, as well as an eavesdropper. Alice and Bob possess the quantum systems A and B, respectively, and the eavesdropper possesses a system E, such that the global state ω_{ABE} is consistent with the reduced state ρ_{AB} of Alice and Bob. The conditional mutual information $I(A;B|E)_\omega$ corresponds to the correlations between Alice and Bob from the perspective of the eavesdropper. The squashed entanglement $E_{\text{sq}}(A;B)_\rho$ is then an optimization over all possible global states ω_{ABE} such that

$\text{Tr}_E[\omega_{ABE}] = \rho_{AB}$, and this optimization corresponds to the worst possible scenario in which the eavesdropper attempts to "squash down" the correlations of Alice and Bob, i.e., to reduce the value of $I(A;B|E)_\omega$ as much as possible. This cryptographic perspective actually allows us to write the squashed entanglement in an alternative way, which we do in Proposition 9.37 below.

We begin by establishing some basic properties of squashed entanglement. As we will see, the squashed entanglement possesses *all* of the desired properties of an entanglement measure stated at the beginning of Section 9.1.

Proposition 9.32 Properties of Squashed Entanglement

The squashed entanglement $E_{\text{sq}}(A;B)_\rho$ has the following properties.

1. *Non-negativity*: For every bipartite state ρ_{AB},
$$E_{\text{sq}}(A;B)_\rho \geq 0. \tag{9.4.5}$$
2. *Faithfulness*: We have $E_{\text{sq}}(A;B)_\sigma = 0$ if and only if σ_{AB} is a separable state.
3. *Convexity*: Let $\mathcal{X}$ be a finite alphabet, $p : \mathcal{X} \to [0,1]$ a probability distribution on $\mathcal{X}$, and $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ a set of states. Then,
$$E_{\text{sq}}(A;B)_{\overline{\rho}} \leq \sum_{x\in\mathcal{X}} p(x) E_{\text{sq}}(A;B)_{\rho^x}, \tag{9.4.6}$$
where $\overline{\rho}_{AB} = \sum_{x\in\mathcal{X}} p(x)\rho^x_{AB}$.
4. *Monogamy*: For every state $\rho_{A_1A_2B_1B_2}$, the overall squashed entanglement is not smaller than the sum of the individual squashed entanglements:
$$E_{\text{sq}}(A_1A_2;B_1B_2)_\rho \geq E_{\text{sq}}(A_1;B_1)_\rho + E_{\text{sq}}(A_1;B_2)_\rho + E_{\text{sq}}(A_2;B_1)_\rho + E_{\text{sq}}(A_2;B_2)_\rho. \tag{9.4.7}$$
5. *Additivity*: For a tensor-product state $\sigma_{A_1A_2B_1B_2} = \omega_{A_1B_1} \otimes \tau_{A_2B_2}$, the following additivity identity holds:
$$E_{\text{sq}}(A_1A_2;B_1B_2)_\sigma = E_{\text{sq}}(A_1;B_1)_\omega + E_{\text{sq}}(A_2;B_2)_\tau. \tag{9.4.8}$$

PROOF:

1. This follows immediately from the fact that the conditional mutual information of an arbitrary state is non-negative (Theorem 7.6).

2. The statement "if σ_{AB} is a separable state, then $E_{\text{sq}}(A;B)_\sigma = 0$" follows from the line of reasoning in (9.1.157)–(9.1.159) used to motivate the definition of squashed entanglement. For a proof of the converse statement, please consult the Bibliographic Notes in Section 9.6.

3. Let ω^x_{ABE} denote an arbitrary extension of ρ^x_{AB}. Then

$$\omega_{ABEX} \coloneqq \sum_{x\in\mathcal{X}} p(x)\omega^x_{ABE} \otimes |x\rangle\langle x|_X \tag{9.4.9}$$

is a particular extension of $\overline{\rho}_{AB}$. It follows that

$$2\cdot E_{\text{sq}}(A;B)_{\overline{\rho}} \leq I(A;B|EX)_\omega \tag{9.4.10}$$

$$= \sum_{x\in\mathcal{X}} p(x) I(A;B|E)_{\omega^x}, \tag{9.4.11}$$

where to obtain the equality we used the direct-sum property of conditional mutual information (see Proposition 7.9). Since the inequality holds for arbitrary extensions of ρ^x_{AB}, we conclude the desired inequality.

4. To see the inequality in (9.4.7), let $\omega_{A_1A_2B_1B_2E}$ be an arbitrary extension of $\rho_{A_1A_2B_1B_2}$. Then by two applications of the chain rule for conditional mutual information, we find that

$$I(A_1A_2;B_1B_2|E)_\omega$$

$$= I(A_1;B_1B_2|E)_\omega + I(A_2;B_1B_2|EA_1)_\omega \tag{9.4.12}$$

$$= I(A_1;B_1|E)_\omega + I(A_1;B_2|EB_1)_\omega + I(A_2;B_2|EA_1)_\omega + I(A_2;B_1|EA_1B_2)_\omega \tag{9.4.13}$$

$$\geq 2\left[E_{\text{sq}}(A_1;B_1)_\rho + E_{\text{sq}}(A_1;B_2)_\rho + E_{\text{sq}}(A_2;B_1)_\rho + E_{\text{sq}}(A_2;B_2)_\rho\right]. \tag{9.4.14}$$

The inequality follows because $\omega_{A_1B_1E}$ is a particular extension of the reduced state $\rho_{A_1B_1}$, the state $\omega_{A_1B_1B_2E}$ is a particular extension of the reduced state $\rho_{A_1B_2}$, the state $\omega_{A_1A_2B_2E}$ is a particular extension of the reduced state $\rho_{A_2B_2}$, and $\omega_{A_1A_2B_1B_2E}$ is a particular extension of the reduced state $\rho_{A_2B_1}$. Since the extension $\omega_{A_1A_2B_1B_2E}$ is arbitrary, optimizing over all such extensions on the left-hand side of the inequality above gives (9.4.7).

5. To see the equality in (9.4.8), first consider that for a tensor-product state $\omega_{A_1B_1} \otimes \tau_{A_2B_2}$, the reduced state on systems A_1B_2 is the product state $\omega_{A_1} \otimes \tau_{B_2}$, and the reduced state on systems A_2B_1 is the product state $\omega_{A_2} \otimes \tau_{B_1}$. Thus, faithfulness implies that $E_{\text{sq}}(A_1; B_2)_\sigma = E_{\text{sq}}(A_2; B_1)_\sigma = 0$, and then the monogamy inequality in (9.4.7) implies that

$$E_{\text{sq}}(A_1A_2; B_1B_2)_\sigma \geq E_{\text{sq}}(A_1; B_1)_\omega + E_{\text{sq}}(A_2; B_2)_\tau. \tag{9.4.15}$$

Now let $\omega_{A_1B_1E_1}$ be an extension of $\omega_{A_1B_1}$, and let $\tau_{A_2B_2E_2}$ be an extension of $\tau_{A_2B_2}$. Then $\omega_{A_1B_1E_1} \otimes \tau_{A_2B_2E_2}$ is an extension of $\omega_{A_1B_1} \otimes \tau_{A_2B_2}$. We then have that

$$2 \cdot E_{\text{sq}}(A_1A_2; B_1B_2)_\sigma \leq I(A_1A_2; B_1B_2|E_1E_2)_{\omega\otimes\tau} \tag{9.4.16}$$

$$= I(A_1; B_1|E_1)_\omega + I(A_2; B_2|E_2)_\tau, \tag{9.4.17}$$

where the equality follows from the additivity of conditional mutual information with respect to tensor-product states (Proposition 7.9). Since the extensions $\omega_{A_1B_1E_1}$ and $\tau_{A_2B_2E_2}$ are arbitrary, the following inequality holds:

$$E_{\text{sq}}(A_1A_2; B_1B_2)_\sigma \leq E_{\text{sq}}(A_1; B_1)_\omega + E_{\text{sq}}(A_2; B_2)_\tau. \tag{9.4.18}$$

We then conclude (9.4.8) by combining (9.4.15) and (9.4.18). ■

Let us now prove that the squashed entanglement is indeed an entanglement measure, as per Definition 9.1.

Theorem 9.33 Selective LOCC Monotonicity of Squashed Entanglement

The squashed entanglement is a selective LOCC monotone, so that (9.1.14) holds with E set to E_{sq}.

Proof: We prove that the conditions of Lemma 9.2 hold and then we apply it. The first part of the proof follows from the fact that conditional mutual information does not increase under the action of local channels (see Proposition 7.9): for a tripartite state ξ_{ABE} and local channels $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$,

$$I(A; B|E)_\xi \geq I(A'; B'|E)_\zeta, \tag{9.4.19}$$

where $\zeta_{A'B'E} \coloneqq (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\xi_{ABE})$. After incorporating optimizations, it then follows immediately that squashed entanglement does not increase under the action of local channels:

$$E_{\mathrm{sq}}(A;B)_\rho \geq E_{\mathrm{sq}}(A';B')_\kappa, \tag{9.4.20}$$

where $\kappa_{A'B'} := (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{AB})$.

Also, the squashed entanglement is invariant under classical communication, which follows from Lemma 9.34 below. Now applying Lemma 9.2, we conclude the statement of the theorem. ■

Lemma 9.34 Invariance of Squashed Entanglement Under Classical Communication

Let ρ_{XAB} be a classical–quantum state of the following form:

$$\rho_{XAB} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}, \tag{9.4.21}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\rho^x_{AB}\}_{x\in\mathcal{X}}$ is a set of states. Then

$$E_{\mathrm{sq}}(XA;B)_\rho = E_{\mathrm{sq}}(A;BX)_\rho = \sum_{x\in\mathcal{X}} p(x)E_{\mathrm{sq}}(A;B)_{\rho^x}. \tag{9.4.22}$$

Proof: Note that discarding or appending a classical state $|x\rangle\langle x|_{X_A}$ is a local channel, so that $E_{\mathrm{sq}}(A;B)_{\rho^x} = E_{\mathrm{sq}}(XA;B)_{|x\rangle\langle x|\otimes\rho^x}$. In Proposition 9.32, we proved that the squashed entanglement is a convex function. Using this fact, it follows that

$$\sum_{x\in\mathcal{X}} p(x)E_{\mathrm{sq}}(A;B)_{\rho^x} = \sum_{x\in\mathcal{X}} p(x)E_{\mathrm{sq}}(XA;B)_{|x\rangle\langle x|\otimes\rho^x} \geq E_{\mathrm{sq}}(XA;B)_\rho. \tag{9.4.23}$$

Similarly, we have that

$$\sum_{x\in\mathcal{X}} p(x)E_{\mathrm{sq}}(A;B)_{\rho^x} \geq E_{\mathrm{sq}}(A;BX)_\rho. \tag{9.4.24}$$

So it suffices to establish the following inequality:

$$\sum_{x\in\mathcal{X}} p(x)E_{\mathrm{sq}}(A;B)_{\rho^x} \leq \min\{E_{\mathrm{sq}}(XA;B)_\rho,\ E_{\mathrm{sq}}(A;BX)_\rho\}. \tag{9.4.25}$$

To this end, let ω_{XABE} be an arbitrary extension of ρ_{XAB}. After the action of a local completely dephasing channel $\Delta_X(\cdot) \coloneqq \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X(\cdot)|x\rangle\langle x|_X$, it follows that the state $\theta_{XABE} \coloneqq \Delta_X(\omega_{XABE})$ has the following form:

$$\theta_{XABE} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_A} \otimes \theta^x_{ABE}, \tag{9.4.26}$$

where θ^x_{ABE} is an extension of ρ^x_{AB}. To see this, let $|\phi^x\rangle_{ABR}$ purify ρ^x_{AB} for each $x \in \mathcal{X}$, and consider that

$$|\varphi\rangle_{XX_EABR} \coloneqq \sum_{x\in\mathcal{X}} \sqrt{p(x)}|x\rangle_X|x\rangle_{X_E}|\phi^x\rangle_{ABR} \tag{9.4.27}$$

is a purification of ρ_{XAB} with purifying systems $X_E R$. By applying Proposition 4.4, we conclude that the extension ω_{XABE} can be realized by the action of a quantum channel $\mathcal{N}_{X_E R\to E}$ on systems $X_E R$ of φ_{XX_EABR}:

$$\omega_{XABE} = \mathcal{N}_{X_E R\to E}(\varphi_{XX_EABR}). \tag{9.4.28}$$

The conclusion stated in (9.4.26) then follows because

$$\theta_{XABE} = (\Delta_X \otimes \mathcal{N}_{X_E R\to E})(\varphi_{XX_EABR}) = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \theta^x_{ABE}, \tag{9.4.29}$$

where $\theta^x_{ABE} \coloneqq \mathcal{N}_{X_E R\to E}(|x\rangle\langle x|_{X_E} \otimes \phi^x_{ABR})$. We then find that

$$2\sum_{x\in\mathcal{X}} p(x)E_{\text{sq}}(A;B)_{\rho^x} \le \sum_x p(x)I(A;B|E)_{\theta^x} \tag{9.4.30}$$

$$= I(A;B|EX)_\theta \tag{9.4.31}$$

$$= I(XA;B|E)_\theta - I(X;B|E)_\theta \tag{9.4.32}$$

$$\le I(XA;B|E)_\theta \tag{9.4.33}$$

$$\le I(XA;B|E)_\omega. \tag{9.4.34}$$

The last equality follows from the chain rule for conditional mutual information and the second-to-last inequality from non-negativity of conditional mutual information: $I(X;B|E)_\rho \ge 0$. The final inequality holds because the conditional mutual information does not increase under the action of a local channel on system X (see Proposition 7.9). Since the inequality holds for an arbitrary extension of ρ_{XAB}, we conclude that

$$\sum_{x\in\mathcal{X}} p(x)E_{\text{sq}}(A;B)_{\rho^x} \le E_{\text{sq}}(XA;B)_\rho. \tag{9.4.35}$$

A proof for the other inequality

$$\sum_{x\in\mathcal{X}} p(x) E_{\text{sq}}(A;B)_{\rho^x} \leq E_{\text{sq}}(A;BX)_\rho \tag{9.4.36}$$

follows similarly. ■

In Section 9.1.1, we introduced the entanglement of formation as an entanglement measure. The following inequality relates the entanglement of formation to the squashed entanglement:

Proposition 9.35 Squashed Entanglement and Entanglement of Formation

The entanglement of formation is never smaller than the squashed entanglement:

$$E_{\text{sq}}(A;B)_\rho \leq E_F(A;B)_\rho, \tag{9.4.37}$$

for every bipartite state ρ_{AB}.

PROOF: We already alluded to this inequality when introducing squashed entanglement in Section 9.1.1.4. If the extension ω_{ABE} is restricted to be of the form

$$\omega_{ABE} = \sum_x p(x) \phi^x_{AB} \otimes |x\rangle\langle x|, \tag{9.4.38}$$

where $p(x)$ is a probability distribution and $\{\phi^x_{AB}\}_x$ is a set of pure states satisfying $\rho_{AB} = \sum_x p(x)\phi^x_{AB}$, then it follows that

$$E_{\text{sq}}(A;B)_\rho \leq \frac{1}{2} I(A;B|X)_\omega = H(A|X)_\omega. \tag{9.4.39}$$

Since the inequality holds for all such extensions, the inequality in (9.4.37) follows by applying the definition of entanglement of formation in (9.1.42). ■

For pure bipartite states, the entanglement of formulation is simply the entropy of the reduced state of one of the subsystems. It turns out that the squashed entanglement reduces to the same quantity for pure bipartite states.

Proposition 9.36 Squashed Entanglement for Pure States

Let ψ_{AB} be a pure bipartite state. Then, the squashed entanglement of ψ_{AB} is equal to the entropy of its reduced state on A:

$$E_{\text{sq}}(A;B)_\psi = H(A)_\psi. \tag{9.4.40}$$

PROOF: Every extension ω_{ABE} of the pure bipartite state ψ_{AB} is a product state of the form $\omega_{ABE} = \psi_{AB} \otimes \rho_E$ for some state ρ_E. By Proposition 7.9, it follows that

$$I(A;B|E)_{\psi\otimes\rho} = I(A;B)_\psi = 2H(A)_\psi, \tag{9.4.41}$$

where the second equality holds because $H(AB)_\psi = 0$ and $H(A)_\psi = H(B)_\psi$ for a pure bipartite state. We thus have $E_{\text{sq}}(A;B)_\psi = H(A)_\psi$. ■

An immediate consequence of Proposition 9.36 is that, for the maximally entangled state $|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}|i,i\rangle_{AB}$ of Schmidt rank d, the squashed entanglement is equal to $\log_2 d$:

$$E_{\text{sq}}(A;B)_\Phi = \log_2 d. \tag{9.4.42}$$

Let us now return to the discussion after Definition 9.31 on the cryptographic interpretation of squashed entanglement. We stated that the quantum conditional mutual information $I(A;B|E)_\omega$ can be interpreted as the amount of correlations between Alice (A) and Bob (B) from the point of view of an eavesdropper (E), where ω_{ABE} is the joint state shared by all three parties. If the eavesdropper wants to reduce, or "squash" these correlations as much as possible, then we optimize with respect to every state ω_{ABE} that is consistent with the state ρ_{AB} shared by Alice and Bob, leading to the squashed entanglement $E_{\text{sq}}(A;B)_\rho$. Now, recall Proposition 4.4, which states that for every extension ω_{ABE} of a given state ρ_{AB} there exists a quantum channel $\mathcal{S}_{E'\to E}$ such that $\mathcal{S}_{E'\to E}(\psi^\rho_{ABE'}) = \omega_{ABE}$, where $\psi^\rho_{ABE'}$ is a purification of ρ_{AB}. We therefore immediately have the following.

Proposition 9.37 Other Representations of Squashed Entanglement

Let ρ_{AB} be a bipartite quantum state, and let $\psi^\rho_{ABE'}$ be a purification of it. Then the squashed entanglement $E_{\text{sq}}(A;B)_\rho$ can be written as

$$E_{\text{sq}}(A;B)_\rho = \frac{1}{2}\inf_{\mathcal{S}_{E'\to E}}\{I(A;B|E)_\omega : \omega_{ABE} = \mathcal{S}_{E'\to E}(\psi^\rho_{ABE'})\}, \tag{9.4.43}$$

where the infimum is with respect to every quantum channel $\mathcal{S}_{E'\to E}$. Another representation of squashed entanglement is

$$E_{\text{sq}}(A;B)_\rho = \frac{1}{2}\inf_{\mathcal{V}_{E'\to EF}}\left\{H(B|E)_\theta + H(B|F)_\theta : \theta_{BEF} := \mathcal{V}_{E'\to EF}(\psi^\rho_{ABE'})\right\}, \tag{9.4.44}$$

where the infimum is with respect to every isometric channel $\mathcal{V}_{E'\to EF}$.

The act of "squashing" the correlations between Alice and Bob can thus be thought of explicitly in terms of an eavesdropper applying the channel $\mathcal{S}_{E'\to E}$ to their purifying system E' of ρ_{AB}. For this reason, we call $\mathcal{S}_{E'\to E}$ a *squashing channel*.

Proof: The equality (9.4.43) follows immediately from Proposition 4.4.

In order to establish (9.4.44), let $\mathcal{S}_{E'\to E}$ be an arbitrary squashing channel, and let $\mathcal{V}_{E'\to EF}$ be an isometric extension of $\mathcal{S}_{E'\to E}$, so that $\mathcal{S}_{E'\to E} = \text{Tr}_F \circ \mathcal{V}_{E'\to EF}$, where the system F satisfies $d_F \geq \text{rank}(\Gamma^{\mathcal{S}}_{EE'})$. Consider that $\theta_{ABEF} := \mathcal{V}_{E'\to EF}(\psi^\rho_{ABE'})$ is a purification of both ω_{ABE} and θ_{BEF}, i.e., $\omega_{ABE} = \mathcal{S}_{E'\to E}(\psi^\rho_{ABE'}) = \text{Tr}_F[\theta_{ABEF}]$ and $\theta_{BEF} = \text{Tr}_A[\theta_{ABEF}]$. Then, using (9.4.4), it follows that

$$I(A;B|E)_\omega = H(B|E)_\omega - H(B|AE)_\omega \tag{9.4.45}$$

$$= H(B|E)_\theta - H(B|AE)_\theta. \tag{9.4.46}$$

Now, since θ_{ABEF} is a pure state, we have from duality of conditional entropy that

$$-H(B|AE)_\theta = H(B|F)_\theta. \tag{9.4.47}$$

Therefore,

$$I(A;B|E)_\omega = H(B|E)_\theta + H(B|F)_\theta. \tag{9.4.48}$$

We conclude that (9.4.44) holds because the squashing channel $\mathcal{S}_{E'\to E}$ is arbitrary in the development above. ■

We now establish an explicit uniform continuity bound for the squashed entanglement:

Proposition 9.38 Uniform Continuity of Squashed Entanglement

Let ρ_{AB} and σ_{AB} be bipartite states such that the following fidelity bound holds

$$F(\rho_{AB}, \sigma_{AB}) \geq 1 - \varepsilon, \tag{9.4.49}$$

for $\varepsilon \in [0, 1]$. Then the following bound applies to their squashed entanglements:

$$\left| E_{\text{sq}}(A; B)_\rho - E_{\text{sq}}(A; B)_\sigma \right| \leq \sqrt{\varepsilon} \log_2 \min\{d_A, d_B\} + g_2(\sqrt{\varepsilon}), \tag{9.4.50}$$

where

$$g_2(\delta) \coloneqq (\delta + 1) \log_2(\delta + 1) - \delta \log_2 \delta. \tag{9.4.51}$$

Proof: Due to Uhlmann's theorem (Theorem 6.8) and Proposition 4.4, for an arbitrary extension ρ_{ABE} of ρ_{AB}, there exists an extension σ_{ABE} of σ_{AB} such that

$$F(\rho_{ABE}, \sigma_{ABE}) = F(\rho_{AB}, \sigma_{AB}) \geq 1 - \varepsilon. \tag{9.4.52}$$

By the relation between trace distance and fidelity (Theorem 6.14), it follows that

$$\frac{1}{2} \|\rho_{ABE} - \sigma_{ABE}\|_1 \leq \sqrt{\varepsilon}. \tag{9.4.53}$$

Then, applying the uniform continuity of conditional mutual information (Proposition 7.10), we find that

$$2E_{\text{sq}}(A; B)_\sigma \leq I(A; B|E)_\sigma \tag{9.4.54}$$

$$\leq I(A; B|E)_\rho + 2\sqrt{\varepsilon} \log_2 \min\{d_A, d_B\} + 2g_2(\sqrt{\varepsilon}). \tag{9.4.55}$$

Since the extension ρ_{ABE} is arbitrary, it follows that

$$E_{\text{sq}}(A; B)_\sigma \leq E_{\text{sq}}(A; B)_\rho + \sqrt{\varepsilon} \log_2 \min\{d_A, d_B\} + g_2(\sqrt{\varepsilon}). \tag{9.4.56}$$

The other bound,

$$E_{\text{sq}}(A; B)_\rho \leq E_{\text{sq}}(A; B)_\sigma + \sqrt{\varepsilon} \log_2 \min\{d_A, d_B\} + g_2(\sqrt{\varepsilon}), \tag{9.4.57}$$

follows from a similar proof. ■

9.5 Summary

In this chapter, we studied entanglement measures for quantum states and quantum channels. The defining property of an entanglement measure for states is monotonicity under local operations and classical communication (LOCC): a function $E : \mathcal{H}_{AB} \to \mathbb{R}$ is an entanglement measure if $E(\rho_{AB}) \geq E(\mathcal{L}(\rho_{AB}))$ for every bipartite state ρ_{AB} and every LOCC channel $\mathcal{L}$. LOCC monotonicity can be thought of as a special kind of data-processing inequality, and it is a core concept in entanglement theory in the same way that the data-processing inequality is the core concept behind generalized divergence.

An important type of state entanglement measure for our purposes in this book is a divergence-based measure, in which the entanglement in a given bipartite quantum state is quantified by its divergence with the set of separable states. As our divergence, we take a generalized divergence $\boldsymbol{D} : \mathrm{D}(\mathcal{H}) \times \mathrm{L}_+(\mathcal{H}) \to \mathbb{R}$, and we call the resulting quantity *generalized divergence of entanglement*. Due to the data-processing inequality (which holds for a generalized divergence by definition), we immediately obtain LOCC monotonicity for the generalized divergence of entanglement, thus making it an entanglement measure. We also consider the divergence with the larger set of PPT′ operators that contains all separable states, and call the resulting quantity *generalized Rains divergence*.

9.6 Bibliographic Notes

Entanglement theory has a long history, with one of the seminal papers being that of Bennett et al. (1996c) (see also Bennett et al. (1996b,a) for earlier works). Bennett et al. (1996c) introduced the resource theory of entanglement, with the separable states as the free states and LOCC channels as the free channels, along with the related operational notions of distillable entanglement and entanglement cost. The review of Horodecki et al. (2009b) is a useful resource for gaining an understanding of notable accomplishments in the area. See also the review of Plenio and Virmani (2007), which focuses specifically on entanglement measures.

The axiomatic approach to defining an entanglement measure that we have taken in this chapter was proposed by Bennett et al. (1996c); Vedral et al. (1997); Vedral and Plenio (1998); Vidal (2000), with LOCC monotonicity emerging as the defining property of an entanglement measure Horodecki et al. (2009b). Vidal (2000) and Horodecki (2005) established simplified conditions for proving that a function is a selective LOCC monotone. Lemma 9.2 is in this spirit.

Entanglement of formation was defined by Bennett et al. (1996c), and they showed that it is a selective LOCC monotone. They also showed that it is an upper bound on entanglement cost, and thus an upper bound on distillable entanglement (we did not define these concepts here, but we will define distillable entanglement in detail in Chapter 13). The uniform continuity bound for entanglement of formation in Proposition 9.4 was given by Winter (2016). The faithfulness bounds for entanglement of formation in Proposition 9.5 were given by Li and Winter (2018) (see also Nielsen (2000)). The non-additivity of entanglement of formation was established by Hastings (2009), which built upon an earlier result of Shor (2004) connecting various additivity conjectures in quantum information theory. The formula in (9.1.93) for the entanglement of formation of two-qubit states was determined by Wootters (1998).

Negavitity and log-negativity were defined by Zyczkowski et al. (1998) (see also (Vidal and Werner, 2002) and (Plenio, 2005)). Vidal and Werner (2002) showed that log-negativity is monotone under LOCC. They also proved that it is additive. See also (Plenio, 2005) for a proof of selective LOCC monotonicity of log-negativity, and for a proof of the fact that log-negativity is not convex. The semi-definite programming approach, which we have taken here to prove selective LOCC monotonicity of log-negativity, is based on Wang and Duan (2016a).

The relative entropy of entanglement was defined by Vedral et al. (1997); Vedral and Plenio (1998), who also proved many of its properties, such as LOCC monotonicity and convexity. They also proposed concepts closely related to the generalized divergence of entanglement. The fact that trace distance of entanglement is not a selective LOCC monotone was shown by Qiao et al. (2018). Optimization over the set of separable states has been shown by Gurvits (2004); Gharibian (2010) to be NP-hard in general (see also (Ioannou, 2007; Shi and Wu, 2012)). A closed-form formula for the relative entropy of entanglement for two-qubit states was derived by Miranowicz and Ishizaka (2008). The max-relative entropy of entanglement was defined by Datta (2009b), the hypothesis testing relative entropy of entanglement by Brandao and Datta (2011), and the sandwiched Rényi relative entropy of entanglement by Wilde et al. (2017). The fact that the relative entropy of entanglement is invariant under classical communication (Proposition 9.17) was stated by Horodecki (2005) and proven by Kaur and Wilde (2017). Our proof of selective separable monotonicity in Proposition 9.19, for the Rényi relative entropies, is based on the approach from Wang and Wilde (2020). The proof of Property 1. of Proposition 9.20 follows the approach from Plenio et al. (2000), and the proof of Property 2. follows the approach of Tomamichel et al. (2017, Proposition 10). Property 2. of Proposition 9.20 was first established by Vedral and Plenio (1998). The cone program formulations of max-relative entropy of entanglement of a bipartite state, as well as max-relative entropy of entanglement of a quantum channel, were given by Berta and Wilde (2018).

The relative entropy with the set of PPT states was defined by Rains (1999a) in the context of entanglement distillation (see Chapter 13). Rains (2001) then modified the quantity to obtain a tighter bound on the distillable entanglement. After this development, Audenaert et al. (2002) defined the set PPT′ and showed that this improved bound can be written as the relative entropy with the set of PPT′ operators, which we refer to here as the Rains relative entropy. The fact that the set PPT' is preserved by completely PPT preserving channels (Property 1. of Lemma 9.14) was shown by Tomamichel et al. (2017). The sandwiched Rains relative entropy of a bipartite state was defined by Tomamichel et al. (2017), as well as the generalized Rains divergence, both as generalizations of the Rains relative entropy of Rains (2001); Audenaert et al. (2002). The semi-definite program formulation of the max-Rains relative entropy of a bipartite state was given by Wang and Duan (2016a); Wang et al. (2019b), who also recognized that the semi-definite programming bound from Wang and Duan (2016a) is equal to the max-Rains relative entropy. Wang and Duan (2016a) proved that the max-Rains relative entropy is a selective PPT monotone (Proposition 9.28) and that it is additive (Proposition 9.29). Miranowicz and Ishizaka (2008) proved that the Rains relative entropy is equal to the relative entropy of entanglement for two-qubit states.

The squashed entanglement of a bipartite state was defined by Christandl and Winter (2004), who established several of its key properties mentioned in Propositions 9.32, 9.33, and 9.36, including non-negativity, vanishing on separable states, convexity, superadditivity in general, additivity for tensor-product states, LOCC monotonicity, reduction for pure states, and the squashing channel representation in (9.4.43). The faithfulness of squashed entanglement was established by Brandao et al. (2011). A function related to squashed entanglement was discussed by Tucci (1999, 2002). Our discussions motivating squashed entanglement are related to those presented by Tucci (1999, 2002). The representation of squashed entanglement in (9.4.44) is due to Takeoka et al. (2014). Uniform continuity of the squashed entanglement of a bipartite state was established by Alicki and Fannes

(2004), and the explicit bound given here is due to Shirokov (2017).

Appendix 9.A Semi-Definite Programs for Negativity

Here we prove that the quantity $\|T_B(\rho_{AB})\|_1$ has the following primal and dual SDP formulations:

$$\|T_B(\rho_{AB})\|_1 = \sup_{R_{AB}} \{\mathrm{Tr}[R_{AB}\rho_{AB}] : -\mathbb{1}_{AB} \leq T_B(R_{AB}) \leq \mathbb{1}_{AB}\}, \quad (9.A.1)$$

$$= \inf_{K_{AB},L_{AB}\geq 0} \{\mathrm{Tr}[K_{AB} + L_{AB}] : T_B(K_{AB} - L_{AB}) = \rho_{AB}\}, \quad (9.A.2)$$

where the optimization in the first line is with respect to Hermitian R_{AB}. Since this is the core quantity underlying both the negativity and the log-negativity, it follows that these entanglement measures can be computed by means of semi-definite programs. To see the first equality, consider that

$$\|T_B(\rho_{AB})\|_1 = \|T_B(\rho_{AB})\|_1 \quad (9.A.3)$$

$$= \sup_{\|R_{AB}\|_\infty \leq 1} \mathrm{Tr}[R_{AB} T_B(\rho_{AB})] \quad (9.A.4)$$

$$= \sup_{\|R_{AB}\|_\infty \leq 1} \mathrm{Tr}[T_B(R_{AB})\rho_{AB}] \quad (9.A.5)$$

$$= \sup_{\|T_B(R_{AB})\|_\infty \leq 1} \mathrm{Tr}[R_{AB}\rho_{AB}] \quad (9.A.6)$$

$$= \sup_{R_{AB}} \{\mathrm{Tr}[R_{AB}\rho_{AB}] : -\mathbb{1}_{AB} \leq T_B(R_{AB}) \leq \mathbb{1}_{AB}\}. \quad (9.A.7)$$

The second equality follows from Hölder duality (see (2.2.97)), and since $T_B(\rho_{AB})$ is Hermitian, it suffices to optimize over Hermitian R_{AB}. The third equality follows because the partial transpose is its own Hilbert–Schmidt adjoint. The fourth equality follows from the substitution $R_{AB} \to T_B(R_{AB})$. The final equality follows because the inequality $\|T_B(R_{AB})\|_\infty \leq 1$ is equivalent to $-\mathbb{1}_{AB} \leq T_B(R_{AB}) \leq \mathbb{1}_{AB}$ for a Hermitian operator $T_B(R_{AB})$.

Now consider that the set of Hermitian operators is equivalent to the set of operators formed as differences of positive semi-definite operators. So this implies that

$$\|\mathrm{T}_B(\rho_{AB})\|_1 = \sup_{P_{AB},Q_{AB}\geq 0} \{\mathrm{Tr}[(P_{AB}-Q_{AB})\,\rho_{AB}] : -\mathbb{1}_{AB} \leq \mathrm{T}_B(P_{AB}-Q_{AB}) \leq \mathbb{1}_{AB}\}\,. \quad (9.A.8)$$

Then by setting

$$X = \begin{pmatrix} P_{AB} & 0 \\ 0 & Q_{AB} \end{pmatrix}, \quad A = \begin{pmatrix} \rho_{AB} & 0 \\ 0 & -\rho_{AB} \end{pmatrix}, \quad B = \begin{pmatrix} \mathbb{1}_{AB} & 0 \\ 0 & \mathbb{1}_{AB} \end{pmatrix}, \quad (9.A.9)$$

$$\Phi(X) = \begin{pmatrix} \mathrm{T}_B(P_{AB}-Q_{AB}) & 0 \\ 0 & -\mathrm{T}_B(P_{AB}-Q_{AB}) \end{pmatrix}, \quad (9.A.10)$$

this primal SDP is now in the standard form of (2.4.3). Then, setting

$$Y = \begin{pmatrix} K_{AB} & 0 \\ 0 & L_{AB} \end{pmatrix}, \quad (9.A.11)$$

we can calculate the Hilbert–Schmidt adjoint of Φ as

$$\begin{aligned}
&\mathrm{Tr}[Y\Phi(X)] \\
&= \mathrm{Tr}\left[\begin{pmatrix} K_{AB} & 0 \\ 0 & L_{AB} \end{pmatrix}\begin{pmatrix} \mathrm{T}_B(P_{AB}-Q_{AB}) & 0 \\ 0 & -\mathrm{T}_B(P_{AB}-Q_{AB}) \end{pmatrix}\right] && (9.A.12)\\
&= \mathrm{Tr}[K_{AB}(\mathrm{T}_B(P_{AB}-Q_{AB}))] - \mathrm{Tr}[L_{AB}(\mathrm{T}_B(P_{AB}-Q_{AB}))] && (9.A.13)\\
&= \mathrm{Tr}[\mathrm{T}_B(K_{AB})(P_{AB}-Q_{AB}))] - \mathrm{Tr}[\mathrm{T}_B(L_{AB})(P_{AB}-Q_{AB})] && (9.A.14)\\
&= \mathrm{Tr}[\mathrm{T}_B(K_{AB}-L_{AB})P_{AB}] + \mathrm{Tr}[\mathrm{T}_B(L_{AB}-K_{AB})Q_{AB}] && (9.A.15)\\
&= \mathrm{Tr}\left[\begin{pmatrix} \mathrm{T}_B(K_{AB}-L_{AB}) & 0 \\ 0 & -\mathrm{T}_B(K_{AB}-L_{AB}) \end{pmatrix}\begin{pmatrix} P_{AB} & 0 \\ 0 & Q_{AB} \end{pmatrix}\right], && (9.A.16)
\end{aligned}$$

so that

$$\Phi^\dagger(Y) = \begin{pmatrix} \mathrm{T}_B(K_{AB}-L_{AB}) & 0 \\ 0 & -\mathrm{T}_B(K_{AB}-L_{AB}) \end{pmatrix}. \quad (9.A.17)$$

Then, plugging into the standard form for the dual SDP in (2.4.4) and simplifying a bit, we find that it is given by

$$\begin{aligned}
&\inf_{K_{AB},L_{AB}\geq 0}\left\{\begin{array}{c} \mathrm{Tr}[K_{AB}+L_{AB}] : \mathrm{T}_B(K_{AB}-L_{AB}) \geq \rho_{AB}, \\ -\mathrm{T}_B(K_{AB}-L_{AB}) \geq -\rho_{AB} \end{array}\right\} \\
&\qquad = \inf_{K_{AB},L_{AB}\geq 0}\{\mathrm{Tr}[K_{AB}+L_{AB}] : \mathrm{T}_B(K_{AB}-L_{AB}) = \rho_{AB}\}\,. \quad (9.A.18)
\end{aligned}$$

Strong duality holds according to Theorem 2.28. Indeed, setting K_{AB} and L_{AB} to the respective positive and negative parts of $\mathrm{T}_B(\rho_{AB})$ is feasible for the dual, while setting $R_{AB} = \mathbb{1}_{AB}/2$ is strictly feasible for the primal.

Chapter 10

Entanglement Measures for Quantum Channels

So far we have considered entanglement measures for quantum states. We now consider entanglement measures for channels. Using the general principle discussed in Section 7.11.2 for constructing channel quantities out of state quantitites, we arrive at the following definition for the entanglement of a channel.

...

10.1 Definition and Basic Properties

...

Definition 10.1 Entanglement of a Quantum Channel

From an entanglement measure E defined on bipartite quantum states, we define the corresponding entanglement of a quantum channel $\mathcal{N}_{A\to B}$ as follows:

$$E(\mathcal{N}) \coloneqq \sup_{\rho_{RA}} E(R;B)_\omega, \tag{10.1.1}$$

where $\omega_{RB} \coloneqq \mathcal{N}_{A\to B}(\rho_{RA})$, and the optimization is with respect to every bipartite state ρ_{RA}, with system R arbitrarily large, yet finite.

REMARK: Note that it suffices to optimize (10.1.1) with respect to pure states ψ_{RA}, with the dimension of R equal to the dimension of A, when calculating the entanglement of a channel, so that

$$E(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} E(R;B)_\omega, \tag{10.1.2}$$

where $\omega_{RB} \coloneqq \mathcal{N}_{A\to B}(\psi_{RA})$. This follows from the fact that an entanglement measure for states is, by definition, monotone under LOCC channels. It is therefore monotone under a local partial trace channel. In particular, consider a mixed state ρ_{RA}, with the dimension of R not necessarily equal to the dimension of A. Let $\omega_{RB} = \mathcal{N}_{A\to B}(\omega_{RA})$. Then, if we take a purification $\phi_{R'RA}$ of ρ_{RA}, we obtain

$$\begin{aligned} E(R;B)_\omega &= E(\mathcal{N}_{A\to B}(\rho_{RA})) && (10.1.3)\\ &= E(\mathcal{N}_{A\to B}(\mathrm{Tr}_{R'}[\phi_{R'RA}])) && (10.1.4)\\ &= E(\mathrm{Tr}_{R'}[\mathcal{N}_{A\to B}(\phi_{R'RA})]) && (10.1.5)\\ &\leq E(\mathcal{N}_{A\to B}(\phi_{R'RA})) && (10.1.6)\\ &= E(R'R;B)_\tau, && (10.1.7)\end{aligned}$$

where $\tau_{R'RB} = \mathcal{N}_{A\to B}(\phi_{R'RA})$ and to obtain the inequality we used the fact that E is monotone under the partial trace channel $\mathrm{Tr}_{R'}$. This demonstrates that it suffices to optimize with respect to pure states when calculating the entanglement of a channel. Furthermore, by the Schmidt decomposition theorem (Theorem 2.2), the dimension of the purifying system $R'R$ need not exceed the dimension of A.

Note that in the definition above the channel $\mathcal{N}_{A\to B}$ acts locally on the state ψ_{RA} to produce the state $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$. We can thus view $\mathcal{N}_{A\to B}$ as an LOCC channel, which means that $E(R;B)_\omega \leq E(R;A)_\psi$, by the definition of an entanglement measure for states. In other words, by sending one share of a bipartite state through the channel $\mathcal{N}$, the entanglement can only stay the same or go down. The quantity $E(\mathcal{N})$ thus indicates how well entanglement is preserved when one share of it is sent through the channel $\mathcal{N}$.

Let us consider three examples of entanglement measures for quantum channels, defined using entanglement measures for bipartite quantum states.

1. The *generalized divergence of entanglement* of a channel $\mathcal{N}_{A\to B}$, defined for every generalized divergence $\boldsymbol{D}$ as

$$\begin{aligned} \boldsymbol{E}(\mathcal{N}) &\coloneqq \sup_{\psi_{RA}} \boldsymbol{E}(R;B)_\omega && (10.1.8)\\ &= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\Vert\sigma_{RB}), && (10.1.9)\end{aligned}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and the optimization is with respect to pure states

ψ_{RA}, with the dimension of R equal to the dimension of A. We investigate this entanglement measure in Section 10.3.

2. The *generalized Rains divergence* of a channel $\mathcal{N}_{A\to B}$, defined for every generalized divergence $\boldsymbol{D}$ as

$$\boldsymbol{R}(\mathcal{N}) := \sup_{\psi_{SA}} \boldsymbol{R}(A;B)_{\omega} \tag{10.1.10}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in \mathrm{PPT}'(S:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}), \tag{10.1.11}$$

where $\omega_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$, and the optimization is with respect to pure states ψ_{SA}, with the dimension of S equal to the dimension of A. We investigate this entanglement measure in Section 10.4.

3. The *squashed entanglement* of a channel $\mathcal{N}_{A\to B}$, defined as

$$E_{\mathrm{sq}}(\mathcal{N}) := \sup_{\psi_{RA}} E_{\mathrm{sq}}(R;B)_{\omega}, \tag{10.1.12}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, and the optimization is with respect to pure states ψ_{RA}, with the dimension of R equal to the dimension of A. We investigate this entanglement measure in Section 10.5.

REMARK: Instead of defining an entanglement measure for channels via an entanglement measure for states, consider that the channel analogue of a separable state is an entanglement-breaking channel, which follows from the discussion in Section 4.4.6. Another way to construct an entanglement measure for quantum channels is through the generalized channel divergence (Definition 7.81) between the channel and the set of entanglement-breaking channels:

$$\boldsymbol{E}'(\mathcal{N}) := \inf_{\mathcal{M}\in \mathrm{EB}(A\to B)} \boldsymbol{D}(\mathcal{N}\|\mathcal{M}), \tag{10.1.13}$$

where $\mathrm{EB}(A\to B)$ denotes the set of entanglement-breaking channels taking system A to system B. Now, using the expression for the generalized channel divergence in (7.11.2), we obtain

$$\boldsymbol{E}'(\mathcal{N}) = \inf_{\mathcal{M}\in \mathrm{EB}(A\to B)} \sup_{\psi_{RA}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})), \tag{10.1.14}$$

where the optimization is with respect to pure states ψ_{RA} is such that the dimension of R is equal to the dimension of A.

Now, because entanglement-breaking channels and separable states (with maximally mixed reduced state) are in one-to-one correspondence (see Section 4.4.6), we find that the generalized divergence of entanglement of $\mathcal{N}$ is bounded from above as follows:

$$\boldsymbol{E}(\mathcal{N}) \leq \sup_{\psi_{RA}} \inf_{\mathcal{M}\in \mathrm{EB}(A\to B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{M}_{A\to B}(\psi_{RA})). \tag{10.1.15}$$

> The right-hand side of the above inequality and the quantity $\boldsymbol{E}'(\mathcal{N})$ differ in the order of the infimum and supremum. From the discussion in Section 2.3, in particular (2.3.14), we conclude that $\boldsymbol{E}(\mathcal{N}) \leq \boldsymbol{E}'(\mathcal{N})$ for all quantum channels $\mathcal{N}$. For the rest of this chapter, and throughout the rest of this book, we thus stick with the definition of a channel entanglement measure given in Definition 10.1.

Many properties of state entanglement measures carry over, or have an analogue, to the corresponding channel entanglement measure.

Proposition 10.2 Properties of Entanglement Measures for Channels

Let E be an entanglement measure for states, as defined in Definition 9.1, and consider the corresponding channel entanglement measure defined in Definition 10.1.

1. *Faithfulness*: If E vanishes for all separable states, then $E(\mathcal{N}) = 0$ for all entanglement breaking channels. If E is faithful (vanishing if and only if the input state is separable), then $E(\mathcal{N}) = 0$ implies that $\mathcal{N}$ is entanglement breaking.

2. *Convexity*: If E is a convex entanglement measure for states, then the corresponding channel entanglement measure is convex: for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0, 1]$, and set $\{\mathcal{N}^x_{A\to B}\}_{x\in\mathcal{X}}$ of quantum channels,
$$E\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^x\right) \leq \sum_{x\in\mathcal{X}} p(x)E(\mathcal{N}^x). \tag{10.1.16}$$

3. *Superadditivity*: If E is superadditive, meaning that
$$E(A_1A_2; B_1B_2)_{\rho\otimes\tau} \geq E(A_1; B_1)_\rho + E(A_2; B_2)_\tau \tag{10.1.17}$$
for all states $\rho_{A_1B_1}$ and $\tau_{A_2B_2}$, then the channel entanglment measure is also superadditive: for every two channels $\mathcal{N}_{A_1\to B_1}$ and $\mathcal{M}_{A_2\to B_2}$,
$$E(\mathcal{N} \otimes \mathcal{M}) \geq E(\mathcal{N}) + E(\mathcal{M}). \tag{10.1.18}$$

Proof:

1. Let $\mathcal{N}$ be an entanglement breaking channel. This means that ω_{RB} =

$\mathcal{N}_{A\to B}(\psi_{RA})$ is separable for every pure state ψ_{RA}. Therefore, since E vanishes for all separable states, we have $E(R;B)_\omega = 0$ for every pure state ψ_{RA}, so that $E(\mathcal{N}) = 0$.

Now, suppose that E is a faithful state entanglement measure, and let $E(\mathcal{N}) = 0$. Since E is faithful, it is non-negative for all input states, which implies that $E(R;B)_\omega = 0$ for every state $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, i.e., for every pure state ψ_{RA}. Furthermore, by faithfulness of E for states, it holds that $\mathcal{N}_{A\to B}(\psi_{RA})$ is separable for every pure state ψ_{RA}. Therefore, $\mathcal{N}$ is entanglement breaking.

2. Let ψ_{RA} be an arbitrary pure state, and let

$$\omega_{RB} = \left(\sum_{x\in\mathcal{X}} \mathcal{N}^x_{A\to B}\right)(\psi_{RA}) = \sum_{x\in\mathcal{X}} p(x)\omega^x_{RB}, \tag{10.1.19}$$

where $\omega^x_{RB} = \mathcal{N}^x_{A\to B}(\psi_{RA})$. Then, by the convexity of E for states,

$$E(R;B)_\omega \le \sum_{x\in\mathcal{X}} p(x)E(R;B)_{\omega^x} \le \sum_{x\in\mathcal{X}} p(x)E(\mathcal{N}^x), \tag{10.1.20}$$

where for the last inequality we used the definition of the channel entanglement measure. Therefore, for every pure state ψ_{RA}, we have

$$E(R;B)_\omega \le \sum_{x\in\mathcal{X}} p(x)E(\mathcal{N}^x). \tag{10.1.21}$$

Thus,

$$E\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^x\right) = \sup_{\psi_{RA}} E(R;B)_\omega \le \sum_{x\in\mathcal{X}} p(x)E(\mathcal{N}^x), \tag{10.1.22}$$

as required.

3. By restricting the optimization in the definition of $E(\mathcal{N}\otimes\mathcal{M})$ to pure product states $\phi_{R_1A_1}\otimes\varphi_{R_2A_2}$, letting $\xi_{R_1B_1} = \mathcal{N}_{A_1\to B_1}(\varphi_{R_1A_1})$, $\tau_{R_2B_2} = \mathcal{M}_{A_2\to B_2}(\varphi_{R_2A_2})$, and using superadditivity of the state entanglement measure E, we obtain

$$E(\mathcal{N}\otimes\mathcal{M}) = \sup_{\psi_{RA_1A_2}} E(R;B_1B_2)_\omega \tag{10.1.23}$$

$$\ge \sup_{\phi_{R_1A_2}\otimes\varphi_{R_2A_2}} E(R_1R_2;B_1B_2)_{\xi\otimes\tau} \tag{10.1.24}$$

$$\ge \sup_{\phi_{R_1A_1}} E(R_1;B_1)_\xi + \sup_{\varphi_{R_2A_2}} E(R_2;B_2)_\tau \tag{10.1.25}$$

$$= E(\mathcal{N}) + E(\mathcal{M}), \tag{10.1.26}$$

as required. ■

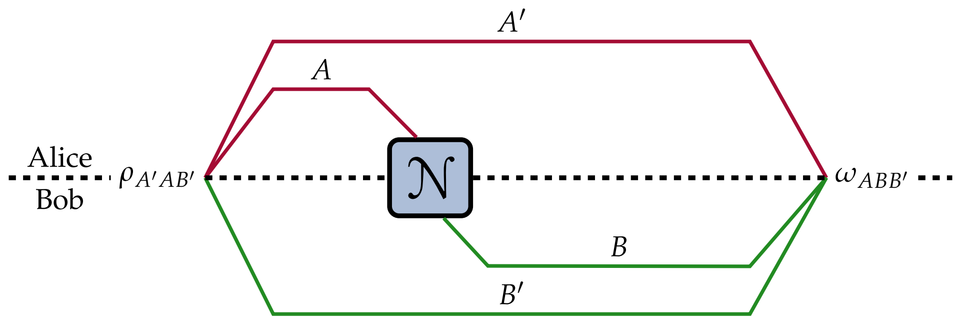


FIGURE 10.1: Starting from a state $\rho_{A'AB'}$, Alice sends the system A through the channel $\mathcal{N}_{A\to B}$ to Bob, resulting in the state $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. The difference between the final and initial entanglement (as quantified by an entanglement measure for states), optimized over all initial states $\rho_{A'AB'}$, is equal to the amortized entanglement of $\mathcal{N}$; see Definition 10.3.

10.2 Amortized Entanglement

There is another way to define the entanglement of a quantum channel from an entanglement measure on quantum states, and this method turns out to be useful in the feedback-assisted communciation protocols that we consider in Part III. To see how this measure is defined, consider the situation shown in Figure 10.1. In this setup, Alice and Bob each have access to systems A' and B', respectively. Alice also possesses the system A, and she passes it through the channel $\mathcal{N}_{A\to B}$ to Bob. The initial joint state $\rho_{A'AB'}$ then becomes $\omega_{ABB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. The systems A' and B' can be thought of as "memory systems" that hold auxiliary information. In feedback-assisted quantum communication protocols, these memory systems can hold the results of previous rounds of communication between Alice and Bob for the purpose of deciding what local operations to perform in subsequent rounds. By taking the difference between the final entanglement in the state $\omega_{A'BB'}$ and the initial entanglement in the state $\rho_{A'AB'}$, and optimizing over all initial states $\rho_{A'AB'}$, we arrive at what is called the *amortized entanglement*.

Definition 10.3 Amortized Entanglement of a Quantum Channel

From an entanglement measure E defined on bipartite quantum states, we define the *amortized entanglement* of a quantum channel $\mathcal{N}_{A\to B}$ as follows:

$$E^{\mathcal{A}}(\mathcal{N}) := \sup_{\rho_{A'AB'}} \left\{ E(A';BB')_\omega - E(A'A;B')_\rho \right\}, \tag{10.2.1}$$

where $\omega_{A'BB'} := \mathcal{N}_{A\to B}(\rho_{A'AB'})$ and the optimization is with respect to states $\rho_{A'AB'}$. The systems A' and B' have arbitrarily large, yet finite dimensions.

Due to the fact that the systems A' and B' can be arbitrarily large, it is not necessarily the case that the supremum above can be achieved. Thus, in general, it might be difficult to compute a channel's amortized entanglement.

For every entanglement measure E that is equal to zero for all separable states, we always have that the entanglement of the channel never exceeds the amortized entanglement of the channel:

Lemma 10.4

For a given quantum channel $\mathcal{N}$ and entanglement measure E that is equal to zero for all separable states, the channel's entanglement does not exceed its amortized entanglement:

$$E(\mathcal{N}) \leq E^{\mathcal{A}}(\mathcal{N}). \tag{10.2.2}$$

PROOF: By choosing the input state $\rho_{A'AB'}$ in the optimization for amortized entanglement to have a trivial (one-dimensional) system B' (so that $\rho_{A'AB'}$ is trivially a separable state between Alice and Bob), we find that $E(A';BB')_\omega = E(A';B)_\omega$ and $E(AA';B')_\rho = 0$. Since such a state is an arbitrary state to consider for optimizing the channel's entanglement, the inequality follows. ■

Whether the inequality reverse to the one in (10.2.2) holds, which would imply that $E(\mathcal{N}) = E^{\mathcal{A}}(\mathcal{N})$, depends on the entanglement measure E. In Section 10.6 below, we show that this so-called "amortization collapse" occurs for some entanglement measures.

The amortized entanglement of a channel has several interesting properties, which we list in some detail in this section. These include convexity, faithfulness, and (sub)additivity.

Proposition 10.5 Properties of Amortized Entanglement of a Quantum Channel

1. *Dimension bound*: Let E be a subadditive entanglement measure for states, and let $\mathcal{N}_{A\to B}$ be a quantum channel. Then,

$$E^{\mathcal{A}}(\mathcal{N}) \leq \min\{E(A;A')_{\Phi}, E(B;B')_{\Phi}\}, \tag{10.2.3}$$

where A' has the same dimension as A, B' has the same dimension as B, and Φ^+ is a maximally entangled state of systems AA' or BB'.

2. *Faithfulness*: Let E be an entanglement measure that is equal to zero for all separable states. If a channel $\mathcal{N}$ is entanglement-breaking, then its amortized entanglement $E^{\mathcal{A}}(\mathcal{N})$ is equal to zero. If the entanglement measure E is faithful (equal to zero if and only if the state is separable) and the amortized entanglement $E^{\mathcal{A}}(\mathcal{N})$ of a channel $\mathcal{N}$ is equal to zero, then the channel $\mathcal{N}$ is entanglement breaking.

3. *Convexity*: Let E be a convex entanglement measure for states. Then, the amortized entanglement $E^{\mathcal{A}}$ of a channel is convex: for every finite alphabet $\mathcal{X}$, probability distribution $p : \mathcal{X} \to [0,1]$, and set $\{\mathcal{N}^x_{A\to B}\}_{x\in\mathcal{X}}$ of quantum channels,

$$E^{\mathcal{A}}\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^x\right) \leq \sum_{x\in\mathcal{X}} p(x)E^{\mathcal{A}}(\mathcal{N}^x). \tag{10.2.4}$$

4. *Subdditivity and additivity*: For every entanglement measure E, the amortized entanglement $E^{\mathcal{A}}$ of a channel is subadditive, meaning that for every two quantum channels $\mathcal{N}$ and $\mathcal{M}$,

$$E^{\mathcal{A}}(\mathcal{N}\otimes\mathcal{M}) \leq E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}). \tag{10.2.5}$$

If E is an additive entanglement measure, then the amortized entanglement $E^{\mathcal{A}}$ is additive, meaning that

$$E^{\mathcal{A}}(\mathcal{N}\otimes\mathcal{M}) = E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}) \tag{10.2.6}$$

for every two quantum channels $\mathcal{N}$ and $\mathcal{M}$.

Proof:

1. To prove (10.2.3), we use the fact that $\mathcal{N}$ can be simulated via teleporation. Specifically, from (5.1.33) and (5.1.34), we can represent the action of $\mathcal{N}_{A\to B}$ on every state $\rho_{A'AB'}$ in the following two ways:

$$\mathcal{N}_{A\to B}(\rho_{A'AB'}) = \mathcal{N}_{B_i'\to B}\left(\mathcal{T}_{AA_iB_i\to B_i'}(\rho_{A'AB'}\otimes\Phi_{A_iB_i})\right), \tag{10.2.7}$$

$$\mathcal{N}_{A\to B}(\rho_{A'AB'}) = \mathcal{T}_{A_o'A_oB_o\to B}\left(\mathcal{N}_{A\to A_o'}(\rho_{A'AB'})\otimes\Phi^+_{A_oB_o}\right), \tag{10.2.8}$$

where A_i, B_i, B_i' are auxiliary systems such that $d_{A_i} = d_{B_i} = d_{B_i'} = d_A$ (i.e., systems with the same dimension as the input system A of the channel) and A_o, A_o', B_o are auxiliary systems such that $d_{A_o} = d_{A_o'} = d_{B_o} = d_B$ (i.e., systems with the same dimension as the output system B of the channel). The teleportation channel $\mathcal{T}$ is given by (5.1.26). Since $\mathcal{T}$ is an LOCC channel, and $\mathcal{N}$ is a local channel, by LOCC monotonicity of the entanglement measure E we obtain

$$E(A';BB')_\omega \leq E(A'AA_i;B'B_i)_{\rho\otimes\Phi}, \tag{10.2.9}$$

$$E(A';BB')_\omega \leq E(A'AA_o;B'B_o)_{\rho\otimes\Phi}, \tag{10.2.10}$$

where $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Then, by subadditivity of E, we find that

$$E(A';BB')_\omega \leq E(A'A;B')_\rho + E(A_i;B_i)_\Phi, \tag{10.2.11}$$

$$E(A';BB')_\omega \leq E(A'A;B')_\rho + E(A_o;B_o)_\Phi. \tag{10.2.12}$$

Since the state $\rho_{A'AB'}$ is arbitrary, we conclude that

$$E^{\mathcal{A}}(\mathcal{N}) \leq E(A_i;B_i)_\Phi = E(A;A')_\Phi, \tag{10.2.13}$$

$$E^{\mathcal{A}}(\mathcal{N}) \leq E(A_o;B_o)_\Phi = E(B;B')_\Phi, \tag{10.2.14}$$

where for the last equality in each case we used $d_{A_i} = d_{B_i} = d_{A'} = d_A$ and $d_{A_o} = d_{B_o} = d_{B'} = d_B$. We thus have

$$E^{\mathcal{A}}(\mathcal{N}) \leq \min\{E(A;A')_\Phi, E(B;B')_\Phi\}, \tag{10.2.15}$$

as required.

2. Let $\mathcal{N}$ be an entanglement breaking channel. For every state $\rho_{A'AB'}$, let $\omega_{ABB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Recall from Section 4.4.6, specifically Theorem 4.15,

that every entanglement breaking channel can be represented as a measurement of the input system followed by the preparation of a state on the output system conditioned on the outcome of the measurement. As such, every entanglement breaking channel can be simulated by an LOCC channel. Therefore, by the monotonicity of the entanglement measure E under LOCC,

$$E(A';BB')_\omega \leq E(A'A;B')_\rho, \tag{10.2.16}$$

which means that

$$E(A';BB')_\omega - E(A'A;B')_\rho \leq E(A'A;B')_\rho - E(A'A;B')_\rho = 0 \tag{10.2.17}$$

for every state $\rho_{A'AB'}$. Therefore, $E^{\mathcal{A}}(\mathcal{N}) \leq 0$. On the other hand, because E vanishes for all separable states, it holds that $E(\mathcal{N}) \geq 0$. Therefore, by Lemma 10.4, $E^{\mathcal{A}}(\mathcal{N}) \geq 0$, and we conclude that $E^{\mathcal{A}}(\mathcal{N}) = 0$.

Now, let E be a faithful entanglement measure, meaning that it vanishes if and only if the input state is separable, and suppose that $E^{\mathcal{A}}(\mathcal{N}) = 0$. By Lemma 10.4, we have that $0 = E^{\mathcal{A}}(\mathcal{N}) \geq E(\mathcal{N})$, which in turn implies that $E(\mathcal{N}) = 0$ because $E(\mathcal{N}) \geq 0$ for all channels $\mathcal{N}$. Therefore, by Proposition 10.2, we conclude that $\mathcal{N}$ is entanglement breaking.

3. Let $\rho_{A'AB'}$ be an arbitrary state, and let

$$\omega_{A'BB'} = \left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^x_{A\to B}\right)(\rho_{A'AB'}) = \sum_{x\in\mathcal{X}} p(x)\omega^x_{A'BB'}, \tag{10.2.18}$$

where $\omega^x_{A'BB'} = \mathcal{N}^x_{A\to B}(\rho_{A'AB'})$ for all $x \in \mathcal{X}$. Then, by convexity of the entanglement measure E, we obtain

$$E(A';BB')_\omega \leq \sum_{x\in\mathcal{X}} p(x)E(A';BB')_{\omega^x}. \tag{10.2.19}$$

Also, $E(A'A;B')_\rho = \sum_{x\in\mathcal{X}} p(x)E(A'A;B')_\rho$. Therefore, by the definition of amortized entanglement, we find that

$$E(A';BB')_\omega - E(A'A;B')_\rho \tag{10.2.20}$$
$$\leq \sum_{x\in\mathcal{X}} p(x)\left(E(A';BB')_{\omega^x} - E(A'A;B')_\rho\right) \tag{10.2.21}$$
$$\leq \sum_{x\in\mathcal{X}} p(x)E^{\mathcal{A}}(\mathcal{N}^x). \tag{10.2.22}$$

Since the state $\rho_{A'AB'}$ is arbitrary, by optimizing over every state $\rho_{A'AB'}$ on the left-hand side of the inequality above, we obtain

$$E^{\mathcal{A}}\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^x\right) \leq \sum_{x\in\mathcal{X}} p(x)E^{\mathcal{A}}(\mathcal{N}^x), \tag{10.2.23}$$

as required.

4. Let A_1 and B_1 denote the respective input and output systems for the quantum channel $\mathcal{N}$, and let A_2 and B_2 denote the respective input and output quantum systems for the quantum channel $\mathcal{M}$. Let $\rho_{A'A_1A_2B'}$ be an arbitrary state. Let

$$\omega_{A'B_1B_2B'} = (\mathcal{N}_{A_1\to B_1} \otimes \mathcal{M}_{A_2\to B_2})(\rho_{A'A_1A_2B'}) \tag{10.2.24}$$
$$= \mathcal{N}_{A_1\to B_1}(\tau_{A'A_1B_2B'}), \tag{10.2.25}$$

where

$$\tau_{A'A_1B_2B'} \coloneqq \mathcal{M}_{A_2\to B_2}(\rho_{A'A_1A_2B'}). \tag{10.2.26}$$

Observe that the state $\omega_{A'B_1B_2B'}$ is both an example of an output state in the optimization defining $E^{\mathcal{A}}(\mathcal{N}\otimes\mathcal{M})$ and in the optimization defining $E^{\mathcal{A}}(\mathcal{N})$ (with an appropriate identification of the A', B, and B' systems for the latter). Observe also that $\tau_{A'A_1B_2B'}$ is an example of an output state in the optimization defining $E^{\mathcal{A}}(\mathcal{M})$. Therefore,

$$E(A';B_1B_2B')_\omega - E(A'A_1A_2;B')_\rho$$
$$= E(A';B_1B_2B')_\omega - E(A'A_1;B_2B')_\tau$$
$$+ E(A'A_1;B_2B')_\tau - E(A'A_1A_2;B')_\rho \tag{10.2.27}$$
$$\leq E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}). \tag{10.2.28}$$

Since the state $\rho_{A'A_1A_2B'}$ is arbitrary, we can optimize over all such states on the left-hand side of the inequality above to obtain

$$E^{\mathcal{A}}(\mathcal{N}\otimes\mathcal{M}) \leq E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}). \tag{10.2.29}$$

Now, suppose that E is an additive entanglement measure. Let us restrict the optimization over states $\rho_{A'A_1A_2B'}$ in the definition of $E^{\mathcal{A}}(\mathcal{N}\otimes\mathcal{M})$ such that $A' \equiv A_1'A_2'$, $B' \equiv B_1'B_2'$, and $\rho_{A'A_1A_2B'} = \rho^1_{A_1'A_1B_1'} \otimes \rho^2_{A_2'A_2B_2'}$, for states $\rho^1_{A_1'A_1B_1'}$ and $\rho^2_{A_2'A_2B_2'}$. Then,

$$\omega_{A'B_1B_2B'} = (\mathcal{N}_{A_1\to B_1} \otimes \mathcal{M}_{A_2\to B_2})(\rho_{A'A_1A_2B'}) \tag{10.2.30}$$

$$= \mathcal{N}_{A_1 \to B_1}(\rho^1_{A'_1 A_1 B'_1}) \otimes \mathcal{M}_{A_2 \to B_2}(\rho^2_{A'_2 A_2 B'_2}) \tag{10.2.31}$$

$$=: \omega^1_{A'_1 B_1 B'_1} \otimes \omega^2_{A_2 B_2 B'_2}. \tag{10.2.32}$$

Therefore, using additivity of E, we obtain

$$E^{\mathcal{A}}(\mathcal{N} \otimes \mathcal{M}) = \sup_{\rho_{A' A_1 A_2 B'}} \{E(A'; B_1 B_2 B')_\omega - E(A' A_1 A_2; B')_\rho\} \tag{10.2.33}$$

$$\geq \sup_{\rho^1_{A'_1 A_1 B'_1} \otimes \rho^2_{A'_2 A_2 B'_2}} \{E(A'_1 A'_2; B_1 B_2 B'_1 B'_2)_{\omega^1 \otimes \omega^2} - E(A'_1 A'_2 A_1 A_2; B'_1 B'_2)_{\rho^1 \otimes \rho^2}\} \tag{10.2.34}$$

$$\geq \sup_{\rho^1_{A'_1 A_1 B'_1} \otimes \rho^2_{A'_2 A_2 B'_2}} \{E(A'_1; B_1 B'_1)_{\omega^1} + E(A'_2; B_2 B'_2)_{\omega^2} - E(A'_1 A_1; B'_1)_{\rho^1} - E(A'_2 A_2; B'_2)_{\rho^2}\} \tag{10.2.35}$$

$$= \sup_{\rho^1_{A'_1 A_1 B'_1}} \{E(A'_1; B_1 B'_1)_{\omega^1} - E(A'_1 A_1; B'_1)_{\rho^1}\} + \sup_{\rho^2_{A'_2 A_2 B'_2}} \{E(A'_2; B_2 B'_2)_{\omega^2} - E(A'_2 A_2; B'_2)_{\rho^2}\} \tag{10.2.36}$$

$$= E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}). \tag{10.2.37}$$

Combining this with (10.2.29), we conclude that

$$E^{\mathcal{A}}(\mathcal{N} \otimes \mathcal{M}) = E^{\mathcal{A}}(\mathcal{N}) + E^{\mathcal{A}}(\mathcal{M}), \tag{10.2.38}$$

as required. ■

An immediate consequence of (10.2.5) is the following inequality:

$$\sup_{\mathcal{M}} \left[E^{\mathcal{A}}(\mathcal{N} \otimes \mathcal{M}) - E^{\mathcal{A}}(\mathcal{M})\right] \leq E^{\mathcal{A}}(\mathcal{N}), \tag{10.2.39}$$

where the supremum is with respect to quantum channels $\mathcal{M}$. This inequality demonstrates that no other channel can help to enhance the amortized entanglement of a quantum channel.

10.2.1 Amortized Entanglement and Teleportation Simulation

Teleportation (or, more generally, LOCC, separable, or PPT) simulation of a quantum channel is a key tool that we can use to establish upper bounds on capacities of certain quantum channels when they are assisted by LOCC. Recalling Section 5.1.4, the basic idea behind this tool is that a quantum channel can be simulated by the action of a teleportation protocol, with a maximally entangled resource state shared between the sender A and receiver B. More generally, recalling Definition 4.25, a channel $\mathcal{N}_{A\to B}$ with input system A and output system B is defined to be LOCC-simulable with associated resource state $\omega_{RB'}$ if the following equality holds for all input states ρ_A:

$$\mathcal{N}_{A\to B}(\rho_A) = \mathcal{L}_{ARB'\to B}(\rho_A \otimes \omega_{RB'}), \tag{10.2.40}$$

where $\mathcal{L}_{ARB'\to B}$ is a quantum channel consisting of LOCC between the sender, who has systems A and R, and the receiver, who has system B'. Whenever the underlying state entanglement measure is subadditive, the amortized entanglement of an LOCC-simulable channel can be bounded from above by the entanglement of the resource state. In fact, this is precisely what we did when proving the dimension bound in Proposition 10.5 above. We can therefore understand the dimension bound as being a consequence of the fact that all channels are teleportation simulable, and hence LOCC simulable, by using a maximally entangled resource state.

Proposition 10.6

Let E be a subadditive state entanglement measure (recall (9.1.9)). If a quantum channel $\mathcal{N}_{A\to B}$ is LOCC-simulable with associated resource state $\omega_{RB'}$, i.e.,

$$\mathcal{N}_{A\to B}(\rho_A) = \mathcal{L}_{ARB'\to B}(\rho_A \otimes \omega_{RB'}), \tag{10.2.41}$$

where $\mathcal{L}_{ARB'\to B}$ is an LOCC channel, then the amortized entanglement $E^{\mathcal{A}}(\mathcal{N})$ of $\mathcal{N}$ is bounded from above by the entanglement of the resource state:

$$E^{\mathcal{A}}(\mathcal{N}) \leq E(R;B')_\omega. \tag{10.2.42}$$

Proof: For every state $\rho_{A'AB''}$, we use monotonicity of the state entanglement measure under LOCC, as well as subadditivity of the measure, to obtain

$$\begin{aligned}
&E(A';BB'')_{\mathcal{L}(\rho\otimes\omega)} - E(A'A;B'')_\rho \\
&\leq E(A'AR;B''B')_{\rho\otimes\omega} - E(A'A;B'')_\rho && (10.2.43)\\
&\leq E(A'A;B'')_\rho + E(R;B')_\omega - E(A'A;B'')_\rho && (10.2.44)\\
&= E(R;B')_\omega, && (10.2.45)
\end{aligned}$$

where for the first inequality we made use of LOCC monotonicity and for the second inequality we made use of the assumption of subadditivity. Since the state $\rho_{A'AB''}$ was arbitrary, we conclude (10.2.42). ■

If it happens that a channel $\mathcal{N}_{A\to B}$ is LOCC-simulable with resource state $\omega_{RB'} = \mathcal{N}_{A\to B'}(\rho_{RA})$ for some state ρ_{RA}, then the inequality in (10.2.42) becomes an equality. In Section 5.1.4, we saw an example in which such a situation arises, namely, when the channel $\mathcal{N}$ is group covariant. In this case, the resource state is simply the Choi state of the channel.

Proposition 10.7

Let E be an entanglement measure that is subadditive with respect to states and zero on separable states, and let $E^{\mathcal{A}}$ denote its amortized version. If a channel $\mathcal{N}_{A\to B}$ is LOCC-simulable with associated resource state $\omega_{RB} = \mathcal{N}_{A\to B}(\rho_{RA})$ for some input state ρ_{RA}, then the following equality holds

$$E^{\mathcal{A}}(\mathcal{N}) = E(R;B)_{\omega}. \tag{10.2.46}$$

Proof: From Proposition 10.6, we have that $E^{\mathcal{A}}(\mathcal{N}) \leq E(R;B)_{\omega}$. For the reverse inequality, we take $\rho_{A'AB''} = \rho_{RA}$ in the optimization that defines $E^{\mathcal{A}}(\mathcal{N})$, where we identify $A' \equiv R$ and $B'' \equiv \varnothing$ (i.e., B'' is a trivial one-dimensional system). Then, $\mathcal{N}_{A\to B}(\rho_{A'AB''}) = \mathcal{N}_{A\to B}(\rho_{RA})$, which is the resource state. Furthermore, since B'' is a one-dimensional system, the state $\rho_{A'AB''}$ is trivially separable, so that $E(A'A;B'')_{\rho} = 0$. Therefore, $E^{\mathcal{A}}(\mathcal{N}) \geq E(A';B)_{\omega} \equiv E(R;B)_{\omega}$. ■

10.3 Generalized Divergence of Entanglement

In this section, we examine the generalized divergence of entanglement of quantum channels, which is a channel entanglement measure that arises from the generalized divergence of entanglement for quantum states that we considered in Section 9.2.

Definition 10.8 Generalized Divergence of Entanglement

Let $\boldsymbol{D}$ be a generalized divergence (see Definition 7.15). For every quantum channel $\mathcal{N}_{A\to B}$, we define the *generalized divergence of entanglement of* $\mathcal{N}$ as

$$\boldsymbol{E}(\mathcal{N}) := \sup_{\psi_{RA}} \boldsymbol{E}(R;B)_\omega \tag{10.3.1}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.3.2}$$

where $\omega_{RB} := \mathcal{N}_{A\to B}(\psi_{RA})$. The supremum is with respect to every pure state ψ_{RA}, with the dimension of R equal to the dimension of A.

In the remark immediately after Definition 10.1, we stated how it suffices to optimize with respect to pure bipartite states (with equal dimension for each subsystem) when calculating the generalized divergence of entanglement of a quantum channel.

We can write the generalized divergence of entanglement of $\mathcal{N}_{A\to B}$ in the following alternate form:

$$\boldsymbol{E}(\mathcal{N}) = \sup_{\rho_A} \boldsymbol{E}(\mathcal{N}_{A\to B}, \rho_A), \tag{10.3.3}$$

$$\boldsymbol{E}(\mathcal{N}_{A\to B}, \rho_A) := \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sigma_{AB}). \tag{10.3.4}$$

This is indeed true because we can write every purification of ρ_A as $(V_{A'}\sqrt{\rho_{A'}}\otimes \mathbb{1}_A)|\Gamma\rangle_{A'A}$ for some isometry $V_{A'}$ (see (2.2.40) and Theorem 2.3), that the set of separable states is invariant under local isometries, and that generalized divergences are invariant under local isometries (see Proposition 7.16).

As with the state quantities, we are interested in the following generalized divergences of entanglement of a quantum channel $\mathcal{N}_{A\to B}$:

1. The *relative entropy of entanglement of* $\mathcal{N}$,

$$E_R(\mathcal{N}) := \sup_{\psi_{SA}} E_R(S;B)_\omega \tag{10.3.5}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\text{SEP}(S:B)} D(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}), \tag{10.3.6}$$

where $\omega_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$.

2. The *ε-hypothesis testing relative entropy of entanglement of* $\mathcal{N}$,

$$E_R^{\varepsilon}(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} E_R^{\varepsilon}(R;B)_{\omega} \tag{10.3.7}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.3.8}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$.

3. The *sandwiched Rényi relative entropy of entanglement of* $\mathcal{N}$,

$$\widetilde{E}_{\alpha}(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} \widetilde{E}_{\alpha}(R;B)_{\omega} \tag{10.3.9}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \widetilde{D}_{\alpha}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.3.10}$$

where $\omega_{RB} \in \mathcal{N}_{A\to B}(\psi_{RA})$ and $\alpha \in [1/2, 1) \cup (1, \infty)$. It follows from Proposition 7.31 that $\widetilde{E}_{\alpha}(\mathcal{N})$ is monotonically increasing in α. Also, in Appendix 10.A, we prove that

$$E_R(\mathcal{N}) = \lim_{\alpha\to 1^+} \widetilde{E}_{\alpha}(\mathcal{N}) \tag{10.3.11}$$

for every quantum channel $\mathcal{N}$.

4. The *max-relative entropy of entanglement of* $\mathcal{N}$,

$$E_{\max}(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} E_{\max}(R;B)_{\omega} \tag{10.3.12}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} D_{\max}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.3.13}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$. In Appendix 10.A, we prove that

$$E_{\max}(\mathcal{N}) = \lim_{\alpha\to\infty} \widetilde{E}_{\alpha}(\mathcal{N}) \tag{10.3.14}$$

for every quantum channel $\mathcal{N}$.

The generalized divergence of entanglement of a quantum channel satisfies all of the general properties of a channel entanglement measure shown in Proposition 10.2 except for the superadditivity property, which holds when the generalized divergence of a bipartite state is superadditive. However, as shown in Proposition 9.16, the generalized divergence of entanglement of a bipartite state is generally only subadditive. Thus, neither the superadditivity nor the subadditivity of the generalized divergence of entanglement of a channel immediately follows. In

Section 10.6 below, we show that the max-relative entropy of entanglement of a quantum channel is subadditive, meaning that

$$E_{\max}(\mathcal{N} \otimes \mathcal{M}) \leq E_{\max}(\mathcal{N}) + E_{\max}(\mathcal{M}) \tag{10.3.15}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$.

The amortized generalized divergence of entanglement, defined according to Definition 10.3 as

$$\boldsymbol{E}^{\mathcal{A}}(\mathcal{N}) := \sup_{\rho_{A'AB'}} \{\boldsymbol{E}(A'; BB')_{\omega} - \boldsymbol{E}(A'A; B)_{\rho}\}, \tag{10.3.16}$$

where $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$, satisfies all of the properties stated in Proposition 10.5. In particular, it is subadditive. We show in Section 10.6 below that $E_{\max}(\mathcal{N}) = E^{\mathcal{A}}_{\max}(\mathcal{N})$ for every quantum channel $\mathcal{N}$, and this is what leads to the subadditivity statement in (10.3.15).

For covariant channels, the optimization over pure input states in the generalized divergence of entanglement can be simplified, as we now show. This simplification is similar to the simplification that occurs for the generalized channel divergence for jointly covariant channels (see Proposition 7.84).

Proposition 10.9 Generalized Divergence of Entanglement for Covariant Channels

Let $\mathcal{N}_{A\to B}$ be a G-covariant quantum channel for a finite group G (recall Definition 4.18). Then, for every pure state ψ_{RA}, with the dimension of R equal to the dimension of A, we have that

$$\boldsymbol{E}(R; B)_{\omega} \leq \boldsymbol{E}(R; B)_{\overline{\omega}}, \tag{10.3.17}$$

where $\omega_{RB} := \mathcal{N}_{A\to B}(\psi_{RA})$, $\overline{\omega}_{RB} := \mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})$,

$$\overline{\rho}_A = \frac{1}{|G|} \sum_{g\in G} U^g_A \psi_A U^{g\dagger}_A =: \mathcal{T}_G(\psi_A), \tag{10.3.18}$$

and $\phi^{\overline{\rho}}_{RA}$ is a purification of $\overline{\rho}_A$. Consequently,

$$\boldsymbol{E}(\mathcal{N}) = \sup_{\phi_{RA}} \{\boldsymbol{E}(R; B)_{\omega} : \phi_A = \mathcal{T}_G(\phi_A)\}, \tag{10.3.19}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})$. In other words, in order to calculate $\boldsymbol{E}(\mathcal{N})$, it suffices to optimize with respect to pure states ϕ_{RA} such that the reduced state ϕ_A is invariant under the channel $\mathcal{T}_G$ defined in (10.3.18).

Remark: Using (10.3.3), we can write (10.3.17) as

$$\boldsymbol{E}(\mathcal{N},\rho) \leq \boldsymbol{E}(\mathcal{N},\mathcal{T}_G(\rho)), \tag{10.3.20}$$

which holds for every state ρ acting on the input space of the channel $\mathcal{N}$. We can write (10.3.19) as

$$\boldsymbol{E}(\mathcal{N}) = \sup_{\rho}\{\boldsymbol{E}(\mathcal{N},\rho) : \rho = \mathcal{T}_G(\rho)\}. \tag{10.3.21}$$

Proof: The proof is similar to the proof of Proposition 7.84. Let ψ_{RA} be an arbitrary pure state, and let $\overline{\rho}_A = \mathcal{T}_G(\psi_A)$. Furthermore, let $\phi^{\overline{\rho}}_{RA}$ be a purification of $\overline{\rho}_A$. Let us also consider the following purification of $\overline{\rho}_A$:

$$|\psi^{\overline{\rho}}\rangle_{R'RA} \coloneqq \frac{1}{\sqrt{|G|}}\sum_{g\in G}|g\rangle_{R'}\otimes(\mathbb{1}_R\otimes U^g_A)|\psi\rangle_{RA}, \tag{10.3.22}$$

where $\{|g\rangle_{R'}\}_{g\in G}$ is an orthonormal basis for $\mathcal{H}_{R'}$ indexed by the elements of G. Since all purifications of a state can be mapped to each other by isometries on the purifying systems, there exists an isometry $W_{R\to R'R}$ such that $|\psi^{\overline{\rho}}\rangle_{R'RA} = W_{R\to R'R}|\phi^{\overline{\rho}}\rangle_{RA}$. Then, because the set SEP of separable states is invariant under local isometries, for every state $\sigma_{RB}\in \text{SEP}(R:B)$ we have that $\tau_{R'RB} \coloneqq \mathcal{W}_{R\to R'R}(\sigma_{RB}) \in \text{SEP}(R'R\!:\!B)$. Therefore,

$$\begin{aligned}
&\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{\overline{\rho}}_{R'RA})\|\tau_{R'RB}) && (10.3.23)\\
&\quad = \boldsymbol{D}(\mathcal{N}_{A\to B}(\mathcal{W}_{R\to R'R}(\phi^{\overline{\rho}}_{RA}))\|\mathcal{W}_{R\to R'A'}(\sigma_{RB})) && (10.3.24)\\
&\quad = \boldsymbol{D}(\mathcal{W}_{R\to R'R}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA}))\|\mathcal{W}_{R\to R'R}(\sigma_{RB})) && (10.3.25)\\
&\quad = \boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})\|\sigma_{RB}), && (10.3.26)
\end{aligned}$$

where, to obtain the last equality, we used the fact that any generalized divergence is isometrically invariant (recall Proposition 7.16). Now, if we apply the dephasing channel $X\mapsto\sum_{g\in G}|g\rangle\!\langle g|X|g\rangle\!\langle g|$ to the R' system, then by the data-processing inequality for the generalized divergence $\boldsymbol{D}$, we obtain

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi^{\overline{\rho}}_{R'RA})\|\tau_{R'RB})$$

$$\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes(\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{RA})\middle\|\sum_{g\in G}p(g)|g\rangle\!\langle g|_{R'}\otimes\tau_{RB}^g\right) \tag{10.3.27}$$

$$= \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes((\mathcal{V}_B^g)^\dagger\circ\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g)(\psi_{RA})\middle\| \sum_{g\in G}p(g)|g\rangle\!\langle g|_{R'}\otimes V_B^{g\dagger}\tau_{RB}^g V_B^g\right), \tag{10.3.28}$$

where to obtain the last line we applied the unitary channel given by the unitary $\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes V_B^{g\dagger}$ and we used the fact that generalized divergences are invariant under unitaries. Furthermore, we wrote the action of the dephasing channel on $\tau_{R'RB}$ as $\sum_{g\in G}p(g)|g\rangle\!\langle g|_{R'}\otimes\tau_{RB}^g$, where $p: G\to[0,1]$ is a probability distribution and $\{\tau_{RB}^g\}_{g\in G}$ is a set of states. This operator is in the set $\text{SEP}(R'R:B)$ because the set SEP is closed under local channels. Next, due to the covariance of $\mathcal{N}$, we have that $(\mathcal{V}_B^g)^\dagger\circ\mathcal{N}\circ\mathcal{U}_A^g=\mathcal{N}$, so that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{R'RA}^{\overline{\rho}})\|\tau_{R'RB}) \tag{10.3.29}$$

$$\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes\mathcal{N}_{A\to B}(\psi_{RA})\middle\|\sum_{g\in G}p(g)|g\rangle\!\langle g|_{R'}\otimes V_B^{g\dagger}\tau_{RB}^g V_B^g\right) \tag{10.3.30}$$

$$\geq \boldsymbol{D}\left(\mathcal{N}_{A\to B}(\psi_{RA})\middle\|\sum_{g\in G}p(g)V_B^{g\dagger}\tau_{RB}^g V_B^g\right), \tag{10.3.31}$$

where to obtain the last inequality we used the data-processing inequality for $\boldsymbol{D}$ under the channel $\text{Tr}_{R'}$. Now, observe that the state $\sum_{g\in G}p(g)|g\rangle\!\langle g|_{R'}\otimes V_B^{g\dagger}\tau_{RB}^g V_B^g$ is in the set $\text{SEP}(R'R:B)$. This is due to the fact that $\sum_{g\in G}|g\rangle\!\langle g|_{R'}\otimes V_B^{g\dagger}$ is a controlled unitary, and since register R' is classical, this controlled unitary can be implemented as an LOCC channel. Also, the set SEP is closed under LOCC channels. It follows then that $\sum_{g\in G}p(g)V_B^{g\dagger}\tau_{RB}V_B^g\in\text{SEP}(R:B)$ because we obtain it from the previous separable state by applying a local partial trace over R'. By taking the infimum over every state $\tau_{RB}\in\text{SEP}(R:B)$ in (10.3.31), we have that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\phi_{RA}^{\overline{\rho}})\|\sigma_{RB})=\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{R'RA}^{\overline{\rho}})\|\tau_{R'RB}) \tag{10.3.32}$$

$$\geq \inf_{\tau_{RB} \in \text{SEP}(R:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA}) \| \tau_{RB}) \tag{10.3.33}$$

$$= \boldsymbol{E}(R;B)_{\omega}, \tag{10.3.34}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$. This inequality holds for every state ψ_{RA} and every state $\sigma_{RB} \in \text{SEP}(R:B)$. Therefore, optimizing over all $\sigma_{RB} \in \text{SEP}(R:B)$ leads to

$$\inf_{\sigma_{RB} \in \text{SEP}(R:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA}) \| \sigma_{RB}) = \boldsymbol{E}(R;B)_{\overline{\omega}} \geq \boldsymbol{E}(R;B)_{\omega}, \tag{10.3.35}$$

where $\overline{\omega}_{RB} = \mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})$. This is precisely the inequality in (10.3.17).

Next, by construction, the state $\phi^{\overline{\rho}}_{RA}$ is such that its reduced state on A is invariant under the channel $\mathcal{T}_G$. Optimizing over all such states leads to

$$\sup_{\phi_{RA}} \{\boldsymbol{E}(R;B)_{\overline{\omega}} : \phi_A = \mathcal{T}_G(\phi_A), \overline{\omega}_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\} \geq \boldsymbol{E}(R;B)_{\omega}. \tag{10.3.36}$$

Since this inequality holds for every pure state ψ_{RA}, we finally obtain

$$\sup_{\phi_{RA}} \{\boldsymbol{E}(R;B)_{\omega} : \phi_A = \mathcal{T}_G(\phi_A), \omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})\} \geq \boldsymbol{E}(\mathcal{N}). \tag{10.3.37}$$

Since the reverse inequality trivially holds, we obtain (10.3.19). ■

We saw in Section 9.2.1 that both the max-relative entropy of entanglement and the hypothesis testing relative entropy of entanglement can be formulated as cone programs. We now show that the max-relative entropy of entanglement for channels can also be formulated as a cone program.

Proposition 10.10 Cone Program for the Max-Relative Entropy of Entanglement of a Quantum Channel

Let $\mathcal{N}_{A\to B}$ be a quantum channel. Then

$$E_{\max}(\mathcal{N}) = \log_2 \Sigma_{\max}(\mathcal{N}), \tag{10.3.38}$$

where

$$\Sigma_{\max}(\mathcal{N}) := \inf_{Y_{SB} \in \widehat{\text{SEP}}} \left\{ \|\text{Tr}_B[Y_{SB}]\|_\infty : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\}, \tag{10.3.39}$$

and $\Gamma^{\mathcal{N}}_{SB}$ is the Choi operator of the channel $\mathcal{N}_{A\to B}$.

Proof: Using the definition of a channel's max-relative entropy of entanglement, and the cone program formulation of the max-relative entropy of entanglement for states from Proposition 9.21, we have

$$E_{\max}(\mathcal{N}) = \sup_{\psi_{SA}} E_{\max}(S;B)_{\omega} \tag{10.3.40}$$

$$= \sup_{\psi_{SA}} \inf_{X_{SB}\in\widehat{\mathrm{SEP}}} \log_2\{\mathrm{Tr}[X_{SB}] : \mathcal{N}_{A\to B}(\psi_{SA}) \leq X_{SB}\}, \tag{10.3.41}$$

where $\omega_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$. Now, recall from (2.2.40) that an arbitrary pure bipartite state ψ_{SA} can be written as $Z_S\Gamma_{SA}Z_S^\dagger$, where $\Gamma_{SA} = |\Gamma\rangle\langle\Gamma|$, $|\Gamma\rangle_{SA} = \sum_{i=0}^{d_A-1}|i,i\rangle_{SA}$, and Z_S is an operator satisfying $\mathrm{Tr}[Z_S^\dagger Z_S] = 1$. Then

$$\mathcal{N}_{A\to B}(\psi_{SA}) = \mathcal{N}_{A\to B}(Z_S\Gamma_{SA}Z_S^\dagger) \tag{10.3.42}$$

$$= Z_S\mathcal{N}_{A\to B}(\Gamma_{SA})Z_S^\dagger \tag{10.3.43}$$

$$= Z_S\Gamma_{SB}^{\mathcal{N}}Z_S^\dagger, \tag{10.3.44}$$

where $\Gamma_{SB}^{\mathcal{N}} = \mathcal{N}_{A\to B}(\Gamma_{SA})$ is the Choi operator of $\mathcal{N}_{A\to B}$. Since the set of operators Z_S satisfying $Z_S^\dagger Z_S > 0$ and $\mathrm{Tr}[Z_S^\dagger Z_S] = 1$ is dense in the set of all operators satisfying $\mathrm{Tr}[Z_S^\dagger Z_S] = 1$, we find that

$$E_{\max}(\mathcal{N}) = \log_2 \sup_{Z_S} \inf_{X_{SB}\in\widehat{\mathrm{SEP}}} \left\{\mathrm{Tr}[X_{SB}] : Z_S\Gamma_{SB}^{\mathcal{N}}Z_S^\dagger \leq X_{SB},\ Z_S^\dagger Z_S > 0,\ \mathrm{Tr}[Z_S^\dagger Z_S] = 1\right\}. \tag{10.3.45}$$

Let us now make a change of variable, defining the variable Y_{SB} according to the relation $X_{SB} = Z_SY_{SB}Z_S^\dagger$. Then, since

$$Z_S\Gamma_{SB}^{\mathcal{N}}Z_S^\dagger \leq X_{SB} = Z_SY_{SB}Z_S^\dagger \quad\Longleftrightarrow\quad \Gamma_{SB}^{\mathcal{N}} \leq Y_{SB}, \tag{10.3.46}$$

$$X_{SB} \in \widehat{\mathrm{SEP}} \quad\Longleftrightarrow\quad Y_{SB} \in \widehat{\mathrm{SEP}}, \tag{10.3.47}$$

we find that

$$\text{Eq. (10.3.45)}$$

$$= \sup_{Z_S} \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \left\{\mathrm{Tr}[Z_SY_{SB}Z_S^\dagger] : \Gamma_{SB}^{\mathcal{N}} \leq Y_{SB},\ Z_S^\dagger Z_S > 0,\ \mathrm{Tr}[Z_S^\dagger Z_S] = 1\right\}$$

$$= \sup_{Z_S} \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \left\{\mathrm{Tr}[Z_S^\dagger Z_SY_{SB}] : \Gamma_{SB}^{\mathcal{N}} \leq Y_{SB},\ Z_S^\dagger Z_S > 0,\ \mathrm{Tr}[Z_S^\dagger Z_S] = 1\right\}$$

$$= \sup_{\rho_S} \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \left\{ \mathrm{Tr}[\rho_S Y_{SB}] : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\}, \qquad (10.3.48)$$

where in the last line we made the substitution $\rho_S = Z_S^\dagger Z_S$, so that the optimization is with respect to density operators. Furthermore, we have employed the fact that the set of density operators satisfying $\rho_S > 0$ is dense in the set of all density operators. Now observing that the objective function is linear in ρ_S and Y_{SB}, the set of density operators is compact and convex, and the set of separable operators is convex, the Sion minimax theorem (Theorem 2.24) applies, such that we can exchange the optimizations to find that

$$\sup_{\rho_S} \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \left\{ \mathrm{Tr}[\rho_S Y_{SB}] : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\}$$

$$= \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \sup_{\rho_S} \left\{ \mathrm{Tr}[\rho_S Y_{SB}] : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\} \qquad (10.3.49)$$

$$= \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \sup_{\rho_S} \left\{ \mathrm{Tr}[\rho_S \mathrm{Tr}_B[Y_{SB}]] : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\} \qquad (10.3.50)$$

$$= \inf_{Y_{SB}\in\widehat{\mathrm{SEP}}} \left\{ \|\mathrm{Tr}_B[Y_{SB}]\|_\infty : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB} \right\} \qquad (10.3.51)$$

$$= \Sigma_{\max}(\mathcal{N}). \qquad (10.3.52)$$

The second equality follows from partial trace, and the third follows because $\|X\|_\infty = \sup_\rho \mathrm{Tr}[X\rho]$ for positive semi-definite operators X, where the optimization is with respect to density operators (see (2.2.108)). ■

10.4 Generalized Rains Divergence

We now examine the generalized Rains divergence of quantum channels, which is a channel entanglement measure that arises from the generalized Rains divergence for bipartite quantum states that we considered in Section 9.3.

Definition 10.11 Generalized Rains Information of a Quantum Channel

Let $\boldsymbol{D}$ be a generalized divergence (see Definition 7.15. For every quantum channel $\mathcal{N}_{A\to B}$, we define the *generalized Rains information of* $\mathcal{N}$ as

$$\boldsymbol{R}(\mathcal{N}) := \sup_{\psi_{SA}} \boldsymbol{R}(S;B)_\omega \qquad (10.4.1)$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB} \in \mathrm{PPT}'(S:B)} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{SA}) \| \sigma_{SB}), \tag{10.4.2}$$

where $\omega_{SB} \coloneqq \mathcal{N}_{A\to B}(\psi_{SA})$. The supremum is with respect to every pure state ψ_{SA}, with the dimension of S the same as the dimension of A.

In the remark immediately after Definition 10.1 we show that it suffices to optimize with respect to pure bipartite states (with equal dimension for each subsystem) when calculating the generalized Rains divergence of a quantum channel.

We can write the generalized Rains divergence of $\mathcal{N}_{A\to B}$ in the following alternate form:

$$\boldsymbol{R}(\mathcal{N}) = \sup_{\rho_A} \boldsymbol{R}(\mathcal{N}_{A\to B}, \rho_A), \tag{10.4.3}$$

$$\boldsymbol{R}(\mathcal{N}_{A\to B}, \rho_A) \coloneqq \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} \boldsymbol{D}(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A} \| \sigma_{AB}). \tag{10.4.4}$$

This is indeed true because we can write every purification of ρ_A as $(V_S\sqrt{\rho_S} \otimes \mathbb{1}_A)|\Gamma\rangle_{SA}$ for some isometry V_S (see (2.2.40) and Theorem 2.3), the set of PPT′ operators is invariant under local isometries, and the generalized divergences are invariant under local isometries (see Proposition 7.16).

As with the state quantities, we are interested in the following generalized Rains information quantities for every quantum channel $\mathcal{N}_{A\to B}$. For every case below, we define $\omega_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$.

1. The *Rains information of* $\mathcal{N}$,

$$R(\mathcal{N}) \coloneqq \sup_{\psi_{SA}} R(S;B)_\omega \tag{10.4.5}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB} \in \mathrm{PPT}'(S:B)} D(\mathcal{N}_{A\to B}(\psi_{SA}) \| \sigma_{SB}). \tag{10.4.6}$$

2. The *ε-hypothesis testing Rains information of* $\mathcal{N}$,

$$R^\varepsilon_H(\mathcal{N}) \coloneqq \sup_{\psi_{SA}} R^\varepsilon_H(S;B)_\omega \tag{10.4.7}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB} \in \mathrm{PPT}'(S:B)} D^\varepsilon_H(\mathcal{N}_{A\to B}(\psi_{SA}) \| \sigma_{SB}), \tag{10.4.8}$$

3. The *sandwiched Rényi Rains information of* $\mathcal{N}$,

$$\widetilde{R}_\alpha(\mathcal{N}) := \sup_{\psi_{SA}} \widetilde{R}_\alpha(S;B)_\omega \tag{10.4.9}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\mathrm{PPT}'(S:B)} \widetilde{D}_\alpha(\mathcal{N}_{A'\to B}(\psi_{SA})\|\sigma_{SB}), \tag{10.4.10}$$

where $\alpha \in [1/2, 1) \cup (1, \infty)$. It follows from Proposition 7.31 that $\widetilde{R}_\alpha(\mathcal{N})$ is monotonically increasing in α. Also, in Appendix 10.A, we prove that

$$R(\mathcal{N}) = \lim_{\alpha\to 1} \widetilde{R}_\alpha(\mathcal{N}) \tag{10.4.11}$$

for every quantum channel $\mathcal{N}$.

4. The *max-Rains information of* $\mathcal{N}$,

$$R_{\max}(\mathcal{N}) := \sup_{\psi_{SA}} R_{\max}(S;B)_\omega \tag{10.4.12}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\mathrm{PPT}'(S:B)} D_{\max}(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}). \tag{10.4.13}$$

In Appendix 10.A, we prove that

$$R_{\max}(\mathcal{N}) = \lim_{\alpha\to\infty} \widetilde{R}_\alpha(\mathcal{N}) \tag{10.4.14}$$

for every quantum channel $\mathcal{N}$.

The generalized Rains divergence of a quantum channel satisfies all of the properties of a channel entanglement measure laid out in Proposition 10.2, except for faithfulness and superadditivity. Faithfulness generally does not hold because the generalized Rains divergence of a bipartite quantum state is not faithful. Superadditivity does not hold in general because the Rains divergence of a bipartite quantum state is generally only subadditive. The max-Rains information of a quantum channel, however, is additive, meaning that

$$R_{\max}(\mathcal{N}\otimes\mathcal{M}) = R_{\max}(\mathcal{N}) + R_{\max}(\mathcal{M}) \tag{10.4.15}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. We defer a proof of this to Section 10.6 below.

The amortized generalized Rains divergence $\boldsymbol{R}^{\mathcal{A}}(\mathcal{N})$, defined according to Definition 10.3 as

$$\boldsymbol{R}^{\mathcal{A}}(\mathcal{N}) := \sup_{\rho_{A'AB'}} \{\boldsymbol{R}(A';BB')_\omega - \boldsymbol{R}(A'A;B)_\rho\}, \tag{10.4.16}$$

where $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$, satisfies all of the properties stated in Proposition 10.5 except for faithfulness, because the generalized Rains divergence of a bipartite quantum state is not faithful. In particular, due to additivity of the max-Rains relative entropy (Proposition 9.29), we immediately obtain additivity of the amortized max-Rains information of a quantum channel, i.e.,

$$R^{\mathcal{A}}_{\max}(\mathcal{N} \otimes \mathcal{M}) = R^{\mathcal{A}}_{\max}(\mathcal{N}) + R^{\mathcal{A}}_{\max}(\mathcal{M}) \tag{10.4.17}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. We show in Section 10.6 below that

$$R_{\max}(\mathcal{N}) = R^{\mathcal{A}}_{\max}(\mathcal{N}) \tag{10.4.18}$$

for every quantum channel $\mathcal{N}$, and it is this fact that leads to the additivity statement in (10.4.15).

For covariant channels, the optimization over pure input states in the generalized Rains divergence simplifies in the same way as it does for the generalized divergence of entanglement.

Proposition 10.12 Generalized Rains Information for Covariant Channels

Let $\mathcal{N}_{A\to B}$ be a G-covariant quantum channel for a finite group G (recall Definition 4.18). Then, for every pure state ψ_{RA}, with the dimension of R equal to the dimension of A, we have that

$$\boldsymbol{R}(S;B)_\omega \leq \boldsymbol{R}(S;B)_{\overline{\omega}}, \tag{10.4.19}$$

where $\omega_{RB} \coloneqq \mathcal{N}_{A\to B}(\psi_{RA})$, $\overline{\omega}_{RB} \coloneqq \mathcal{N}_{A\to B}(\phi^{\overline{\rho}}_{RA})$,

$$\overline{\rho}_A = \frac{1}{|G|}\sum_{g\in G} U^g_A \rho_A U^{g\dagger}_A \eqqcolon \mathcal{T}_G(\rho_A), \tag{10.4.20}$$

$\rho_A = \psi_A = \operatorname{Tr}_S[\psi_{SA}]$, and $\phi^{\overline{\rho}}_{SA}$ a purification of $\overline{\rho}_A$. Consequently,

$$\boldsymbol{R}(\mathcal{N}) = \sup_{\phi_{SA}}\{\boldsymbol{R}(S;B)_\omega : \phi_A = \mathcal{T}_G(\phi_A), \omega_{SB} = \mathcal{N}_{A\to B}(\phi_{SA})\}. \tag{10.4.21}$$

In other words, in order to calculate $\boldsymbol{R}(\mathcal{N})$, it suffices to optimize with respect to pure states ϕ_{SA} such that the reduced state ϕ_A is invariant under the channel $\mathcal{T}_G$ defined in (10.4.20).

REMARK: Using (10.4.3), we can write (10.4.19) as

$$\boldsymbol{R}(\mathcal{N},\rho) \leq \boldsymbol{R}(\mathcal{N},\mathcal{T}_G(\rho)), \tag{10.4.22}$$

which holds for every state ρ acting on the input space of the channel $\mathcal{N}$. We can write (10.4.21) as

$$\boldsymbol{R}(\mathcal{N}) = \sup_R \{\boldsymbol{R}(\mathcal{N},\rho) : \rho = \mathcal{T}_G(\rho)\}. \tag{10.4.23}$$

PROOF: The proof is identical to the proof of Proposition 10.9, with the exception that the set PPT' is involved rather than the set SEP. The LOCC channels discussed there preserve the set PPT', and this is the main reason why the same proof applies. ■

Using the SDP formulation of max-Rains relative entropy in Proposition 9.27, we arrive at an SDP formulation for the max-Rains information of a quantum channel.

Proposition 10.13 SDPs for Max-Rains Information of a Quantum Channel

Let $\mathcal{N}_{A\to B}$ be a quantum channel. Then

$$R_{\max}(\mathcal{N}) = \log_2 \Gamma_{\max}(\mathcal{N}), \tag{10.4.24}$$

where

$$\Gamma_{\max}(\mathcal{N})$$
$$= \inf_{Y_{SB},V_{SB}\geq 0} \{\|\mathrm{Tr}_B[V_{SB}+Y_{SB}]\|_\infty : \mathrm{T}_B(V_{SB}-Y_{SB}) \geq \Gamma^{\mathcal{N}}_{SB}\} \tag{10.4.25}$$
$$= \sup_{\rho_S\geq 0} \{\mathrm{Tr}[\Gamma^{\mathcal{N}}_{SB}X_{SB}] : \mathrm{Tr}[\rho_S] \leq 1,\ -\rho_S\otimes \mathbb{1}_B \leq \mathrm{T}_B(X_{SB}) \leq \rho_S\otimes\mathbb{1}_B\}. \tag{10.4.26}$$

PROOF: To arrive at (10.4.26), consider that, with $\omega_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$,

$$2^{R_{\max}(\mathcal{N})}$$
$$= \sup_{\psi_{SA}} 2^{R_{\max}(S;B)_\omega} \tag{10.4.27}$$
$$= \sup \{\mathrm{Tr}[\mathcal{N}_{A\to B}(\psi_{SA})X_{SB}] : \|\mathrm{T}_B(X_{SB})\|_\infty \leq 1, X_{SB}\geq 0\}, \tag{10.4.28}$$

where the last equality follows from (9.3.47)–(9.3.49). Recall from (2.2.40) that an arbitrary pure bipartite state ψ_{SA} can be written as $Z_S \Gamma_{SA} Z_S^\dagger$, where $\Gamma_{SA} = |\Gamma\rangle\langle\Gamma|_{SA}$, $|\Gamma\rangle_{SA} = \sum_{i=0}^{d_A-1} |i,i\rangle_{SA}$, and Z_S is an operator satisfying $\text{Tr}[Z_S^\dagger Z_S] = 1$. Then

$$\text{Tr}[\mathcal{N}_{A\to B}(\psi_{SA})X_{SB}] = \text{Tr}[\mathcal{N}_{A\to B}(Z_S\Gamma_{SA}Z_S^\dagger)X_{SB}] \tag{10.4.29}$$

$$= \text{Tr}[Z_S\mathcal{N}_{A\to B}(\Gamma_{SA})Z_S^\dagger X_{SB}] \tag{10.4.30}$$

$$= \text{Tr}[\Gamma_{SB}^{\mathcal{N}} Z_S^\dagger X_{SB} Z_S], \tag{10.4.31}$$

where $\Gamma_{SB}^{\mathcal{N}} = \mathcal{N}_{A\to B}(\Gamma_{SA})$ denotes the Choi operator of the channel $\mathcal{N}_{A\to B}$. Since the set of operators Z_S satisfying $Z_S^\dagger Z_S > 0$ and $\text{Tr}[Z_S^\dagger Z_S] = 1$ is dense in the set of all operators satisfying $\text{Tr}[Z_S^\dagger Z_S] = 1$, we find that

$$2^{R_{\max}(\mathcal{N})} = \sup\{\text{Tr}[\Gamma_{SB}^{\mathcal{N}} Z_S^\dagger X_{SB} Z_S] : \|\text{T}_B(X_{SB})\|_\infty \le 1, \\ X_{SB} \ge 0, Z_S^\dagger Z_S > 0,\ \text{Tr}[Z_S^\dagger Z_S] = 1\}. \tag{10.4.32}$$

Consider that $X_{SB} \ge 0 \Leftrightarrow Z_S^\dagger X_{SB} Z_S \ge 0$ and

$$\|\text{T}_B(X_{SB})\|_\infty \le 1$$

$$\Longleftrightarrow \quad -\mathbb{1}_{SB} \le \text{T}_B(X_{SB}) \le \mathbb{1}_{SB} \tag{10.4.33}$$

$$\Longleftrightarrow \quad -Z_S^\dagger Z_S \otimes \mathbb{1}_B \le Z_S^\dagger \text{T}_B(X_{SB}) Z_S \le Z_S^\dagger Z_S \otimes \mathbb{1}_B \tag{10.4.34}$$

$$\Longleftrightarrow \quad -Z_S^\dagger Z_S \otimes \mathbb{1}_B \le \text{T}_B(Z_S^\dagger X_{SB} Z_S) \le Z_S^\dagger Z_S \otimes \mathbb{1}_B. \tag{10.4.35}$$

We now set $X'_{SB} := Z_S^\dagger X_{SB} Z_S$ and $\rho_S = Z_S^\dagger Z_S > 0$ and rewrite as follows:

$$2^{R_{\max}(\mathcal{N})} = \sup\{\text{Tr}[\Gamma_{SB}^{\mathcal{N}} X'_{SB}] : -\rho_S \otimes \mathbb{1}_B \le \text{T}_B(X'_{SB}) \le \rho_S \otimes \mathbb{1}_B, \\ X'_{SB} \ge 0,\ \rho_S > 0, \text{Tr}[\rho_S] = 1\}, \tag{10.4.36}$$

which is the equality in (10.4.24) and (10.4.26), after observing that the set $\{\rho_S : \rho_S > 0, \text{Tr}[\rho_S] = 1\}$ is dense in the set $\{\rho_S : \rho_S \ge 0, \text{Tr}[\rho_S] = 1\}$.

To arrive at the equality in (10.4.25), we employ the dual formulation of the max-Rains relative entropy in (9.3.49). Consider that

$$2^{R_{\max}(\mathcal{N})}$$

$$= \sup_{\psi_{SA}} 2^{R_{\max}(S;B)_\omega} \tag{10.4.37}$$

$$= \sup_{\psi_{SA}} \inf_{K_{SB}, L_{SB} \geq 0} \{\text{Tr}[K_{SB} + L_{SB}] : \text{T}_B(K_{SB} - L_{SB}) \geq \mathcal{N}_{A\to B}(\psi_{SA})\}. \quad (10.4.38)$$

Making the same observations as we did previously, we have that $\mathcal{N}_{A\to B}(\psi_{SA}) = Z_S \Gamma^{\mathcal{N}}_{SB} Z_S^\dagger$, as well as

$$\text{T}_B(K_{SB} - L_{SB}) \geq \mathcal{N}_{A\to B}(\psi_{SA}) \quad \Longleftrightarrow \quad \text{T}_B\big(K'_{SB} - L'_{SB}\big) \geq \Gamma^{\mathcal{N}}_{SB}, \quad (10.4.39)$$

where K'_{SB} and L'_{SB} are such that $K_{SB} = Z_S K'_{SB} Z_S^\dagger$ and $L_{SB} = Z_S L'_{SB} Z_S^\dagger$, respectively. Then $K_{SB}, L_{SB} \geq 0 \Longleftrightarrow K'_{SB}, L'_{SB} \geq 0$, and we find that

$$2^{R_{\max}(\mathcal{N})} = \sup_{Z_S} \inf_{K'_{SB}, L'_{SB} \geq 0} \{\text{Tr}[Z_S K'_{SB} Z_S^\dagger + Z_S L'_{SB} Z_S^\dagger] : \text{T}_B\big(K'_{SB} - L'_{SB}\big) \geq \Gamma^{\mathcal{N}}_{SB},$$
$$Z_S^\dagger Z_S > 0,\ \text{Tr}[Z_S^\dagger Z_S] = 1\}. \quad (10.4.40)$$

Employing cyclicity of trace, setting $\rho_S = Z_S^\dagger Z_S$, and exploiting the fact that the set $\{\rho_S : \rho_S > 0, \text{Tr}[\rho_S] = 1\}$ is dense in the set $\{\rho_S : \rho_S \geq 0, \text{Tr}[\rho_S] = 1\}$, we find that

$$2^{R_{\max}(\mathcal{N})} = \sup_{\rho_S} \inf_{K'_{SB}, L'_{SB}} \{\text{Tr}[\rho_S(K'_{SB} + L'_{SB})] : K'_{SB}, L'_{SB} \geq 0,$$
$$\text{T}_B\big(K'_{SB} - L'_{SB}\big) \geq \Gamma^{\mathcal{N}}_{SB},\ \rho_S \geq 0,\ \text{Tr}[\rho_S] = 1\}. \quad (10.4.41)$$

The function that we are optimizing is linear in ρ_S and jointly convex in K'_{SB} and L'_{SB} (the set with respect to which the infimum is performed is also compact), so that the minimax theorem (Theorem 2.24) applies and we can exchange sup with inf to find that

$$2^{R_{\max}(\mathcal{N})} = \inf_{K'_{SB}, L'_{SB}} \sup_{\rho_S} \{\text{Tr}[\rho_S(K'_{SB} + L'_{SB})] : K'_{SB}, L'_{SB} \geq 0,$$
$$\text{T}_B\big(K'_{SB} - L'_{SB}\big) \geq \Gamma^{\mathcal{N}}_{SB},\ \rho_S \geq 0,\ \text{Tr}[\rho_S] = 1\}. \quad (10.4.42)$$

For fixed K'_{SB} and L'_{SB}, consider that

$$\sup_{\rho_S} \{\text{Tr}[\rho_S(K'_{SB} + L'_{SB})] : \rho_S \geq 0,\ \text{Tr}[\rho_S] = 1\}$$
$$= \sup_{\rho_S} \{\text{Tr}[\rho_S \text{Tr}_B[K'_{SB} + L'_{SB}]] : \rho_S \geq 0,\ \text{Tr}[\rho_S] = 1\} \quad (10.4.43)$$
$$= \left\|\text{Tr}_B[K'_{SB} + L'_{SB}]\right\|_\infty, \quad (10.4.44)$$

where for the last line we used (2.2.108). Substituting back in, we find that

$$2^{R_{\max}(\mathcal{N})} = \inf_{K'_{SB},L'_{SB}} \{\left\|\mathrm{Tr}_B[K'_{SB} + L'_{SB}]\right\|_\infty : K'_{SB}, L'_{SB} \geq 0,$$
$$\mathrm{T}_B(K'_{SB} - L'_{SB}) \geq \Gamma^{\mathcal{N}}_{SB}\}, \quad (10.4.45)$$

as claimed in (10.4.26).

According to Theorem 2.28, strong duality holds by picking V_{SB} and Y_{SB} equal to the positive and negative parts of $\mathrm{T}_B(\Gamma^{\mathcal{N}}_{SB})$, respectively, which are feasible for (10.4.26). Furthermore, we can pick $\rho_S = \mathbb{1}_S/(2d_S)$ and $X_{SB} = \mathbb{1}_{SB}/(3d_S)$, which are strictly feasible for (10.4.25). ■

10.5 Squashed Entanglement

We now move on to the squashed entanglement of a quantum channel, which is a channel entanglement measure that arises from the squashed entanglement of a bipartite state, the latter defined in Section 9.4.

Definition 10.14 Squashed Entanglement of a Quantum Channel

For every quantum channel $\mathcal{N}_{A\to B}$, we define the *squashed entanglement of* $\mathcal{N}$ as

$$E_{\mathrm{sq}}(\mathcal{N}) = \sup_{\psi_{RA}} E_{\mathrm{sq}}(R;B)_\omega \quad (10.5.1)$$

$$= \frac{1}{2} \sup_{\psi_{RA}} \inf_{\tau_{RBE}} \{I(R;B|E)_\tau : \mathrm{Tr}_E[\tau_{RBE}] = \omega_{RB}\}, \quad (10.5.2)$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$ and the quantum conditional mutual information $I(R;B|E)_\tau$ is defined as

$$I(R;B|E)_\tau = H(R|E)_\tau + H(B|E)_\tau - H(RB|E)_\tau. \quad (10.5.3)$$

We can write the squashed entanglement of a quantum channel in the following alternate form:

$$E_{\mathrm{sq}}(\mathcal{N}) = \sup_{\rho_A} E_{\mathrm{sq}}(\mathcal{N},\rho), \quad (10.5.4)$$

$$E_{\text{sq}}(\mathcal{N}, \rho) \coloneqq E_{\text{sq}}(R; B)_{\omega}. \tag{10.5.5}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi^{\rho}_{RA})$, with ψ^{ρ}_{RA} some purification of ρ_A. This is indeed true because the squashed entanglement of a bipartite state is invariant under isometries and because all purifications of a given state are related to each other by local isometries acting on the purifying system.

Let us now recall the following alternate expression for the squashed entanglement of a bipartite state from (9.4.44):

$$E_{\text{sq}}(A; B)_{\rho} = \frac{1}{2} \inf_{\mathcal{V}_{E'\to EF}} \left\{ H(B|E)_{\theta} + H(B|F)_{\theta} : \theta_{BEF} = \mathcal{V}_{E'\to EF}(\psi^{\rho}_{ABE'}), \right\}, \tag{10.5.6}$$

where $\psi^{\rho}_{ABE'}$ is some purification of ρ_{AB}. Now, given an input state ρ_A of a quantum channel $\mathcal{N}_{A\to B}$, let ϕ_{RA} be a purification of ρ_A. Then, for the state $\omega_{RB} \coloneqq \mathcal{N}_{A\to B}(\phi_{RA})$, we can take a purification to be $\psi_{RBE'} = \mathcal{U}^{\mathcal{N}}_{A\to BE'}(\phi_{RA})$, where $\mathcal{U}^{\mathcal{N}}$ is an isometric channel that extends $\mathcal{N}$. Then, for every isometric channel $\mathcal{V}_{E'\to EF}$, we can define the state

$$\theta_{BEF} = \mathcal{V}_{E'\to EF}(\psi_{BE'}) = (\mathcal{V}_{E'\to EF} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE'})(\rho_A), \tag{10.5.7}$$

where $\psi_{BE'} = \text{Tr}_R[\psi_{RBE'}]$. We then have

$$E_{\text{sq}}(\mathcal{N}, \rho_A) = E(R; B)_{\omega} = \frac{1}{2} \inf_{\mathcal{V}_{E'\to EF}} \left\{ H(B|E)_{\theta} + H(B|F)_{\theta} : \theta_{BEF} = (\mathcal{V}_{E'\to EF} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE'})(\rho_A) \right\} \tag{10.5.8}$$

for every state ρ_A. We thus obtain the following expression for the squashed entanglement of a channel:

$$E_{\text{sq}}(\mathcal{N}) = \frac{1}{2} \sup_{\rho_A} \inf_{\mathcal{V}_{E'\to EF}} \left\{ H(B|E)_{\theta} + H(B|F)_{\theta} : \theta_{BEF} = (\mathcal{V}_{E'\to EF} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE'})(\rho_A) \right\}. \tag{10.5.9}$$

The squashed entanglement of a quantum channel, as well as its amortized version $E^{\mathcal{A}}_{\text{sq}}$ defined as

$$E^{\mathcal{A}}_{\text{sq}}(\mathcal{N}) \coloneqq \sup_{\rho_{A'AB'}} \{E_{\text{sq}}(A'; BB')_{\omega} - E_{\text{sq}}(A'A; B')_{\rho}\}, \tag{10.5.10}$$

with $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$, satisfy all of the properties stated in Proposition 10.2 and Proposition 10.5, respectively. In particular, because the squashed entanglement for states is additive (see (9.4.8)), we immediately have that the amortized squashed entanglement of a channel is additive, i.e.,

$$E_{\text{sq}}^{\mathcal{A}}(\mathcal{N} \otimes \mathcal{M}) = E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) + E_{\text{sq}}^{\mathcal{A}}(\mathcal{M}) \tag{10.5.11}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. In Section 10.6 below, we prove that $E_{\text{sq}}(\mathcal{N}) = E_{\text{sq}}^{\mathcal{A}}(\mathcal{N})$ for every quantum channel $\mathcal{N}$, which then implies the additivity of the squashed entanglement of a channel, i.e.,

$$E_{\text{sq}}(\mathcal{N} \otimes \mathcal{M}) = E_{\text{sq}}(\mathcal{N}) + E_{\text{sq}}(\mathcal{M}) \tag{10.5.12}$$

for all channels $\mathcal{N}$ and $\mathcal{M}$.

Proposition 10.15

Let $\mathcal{N}_{A\to B}$ be a quantum channel. The function $\rho \mapsto E_{\text{sq}}(\mathcal{N}, \rho)$, where $E_{\text{sq}}(\mathcal{N}, \rho)$ is defined in (10.5.5), is concave: for $\mathcal{X}$ a finite alphabet, $p : \mathcal{X} \to [0, 1]$ a probability distribution on $\mathcal{X}$, and $\{\rho_A^x\}_{x\in\mathcal{X}}$ a set of states, the following inequality holds

$$E_{\text{sq}}\left(\mathcal{N}, \sum_{x\in\mathcal{X}} p(x)\rho_A^x\right) \geq \sum_{x\in\mathcal{X}} p(x) E_{\text{sq}}(\mathcal{N}, \rho_A^x). \tag{10.5.13}$$

PROOF: In order to prove this, we make use of the expression for $E_{\text{sq}}(\mathcal{N}, \rho_A)$ in (10.5.8).

For every state ρ_A^x, with $x \in \mathcal{X}$, define the state

$$\theta_{BEF}^x := (\mathcal{V}_{E'\to EF} \circ \mathcal{U}_{A\to BE'}^{\mathcal{N}})(\rho_A^x), \tag{10.5.14}$$

where $\mathcal{V}_{E'\to EF}$ is an arbitrary isometric channel. Then, using (10.5.8), we have

$$E_{\text{sq}}(\mathcal{N}, \rho_A^x) = E(R; B)_{\omega^x} \leq \frac{1}{2}(H(B|E)_{\theta^x} + H(B|F)_{\theta^x},) \tag{10.5.15}$$

Now, let

$$\overline{\rho}_A := \sum_{x\in\mathcal{X}} p(x)\rho_A^x, \tag{10.5.16}$$

$$\overline{\theta}_{BEF} \coloneqq \sum_{x\in\mathcal{X}} p(x)\theta^x_{BEF} \tag{10.5.17}$$
$$= \sum_{x\in\mathcal{X}} p(x)(\mathcal{V}_{E'\to EF}\circ\mathcal{U}^{\mathcal{N}}_{A\to BE'})(\rho^x_A) \tag{10.5.18}$$
$$= (\mathcal{V}_{E'\to EF}\circ\mathcal{U}^{\mathcal{N}}_{A\to BE'})(\overline{\rho}_A). \tag{10.5.19}$$

Using concavity of conditional entropy (see (7.2.120)), we obtain

$$\sum_{x\in\mathcal{X}} p(x)E_{\text{sq}}(\mathcal{N},\rho^x_A) \leq \frac{1}{2}\sum_{x\in\mathcal{X}} p(x)\left(H(B|E)_{\theta^x} + H(B|F)_{\theta^x}\right) \tag{10.5.20}$$
$$\leq \frac{1}{2}(H(B|E)_{\overline{\theta}} + H(B|F)_{\overline{\theta}}). \tag{10.5.21}$$

Finally, since the isometric channel $\mathcal{V}_{E'\to EF}$ is arbitrary, taking the infimum over all such channels on the right-hand side of the inequality above and using (10.5.19) gives us

$$\sum_{x\in\mathcal{X}} p(x)E_{\text{sq}}(\mathcal{N},\rho^x_A) \leq \frac{1}{2}\inf_{\mathcal{V}_{E'\to EF}}\{H(B|E)_{\overline{\theta}} + H(B|F)_{\overline{\theta}}\} \tag{10.5.22}$$
$$= E_{\text{sq}}(\mathcal{N},\overline{\rho}_A), \tag{10.5.23}$$

which is what we set out to prove. ■

10.6 Amortization Collapses

In Lemma 10.4, we proved the following relation between the entanglement $E(\mathcal{N})$ of a channel $\mathcal{N}$ and its amortized entanglement $E^{\mathcal{A}}(\mathcal{N})$:

$$E(\mathcal{N}) \leq E^{\mathcal{A}}(\mathcal{N}). \tag{10.6.1}$$

In general, therefore, amortization can yield a larger value for the entanglement of a channel than the usual channel entanglement measure.

For which entanglement measures does the reverse inequality hold? In this section, we investigate this question, and we prove that three of the channel entanglement measures that we have considered in this chapter — max-relative entropy of entanglement, max-Rains information, and squashed entanglement — satisfy the reverse inequality. Thus, for these three entanglement measures, amortization does not yield a higher entanglement value than the usual channel entanglement measure. This so-called "amortization collapse" is important because it immediately implies additivity of the usual channel entanglement measure.

10.6.1 Max-Relative Entropy of Entanglement

We start by proving that the amortization collapse occurs for max-relative entropy of entanglement. The key tools in the proof are Propositions 9.21 and 10.10, which provide cone programs for both the max-relative entropy of entanglement for bipartite states and the max-relative entropy of entanglement for quantum channels. Let us recall these now:

$$E_{\max}(A;B)_\rho = \log_2 G_{\max}(A;B)_\rho \tag{10.6.2}$$
$$= \log_2 \inf_{X_{AB}\in\widehat{\text{SEP}}} \{\text{Tr}[X_{AB}] : \rho_{AB} \leq X_{AB}\}, \tag{10.6.3}$$

$$E_{\max}(\mathcal{N}) = \log_2 \Sigma_{\max}(\mathcal{N}) \tag{10.6.4}$$
$$= \log_2 \inf_{Y_{SB}\in\widehat{\text{SEP}}} \left\{\|\text{Tr}_B[Y_{SB}]\|_\infty : \Gamma^{\mathcal{N}}_{SB} \leq Y_{SB}\right\}, \tag{10.6.5}$$

where ρ_{AB} is a bipartite state and $\mathcal{N}$ is a quantum channel with Choi operator $\Gamma^{\mathcal{N}}_{SB}$.

Theorem 10.16 Amortization Collapse for Max-Relative Entropy of Entanglement

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For every state $\rho_{A'AB'}$,

$$E_{\max}(A';BB')_\omega \leq E_{\max}(\mathcal{N}) + E_{\max}(A'A;B')_\rho, \tag{10.6.6}$$

where $\omega_{A'BB'} := \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Consequently,

$$E^{\mathcal{A}}_{\max}(\mathcal{N}) \leq E_{\max}(\mathcal{N}), \tag{10.6.7}$$

and thus (by Lemma 10.4), we have that

$$E^{\mathcal{A}}_{\max}(\mathcal{N}) = E_{\max}(\mathcal{N}) \tag{10.6.8}$$

for every quantum channel $\mathcal{N}$.

Proof: Using the cone program formulations in (10.6.2)–(10.6.5), we find that the inequality in (10.6.6) is equivalent to

$$G_{\max}(A';BB')_{\omega} = \Sigma_{\max}(\mathcal{N}) \cdot G_{\max}(A'A;B')_{\rho}. \tag{10.6.9}$$

We now set out to prove this inequality.

Using (10.6.3), we find that

$$G_{\max}(A'A;B')_{\rho} = \inf \operatorname{Tr}[C_{A'AB'}], \tag{10.6.10}$$

subject to the constraints

$$C_{A'AB'} \in \widehat{\mathrm{SEP}}(A'A\!:\!B'), \tag{10.6.11}$$

$$C_{A'AB'} \geq \rho_{A'AB'}, \tag{10.6.12}$$

and

$$G_{\max}(A';BB')_{\omega} = \inf \operatorname{Tr}[D_{A'BB'}], \tag{10.6.13}$$

subject to the constraints

$$D_{A'BB'} \in \widehat{\mathrm{SEP}}(A'\!:\!BB'), \tag{10.6.14}$$

$$D_{A'BB'} \geq \mathcal{N}_{A\to B}(\rho_{A'AB'}). \tag{10.6.15}$$

Furthermore, (10.6.5) gives that

$$\Sigma_{\max}(\mathcal{N}) = \inf \left\| \operatorname{Tr}_B[Y_{SB}] \right\|_{\infty}, \tag{10.6.16}$$

subject to the constraints

$$Y_{SB} \in \widehat{\mathrm{SEP}}(S\!:\!B), \tag{10.6.17}$$

$$Y_{SB} \geq \Gamma^{\mathcal{N}}_{SB}. \tag{10.6.18}$$

With these optimizations in place, we can now establish the inequality in (10.6.9) by making a judicious choice for $D_{A'BB'}$. Let $C_{A'AB'}$ be an arbitrary operator to consider in the optimization for $G_{\max}(A'A;B')_{\rho}$ (i.e., satisfying (10.6.11)–(10.6.12)), and let Y_{SB} be an arbitrary operator to consider in the optimization for $\Sigma_{\max}(\mathcal{N})$ (i.e., satisfying (10.6.17)–(10.6.18)). Let $|\Gamma\rangle_{SA} = \sum_{i=0}^{d_A-1} |i,i\rangle_{SA}$. Pick

$$D_{A'BB'} = \langle\Gamma|_{SA} C_{A'AB'} \otimes Y_{SB} |\Gamma\rangle_{SA}. \tag{10.6.19}$$

We need to prove that $D_{A'BB'}$ is feasible for $G_{\max}(A';BB')_\omega$. To this end, consider that

$$\langle\Gamma|_{SA}C_{A'AB'}\otimes Y_{SB}|\Gamma\rangle_{SA} \geq \langle\Gamma|_{SA}\rho_{A'AB'}\otimes\Gamma^{\mathcal{N}}_{SB}|\Gamma\rangle_{SA} \tag{10.6.20}$$

$$= \mathcal{N}_{A\to B}(\rho_{A'AB'}), \tag{10.6.21}$$

which follows from (10.6.12), (10.6.18), and (4.2.6). Now, since $C_{A'AB'} \in \widehat{\text{SEP}}(A'A:B')$, it can be written as $\sum_{x\in\mathcal{X}} P^x_{A'A}\otimes Q^x_{B'}$ for some finite alphabet $\mathcal{X}$ and for sets $\{P^x_{A'A}\}_{x\in\mathcal{X}}$, $\{Q^x_{B'}\}_{x\in\mathcal{X}}$ of positive semi-definite operators. Similarly, since $Y_{SB}\in\widehat{\text{SEP}}(S:B)$, it can be written as $\sum_{y\in\mathcal{Y}} L^y_S\otimes M^y_B$, for some finite alphabet $\mathcal{Y}$ and sets $\{L^y_S\}_{y\in\mathcal{Y}}$, $\{M^y_B\}_{y\in\mathcal{Y}}$ of positive semi-definite operators. Then, using (2.2.42) and (2.2.43), we have that

$$\langle\Gamma|_{SA}C_{A'AB'}\otimes Y_{SB}|\Gamma\rangle_{SA}$$

$$= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}}\langle\Gamma|_{SA}P^x_{A'A}\otimes Q^x_{B'}\otimes L^y_S\otimes M^y_B|\Gamma\rangle_{SA} \tag{10.6.22}$$

$$= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}}\langle\Gamma|_{SA}P^x_{A'A}\text{T}_A(L^y_A)\otimes Q^x_{B'}\otimes\mathbb{1}_S\otimes M^y_B|\Gamma\rangle_{SA} \tag{10.6.23}$$

$$= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}}\text{Tr}_A[P^x_{A'A}\text{T}_A(L^y_A)]\otimes Q^x_{B'}\otimes M^y_B \tag{10.6.24}$$

$$\in \widehat{\text{SEP}}(A':BB'). \tag{10.6.25}$$

For the second equality, we used the transpose trick from (2.2.42), and for the third, we used (2.2.43). The last statement follows because

$$\text{Tr}_A[P^x_{A'A}\text{T}_A(L^y_A)] = \text{Tr}_A\left[\sqrt{\text{T}_A(L^y_A)}P^x_{A'A}\sqrt{\text{T}_A(L^y_A)}\right] \tag{10.6.26}$$

is positive semi-definite for each $x\in\mathcal{X}$ and $y\in\mathcal{Y}$. Thus, $D_{A'BB'}$ is feasible for $G_{\max}(A';BB')_\omega$. Finally, using (2.2.43) again, consider that

$$G_{\max}(A';BB')_\omega \leq \text{Tr}[D_{A'BB'}] \tag{10.6.27}$$

$$= \text{Tr}[\langle\Gamma|_{SA}C_{A'AB'}\otimes Y_{SB}|\Gamma\rangle_{SA}] \tag{10.6.28}$$

$$= \text{Tr}[C_{A'AB'}\text{T}_A(Y_{AB})] \tag{10.6.29}$$

$$= \text{Tr}[C_{A'AB'}\text{T}_A(\text{Tr}_B[Y_{AB}])] \tag{10.6.30}$$

For the second equality, we used the transpose trick from (2.2.42). Since $C_{A'AB'}$ and Y_{SB} are positive semi-definite (this follows from (10.6.12) and (10.6.18), respectively), using (2.2.97) we have that

$$\begin{aligned}\mathrm{Tr}[C_{A'AB'}\mathrm{T}_A(\mathrm{Tr}_B[Y_{AB}])] &= |\mathrm{Tr}[C_{A'AB'}\mathrm{T}_A(\mathrm{Tr}_B[Y_{AB}])]| && (10.6.31)\\ &\leq \|C_{A'AB'}\|_1 \, \|\mathrm{T}_A(\mathrm{Tr}_B[Y_{AB}])\|_\infty && (10.6.32)\\ &= \mathrm{Tr}[C_{A'AB'}] \, \|\mathrm{T}_A(\mathrm{Tr}_B[Y_{AB}])\|_\infty && (10.6.33)\\ &= \mathrm{Tr}[C_{A'AB'}] \, \|\mathrm{Tr}_B[Y_{AB}]\|_\infty , && (10.6.34)\end{aligned}$$

where for the last equality we used the fact that the spectrum of an operator is invariant under the action of a full transpose (note, in this case, that T_A is a full transpose because the operator $\mathrm{Tr}_B[Y_{AB}]$ acts only on A). Therefore,

$$G_{\max}(A';BB')_\omega \leq \mathrm{Tr}[C_{A'AB'}] \, \|\mathrm{Tr}_B[Y_{AB}]\|_\infty . \tag{10.6.35}$$

Since this inequality holds for all $C_{A'AB'}$ satisfying (10.6.11)–(10.6.12) and for all Y_{SB} satisfying (10.6.17)–(10.6.18), we conclude (10.6.9) after taking an infimum with respect to all such operators.

Having shown that

$$\begin{aligned}&E_{\max}(A';BB')_\omega \leq E_{\max}(\mathcal{N}) + E_{\max}(A'A;B')_\rho\\ &\qquad\implies \quad E_{\max}(A';BB')_\omega - E_{\max}(A'A;B')_\rho \leq E_{\max}(\mathcal{N}) \qquad (10.6.36)\end{aligned}$$

for every state $\rho_{A'AB'}$, it follows immediately from the definition of $E^{\mathcal{A}}_{\max}(\mathcal{N})$ that $E^{\mathcal{A}}_{\max}(\mathcal{N}) \leq E_{\max}(\mathcal{N})$. Combined with the result of Lemma 10.4, which is the reverse inequality, we obtain $E^{\mathcal{A}}_{\max}(\mathcal{N}) = E_{\max}(\mathcal{N})$. ■

With the equality $E^{\mathcal{A}}_{\max}(\mathcal{N}) = E_{\max}(\mathcal{N})$ in hand, the subadditivity of max-relative entropy of entanglement of quantum channels immediately follows.

Corollary 10.17 Subadditivity of Max-Relative Entropy of Entanglement of a Quantum Channel

The max-relative entropy of entanglement of a quantum channel is subadditive: for every two quantum channels $\mathcal{N}$ and $\mathcal{M}$,

$$E_{\max}(\mathcal{N} \otimes \mathcal{M}) \leq E_{\max}(\mathcal{N}) + E_{\max}(\mathcal{M}). \tag{10.6.37}$$

PROOF: Given that the amortized entanglement of a quantum channel is always subadditive, regardless of whether or not the underlying state entanglement measure is additive (see Proposition 10.5), we have that

$$E^{\mathcal{A}}_{\max}(\mathcal{N} \otimes \mathcal{M}) \leq E^{\mathcal{A}}_{\max}(\mathcal{N}) + E^{\mathcal{A}}_{\max}(\mathcal{M}) \tag{10.6.38}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. Using (10.6.8), we immediately obtain the desired result. ■

10.6.2 Max-Rains Information

Let us now prove that the amortization collapse also occurs for the max-Rains information of a quantum channel. The key tools needed for the proof are Propositions 9.27 and 10.13, which provide semi-definite programs for both the max-Rains relative entropy for bipartite states and the max-Rains information for quantum channels. Let us recall these now:

$$2^{R_{\max}(A;B)_\rho} = W_{\max}(A;B)_\rho \tag{10.6.39}$$

$$= \inf_{K_{AB},L_{AB}\geq 0} \{\mathrm{Tr}[K_{AB} + L_{AB}] : \mathrm{T}_B[K_{AB} - L_{AB}] \geq \rho_{AB}\}, \tag{10.6.40}$$

$$2^{R_{\max}(\mathcal{N})} = \Gamma_{\max}(\mathcal{N}) \tag{10.6.41}$$

$$= \inf_{Y_{SB},V_{SB}\geq 0} \{\|\mathrm{Tr}_B[V_{SB} + Y_{SB}]\|_\infty : \mathrm{T}_B[V_{SB} - Y_{SB}] \geq \Gamma^{\mathcal{N}}_{SB}\}, \tag{10.6.42}$$

where ρ_{AB} is a bipartite state and $\mathcal{N}$ is a quantum channel with Choi operator $\Gamma^{\mathcal{N}}_{SB}$.

Theorem 10.18 Amortization Collapse for Max-Rains Information

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For every state $\rho_{A'AB'}$,

$$R_{\max}(A';BB')_\omega \leq R_{\max}(\mathcal{N}) + R_{\max}(A'A;B')_\rho, \tag{10.6.43}$$

where $\omega_{A'BB'} := \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Consequently,

$$R^{\mathcal{A}}_{\max}(\mathcal{N}) \leq R_{\max}(\mathcal{N}), \tag{10.6.44}$$

and thus (by Lemma 10.4), we have that

$$R^{\mathcal{A}}_{\max}(\mathcal{N}) = R_{\max}(\mathcal{N}) \tag{10.6.45}$$

for every quantum channel $\mathcal{N}$.

Proof: The proof given below is conceptually similar to the proof of Theorem 10.16, but it has some key differences.

Using the semi-definite program formulations in (10.6.39)–(10.6.42), we find that the inequality in (10.6.43) is equivalent to

$$W_{\max}(A';BB')_\omega \leq \Gamma_{\max}(\mathcal{N}) \cdot W_{\max}(A'A;B')_\rho, \tag{10.6.46}$$

and so we aim to prove this one.

Using (10.6.40), we have that

$$W(A'A;B')_\rho = \inf \operatorname{Tr}[C_{A'AB'} + D_{A'AB'}], \tag{10.6.47}$$

subject to the constraints

$$C_{A'AB'},\ D_{A'AB'} \geq 0, \tag{10.6.48}$$
$$\mathrm{T}_{B'}(C_{A'AB'} - D_{A'AB'}) \geq \rho_{A'AB'}, \tag{10.6.49}$$

and we also have

$$W(A';BB')_\omega = \inf \operatorname{Tr}[G_{A'BB'} + F_{A'BB'}], \tag{10.6.50}$$

subject to the constraints

$$G_{A'BB'}, F_{A'BB'} \geq 0, \tag{10.6.51}$$
$$\mathcal{N}_{A\to B}(\rho_{A'AB'}) \leq \mathrm{T}_{BB'}[G_{A'BB'} - F_{A'BB'}]. \tag{10.6.52}$$

Using (10.6.42), we have that

$$\Gamma_{\max}(\mathcal{N}) = \inf \left\|\operatorname{Tr}_B[V_{SB} + Y_{SB}]\right\|_\infty, \tag{10.6.53}$$

subject to the constraints

$$Y_{SB}, V_{SB} \geq 0, \tag{10.6.54}$$
$$\mathrm{T}_B[V_{SB} - Y_{SB}] \geq \Gamma^{\mathcal{N}}_{SB}. \tag{10.6.55}$$

With these SDP formulations in place, we can now establish the inequality in (10.6.46) by making judicious choices for $G_{A'BB'}$ and $F_{A'BB'}$. Let $C_{A'AB'}$ and $D_{A'AB'}$ be arbitrary operators in the optimization for $W_{\max}(A'A;B')_\rho$, and let Y_{SB} and V_{SB}

be arbitrary operators in the optimization for $\Gamma_{\max}(\mathcal{N})$. Let $|\Gamma\rangle_{SA} = \sum_{i=0}^{d_A-1} |i,i\rangle_{SA}$. Pick

$$G_{A'BB'} = \langle\Gamma|_{SA} C_{A'AB'} \otimes V_{SB} + D_{A'AB'} \otimes Y_{SB}|\Gamma\rangle_{SA}, \tag{10.6.56}$$

$$F_{A'BB'} = \langle\Gamma|_{SA} C_{A'AB'} \otimes Y_{SB} + D_{A'AB'} \otimes V_{SB}|\Gamma\rangle_{SA}. \tag{10.6.57}$$

Note that $G_{A'BB'}, F_{A'BB'} \geq 0$ because $C_{A'AB'}$, $D_{A'AB'}$, Y_{SB}, $V_{SB} \geq 0$. Using (10.6.49) and (10.6.55), consider that

$$\begin{aligned} &\mathrm{T}_{BB'}[G_{A'BB'} - F_{A'BB'}] \\ &= \mathrm{T}_{BB'}\left[\langle\Gamma|_{SA}(C_{A'AB'} - D_{A'AB'}) \otimes (V_{SB} - Y_{SB})|\Gamma\rangle_{SA}\right] && (10.6.58) \\ &= \langle\Gamma|_{SA}\mathrm{T}_{B'}[C_{A'AB'} - D_{A'AB'}] \otimes \mathrm{T}_B[V_{SB} - Y_{SB}]|\Gamma\rangle_{SA} && (10.6.59) \\ &\geq \langle\Gamma|_{SA}\rho_{A'AB'} \otimes \Gamma^{\mathcal{N}}_{SB}|\Gamma\rangle_{SA} && (10.6.60) \\ &= \mathcal{N}_{A\to B}(\rho_{A'AB'}), && (10.6.61) \end{aligned}$$

where the last equality follows from (4.2.6). Our choices of $G_{A'BB'}$ and $F_{A'BB'}$ are thus feasible points for $W_{\max}(A';BB')_\omega$. Using this, along with (2.2.42) and (2.2.43), we obtain

$$\begin{aligned} &W_{\max}(A';BB')_\omega \\ &\leq \mathrm{Tr}[G_{A'BB'} + F_{A'BB'}] && (10.6.62) \\ &= \mathrm{Tr}[\langle\Gamma|_{SA}(C_{A'AB'} + D_{A'AB'}) \otimes (V_{SB} + Y_{SB})|\Gamma\rangle_{SA}] && (10.6.63) \\ &= \mathrm{Tr}[(C_{A'AB'} + D_{A'AB'})\mathrm{T}_A(V_{AB} + Y_{AB})] && (10.6.64) \\ &= \mathrm{Tr}[(C_{A'AB'} + D_{A'AB'})\mathrm{T}_A(\mathrm{Tr}_B[V_{AB} + Y_{AB}])]. && (10.6.65) \end{aligned}$$

The second equality follows from the transpose trick from (2.2.42). Now, since $C_{A'AB'}, D_{A'AB'} \geq 0$ (recall (10.6.48)), and $V_{AB}, Y_{AB} \geq 0$ (recall (10.6.54)), we can use (2.2.97) to conclude that

$$\begin{aligned} &\mathrm{Tr}[(C_{A'AB'} + D_{A'AB'})\mathrm{T}_A(\mathrm{Tr}_B[V_{AB} + Y_{AB}])] && (10.6.66) \\ &\quad = |\mathrm{Tr}[(C_{A'AB'} + D_{A'AB'})\mathrm{T}_A(\mathrm{Tr}_B[V_{AB} + Y_{AB}])]| && (10.6.67) \\ &\quad \leq \|C_{A'AB'} + D_{A'AB'}\|_1 \, \|\mathrm{T}_A(\mathrm{Tr}_B[V_{AB} + Y_{AB}])\|_\infty && (10.6.68) \\ &\quad = \mathrm{Tr}[C_{A'AB'} + D_{A'AB'}] \, \|\mathrm{T}_A(\mathrm{Tr}_B[V_{AB} + Y_{AB}])\|_\infty && (10.6.69) \\ &\quad = \mathrm{Tr}[C_{A'AB'} + D_{A'AB'}] \, \|\mathrm{Tr}_B[V_{AB} + Y_{AB}]\|_\infty, && (10.6.70) \end{aligned}$$

where the final equality follows because the spectrum of an operator is invariant under the action of a (full) transpose (note, in this case, that T_A is a full transpose because the operator $\mathrm{Tr}_B[V_{AB} + Y_{AB}]$ acts only on system A). We thus have

$$W_{\max}(A';BB')_\omega \leq \mathrm{Tr}[C_{A'AB'} + D_{A'AB'}] \, \|\mathrm{T}_A[\mathrm{Tr}_B[V_{AB} + Y_{AB}]]\|_\infty \tag{10.6.71}$$

Since this inequality holds for all $C_{A'AB'}$ and $D_{A'AB'}$ satisfying (10.6.49) and for all V_{AB} and Y_{AB} satisfying (10.6.55), we conclude the inequality in (10.6.46).

Having shown that

$$R_{\max}(A';BB')_{\omega} \leq R_{\max}(\mathcal{N}) + R_{\max}(A'A;B')_{\rho},$$
$$\Longrightarrow \quad R_{\max}(A';BB')_{\omega} - R_{\max}(A'A;B')_{\rho} \leq R_{\max}(\mathcal{N}) \quad (10.6.72)$$

for every state $\rho_{A'AB'}$, it immediately follows from the definition of amortized entanglement of a channel that

$$R^{\mathcal{A}}_{\max}(\mathcal{N}) \leq R_{\max}(\mathcal{N}). \quad (10.6.73)$$

Thus, by Lemma 10.4, which proves the reverse inequality, we obtain

$$R^{\mathcal{A}}_{\max}(\mathcal{N}) = R_{\max}(\mathcal{N}) \quad (10.6.74)$$

for every quantum channel $\mathcal{N}$. ■

With the equality $R^{\mathcal{A}}_{\max}(\mathcal{N}) = R_{\max}(\mathcal{N})$ in hand, along with additivity of max-Rains relative entropy for bipartite states, additivity of max-Rains information of a quantum channel immediately follows.

Corollary 10.19 Additivity of Max-Rains Information of a Channel

The max-Rains information of a quantum channel is additive: for every two quantum channels $\mathcal{N}$ and $\mathcal{M}$,

$$R_{\max}(\mathcal{N} \otimes \mathcal{M}) = R_{\max}(\mathcal{N}) + R_{\max}(\mathcal{M}). \quad (10.6.75)$$

Proof: The additivity of max-Rains relative entropy for bipartite states (see Proposition 9.29), along with Proposition 10.5, implies that the amortized max-Rains information of a quantum channel is additive, meaning that

$$R^{\mathcal{A}}_{\max}(\mathcal{N} \otimes \mathcal{M}) = R^{\mathcal{A}}_{\max}(\mathcal{N}) + R^{\mathcal{A}}_{\max}(\mathcal{M}) \quad (10.6.76)$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. Then, from (10.6.45), we obtain the desired result. ■

10.6.3 Squashed Entanglement

Finally, let us prove that the amortization collapse occurs for the squashed entanglement of a quantum channel.

Theorem 10.20 Amortization Collapse for Squashed Entanglement

The squashed entanglement of a channel does not increase under amortization, i.e.,

$$E_{\text{sq}}(\mathcal{N}) = E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) \tag{10.6.77}$$

for every quantum channel $\mathcal{N}$.

Proof: The inequality $E_{\text{sq}}(\mathcal{N}) \leq E_{\text{sq}}^{\mathcal{A}}(\mathcal{N})$ has already been shown in Lemma 10.4. We thus prove the reverse inequality.

Let $\rho_{A'AB'}$ be an arbitrary state, and let $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending $\mathcal{N}_{A\to B}$, and let $\varphi_{A'AB'R}$ be a purification of $\rho_{A'AB'}$. Then, the state $\theta_{A'BB'ER} \coloneqq \mathcal{U}^{\mathcal{N}}_{A\to BE}(\varphi_{A'AB'R})$ is a purification of $\omega_{A'BB'}$. As we show in Lemma 10.22 below, the following inequality holds

$$E_{\text{sq}}(A';BB')_\omega = E_{\text{sq}}(A';BB')_\theta \leq E_{\text{sq}}(A'B'R;B)_\theta + E_{\text{sq}}(A'BE;B')_\theta. \tag{10.6.78}$$

Now, squashed entanglement is invariant under the action of an isometric channel, in particular $\mathcal{U}^{\mathcal{N}}_{A\to BE}$, so that $E_{\text{sq}}(A'BE;B')_\theta = E_{\text{sq}}(A'A;B')_\rho$. Also, the state $\theta_{A'BB'ER}$ can be viewed as a particular state in the optimization over states that defines $E_{\text{sq}}(\mathcal{N})$ because $\theta_{A'BB'ER}$ is a purification of $\omega_{A'BB'}$. Therefore,

$$E_{\text{sq}}(A'B'R;B)_\theta \leq E_{\text{sq}}(\mathcal{N}). \tag{10.6.79}$$

Altogether, we thus have

$$\begin{aligned} &E_{\text{sq}}(A';BB')_\omega \leq E_{\text{sq}}(\mathcal{N}) + E_{\text{sq}}(A'A;B')_\rho \\ &\qquad\Longrightarrow \quad E_{\text{sq}}(A';BB')_\omega - E_{\text{sq}}(A'A;B')_\rho \leq E_{\text{sq}}(\mathcal{N}) \end{aligned} \tag{10.6.80}$$

for every state $\rho_{A'AB'}$, which means by definition of amortized entanglement that

$$E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N}), \tag{10.6.81}$$

which is what we sought to prove. ■

With the equality $E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) = E_{\text{sq}}(\mathcal{N})$ in hand, along with additivity of squashed entanglement for bipartite states (Property 4. of Proposition 9.32), additivity of squashed entanglement for channels immediately follows.

Corollary 10.21 Additivity of Squashed Entanglement of a Channel

The squashed entanglement of a quantum channel is additive: for every two quantum channels $\mathcal{N}$ and $\mathcal{M}$,

$$E_{\text{sq}}(\mathcal{N} \otimes \mathcal{M}) = E_{\text{sq}}(\mathcal{N}) + E_{\text{sq}}(\mathcal{M}). \tag{10.6.82}$$

PROOF: The additivity of the squashed entanglement for bipartite states (see Proposition 9.32), along with Proposition 10.5, implies that the amortized squashed entanglement of a quantum channel is additive, meaning that

$$E_{\text{sq}}^{\mathcal{A}}(\mathcal{N} \otimes \mathcal{M}) = E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) + E_{\text{sq}}^{\mathcal{A}}(\mathcal{M}) \tag{10.6.83}$$

for all quantum channels $\mathcal{N}$ and $\mathcal{M}$. Then, from (10.6.77), we obtain the desired result. ■

Lemma 10.22

Let $\psi_{KL_1L_2M_1M_2}$ be a pure state. Then

$$E_{\text{sq}}(K; L_1L_2)_\psi \leq E_{\text{sq}}(KL_2M_2; L_1)_\psi + E_{\text{sq}}(KL_1M_1; L_2)_\psi. \tag{10.6.84}$$

PROOF: We make use of the squashing channel representation of squashed entanglement in (9.4.43), namely,

$$E_{\text{sq}}(A; B)_\rho = \frac{1}{2} \inf_{\mathcal{S}_{E' \to E}} \{I(A; B|E)_\omega : \omega_{ABE} = \mathcal{S}_{E' \to E}(\psi^\rho_{ABE'})\}, \tag{10.6.85}$$

where $\psi^\rho_{ABE'}$ is a purification of ρ and the infimum is with respect to every quantum channel $\mathcal{S}_{E' \to E}$. Let us also recall that

$$I(A; B|E)_\omega = H(A|E)_\omega + H(B|E)_\omega - H(AB|E)_\omega \tag{10.6.86}$$

$$= H(B|E)_\omega - H(B|AE)_\omega, \tag{10.6.87}$$

and that strong subadditivity (Theorem 7.6) is the statement that $I(A; B|E)_\omega \geq 0$. From this we obtain the following two inequalitites:

$$H(AB|E)_\omega \leq H(A|E)_\omega + H(B|E)_\omega, \tag{10.6.88}$$

$$H(B|AE)_\omega \leq H(B|E)_\omega. \tag{10.6.89}$$

Now, the given pure state $\psi_{KL_1L_2M_1M_2}$ can be thought of as a purification of the reduced state $\psi_{KL_1L_2}$ for which the squashed entanglement $E_{\text{sq}}(K; L_1L_2)_\psi$ is evaluated, with the purifying systems being M_1 and M_2. Then, considering an arbitrary product squashing channel $\mathcal{S}^1_{M_1\to M_1'} \otimes \mathcal{S}^2_{M_2\to M_2'}$, and letting

$$\omega_{KL_1L_2M_1'M_2'} = (\mathcal{S}^1_{M_1\to M_1'} \otimes \mathcal{S}^2_{M_2\to M_2'})(\psi_{KL_1L_2M_1M_2}), \tag{10.6.90}$$

we find from (10.6.85) that

$$2E_{\text{sq}}(K; L_1L_2)_\psi \leq I(K; L_1L_2|M_1'M_2')_\omega. \tag{10.6.91}$$

Expanding the quantum conditional mutual information using (10.6.87), we have that

$$I(K; L_1L_2|M_1'M_2')_\omega = H(L_1L_2|M_1'M_2')_\omega - H(L_1L_2|M_1'M_2'K)_\omega. \tag{10.6.92}$$

Now, let $\phi_{KL_1L_2M_1'M_2'R}$ be a purification of $\omega_{KL_1L_2M_1'M_2'}$ with purifying system R. Then, by definition of conditional entropy, and using the fact that $\phi_{KL_1L_2M_1'M_2'R}$ is pure, we obtain[1]

$$H(L_1L_2|M_1'M_2'K)_\omega = H(L_1L_2M_1'M_2'K)_\omega - H(M_1'M_2'K)_\omega \tag{10.6.93}$$
$$= H(R)_\phi - H(L_1L_2R)_\phi \tag{10.6.94}$$
$$= -H(L_1L_2|R)_\phi. \tag{10.6.95}$$

Therefore,

$$2E_{\text{sq}}(K; L_1L_2)_\psi \leq H(L_1L_2|M_1'M_2')_\phi + H(L_1L_2|R)_\phi. \tag{10.6.96}$$

Using the inequality in (10.6.88) followed by two applications of (10.6.89) (with appropriate identification of subsystems in all three cases), we obtain

$$H(L_1L_2|M_1'M_2')_\phi \leq H(L_1|M_1'M_2')_\phi + H(L_2|M_1'M_2')_\phi \tag{10.6.97}$$
$$\leq H(L_1|M_1')_\phi + H(L_2|M_2')_\phi. \tag{10.6.98}$$

[1]The steps in (10.6.93)–(10.6.95) establish a general fact called *duality of conditional entropy*: for every pure state ψ_{ABE}, the following equality holds $H(A|E)_\psi + H(B|E)_\psi = 0$.

Next, using (10.6.88), we conclude that

$$H(L_1L_2|R)_\phi \leq H(L_1|R)_\phi + H(L_2|R)_\phi. \tag{10.6.99}$$

Therefore, continuing from (10.6.96), we have

$$2E_{\text{sq}}(K;L_1L_2)_\psi \leq H(L_1|M_1')_\phi + H(L_2|M_2')_\phi + H(L_1|R)_\phi + H(L_2|R)_\phi. \tag{10.6.100}$$

Now, applying reasoning analogous to that in (10.6.93)–(10.6.95) for the last two terms, we find that

$$\begin{aligned} 2E_{\text{sq}}(K;L_1L_2)_\psi &\leq H(L_1|M_1')_\omega - H(L_2|KL_2M_1'M_2')_\omega \\ &\qquad + H(L_2|M_2')_\omega - H(L_2|KL_1M_1'M_2')_\omega &(10.6.101)\\ &= I(KL_2M_2';L_1|M_1')_\omega + I(KL_1M_1';L_2|M_2')_\omega, &(10.6.102)\end{aligned}$$

where in the last line we used the expression in (10.6.87) for the conditional mutual information.

Now, let us regard $\psi_{KL_1L_2M_1M_2}$ as a purification of the reduced state $\psi_{KL_2M_2L_1}$, for which the squashed entanglement $E_{\text{sq}}(KL_2M_2;L_1)_\psi$ is evaluated, with the purifying system being M_1. Then, the state

$$\tau_{KL_2M_2L_1M_1'} := \mathcal{S}^1_{M_1\to M_1'}(\psi_{KL_2M_1L_1M_1}) \tag{10.6.103}$$

is a particular extension of $\psi_{KL_2M_2L_1}$ in the optimization for $E_{\text{sq}}(KL_2M_2;L_1)_\psi$. Similarly, we can regard $\psi_{KL_1L_2M_1M_2}$ as a purification of the reduced state $\psi_{KL_1M_1L_2}$, for which the squashed entanglement $E_{\text{sq}}(KL_1M_1;L_2)_\psi$ is evaluated, with the purifying system being M_2. Then, the state

$$\sigma_{KL_1M_1L_2M_2'} := \mathcal{S}^2_{M_2\to M_2'}(\psi_{KL_1M_1L_2M_2}) \tag{10.6.104}$$

is a particular extension of $\psi_{KL_1M_1L_2}$ in the optimization for $E_{\text{sq}}(KL_1M_1;L_2)_\psi$. Using all of this, we proceed from (10.6.102) to obtain

$$\begin{aligned} 2E_{\text{sq}}(K;L_1L_2)_\psi &\leq I(KL_2M_2';L_1|M_1')_\omega + I(KL_1M_1';L_2|M_2')_\omega &(10.6.105)\\ &\leq I(KL_2M_2;L_1|M_1')_\tau + I(KL_1M_1;L_2|M_2')_\sigma, &(10.6.106)\end{aligned}$$

where for the second inequality we used the data-processing inequality for conditional mutual information (Proposition 7.9). Since the squashing channels $\mathcal{S}^1_{M_1\to M_1'}$ and $\mathcal{S}^2_{M_2\to M_2'}$ are arbitrary, optimizing over all such channels on the right-hand side of the inequality above leads us to

$$E_{\text{sq}}(K;L_1L_2)_\psi \leq E_{\text{sq}}(KL_2M_2;L_1)_\psi + E_{\text{sq}}(KL_1M_1;L_2)_\psi, \tag{10.6.107}$$

as required. ■

10.7 Summary

... We considered two types of channel entanglement measures. The first type quantifies the entanglement of a bipartite state after one share of it is sent through the given quantum channel, in a manner analogous to the channel information measures defined in Chapter 7. The second type of channel entanglement measure is called *amortized entanglement*, which essentially quantifies the difference in entanglement between a bipartite state and the state obtained after sending one share of it through the given channel. The concept of amortized entanglement turns out to play an important role in feedback-assisted communciation scenarios (as considered in Part III), as it can be used to prove important properties of entanglement measures of the first kind....

10.8 Bibliographic Notes

...The entanglement of a quantum channel was presented by Takeoka et al. (2014); Tomamichel et al. (2017), by employing the squashed entanglement and the Rains relative entropy entanglement measures, respectively. The amortized entanglement of a quantum channel has its roots in early work by Bennett et al. (2003); Leifer et al. (2003), and it was formally defined and its various properties established by Kaur and Wilde (2017). The work by Rigovacca et al. (2018) is related to the notion of amortized entanglement. The connection of amortized entanglement to teleportation simulation of quantum channels was elucidated by Kaur and Wilde (2017). A channel's relative entropy of entanglement was defined by Pirandola et al. (2017), max-relative entropy of entanglement by Christandl and Müller-Hermes (2017), and the hypothesis testing and sandwiched Rényi relative entropy of entanglement by Wilde et al. (2017). Proposition 10.9 is based on (Tomamichel et al., 2017, Proposition 2) and (Wilde et al., 2017, Proposition 14). The generalized Rains information of a quantum channel was defined by Tomamichel et al. (2017), who explicitly considered the Rains information and the sandwiched Rényi Rains information as special cases. Tomamichel et al. (2016) focused on the hypothesis testing Rains information of a quantum channel. Wang et al. (2019b) observed that the quantity defined by Wang and Duan (2016b) is equal to the max-Rains information of a quantum channel. The semi-definite program formulation of the max-Rains information of a quantum channel in Proposition 10.13 is due to Wang and Duan (2016b); Wang et al. (2019b). Proposition 10.12 is due to Tomamichel et al. (2017). Concavity of a channel's unoptimized squashed entanglement (Proposition 10.15) is due to Takeoka et al. (2014). The amortization collapse of max-relative entropy of entanglement and of max-Rains information (Theorems 10.16 and 10.18, respectively) were shown by Berta and Wilde (2018), with the amortization collapse of max-relative entropy implicitly considered by Christandl and Müller-Hermes (2017). Lemma 10.22 is due to Takeoka et al. (2014), and the explicit observation that Lemma 10.22 implies that amortization does not increase the squashed entanglement of a quantum channel (Theorem 10.20) was realized by Kaur and Wilde (2017). Corollary 10.21 was established by Takeoka et al. (2014). ...

10.9 Problems

Appendix 10.A The $\alpha \to 1$ and $\alpha \to \infty$ Limits of the Sandwiched Rényi Entanglement Measures

In this section, we prove the $\alpha \to 1$ and $\alpha \to \infty$ limits of the sandwiched Rényi state and channel entanglement measures that we have considered in this chapter. Specifically, we consider the limits of

$$\widetilde{E}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}), \tag{10.A.1}$$

$$\widetilde{E}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.A.2}$$

$$\widetilde{R}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in\text{PPT}'(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}), \tag{10.A.3}$$

$$\widetilde{R}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{PPT}'(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}), \tag{10.A.4}$$

where $\alpha \in [1/2, 1) \cup (1, \infty)$.

We make consistent use throughout this section of the fact that the sandwiched Rényi relative entropy $\widetilde{D}_\alpha$ is monotonically increasing in α for all $\alpha \in (0,1)\cup(1,\infty)$ (see Proposition 7.31), as well as the fact that $\lim_{\alpha\to 1}\widetilde{D}_\alpha = D$, where D is the quantum relative entropy (see Proposition 7.30). We also use the fact that $\lim_{\alpha\to\infty}\widetilde{D}_\alpha = D_{\max}$ (see Proposition 7.61).

10.A.0.1 $\alpha \to 1$ Limits

We start by showing that

$$\lim_{\alpha\to 1}\widetilde{E}_\alpha(A;B)_\rho = E_R(A;B) = \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}). \tag{10.A.5}$$

The proof of

$$\lim_{\alpha\to 1}\widetilde{R}_\alpha(A;B)_\rho = R(A;B)_\rho = \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D(\rho_{AB}\|\sigma_{AB}) \tag{10.A.6}$$

is analogous, and so we omit it.

When approaching one from above, due to monotonicity in α of $\widetilde{D}_\alpha$ (Proposition 7.31), we have that

$$\lim_{\alpha\to 1^+}\widetilde{D}_\alpha = \inf_{\alpha\in(1,\infty)}\widetilde{D}_\alpha = D. \tag{10.A.7}$$

Therefore, we readily obtain

$$\lim_{\alpha\to 1^+}\widetilde{E}_\alpha(A;B)_\rho = \inf_{\alpha\in(1,\infty)}\inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)}\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.8}$$

$$= \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)}\inf_{\alpha\in(1,\infty)}\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.9}$$

$$= \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}) \tag{10.A.10}$$

$$= E_R(A;B)_\rho. \tag{10.A.11}$$

Now, when approaching one from below, we have

$$\lim_{\alpha\to 1^-}\widetilde{D}_\alpha = \sup_{\alpha\in(0,1)}\widetilde{D}_\alpha = D. \tag{10.A.12}$$

Therefore,

$$\lim_{\alpha\to 1^-}\widetilde{E}_\alpha(A;B)_\rho = \sup_{\alpha\in(1/2,1)}\inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)}\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}). \tag{10.A.13}$$

Now, we apply Theorem 2.25. Specifically, we can apply the theorem in order to exchange the order of the infimum and supremum because the function

$$(\sigma_{AB},\alpha)\mapsto\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.14}$$

is continuous in the first argument and montonically increasing in the second argument (also, the set of separable states is compact). We thus obtain

$$\lim_{\alpha\to1^-}\widetilde{E}_\alpha(A;B)_\rho = \sup_{\alpha\in(1/2,1)}\inf_{\sigma_{AB}\in\text{SEP}(A:B)}\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.15}$$

$$= \inf_{\sigma_{AB}\in\text{SEP}(A:B)}\sup_{\alpha\in(1/2,1)}\widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.16}$$

$$= \inf_{\sigma_{AB}\in\text{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}) \tag{10.A.17}$$

$$= E_R(A;B)_\rho. \tag{10.A.18}$$

This concludes the proof of (10.A.5).

Now, for the channel measure, we show that

$$\lim_{\alpha\to1}\widetilde{E}_\alpha(\mathcal{N}) = E_R(\mathcal{N}) = \sup_{\psi_{RA}}\inf_{\sigma_{RB}\in\text{SEP}(R:B)} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}). \tag{10.A.19}$$

The proof of

$$\lim_{\alpha\to1}\widetilde{R}_\alpha(\mathcal{N}) = R(\mathcal{N}) = \sup_{\psi_{RA}}\inf_{\sigma_{RB}\in\text{PPT}'(R:B)} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.20}$$

is analogous, and so we omit it.

When approaching one from below, we use exactly the same arguments as above to exchange the infimum and supremum, in order to conclude that

$$\lim_{\alpha\to1^-}\widetilde{E}_\alpha(\mathcal{N}) = \sup_{\alpha\in(1/2,1)}\sup_{\psi_{RA}}\inf_{\sigma_{RB}\in\text{SEP}(R:B)}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.21}$$

$$= \sup_{\psi_{RA}}\sup_{\alpha\in(1/2,1)}\inf_{\sigma_{RB}\in\text{SEP}(R:B)}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.22}$$

$$= \sup_{\psi_{RA}}\inf_{\sigma_{RB}\in\text{SEP}(R:B)}\sup_{\alpha\in(1/2,1)}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.23}$$

$$= \sup_{\psi_{RA}}\inf_{\sigma_{RB}\in\text{SEP}(R:B)} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.24}$$

$$= E_R(\mathcal{N}). \tag{10.A.25}$$

Next, when approaching from above, we again use Theorem 2.25. This time, since the function

$$(\psi_{RA},\alpha)\mapsto\inf_{\sigma_{RB}\in\text{SEP}(R:B)}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.26}$$

is continuous in the first argument and monotonically increasing in the second argument, we can exchange the order of the infimum and supremum to obtain

$$\lim_{\alpha\to 1^+} \widetilde{E}_\alpha(\mathcal{N}) = \inf_{\alpha\in(1,\infty)} \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.27}$$

$$= \sup_{\psi_{RA}} \inf_{\alpha\in(1,\infty)} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.28}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} \inf_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.29}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\text{SEP}(R:B)} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.30}$$

$$= E_R(\mathcal{N}). \tag{10.A.31}$$

This concludes the proof of (10.A.19).

10.A.0.2 $\alpha \to \infty$ Limits

We now move on to the $\alpha \to \infty$ limits. We first show that

$$\lim_{\alpha\to\infty} \widetilde{E}_\alpha(A;B)_\rho = E_{\max}(A;B) = \inf_{\sigma_{AB}\in\text{SEP}(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}). \tag{10.A.32}$$

The proof of

$$\lim_{\alpha\to\infty} \widetilde{R}_\alpha(A;B)_\rho = R_{\max}(A;B)_\rho = \inf_{\sigma_{AB}\in\text{PPT}'(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}) \tag{10.A.33}$$

is analogous, and so we omit it.

Note that due to monotonicity in α of $\widetilde{D}_\alpha$, the following equality holds

$$\lim_{\alpha\to\infty} \widetilde{D}_\alpha = \sup_{\alpha\in(1,\infty)} \widetilde{D}_\alpha = D_{\max}. \tag{10.A.34}$$

Therefore,

$$\lim_{\alpha\to\infty} \widetilde{E}_\alpha(A;B)_\rho = \sup_{\alpha\in(1,\infty)} \inf_{\sigma_{AB}\in\text{SEP}(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}). \tag{10.A.35}$$

Now, since the function

$$(\sigma_{AB},\alpha) \mapsto \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.36}$$

is continuous in the first argument, monotonically increasing in the second argument, and because the set of separable states is compact, we can use Theorem 2.25 to change the order of the supremum and infimum to obtain

$$\lim_{\alpha\to\infty} \widetilde{E}_\alpha(A;B)_\rho = \sup_{\alpha\in(1,\infty)} \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.37}$$

$$= \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} \sup_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{10.A.38}$$

$$= \inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D_{\max}(\rho_{AB}\|\sigma_{AB}) \tag{10.A.39}$$

$$= E_{\max}(A;B)_\rho. \tag{10.A.40}$$

This completes the proof of (10.A.32).

Finally, we prove that

$$\lim_{\alpha\to\infty} \widetilde{E}_\alpha(\mathcal{N}) = E_{\max}(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} D_{\max}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}). \tag{10.A.41}$$

The proof of

$$\lim_{\alpha\to\infty} \widetilde{R}_\alpha(\mathcal{N}) = R_{\max}(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{PPT}'(R:B)} D_{\max}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.42}$$

is analogous, and so we omit it.

Using exactly the same argument as in the proof of (10.A.32) above, we obtain

$$\lim_{\alpha\to\infty} \widetilde{E}_\alpha(\mathcal{N}) = \sup_{\alpha\in(1,\infty)} \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.43}$$

$$= \sup_{\psi_{RA}} \sup_{\alpha\in(1,\infty)} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.44}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} \sup_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.45}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_{RB}\in\mathrm{SEP}(R:B)} D_{\max}(\mathcal{N}_{A\to B}(\psi_{RA})\|\sigma_{RB}) \tag{10.A.46}$$

$$= E_{\max}(\mathcal{N}), \tag{10.A.47}$$

where in the third line we used Theorem 2.25 in order to exchange the infimum and supremum based on exactly the same arguments used in the proof of (10.A.32) above. This concludes the proof of (10.A.41).

Part II

Quantum Communication Protocols

We now begin our study of quantum communication protocols. In this part of the book, we focus on point-to-point communication protocols, most of which do not make use of feedback between the sender and receiver. The settings include classical communication, entanglement-assisted classical communication, entanglement distillation, quantum communication, secret key distillation, and private communication. These point-to-point protocols are the most basic communication models in quantum information, and in many cases, they are relevant from a practical perspective. Furthermore, these protocols serve as benchmarks for the usefulness of the feedback-assisted protocols that we consider in Part III, in the sense that the optimal communication rates of any feedback-assisted protocol should not be smaller than the corresponding point-to-point protocol, in order for the feedback-assisted protocol to be deemed useful or advantageous.

Chapter 11

Entanglement-Assisted Classical Communication

The first communication task that we consider is entanglement-assisted classical communication. In this scenario, Alice and Bob are allowed to share an unlimited amount of entanglement prior to communication, and the goal is for Alice to transmit the maximum possible amount of classical information over a given channel $\mathcal{N}$, by using this prior shared entanglement as a resource. We consider this particular setting before all other communication settings because, perhaps unexpectedly, the main information-theoretic results in this setting are much simpler than those in all other communication settings that we consider in this book.

Entanglement is a uniquely quantum phenomenon, and it is natural to ask, when communicating over quantum channels, whether it can be used to provide an advantage for sending classical information. The super-dense coding protocol, described in Section 5.2, is an example of such an advantage. Recall that in this protocol, Alice and Bob share a pair of quantum systems in the maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$, and they are connected by a noiseless qubit channel. With this shared entanglement, along with only one use of the channel, Alice can communicate *two* bits of classical information to Bob. In the case of qudits, using the maximally entangled state $\frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}|i,i\rangle$, Alice can communication $2\log_2 d$ bits to Bob with only one use of a noiseless qudit quantum channel. Does this kind of advantage exist in general? Specifically, supposing that we allow Alice and Bob unlimited shared entanglement, what is the maximum amount of classical information that can be communicated over a given quantum channel $\mathcal{N}$?

The answer to this question is provided by Theorem 11.16, which tells us that the entanglement-assisted classical capacity of a channel $\mathcal{N}$ is equal to the mutual information $I(\mathcal{N})$ of the channel (see (7.11.102)). The strength of this result is that it holds for *all* channels. Entanglement-assisted classical communication is one of the few scenarios in which such a profoundly simple statement—applying to all channels—can be made. Furthermore, the fact that the mutual information $I(\mathcal{N})$ is the optimal rate for entanglement-assisted classical communication for all channels $\mathcal{N}$ makes this communication scenario formally analogous to communication over classical channels. Indeed, the famous result of Shannon from 1948 is that the capacity of a classical channel described by a conditional probability distribution $p_{Y|X}(y|x)$ with input and output random variables X and Y, respectively, is equal to $\max_{p_X} I(X;Y)$, where $I(X;Y)$ is the mutual information between the random variables X and Y and the optimization is performed over all probability distributions p_X corresponding to the input X. Entanglement-assisted classical communication can thus be viewed as a "natural" analogue of classical communication in the quantum setting.

11.1 One-Shot Setting

We begin by considering the one-shot setting for entanglement-assisted classical communication over $\mathcal{N}$, with such a protocol depicted in Figure 11.1. We call this the "one-shot setting" because the channel $\mathcal{N}$ is used only once. This is in contrast to the "asymptotic setting" that we consider in the next section, in which the channel may be used an arbitrarily large number of times.

The protocol depicted in Figure 11.1 is defined by the four elements $(\mathcal{M}, \Psi_{A'B'}, \mathcal{E}_{M'A'\to A}, \mathcal{D}_{BB'\to\widehat{M}})$, in which $\mathcal{M}$ is a message set, $\Psi_{A'B'}$ is an entangled state shared by Alice and Bob, $\mathcal{E}_{M'A'\to A}$ is an encoding channel, and $\mathcal{D}_{BB'\to\widehat{M}}$ is a decoding channel. The triple $(\Psi, \mathcal{E}, \mathcal{D})$, consisting of the resource state and encoding and decoding channels, is called an entanglement-assisted code or, more simply, a code. In what follows, we employ the abbreviation

$$\mathcal{P} \equiv (\Psi, \mathcal{E}, \mathcal{D}), \tag{11.1.1}$$

where the notation $\mathcal{P}$ indicates *protocol*.

Given that there are $|\mathcal{M}|$ messages in the message set, it holds that each message can be uniquely associated with a bit string of size at least $\log_2 |\mathcal{M}|$. The quantity

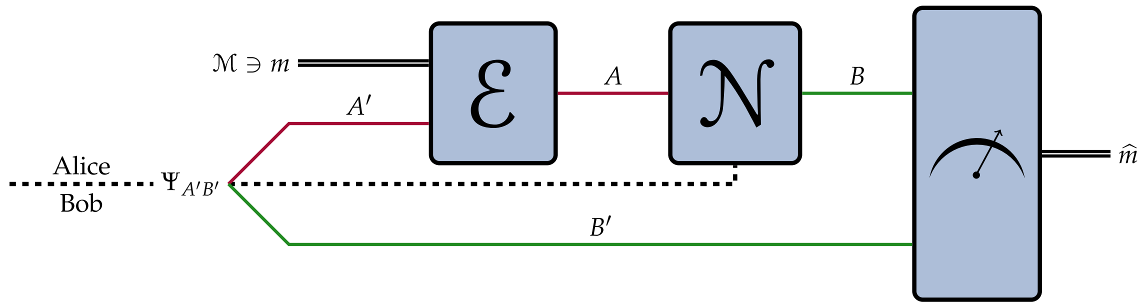


FIGURE 11.1: Depiction of a protocol for entanglement-assisted classical communication over one use of the quantum channel $\mathcal{N}$. Alice and Bob initially share a pair of quantum systems in the state $\Psi_{A'B'}$. Alice, who wishes to send a message m selected from a set $\mathcal{M}$ of messages, first encodes the message into a quantum state on a quantum system A by using an encoding channel $\mathcal{E}$. She then sends the quantum system A through the channel $\mathcal{N}$. After Bob receives the system B, he performs a measurement on both of his systems BB', using the outcome of the measurement to give an estimate $\widehat{m}$ of the message sent by Alice.

$\log_2 |\mathcal{M}|$ thus represents the number of bits that are communicated in the protocol. One of the goals of this section is to obtain upper and lower bounds, in terms of information measures for channels, on the maximum number $\log_2 |\mathcal{M}|$ of bits that can be communicated in an entanglement-assisted classical communication protocol.

The protocol proceeds as follows: let $p : \mathcal{M} \to [0, 1]$ be a probability distribution over the message set. Alice starts by preparing two systems M and M' in the following classically correlated state:

$$\overline{\Phi}^p_{MM'} := \sum_{m \in \mathcal{M}} p(m)|m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'}. \tag{11.1.2}$$

Note that if Alice wishes to send a particular message m deterministically, then she can choose the distribution p to be the degenerate distribution, equal to one for m and zero for all other messages. Alice and Bob share the state $\Psi_{A'B'}$ before communication begins, so that the global state shared between them is

$$\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'}. \tag{11.1.3}$$

Alice then sends the M' and A' registers through the encoding channel $\mathcal{E}_{M'A' \to A}$. Due to the fact that the system M' is classical, this encoding channel realizes a set $\{\mathcal{E}^m_{A' \to A}\}_{m \in \mathcal{M}}$ of quantum channels as follows:

$$\mathcal{E}^m_{A' \to A}(\tau_{A'}) := \mathcal{E}_{M'A' \to A}(|m\rangle\langle m|_{M'} \otimes \tau_{A'}) \tag{11.1.4}$$

for every state $\tau_{A'}$. The global state after the encoding channel is therefore

$$\mathcal{E}_{M'A'\to A}(\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'}) = \sum_{m\in\mathcal{M}} p(m)|m\rangle\!\langle m|_M \otimes \mathcal{E}^m_{A'\to A}(\Psi_{A'B'}). \tag{11.1.5}$$

Alice then transmits the A system through the channel $\mathcal{N}_{A\to B}$, leading to the state

$$\begin{aligned}(\mathcal{N}_{A\to B} \circ \mathcal{E}_{M'A'\to A})(\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'})\\ = \sum_{m\in\mathcal{M}} p(m)|m\rangle\!\langle m|_M \otimes (\mathcal{N}_{A\to B} \circ \mathcal{E}^m_{A'\to A})(\Psi_{A'B'}).\\ = \sum_{m\in\mathcal{M}} p(m)|m\rangle\!\langle m|_M \otimes \tau^m_{BB'},\end{aligned} \tag{11.1.6}$$

where

$$\tau^m_{BB'} \coloneqq (\mathcal{N}_{A\to B} \circ \mathcal{E}^m_{A'\to A})(\Psi_{A'B'}) \quad \forall\, m \in \mathcal{M}. \tag{11.1.7}$$

Bob, whose task is to determine which message Alice sent, applies a decoding channel $\mathcal{D}_{BB'\to\widehat{M}}$ on his system B' and the system B received through the channel $\mathcal{N}$. The decoding channel is a quantum-classical channel (Definition 4.10) associated with a POVM $\{\Lambda^m_{BB'}\}_{m\in\mathcal{M}}$, so that

$$\mathcal{D}_{BB'\to\widehat{M}}(\tau^m_{BB'}) \coloneqq \sum_{\widehat{m}\in\mathcal{M}} \mathrm{Tr}[\Lambda^{\widehat{m}}_{BB'}\tau^m_{BB'}]|\widehat{m}\rangle\!\langle\widehat{m}|_{\widehat{M}}, \tag{11.1.8}$$

for all $m \in \mathcal{M}$. The global state in (11.1.6) thus becomes

$$\omega^p_{M\widehat{M}} \coloneqq (\mathcal{D}_{BB'\to\widehat{M}} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{M'A'\to A})(\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'}) \tag{11.1.9}$$

$$= \sum_{m,\widehat{m}\in\mathcal{M}} p(m)|m\rangle\!\langle m|_M \otimes \mathrm{Tr}[\Lambda^{\widehat{m}}_{BB'}\mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))]|\widehat{m}\rangle\!\langle\widehat{m}|_{\widehat{M}}. \tag{11.1.10}$$

The final decoding measurement by Bob induces the conditional probability distribution $q : \mathcal{M}\times\mathcal{M} \to [0,1]$ defined by

$$q(\widehat{m}|m) \coloneqq \Pr[\widehat{M} = \widehat{m}|M = m] \tag{11.1.11}$$

$$= \mathrm{Tr}[\Lambda^{\widehat{m}}_{BB'}\mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))]. \tag{11.1.12}$$

Bob's strategy is such that if the outcome $\widehat{m}$ occurs from his measurement, then he declares that the message sent was $\widehat{m}$.

The probability that Bob correctly identifies a given message m is then given by $q(m|m)$. The *message error probability of the code* $\mathcal{P} \equiv (\Psi, \mathcal{E}, \mathcal{D})$ *and message* m is then given by

$$\begin{aligned} p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) &\coloneqq 1 - q(m|m) \\ &= \operatorname{Tr}[(\mathbb{1}_{BB'} - \Lambda^m_{BB'})\mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))] \\ &= \sum_{\widehat{m}\in\mathcal{M}\setminus\{m\}} q(\widehat{m}|m). \end{aligned} \tag{11.1.13}$$

The *average error probability of the code* is

$$\begin{aligned} \overline{p}_{\text{err}}(\mathcal{P}; p, \mathcal{N}) &\coloneqq \sum_{m\in\mathcal{M}} p(m) p_{\text{err}}(m; \mathcal{P}) \\ &= \sum_{m\in\mathcal{M}} p(m)(1 - q(m|m)) \\ &= \sum_{m\in\mathcal{M}} \sum_{\widehat{m}\in\mathcal{M}\setminus\{m\}} p(m) q(\widehat{m}|m). \end{aligned} \tag{11.1.14}$$

The *maximal error probability of the code* is

$$p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \coloneqq \max_{m\in\mathcal{M}} p_{\text{err}}(m, \mathcal{P}; \mathcal{N}). \tag{11.1.15}$$

Each of these three error probabilities can be used to assess the *reliability* of the protocol, i.e., how well the encoding and decoding allow Alice to transmit her message to Bob.

Definition 11.1 $(|\mathcal{M}|, \varepsilon)$ Entanglement-Assisted Classical Communication Protocol

Let $(\mathcal{M}, \Psi_{A'B'}, \mathcal{E}_{M'A'\to A}, \mathcal{D}_{BB'\to\widehat{M}})$ be the elements of an entanglement-assisted classical communication protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an $(|\mathcal{M}|, \varepsilon)$ *protocol*, with $\varepsilon \in [0, 1]$, if $p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \leq \varepsilon$.

Lemma 11.2

The following equalities hold

$$\overline{p}_{\text{err}}(\mathcal{P}; p, \mathcal{N}) = \frac{1}{2} \left\| \overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}} \right\|_1, \tag{11.1.16}$$

$$p^*_{\text{err}}(\mathcal{P};\mathcal{N}) = \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1, \tag{11.1.17}$$

where $\overline{\Phi}^p_{MM'}$ and $\omega^p_{M\widehat{M}}$ are defined in (11.1.2) and (11.1.9), respectively. Thus, the error criterion $p^*_{\text{err}}(\mathcal{P};\mathcal{N}) \leq \varepsilon$ is equivalent to $\max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1 \leq \varepsilon$.

Remark: The final criterion above states that the normalized trace distance between the initial and final states of the protocol, maximized over all possible prior probability distributions, does not exceed ε.

Proof: To see this, let us first note that the normalized trace distance in (11.1.17) is equal to the average error probability of the code. Indeed,

$$\begin{aligned}
&\frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1 \\
&= \frac{1}{2}\left\|\sum_{m\in\mathcal{M}} p(m)|m,m\rangle\langle m,m|_{MM'}\right. \\
&\qquad\qquad \left. - \sum_{m,\widehat{m}\in\mathcal{M}} p(m)q(\widehat{m}|m)|m,\widehat{m}\rangle\langle m,\widehat{m}|_{M\widehat{M}}\right\|_1 && (11.1.18)\\
&= \frac{1}{2}\left\|\sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M\right. \\
&\qquad\qquad \left.\otimes\left(|m\rangle\langle m|_{M'} - \sum_{\widehat{m}\in\mathcal{M}} q(\widehat{m}|m)|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}\right)\right\|_1 && (11.1.19)\\
&= \frac{1}{2}\sum_{m\in\mathcal{M}} p(m)\left\||m\rangle\langle m|_{M'} - \sum_{\widehat{m}\in\mathcal{M}} q(\widehat{m}|m)|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}\right\|_1 && (11.1.20)\\
&= \frac{1}{2}\sum_{m\in\mathcal{M}} p(m)\left\|(1-q(m|m))|m\rangle\langle m|\right. \\
&\qquad\qquad \left. - \sum_{\widehat{m}\in\mathcal{M}\setminus\{m\}} q(\widehat{m}|m)|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}\right\|_1 && (11.1.21)
\end{aligned}$$

$$= \frac{1}{2} \sum_{m \in \mathcal{M}} p(m) \left((1 - q(m|m)) + \sum_{\widehat{m} \in \mathcal{M} \setminus \{m\}} q(\widehat{m}|m) \right) \tag{11.1.22}$$

$$= \underbrace{\frac{1}{2} \sum_{m \in \mathcal{M}} p(m)(1 - q(m|m))}_{\overline{p}_{\text{err}}(\mathcal{P};p,\mathcal{N})} + \underbrace{\frac{1}{2} \sum_{m \in \mathcal{M}} \sum_{\widehat{m} \in \mathcal{M} \setminus \{m\}} p(m) q(\widehat{m}|m)}_{\overline{p}_{\text{err}}(\mathcal{P};p,\mathcal{N})} \tag{11.1.23}$$

$$= \overline{p}_{\text{err}}(\mathcal{P}; p, \mathcal{N}), \tag{11.1.24}$$

where the third and fifth equalities follow from (2.2.96) with $\alpha = 1$. Then, if $m^* \in \mathcal{M}$ is the message attaining the maximum error probability p^*_{err}, let $\tilde{p} : \mathcal{M} \to [0, 1]$ be the probability distribution such that $\tilde{p}(m^*) = 1$ and $\tilde{p}(m) = 0$ for all $m \neq m^*$. Using this probability distribution, we obtain

$$p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) = \max_{m \in \mathcal{M}} p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \tag{11.1.25}$$

$$= \sum_{m \in \mathcal{M}} \tilde{p}(m) p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \tag{11.1.26}$$

$$= \overline{p}_{\text{err}}(\mathcal{P}; \tilde{p}, \mathcal{N}) \tag{11.1.27}$$

$$= \frac{1}{2} \left\| \overline{\Phi}^{\tilde{p}}_{MM'} - \omega^{\tilde{p}}_{M\widehat{M}} \right\|_1 \tag{11.1.28}$$

$$\leq \max_{p:\mathcal{M} \to [0,1]} \frac{1}{2} \left\| \overline{\Phi}^{p}_{MM'} - \omega^{p}_{M\widehat{M}} \right\|_1 . \tag{11.1.29}$$

Furthermore, letting p^* be the distribution attaining the maximum average error probability, we find that

$$\max_{p:\mathcal{M} \to [0,1]} \frac{1}{2} \left\| \overline{\Phi}^{p}_{MM'} - \omega^{p}_{M\widehat{M}} \right\|_1 = \overline{p}_{\text{err}}(\mathcal{P}; p^*, \mathcal{N}) \tag{11.1.30}$$

$$= \sum_{m \in \mathcal{M}} p^*(m) p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \tag{11.1.31}$$

$$\leq \sum_{m \in \mathcal{M}} p^*(m) p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \tag{11.1.32}$$

$$= p^*_{\text{err}}(\mathcal{P}; \mathcal{N}), \tag{11.1.33}$$

where the inequality follows from the fact that

$$p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \leq \max_{m \in \mathcal{M}} p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) = p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \tag{11.1.34}$$

for all $m \in \mathcal{M}$. So we find that

$$p^*_{\text{err}}(\mathcal{P};\mathcal{N}) = \max_{p:\mathcal{M}\to[0,1]} \overline{p}_{\text{err}}(\mathcal{P};p) \tag{11.1.35}$$

$$= \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1. \tag{11.1.36}$$

This concludes the proof. ■

Another way to define the error criterion of an $(|\mathcal{M}|,\varepsilon)$ protocol, which is equivalent to the average error probability, is through what is called the *comparator test*. The comparator test is a measurement defined by the two-element POVM $\{\Pi_{M\widehat{M}}, \mathbb{1} - \Pi_{M\widehat{M}}\}$, where $\Pi_{M\widehat{M}}$ is the projection defined as

$$\Pi_{M\widehat{M}} \coloneqq \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\widehat{M}}. \tag{11.1.37}$$

Note that $\text{Tr}[\Pi_{M\widehat{M}}\omega^p_{M\widehat{M}}]$ is equal to the probability that the classical registers M and $\widehat{M}$ in the state $\omega_{M\widehat{M}}$ have the same values. In particular, observe that

$$\begin{aligned}
&\text{Tr}\left[\Pi_{M\widehat{M}}\omega^p_{M\widehat{M}}\right] \\
&\quad = \text{Tr}\left[\left(\sum_{m\in\mathcal{M}} |m,m\rangle\langle m,m|_{M\widehat{M}}\right)\right. \\
&\qquad\qquad \left.\times\left(\sum_{m',\widehat{m}\in\mathcal{M}} p(m')q(\widehat{m}|m')|m',\widehat{m}\rangle\langle m',\widehat{m}|_{M\widehat{M}}\right)\right] && (11.1.38) \\
&\quad = \sum_{m,m',\widehat{m}\in\mathcal{M}} p(m')q(\widehat{m}|m')\delta_{m,m'}\delta_{m,\widehat{m}} && (11.1.39) \\
&\quad = \sum_{m\in\mathcal{M}} p(m)q(m|m) && (11.1.40) \\
&\quad = 1 - \overline{p}_{\text{err}}(\mathcal{P};p,\mathcal{N}). && (11.1.41)
\end{aligned}$$

We can interpret this expression as the *average success probability of the code* $\mathcal{P} \equiv (\Psi,\mathcal{E},\mathcal{D})$, which we denote by

$$\overline{p}_{\text{succ}}(\mathcal{P};p,\mathcal{N}) \coloneqq 1 - \overline{p}_{\text{err}}(\mathcal{P};p,\mathcal{N}). \tag{11.1.42}$$

As mentioned at the beginning of this chapter, our goal is to bound (from above and below) the maximum number $\log_2|\mathcal{M}|$ of transmitted bits in any entanglement-assisted classical communication protocol over $\mathcal{N}$. Given an error probability tolerance of ε, we call the maximum bits of transmitted bits the *one-shot entanglement-assisted classical capacity* of $\mathcal{N}$.

Definition 11.3 One-Shot Entanglement-Assisted Classical Capacity

Given a quantum channel $\mathcal{N}_{A\to B}$ and $\varepsilon \in [0,1]$, the *one-shot ε-error entanglement-assisted classical capacity of* $\mathcal{N}$, denoted by $C^{\varepsilon}_{\text{EA}}(\mathcal{N})$, is defined to be the maximum number $\log_2|\mathcal{M}|$ of transmitted bits among all $(|\mathcal{M}|,\varepsilon)$ entanglement-assisted classical communication protocols over $\mathcal{N}$. In other words,

$$C^{\varepsilon}_{\text{EA}}(\mathcal{N}) := \sup_{(\mathcal{M},\Psi,\mathcal{E},\mathcal{D})} \{\log_2|\mathcal{M}| : p^*_{\text{err}}((\Psi,\mathcal{E},\mathcal{D});\mathcal{N}) \leq \varepsilon\}, \tag{11.1.43}$$

where the optimization is with respect to all protocols $(\mathcal{M},\Psi_{A'B'},\mathcal{E}_{M'A'\to A},\mathcal{D}_{BB'\to\widehat{M}})$ such that $d_{M'} = d_{\widehat{M}} = |\mathcal{M}|$.

In addition to finding, for a given $\varepsilon \in [0,1]$, the maximum number of transmitted bits among all $(|\mathcal{M}|,\varepsilon)$ classical communication protocols over $\mathcal{N}_{A\to B}$, we can consider the following complementary problem: for a given number of messages $|\mathcal{M}|$, find the smallest possible error among all $(|\mathcal{M}|,\varepsilon)$ entanglement-assisted classical communication protocols, which we denote by $\varepsilon^*_{\text{EA}}(|\mathcal{M}|;\mathcal{N})$. In other words, the problem is to determine

$$\varepsilon^*_{\text{EA}}(|\mathcal{M}|;\mathcal{N}) := \inf_{(\Psi,\mathcal{E},\mathcal{D})} \{p^*_{\text{err}}((\Psi,\mathcal{E},\mathcal{D});\mathcal{N}) : d_{M'} = d_{\widehat{M}} = |\mathcal{M}|\}, \tag{11.1.44}$$

where the optimization is over every state $\Psi_{A'B'}$, encoding channel $\mathcal{E}_{M'A'\to A}$, and decoding channel $\mathcal{D}_{BB'\to\widehat{M}}$, such that $d_{M'} = d_{\widehat{M}} = |\mathcal{M}|$. In this chapter, we focus primarily on the problem of optimizing the number of transmitted bits rather than the error, and so our primary quantity of interest is the one-shot capacity $C^{\varepsilon}_{\text{EA}}(\mathcal{N})$.

11.1.1 Protocol Over a Useless Channel

Our first goal is to obtain an upper bound on the one-shot entanglement-assisted classical capacity defined in (11.1.43). To do so, along with the entanglement-assisted classical communication protocol over the actual channel $\mathcal{N}$ described above, we also consider the same protocol but over the *useless* channel depicted

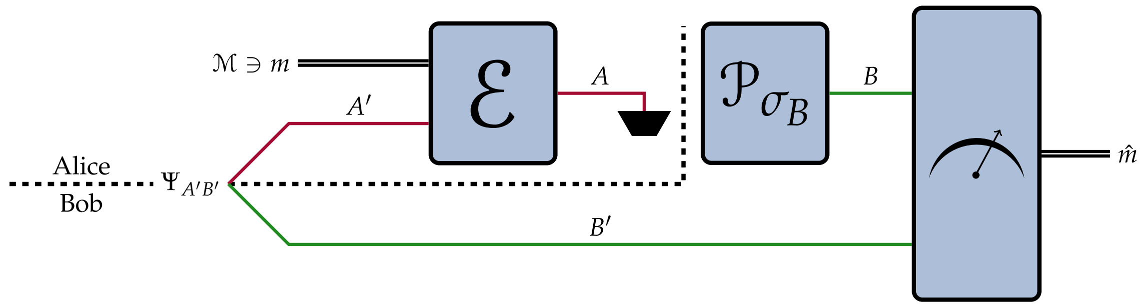


FIGURE 11.2: Depiction of a protocol that is useless for entanglement-assisted classical communication. The state encoding the message m via $\mathcal{E}$ is discarded and replaced by an arbitrary (but fixed) state σ_B.

in Figure 11.2. This useless channel discards the quantum state encoded with the message and replaces it with some arbitrary (but fixed) state σ_B. This replacement channel is useless for communication because the state σ_B does not contain any information about the message m. Intuitively, we can say that such a channel corresponds to "cutting the communication line." As we show in Lemma 11.4, comparing this protocol over the useless channel with the actual protocol allows us to obtain an upper bound on the quantity $\log_2 |\mathcal{M}|$, which we recall represents the number of bits that are transmitted over the channel.

The definition of the useless channel implies that, for every message $m \in \mathcal{M}$,

$$\mathcal{E}^m_{A' \to A}(\Psi_{A'B'}) \mapsto \mathcal{P}_{\sigma_B} \circ \mathrm{Tr}_A \circ \mathcal{E}^m_{A' \to A}(\Psi_{A'B'}) = \sigma_B \otimes \Psi_{B'}, \tag{11.1.45}$$

where $\Psi_{B'} := \mathrm{Tr}_{A'}[\Psi_{A'B'}]$. Making use of the definition of the replacement channel $\mathcal{R}^{\sigma_B}_{A \to B}$ in Definition 4.8, we can write this as

$$\mathcal{E}^m_{A' \to A}(\Psi_{A'B'}) \mapsto \mathcal{R}^{\sigma_B}_{A \to B}(\mathcal{E}_{A' \to A}(\Psi_{A'B'})). \tag{11.1.46}$$

The state at the end of the protocol over the useless channel is then

$$\begin{aligned} \tau^p_{M\widehat{M}} &:= \sum_{m,\widehat{m} \in \mathcal{M}} p(m) \mathrm{Tr}[\Lambda^{\widehat{m}}_{BB'}(\sigma_B \otimes \Psi_{B'})] |m\rangle\langle m|_M \otimes |\widehat{m}\rangle\langle \widehat{m}|_{\widehat{M}} \\ &= \pi^p_M \otimes \sum_{\widehat{m} \in \mathcal{M}} \mathrm{Tr}[\Lambda^{\widehat{m}}_{BB'}(\sigma_B \otimes \Psi_{B'})] |\widehat{m}\rangle\langle \widehat{m}|_{\widehat{M}}, \end{aligned} \tag{11.1.47}$$

where

$$\pi^p_M := \sum_{m \in \mathcal{M}} p(m) |m\rangle\langle m|_M. \tag{11.1.48}$$

Now, recall from (11.1.10) that the state $\omega^p_{M\widehat{M}}$ at the end of the actual protocol over the channel $\mathcal{N}$ is given by

$$\omega^p_{M\widehat{M}} = \sum_{m,\widehat{m}\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \operatorname{Tr}[\Lambda^{\widehat{m}}_{BB'}\mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))]|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}. \tag{11.1.49}$$

It is helpful in what follows to let

$$\overline{\Phi}_{MM'} \coloneqq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'} \tag{11.1.50}$$

be the state in (11.1.2) in which the probability distribution p is the uniform distribution over $\mathcal{M}$.

We now state a lemma that is helpful for placing an upper bound on the number $\log_2|\mathcal{M}|$ of bits communicated in an entanglement-assisted classical communication protocol. This lemma can also be used for the same purpose for unassisted classical communication protocols, as discussed in the next chapter.

Lemma 11.4

Let $\overline{\Phi}_{MM'}$ be the state defined in (11.1.50), and let $\omega_{MM'}$ be a state on the two classical registers M and M' such that $\omega_M = \operatorname{Tr}_{M'}[\omega_{MM'}] = \pi_M = \frac{\mathbb{1}_M}{|\mathcal{M}|}$. If the probability $\operatorname{Tr}[\Pi_{MM'}\omega_{MM'}]$ that the state $\omega_{MM'}$ passes the comparator test defined by the POVM $\{\Pi_{MM'}, \mathbb{1} - \Pi_{MM'}\}$, where $\Pi_{MM'}$ is the projection defined in (11.1.37), satisfies

$$\operatorname{Tr}[\Pi_{MM'}\omega_{MM'}] \geq 1 - \varepsilon, \tag{11.1.51}$$

for some $\varepsilon \in [0,1]$, then

$$\log_2|\mathcal{M}| \leq I^{\varepsilon}_H(M;M')_{\omega}, \tag{11.1.52}$$

where the ε-hypothesis testing mutual information $I^{\varepsilon}_H(M;M')_{\omega}$ is defined in (7.11.88).

PROOF: By assumption, we have that

$$\operatorname{Tr}[\Pi_{MM'}\omega_{MM'}] \geq 1 - \varepsilon, \tag{11.1.53}$$

Now, consider a state $\tau_{MM'}$ of the form $\tau_{MM'} = \omega_M \otimes \sigma_{M'} = \pi_M \otimes \sigma_{M'}$, where $\sigma_{M'}$ is some state.[1] Then,

$$\mathrm{Tr}[\Pi_{MM'}\tau_{MM'}] = \mathrm{Tr}[\Pi_{MM'}(\pi_M \otimes \sigma_{M'})] \tag{11.1.54}$$
$$= \frac{1}{|\mathcal{M}|}\mathrm{Tr}[\Pi_{MM'}(\mathbb{1}_M \otimes \sigma_{M'})] \tag{11.1.55}$$
$$= \frac{1}{|\mathcal{M}|}\mathrm{Tr}[\mathrm{Tr}_M[\Pi_{MM'}]\sigma_{M'}] \tag{11.1.56}$$
$$= \frac{1}{|\mathcal{M}|}\mathrm{Tr}[\mathbb{1}_{M'}]\sigma_{M'}] \tag{11.1.57}$$
$$= \frac{1}{|\mathcal{M}|}, \tag{11.1.58}$$

We thus obtain

$$\log_2 |\mathcal{M}| = -\log_2 \mathrm{Tr}[\Pi_{MM'}\tau_{MM'}] \tag{11.1.59}$$
$$\leq D_H^{\varepsilon}(\omega_{MM'}\|\tau_{MM'}) \tag{11.1.60}$$
$$= D_H^{\varepsilon}(\omega_{MM'}\|\omega_M \otimes \sigma_{M'}), \tag{11.1.61}$$

where the inequality follows from the definition of the hypothesis testing relative entropy in (7.9.1) (i.e., $\Pi_{MM'}$ is a particular measurement operator satisfying (11.1.53), but $D_H^{\varepsilon}(\omega_{MM'}\|\tau_{MM'})$ involves an optimization over all such operators). Since the state $\sigma_{M'}$ is arbitrary, we conclude that

$$\log_2 |\mathcal{M}| \leq \inf_{\sigma_{M'}} D_H^{\varepsilon}(\omega_{MM'}\|\omega_M \otimes \sigma_{M'}) = I_H^{\varepsilon}(M;M')_{\omega}, \tag{11.1.62}$$

which is (11.1.52), as required. ■

The right-hand side of (11.1.52) is an upper bound on the number $\log_2 |\mathcal{M}|$ of bits communicated using an $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over the channel $\mathcal{N}$. Indeed, since the error criterion $p_{\mathrm{err}}^*(\mathcal{P}) \leq \varepsilon$ holds by definition of an $(|\mathcal{M}|, \varepsilon)$ protocol, using (11.1.36) and (11.1.24) we obtain

$$\overline{p}_{\mathrm{err}}(\mathcal{P}; p, \mathcal{N}) \leq \max_{p:\mathcal{M}\to[0,1]} \overline{p}_{\mathrm{err}}(\mathcal{P}; p, \mathcal{N}) \tag{11.1.63}$$

[1] Note that the state in (11.1.47) at the end of the entanglement-assisted classical communication protocol over the useless channel has precisely this form (when p is taken to be the uniform distribution over the message set $\mathcal{M}$).

$$= \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2} \left\| \overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}} \right\|_1 \tag{11.1.64}$$

$$= p^*_{\text{err}}(\mathcal{P};\mathcal{N}) \tag{11.1.65}$$

$$\leq \varepsilon \tag{11.1.66}$$

for every probability distribution p on $\mathcal{M}$. In particular, the inequality above holds with p being the uniform distribution on $\mathcal{M}$, so that $p(m) = \frac{1}{|\mathcal{M}|}$ for all $m \in \mathcal{M}$. Let us define the state

$$\omega_{M\widehat{M}} := \frac{1}{|\mathcal{M}|} \sum_{m,\widehat{m}\in\mathcal{M}} \text{Tr}[\Lambda^{\widehat{m}}_{BB'} \mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))] |m,\widehat{m}\rangle\langle m,\widehat{m}|_{M\widehat{M}}, \tag{11.1.67}$$

which is the state $\omega^p_{M\widehat{M}}$ with p being the uniform distribution over $\mathcal{M}$. Observe that

$$\text{Tr}_{\widehat{M}}[\omega_{M\widehat{M}}] = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}\left[\left(\sum_{\widehat{m}\in\mathcal{M}} \Lambda^{\widehat{m}}_{BB'}\right) \mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))\right] |m\rangle\langle m|_M \tag{11.1.68}$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}[\mathcal{N}_{A\to B}(\mathcal{E}^m_{A'\to A}(\Psi_{A'B'}))] |m\rangle\langle m|_M \tag{11.1.69}$$

$$= \pi_M, \tag{11.1.70}$$

where the last equality follows because the channels $\mathcal{N}_{A\to B}$ and $\mathcal{E}^m_{A'\to A}$ are trace preserving. Finally, since the probability of passing the comparator test is given by (11.1.41), i.e.,

$$\text{Tr}\left[\Pi_{M\widehat{M}} \omega^p_{M\widehat{M}}\right] = 1 - \overline{p}_{\text{err}}(\mathcal{P}; p, \mathcal{N}) \tag{11.1.71}$$

for every probability distribution p, we find that $\text{Tr}[\Pi_{M\widehat{M}}\omega_{M\widehat{M}}] \geq 1 - \varepsilon$. The state $\omega_{M\widehat{M}}$ thus satisfies the condition of Lemma 11.4. We conclude that

$$\log_2 |\mathcal{M}| \leq I^\varepsilon_H(M;\widehat{M})_\omega \tag{11.1.72}$$

for every $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol.

Recall from Section 7.9 that the hypothesis testing relative entropy has an operational meaning as the optimal type-II error exponent in asymmetric hypothesis testing. The quantity $D^\varepsilon_H(\omega_{M\widehat{M}} \| \omega_M \otimes \sigma_{\widehat{M}})$ thus represents the optimal type-II error exponent, subject to the upper bound of ε on the type-I error exponent, for distinguishing between the state resulting from the actual entanglement-assisted classical communication protocol over $\mathcal{N}$ and the state resulting from an entanglement-assisted classical communication protocol over a useless channel, which discards the state encoded with the message and replaces it with the state $\sigma_{\widehat{M}}$. By taking an infimum over all states $\sigma_{\widehat{M}}$, the quantity $I_H^{\varepsilon}(M;\widehat{M})_\omega$ represents the smallest possible minimum type-II error exponent. The bound in (11.1.72) thus establishes a close link between the tasks of reliable communication and hypothesis testing.

Given a particular choice of the encoding and decoding channels, as well as a particular choice of the shared state $\Psi_{A'B'}$, if $p^*_{\text{err}}(\mathcal{P};\mathcal{N}) \leq \varepsilon$, then the quantity $I_H^{\varepsilon}(M;\widehat{M})_\omega$ in (11.1.72) is an upper bound on the maximum number of bits that can be transmitted over the channel $\mathcal{N}$. The optimal value of this upper bound can be realized by finding the state $\sigma_{\widehat{M}}$ defining the useless channel that optimizes the quantity $I_H^{\varepsilon}(M;\widehat{M})_\omega$ in addition to the measurement that achieves the ε-hypothesis testing relative entropy. Importantly, a different choice of encoding, decoding, and of the state $\Psi_{A'B'}$ produces a different value for this upper bound. We would thus like to find an upper bound that applies regardless of which specific protocol is chosen. In other words, we would like an upper bound that is a function of the channel $\mathcal{N}$ only, and this is the topic of the next section.

11.1.2 Upper Bound on the Number of Transmitted Bits

We now give a general upper bound on the number of transmitted bits possible for an arbitrary one-shot entanglement-assisted classical communication protocol for a channel $\mathcal{N}$. This result is stated in Theorem 11.6. The upper bound obtained therein holds independently of the encoding and decoding channels used in the protocol and depends only on the given communication channel $\mathcal{N}$.

Let us start with an arbitrary $(|\mathcal{M}|,\varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}$ corresponding to, as described at the beginning of this chapter, a message set $\mathcal{M}$, a prior shared entangled state $\Psi_{A'B'}$, an encoding channel $\mathcal{E}$, and a decoding channel $\mathcal{D}$. The error criterion $p^*_{\text{err}}(\mathcal{P};\mathcal{N}) \leq \varepsilon$ holds by the definition of an $(|\mathcal{M}|,\varepsilon)$ protocol. Then, by the arguments at the end of the previous section, Lemma 11.4 implies that the inequality $\log_2|\mathcal{M}| \leq I_H^{\varepsilon}(M;\widehat{M})_\omega$ holds. Using this bound on the number $\log_2|\mathcal{M}|$ of transmitted bits, we obtain the following result:

Proposition 11.5 Upper Bound on One-Shot Entanglement-Assisted Classical Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For every $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}_{A\to B}$, with $\varepsilon \in [0,1]$, the number of bits transmitted over $\mathcal{N}$ is bounded from above by the ε-hypothesis testing mutual information of $\mathcal{N}$, defined in (7.11.87), i.e.,

$$\log_2 |\mathcal{M}| \leq I_H^\varepsilon(\mathcal{N}). \tag{11.1.73}$$

Consequently, for the one-shot entanglement-assisted classical capacity, we have

$$C_{\text{EA}}^\varepsilon(\mathcal{N}) \leq I_H^\varepsilon(\mathcal{N}) \tag{11.1.74}$$

for all $\varepsilon \in [0,1]$.

Proof: First, let us apply Lemma 11.4 to conclude that

$$\log_2 |\mathcal{M}| \leq I_H^\varepsilon(M; \widehat{M})_\omega, \tag{11.1.75}$$

where $\omega_{M\widehat{M}}$ is defined in (11.1.67). Using the data-processing inequality for the hypothesis testing mutual information under the action of the decoding channel $\mathcal{D}_{BB'\to\widehat{M}}$ (from Proposition 7.19), we find that

$$I_H^\varepsilon(M; \widehat{M})_\omega \leq I_H^\varepsilon(M; BB')_\theta, \tag{11.1.76}$$

where the state $\theta_{MBB'}$ is the same as that in (11.1.6):

$$\theta_{MBB'} := (\mathcal{N}_{A\to B} \circ \mathcal{E}_{M'A'\to A})(\overline{\Phi}_{MM'} \otimes \Psi_{A'B'}). \tag{11.1.77}$$

Observe that the reduced state $\theta_{MB'}$ is a product state because the channel $\mathcal{N}_{A\to B}$ and encoding $\mathcal{E}_{M'A'\to A}$ are trace preserving:

$$\begin{aligned}
\theta_{MB'} &= \text{Tr}_B[(\mathcal{N}_{A\to B} \circ \mathcal{E}_{M'A'\to A})(\overline{\Phi}_{MM'} \otimes \Psi_{A'B'})] && (11.1.78)\\
&= \text{Tr}_{M'A'}[\overline{\Phi}_{MM'} \otimes \Psi_{A'B'}] && (11.1.79)\\
&= \overline{\Phi}_M \otimes \Psi_{B'} && (11.1.80)\\
&= \theta_M \otimes \theta_{B'}. && (11.1.81)
\end{aligned}$$

Now, by definition, we have that

$$I_H^\varepsilon(M; BB')_\theta = \inf_{\sigma_{BB'}} D_H^\varepsilon(\theta_{MBB'} \| \theta_M \otimes \sigma_{BB'}) \tag{11.1.82}$$

Choosing $\sigma_{BB'}$ to be $\sigma_B \otimes \theta_{B'}$ and optimizing over σ_B only, we find that

$$I_H^{\varepsilon}(M;BB')_\theta \leq \inf_{\sigma_B} D_H^{\varepsilon}(\theta_{MBB'} \| \theta_M \otimes \sigma_B \otimes \theta_{B'}) \tag{11.1.83}$$

$$= \inf_{\sigma_B} D_H^{\varepsilon}(\theta_{MBB'} \| \theta_{MB'} \otimes \sigma_B) \tag{11.1.84}$$

$$= I_H^{\varepsilon}(MB';B)_\theta, \tag{11.1.85}$$

where the first equality follows from the observation in (11.1.81) and the second equality follows by definition. Now, observe that the state $\theta_{MBB'}$ has the form $\mathcal{N}_{A\to B}(\rho_{SA})$, where $S \equiv MB'$ and $\rho_{SA} \equiv \mathcal{E}_{M'A'\to A}(\overline{\Phi}_{MM'} \otimes \Psi_{A'B'})$. This means that we can optimize over every state ρ_{SA} to obtain

$$I_H^{\varepsilon}(MB';B)_\theta \leq \sup_{\rho_{SA}} I_H^{\varepsilon}(S;B)_\xi, \tag{11.1.86}$$

where $\xi_{SB} = \mathcal{N}_{A\to B}(\rho_{SA})$. Note that this optimization over states ρ_{SA} is effectively an optimization over all possible encoding channels $\mathcal{E}_{M'A'\to A}$ that define the $(|\mathcal{M}|, \varepsilon)$ protocol. Now, since it suffices to take pure states when optimizing the ε-hypothesis testing mutual information of bipartite states, following from the same reasoning in (7.11.4)–(7.11.7) in the context of generalized divergences, we find that

$$I_H^{\varepsilon}(MB';B)_\theta \leq \sup_{\psi_{SA}} I_H^{\varepsilon}(S;B)_\zeta \tag{11.1.87}$$

$$= I_H^{\varepsilon}(\mathcal{N}), \tag{11.1.88}$$

where ψ_{SA} is a pure state, with the dimension of S the same as that of A, and $\zeta_{SB} = \mathcal{N}_{A\to B}(\psi_{SA})$. So we have

$$\log_2 |\mathcal{M}| \leq I_H^{\varepsilon}(M;\widehat{M})_\omega \leq I_H^{\varepsilon}(MB';B)_\theta \leq I_H^{\varepsilon}(\mathcal{N}), \tag{11.1.89}$$

as required. ■

The result of Proposition 11.5 can be written explicitly as

$$\log_2 |\mathcal{M}| \leq \sup_{\psi_{RA}} \inf_{\sigma_B} D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B) \tag{11.1.90}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_B} D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{RA}) \| \mathcal{R}^{\sigma_B}_{A\to B}(\psi_{RA})), \tag{11.1.91}$$

in which we explictly see the comparison, via the hypothesis testing relative entropy, between the actual entanglement-assisted classical communication protocol and the protocols over the useless channels $\mathcal{R}^{\sigma_B}_{A\to B}$, with each labeled by the state σ_B. The state ψ_{RA} corresponds to the state after the encoding channel, and optimizing over these states is effectively an optimization over all encoding channels and all shared entangled states. The latter is true since the preparation of the shared state $\Psi_{A'B'}$ can always be incorporated into a larger encoding channel. Specifically, the encoding $\mathcal{E}_{M'A'\to A}(\overline{\Phi}_{MM'} \otimes \Psi_{A'B'})$ can be written as $\mathcal{E}'_{M'\to A}(\overline{\Phi}_{MM'})$, where $\mathcal{E}'_{M'\to A} \coloneqq \mathcal{E}_{M'A'\to A} \circ \mathcal{P}_{\Psi_{A'B'}}$.

As an immediate consequence of Propositions 11.5, 7.70, and 7.71, we have the following two bounds:

Theorem 11.6 One-Shot Upper Bounds for Entanglement-Assisted Classical Communication

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols over the channel $\mathcal{N}$, the following bounds hold,

$$\log_2 |\mathcal{M}| \le \frac{1}{1-\varepsilon}(I(\mathcal{N}) + h_2(\varepsilon)), \tag{11.1.92}$$

$$\log_2 |\mathcal{M}| \le \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{11.1.93}$$

where $I(\mathcal{N})$ is the mutual information of $\mathcal{N}$, as defined in (7.11.102), and $\widetilde{I}_\alpha(\mathcal{N})$ is the sandwiched Rényi mutual information of $\mathcal{N}$, as defined in (7.11.91).

Since the bounds in (11.1.92) and (11.1.93) hold for every $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}$, we have that

$$C^{\varepsilon}_{\text{EA}}(\mathcal{N}) \le \frac{1}{1-\varepsilon}(I(\mathcal{N}) + h_2(\varepsilon)), \tag{11.1.94}$$

$$C^{\varepsilon}_{\text{EA}}(\mathcal{N}) \le \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{11.1.95}$$

for all $\varepsilon \in [0,1)$.

Let us recap the steps we took to arrive at the bounds in (11.1.92) and (11.1.93).

1. We first compared the entanglement-assisted classical communication protocol over $\mathcal{N}$ with the same protocol over a useless channel, by using the hypothesis

testing relative entropy. Lemma 11.4 plays a role in bounding the maximum number of transmitted bits for a particular protocol.

2. We then used the data-processing inequality for the hypothesis testing relative entropy to obtain a quantity that is independent of the decoding channel, as well as being minimized over all useless protocols when compared to the actual protocol. This is done in (11.1.76) and (11.1.82)–(11.1.85) in the proof of Proposition 11.5.

3. Finally, we optimized over all encoding channels (and, effectively, over all shared states $\Psi_{A'B'}$) in (11.1.86)–(11.1.88) to obtain Proposition 11.5, in which the bound is a function solely of the channel and the error probability.

4. Using Propositions 7.70 and 7.71, which relate the hypothesis testing relative entropy to the quantum relative entropy and the sandwiched Rényi relative entropy, respectively, we arrived at Theorem 11.6.

The bounds in (11.1.92) and (11.1.93) are fundamental upper bounds on the number of transmitted bits for *every* entanglement-assisted classical communication protocol. A natural question to ask now is whether the upper bounds in (11.1.92) and (11.1.93) can be achieved. In other words, is it possible to devise protocols such that the number of transmitted bits is equal to the right-hand side of either (11.1.92) or (11.1.93)? We do not know how to, especially if we demand that we exactly attain the right-hand side of either (11.1.92) or (11.1.93). However, when given many uses of a channel (in the asymptotic setting), we can come close to achieving these upper bounds. This motivates finding lower bounds on the number of transmitted bits.

11.1.3 Lower Bound on the Number of Transmitted Bits via Position-Based Coding and Sequential Decoding

To obtain lower bounds on the number of transmitted bits, as discussed in Appendix A, we should devise particular coding schemes. Concretely, we should devise, for all $\varepsilon \in (0,1)$, an entanglement-assisted classical communication protocol $(\mathcal{M}, \Psi_{A'B'}, \mathcal{E}, \mathcal{D})$ that is an $(|\mathcal{M}|, \varepsilon)$ protocol, meaning that the maximal error probability $p^*_{\text{err}}(\mathcal{P};\mathcal{N})$ satisfies $p^*_{\text{err}}(\mathcal{P};\mathcal{N}) \leq \varepsilon$. Recall from (11.1.15) that the maximal error probability is defined as

$$p^*_{\text{err}}(\mathcal{P};\mathcal{N}) = \max_{m\in\mathcal{M}} p_{\text{err}}(m, \mathcal{P};\mathcal{N}), \tag{11.1.96}$$

where for $m \in \mathcal{M}$ the message error probability $p_{\text{err}}(m, \mathcal{P}; \mathcal{N})$ is defined in (11.1.13) as

$$p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) = 1 - q(m|m), \tag{11.1.97}$$

with $q(\widehat{m}|m)$ being the probability of identifying the message sent as $\widehat{m}$, given that the message m was sent.

We make use of a technique called *position-based coding* along with *sequential decoding* to establish the lower bound (11.1.108) in Proposition 11.8 below, which is analogous to the upper bound (11.1.73) in Proposition 11.5. We now give a brief description of position-based coding and sequential decoding, while leaving the details to the proof of Proposition 11.8.

Let us consider an entanglement-assisted classical communication protocol defined by the four elements $(\mathcal{M}, \rho_{A'B'}^{\otimes|\mathcal{M}|}, \mathcal{E}, \mathcal{D})$ and depicted in Figure 11.3. The state shared by Alice and Bob prior to communication is $|\mathcal{M}|$ copies of a state $\rho_{A'B'}$. The encoding $\mathcal{E}$ is defined such that if Alice wishes to send a message $m \in \mathcal{M}$, then she sends her mth A' system through the channel. Specifically, the encoding channels $\mathcal{E}^m_{(A')^{|\mathcal{M}|}\to A}$ are defined as

$$\mathcal{E}^m_{(A')^{|\mathcal{M}|}\to A}(\rho_{A'B'}^{\otimes|\mathcal{M}|}) = \rho_{B_1'} \otimes \cdots \otimes \rho_{AB_m'} \otimes \cdots \otimes \rho_{B_m'} = \text{Tr}_{\bar{A}_m}\left[\rho_{A'B'}^{\otimes|\mathcal{M}|}\right], \tag{11.1.98}$$

where $\bar{A}_m$ indicates all systems A_k except for A_m. This encoding procedure is called position-based coding because the message is encoded into the particular A' system that is sent to Bob. In other words, the message is encoded into the "position" of the A' systems.[2]

The state held by the receiver Bob after Alice sends the A system in (11.1.98) through the channel $\mathcal{N}_{A\to B}$ is

$$\tau^m_{B_1'\cdots B_m'\cdots B_{|\mathcal{M}|}'B} := \rho_{B_1'} \otimes \cdots \otimes \mathcal{N}_{A\to B}(\rho_{AB_m'}) \otimes \cdots \otimes \rho_{B_{|\mathcal{M}|}'}. \tag{11.1.99}$$

Bob, whose task is to determine the message m sent to him, should apply a decoding channel that ideally succeeds with high probability. The sequential decoding strategy consists of Bob performing a sequence of measurements on systems B_i' and

[2]In practice, it would be wasteful for the sender to discard so much entanglement by explicitly using the encoding procedure in (11.1.98). The explicit encoding given should thus be considered a conceptual tool for understanding that the mth system is sent through the channel, and in practice it can be realized simply by sending the mth system through the channel.

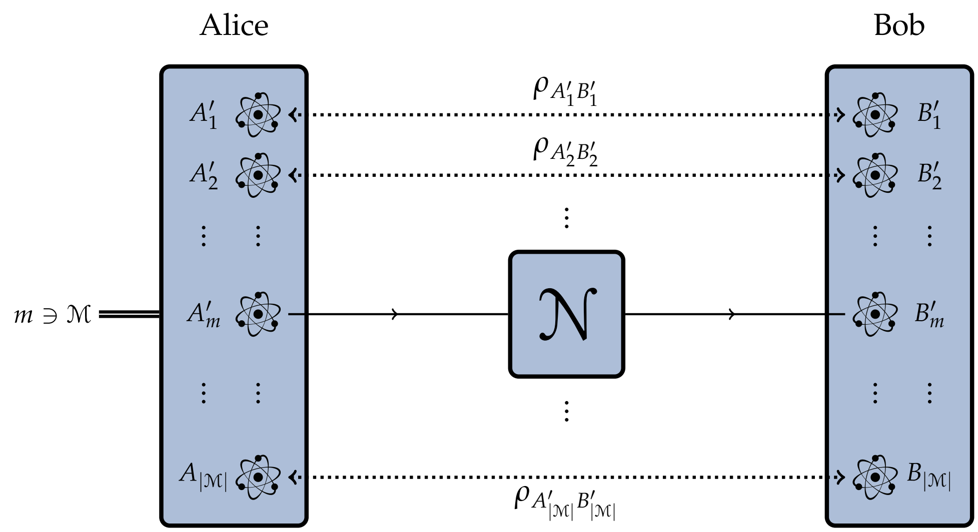


FIGURE 11.3: Schematic depiction of position-based coding. Alice and Bob start with $|\mathcal{M}|$ copies of a state $\rho_{A'B'}$. If Alice wants to send the message $m \in \mathcal{M}$, she sends the mth system through the channel $\mathcal{N}$ to Bob.

B. Each of these measurements is defined by the POVM $\{\Pi_{B'BR}, \mathbb{1}_{B'BR} - \Pi_{B'BR}\}$, where R is an arbitrary (finite-dimensional) reference system held by Bob that he makes use of to help with the decoding. In particular, Bob has identical reference systems $R_1, \ldots, R_{|\mathcal{M}|}$, each associated with his B' systems. The projectors defining the sequential decoding strategy are then

$$P_i \coloneqq \mathbb{1}_{B'_1R_1} \otimes \cdots \otimes \mathbb{1}_{B'_{i-1}R_{i-1}} \otimes \Pi_{B'_iBR_i} \otimes \mathbb{1}_{B'_{i+1}R_{i+1}} \otimes \cdots \otimes \mathbb{1}_{B'_{|\mathcal{M}|}R_{|\mathcal{M}|}} \quad (11.1.100)$$

for all $1 \leq i \leq |\mathcal{M}|$, and they correspond to measuring systems B'_iBR_i with the POVM $\{\Pi_{B'BR}, \mathbb{1}_{B'BR} - \Pi_{B'BR}\}$. This measurement can be thought of intuitively as asking the question "Was the ith message sent?", with the outcome corresponding to P_i being "yes" and the outcome corresponding to

$$\widehat{P}_i \coloneqq \mathbb{1} - P_i \quad (11.1.101)$$

being "no." Bob performs a measurement on the systems B'_1BR_1, followed by a measurement on B'_2BR_2, followed by a measurement on B'_3BR_3, etc., until he obtains the outcome corresponding to "yes." The system number corresponding to this outcome is then his guess for the message. The probability $q(\widehat{m}|m)$ of guessing $\widehat{m}$ given that the message m was sent is, therefore,

$$q(\widehat{m}|m) = \text{Tr}[P_{\widehat{m}}\widehat{P}_{\widehat{m}-1} \cdots \widehat{P}_1 \omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}} \widehat{P}_1 \cdots \widehat{P}_{\widehat{m}-1}P_{\widehat{m}}], \quad (11.1.102)$$

where

$$\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}} := \tau^m_{B'_1\cdots B'_{|\mathcal{M}|}B} \otimes |0,\cdots,0\rangle\langle 0,\cdots,0|_{R_1\cdots R_{|\mathcal{M}|}}. \tag{11.1.103}$$

The message error probability is then

$$p_{\text{err}}(m;\mathcal{P}) = 1 - \text{Tr}[P_m\widehat{P}_{m-1}\cdots\widehat{P}_1\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}\widehat{P}_1\cdots\widehat{P}_{m-1}P_m] \tag{11.1.104}$$

for all $m \in \mathcal{M}$.

Recall that our goal is to place an upper bound on the maximal error probability $p^*_{\text{err}}(\mathcal{P};\mathcal{N})$ of this position-based coding and sequential decoding protocol. We obtain an upper bound on $p_{\text{err}}(m,\mathcal{P};\mathcal{N})$ for each message m from applying the following theorem, called the *quantum union bound*, whose proof can be found in Appendix 11.A. This theorem can be thought of as a quantum generalization of the union bound from probability theory. Indeed, if $A_1,\dots,A_N$ is a sequence of events, then the union bound is as follows:

$$\Pr[(A_1\cap\cdots\cap A_N)^c] = \Pr[A_1^c\cup\cdots\cup A_N^c] \leq \sum_{i=1}^N \Pr[A_i^c], \tag{11.1.105}$$

where the superscript c denotes the complement of an event.

Theorem 11.7 Quantum Union Bound

Let $\{P_i\}_{i=1}^N$ be a set of projectors. For every state ρ and $c>0$,

$$\begin{aligned}1-&\text{Tr}[P_NP_{N-1}\cdots P_1\rho P_1\cdots P_{N-1}P_N]\\ &\leq (1+c)\text{Tr}[(\mathbb{1}-P_N)\rho] + (2+c+c^{-1})\sum_{i=1}^{N-1}\text{Tr}[(\mathbb{1}-P_i)\rho].\end{aligned} \tag{11.1.106}$$

Proof: See Appendix 11.A. ■

Using this theorem, we place the following upper bound on the message error probability $p_{\text{err}}(m,\mathcal{P};\mathcal{N})$:

$$\begin{aligned}p_{\text{err}}(m,\mathcal{P};\mathcal{N}) \leq{}& (1+c)\text{Tr}[\widehat{P}_m\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}]\\ &+ (2+c+c^{-1})\sum_{i=1}^{m-1}\text{Tr}[P_i\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}],\end{aligned} \tag{11.1.107}$$

which holds for all $c > 0$. By making a particular choice for the projectors $P_1, \ldots, P_{|\mathcal{M}|}$, and a particular choice for the constant c, we obtain the following.

Proposition 11.8 Lower Bound on One-Shot Entanglement-Assisted Classical Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$, there exists an $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}_{A\to B}$ such that

$$\log_2 |\mathcal{M}| = \overline{I}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{11.1.108}$$

Consequently, for all $\varepsilon \in (0,1)$ and for all $\eta \in (0,\varepsilon)$,

$$C_{\mathrm{EA}}^{\varepsilon}(\mathcal{N}) \geq \overline{I}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{11.1.109}$$

Here,

$$\overline{I}_H^{\varepsilon}(\mathcal{N}) := \sup_{\psi_{RA}} \overline{I}_H^{\varepsilon}(R;B)_\omega, \tag{11.1.110}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, ψ_{RA} is a pure state with the dimension of R the same as that of A, and

$$\overline{I}_H^{\varepsilon}(A;B)_\rho := D_H^{\varepsilon}(\rho_{AB} \| \rho_A \otimes \rho_B). \tag{11.1.111}$$

Remark: The quantity $\overline{I}_H^{\varepsilon}(A;B)_\rho$ defined in the statement of Proposition 11.8 above is similar to the quantity $I_H^{\varepsilon}(A;B)_\rho$ defined in (7.11.88), except that we do not perform an optimization over states σ_B. The resulting channel quantity $\overline{I}_H^{\varepsilon}(\mathcal{N})$ is then similar to the quantity $I_H^{\varepsilon}(\mathcal{N})$ defined in (7.11.87). The fact that it suffices to optimize over pure states ψ_{RA} in $\overline{I}_H^{\varepsilon}(\mathcal{N})$, with the dimension of R the same as that of A, follows from arguments analogous to those presented in Section 7.11.

Proof: Fix $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$. Starting with the encoded state on Bob's systems as defined in (11.1.99), observe that the state of the systems B and $B'_{\widehat{m}}$ is given by

$$\tau^m_{BB'_{\widehat{m}}} = \begin{cases} \mathcal{N}_{A'\to B}(\rho_{A'B'}) & \text{if } \widehat{m} = m, \\ \mathcal{N}_{A'\to B}(\rho_{A'}) \otimes \rho_{B'} & \text{if } \widehat{m} \neq m. \end{cases} \tag{11.1.112}$$

Recall that the system B results from the system A_m being sent through the channel by Alice. If, along with system B, he measures the system B'_m, then Bob is performing a measurement on the state $\mathcal{N}_{A'\to B}(\rho_{A'B'})$. On the other hand, if Bob measures a system $B'_{\widehat{m}}$, with $\widehat{m} \neq m$, then Bob is performing a measurement on the state $\mathcal{N}_{A'\to B}(\rho_{A'}) \otimes \rho_{B'}$. Bob, knowing the states $\rho_{A'B'}$ as well as the channel $\mathcal{N}$, can devise the measurement $\{\Lambda_{B'B}, \mathbb{1}_{B'B} - \Lambda_{B'B}\}$ that achieves the minimum value of the probability $\mathrm{Tr}[\Lambda_{B'B}(\rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'}))]$ while satisfying

$$\mathrm{Tr}[\Lambda_{B'B}\mathcal{N}_{A'\to B}(\rho_{A'B'})] \geq 1 - (\varepsilon - \eta). \tag{11.1.113}$$

Equivalently, by recalling Definition 7.65, the measurement achieves the ε-hypothesis testing relative entropy

$$D_H^{\varepsilon-\eta}(\mathcal{N}_{A'\to B}(\rho_{A'B'}) \| \rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'})), \tag{11.1.114}$$

meaning that

$$\begin{aligned} &-\log_2 \mathrm{Tr}[\Lambda_{B'B}(\rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'}))] \\ &\quad = D_H^{\varepsilon-\eta}(\mathcal{N}_{A'\to B}(\rho_{A'B'}) \| \rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'})) \\ &\quad = \overline{I}_H^{\varepsilon-\eta}(B'; B)_\xi, \end{aligned} \tag{11.1.115}$$

where $\xi_{B'B} \coloneqq \mathcal{N}_{A'\to B}(\rho_{A'B'})$.

The measurement with POVM $\{\Lambda_{B'B}, \mathbb{1}_{B'B} - \Lambda_{B'B}\}$ forms one part of Bob's decoding strategy. The other part of the decoding strategy is based on the fact that Bob does not know the position corresponding to the system B he receives through the channel from Alice. He therefore does not know which of the systems $B'_1, \dots, B'_{|\mathcal{M}|}$ to measure along with B. As described before the statement of the proposition, the sequential decoding strategy consists of Bob performing a sequence of projective measurements on the systems B'_iBR corresponding to the question "Was the ith message sent?". Let us define the projectors $\{\Pi_{B'BR}, \mathbb{1}_{B'BR} - \Pi_{B'BR}\}$ on which this measurement is based as follows:

$$\Pi_{B'BR} \coloneqq U^\dagger_{B'BR}(\mathbb{1}_{B'B} \otimes |1\rangle\langle 1|_R)U_{B'BR}, \tag{11.1.116}$$

where R is a qubit system and the unitary $U_{B'BR}$ is defined as

$$\begin{aligned} U_{B'BR} \coloneqq \sqrt{\mathbb{1}_{B'B} - \Lambda_{B'B}} \otimes (|0\rangle\langle 0|_R + |1\rangle\langle 1|_R) \\ + \sqrt{\Lambda_{B'B}} \otimes (|1\rangle\langle 0|_R - |0\rangle\langle 1|_R)\,. \end{aligned} \tag{11.1.117}$$

Then, it follows that

$$\mathrm{Tr}[\Pi_{B'BR}(\mathcal{N}_{A'\to B}(\rho_{A'B'}) \otimes |0\rangle\langle 0|_R)]$$

$$= \operatorname{Tr}[(\mathbb{1}_{B'B} \otimes \langle 0|_R)\Pi_{B'BR}(\mathbb{1}_{B'B} \otimes |0\rangle_R)\mathcal{N}_{A'\to B}(\rho_{A'B'})] \tag{11.1.118}$$
$$= \operatorname{Tr}[\Lambda_{B'B}\mathcal{N}_{A'\to B}(\rho_{A'B'})] \tag{11.1.119}$$
$$\geq 1 - (\varepsilon - \eta), \tag{11.1.120}$$

where the second equality holds by the definition of $\Pi_{B'BR}$ in (11.1.116) and the fact that $(\mathbb{1}_{B'B} \otimes \langle 1|_R)U_{B'BR}(\mathbb{1}_{B'B} \otimes |0\rangle_R) = \sqrt{\Lambda_{B'B}}$, which can be seen from (11.1.117). To obtain the inequality, we used (11.1.113). Defining the projection operators $\{P_i\}_{i=1}^{|\mathcal{M}|}$ as in (11.1.100), and defining the state ω as in (11.1.103), it also holds that

$$\operatorname{Tr}[P_i\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}] = \operatorname{Tr}[\Lambda_{B'B}(\rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'}))] \tag{11.1.121}$$

for all $i < m$, and

$$\operatorname{Tr}[\widehat{P}_m\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}] = \operatorname{Tr}[(\mathbb{1}_{B'B} - \Lambda_{B'B})\mathcal{N}_{A'\to B}(\rho_{B'A'})]. \tag{11.1.122}$$

Now, recall that the message error probability $p_{\text{err}}(m, \mathcal{P}; \mathcal{N})$ is defined as in (11.1.104), i.e.,

$$\begin{aligned} &p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \\ &\quad = 1 - \operatorname{Tr}[P_m\widehat{P}_{m-1}\cdots\widehat{P}_1\omega_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}\widehat{P}_1\cdots\widehat{P}_{m-1}P_m], \end{aligned} \tag{11.1.123}$$

and that we can use the quantum union bound (Theorem 11.7) to place an upper bound on this quantity as in (11.1.107), i.e.,

$$\begin{aligned} p_{\text{err}}(m; \mathcal{P}) \leq (1+c)&\operatorname{Tr}[\widehat{P}_m\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}] \\ &+ (2 + c + c^{-1})\sum_{i=1}^{m-1}\operatorname{Tr}[P_i\omega^m_{B'_1\cdots B'_{|\mathcal{M}|}BR_1\cdots R_{|\mathcal{M}|}}] \end{aligned} \tag{11.1.124}$$

for all $c > 0$. Using (11.1.121) and (11.1.122), the inequality in (11.1.113), and the equality in (11.1.115), the upper bound can be simplified so that

$$\begin{aligned} &p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \\ &\quad\leq (1+c)\operatorname{Tr}[(\mathbb{1}_{B'B} - \Lambda_{B'B})\mathcal{N}_{A'\to B}(\rho_{B'A'})] \\ &\qquad\quad + (2 + c + c^{-1})(m-1)\operatorname{Tr}[\Lambda_{B'B}(\rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'})] \end{aligned} \tag{11.1.125}$$
$$\quad\leq (1+c)(\varepsilon - \eta) + (2 + c + c^{-1})|\mathcal{M}|2^{-\overline{I}_H^{\varepsilon-\eta}(B';B)_\xi} \tag{11.1.126}$$

for all $c > 0$, where the second inequality follows because $m - 1 \leq |\mathcal{M}|$. The inequality in (11.1.126) holds for all $m \in \mathcal{M}$, which means that for all $c > 0$,

$$p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \leq (1+c)(\varepsilon - \eta) + (2 + c + c^{-1})|\mathcal{M}|2^{-\overline{I}_H^{\varepsilon-\eta}(B';B)_\xi}. \tag{11.1.127}$$

Let us set $\gamma \equiv \overline{I}_H^{\varepsilon-\eta}(B';B)_\xi$ and solve for the value of $|\mathcal{M}|$ such that

$$(1+c)(\varepsilon-\eta)+(2+c+c^{-1})|\mathcal{M}|2^{-\gamma}=\varepsilon. \tag{11.1.128}$$

We find that

$$|\mathcal{M}|=2^\gamma b(\eta-b\varepsilon), \tag{11.1.129}$$

where $b\equiv\frac{c}{1+c}$. Since b is a variable and our goal is to make $|\mathcal{M}|$ as large as possible for fixed ε and η, let us maximize $|\mathcal{M}|$ with respect to b. Solving $\frac{\partial|\mathcal{M}|}{\partial b}=0$, we find that $b=\frac{\eta}{2\varepsilon}$. This is a permissible value of b since it is required that $b>0$ and $\eta-b\varepsilon\geq 0$. Plugging back into (11.1.129), we find that

$$|\mathcal{M}|=2^\gamma\frac{\eta^2}{4\varepsilon}=2^{\overline{I}_H^{\varepsilon-\eta}(B';B)_\xi-\log_2\left(\frac{4\varepsilon}{\eta^2}\right)}. \tag{11.1.130}$$

Thus, with $|\mathcal{M}|$ given by (11.1.130), we conclude that

$$p_{\text{err}}^*(\mathcal{P})\leq\varepsilon, \tag{11.1.131}$$

and this proves the existence of an $(|\mathcal{M}|,\varepsilon)$ protocol with $|\mathcal{M}|$ given by (11.1.130).

However, (11.1.130) holds for every state $\rho_{A'B'}$, which means that we can take

$$\begin{aligned}\log_2|\mathcal{M}|&=\sup_{\rho_{A'B'}}\overline{I}_H^{\varepsilon-\eta}(B';B)_\xi-\log_2\left(\frac{4\varepsilon}{\eta^2}\right)\\&=\overline{I}_H^{\varepsilon-\eta}(\mathcal{N})-\log_2\left(\frac{4\varepsilon}{\eta^2}\right),\end{aligned} \tag{11.1.132}$$

and have (11.1.131) hold. This is precisely (11.1.108), and since $\varepsilon\in(0,1)$ and $\eta\in(0,\varepsilon)$ are arbitrary, the proof is complete. ■

Let us take note of the following two facts from the proof of Proposition 11.8 given above:

1. Given a particular $\varepsilon\in(0,1)$ and an $\eta\in(0,\varepsilon)$, we can construct a position-based coding and sequential decoding protocol achieving a maximal error probability of $p_{\text{err}}^*(\mathcal{P})\leq\varepsilon$ by taking

$$\log_2|\mathcal{M}|=\widehat{I}_H^{\varepsilon-\eta}(B';B)_\xi-\log_2\left(\frac{4\varepsilon}{\eta^2}\right), \tag{11.1.133}$$

where $\xi_{B'B} = \mathcal{N}_{A'\to B}(\rho_{A'B'})$ and $\rho_{A'B'}$ is the shared state that Alice and Bob start with (see (11.1.130)). Note that this holds for *every* state $\rho_{A'B'}$. We take the supremum at the end of the proof in order to obtain the highest possible number of transmitted bits, and also to obtain a quantity that is a function of the channel $\mathcal{N}$ only. The right-hand side of (11.1.108) can thus be achieved by determining the optimal shared state $\rho_{A'B'}$.

2. Since it suffices to optimize over pure states in order to obtain $\widehat{I}_H^{\varepsilon-\eta}(\mathcal{N})$, we see that the shared state $\rho_{A'B'}$ can be taken to be pure, with the dimension of B' the same as that of A'.

An immediate consequence of Propositions 11.8 and 7.72 is the following theorem.

Theorem 11.9 One-Shot Lower Bounds for Entanglement-Assisted Classical Communication

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For all $\varepsilon \in (0,1)$, $\eta \in (0,\varepsilon)$, and $\alpha \in (0,1)$, there exists an $(|\mathcal{M}|,\varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}_{A\to B}$ such that

$$\log_2|\mathcal{M}| \geq \overline{I}_\alpha(\mathcal{N}) - \frac{\alpha}{1-\alpha}\log_2\!\left(\frac{1}{\varepsilon-\eta}\right) - \log_2\!\left(\frac{4\varepsilon}{\eta^2}\right). \tag{11.1.134}$$

Here,

$$\overline{I}_\alpha(\mathcal{N}) \coloneqq \sup_{\psi_{RA}} \overline{I}_\alpha(R;B)_\omega, \tag{11.1.135}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, ψ_{RA} is a pure state, R has dimension equal to that of A, and

$$\overline{I}_\alpha(A;B)_\rho \coloneqq D_\alpha(\rho_{AB}\|\rho_A\otimes\rho_B). \tag{11.1.136}$$

Remark: The quantity $\overline{I}_\alpha(A;B)_\rho$ defined in the statement of Theorem 11.9 above is similar to the quantity $I_\alpha(A;B)_\rho$ defined in (7.11.90), except that we do not perform an optimization over states σ_B. The resulting channel quantity $\overline{I}_\alpha(\mathcal{N})$ is then similar to the quantity $I_\alpha(\mathcal{N})$ defined in (7.11.89). The fact that it suffices to optimize over pure states ψ_{RA} in $\overline{I}_\alpha(\mathcal{N})$, with the dimension of R the same as that of A, follows from arguments analogous to those presented in Section 7.11.

Proof: From Proposition 11.8, we know that for all $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$ there exists an $(|\mathcal{M}|,\varepsilon)$ entanglement-assisted classical communication protocol such

that

$$\log_2 |\mathcal{M}| = \overline{I}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{11.1.137}$$

Proposition 7.72 relates the hypothesis testing relative entropy to the Petz–Rényi relative entropy according to

$$D_H^{\varepsilon}(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right) \tag{11.1.138}$$

for all $\alpha \in (0,1)$, which implies that

$$\overline{I}_H^{\varepsilon}(\mathcal{N}) \geq \overline{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right). \tag{11.1.139}$$

Combining this inequality with (11.1.137), we obtain the desired result. ■

Since the inequality in (11.1.134) holds for all $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols, we have that

$$C_{\text{EA}}^{\varepsilon}(\mathcal{N}) \geq \overline{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon-\eta}\right) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{11.1.140}$$

for all $\alpha \in (0,1)$, where $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$.

11.2 Entanglement-Assisted Classical Capacity of a Quantum Channel

Let us now consider the asymptotic setting of entanglement-assisted classical communication. In this scenario, depicted in Figure 11.4, instead of encoding the message into one quantum system and consequently using the channel $\mathcal{N}$ only once, Alice encodes the message into $n \geq 1$ quantum systems $A_1, \ldots, A_n$, all with the same dimension as that of A, and sends each one of these through the channel $\mathcal{N}$. We call this the asymptotic setting because the number n can be arbitrarily large.

The analysis of the asymptotic setting is almost exactly the same as that of the one-shot setting. This is due to the fact that n independent uses of the channel $\mathcal{N}$ can be regarded as a single use of the tensor-product channel $\mathcal{N}^{\otimes n}$. So the only change that needs to be made is to replace $\mathcal{N}$ with $\mathcal{N}^{\otimes n}$ and to define the states and

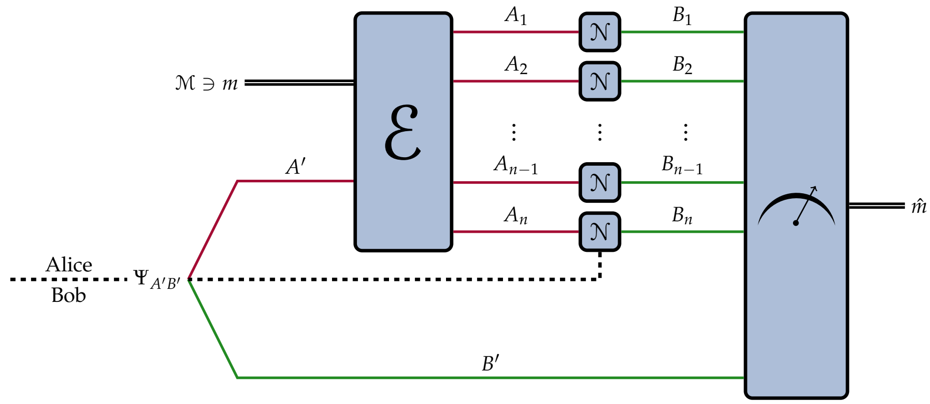


FIGURE 11.4: The most general entanglement-assisted classical communication protocol over a multiple number $n \geq 1$ uses of a quantum channel $\mathcal{N}$. Alice and Bob intially share a pair of quantum systems in the state $\Psi_{A'B'}$. Alice, who wishes to send a message m from a set $\mathcal{M}$ of messages, first encodes the message into a quantum state on n quantum systems using an encoding channel $\mathcal{E}$. She then sends each quantum system through the channel $\mathcal{N}$. After Bob receives the systems, he performs a joint measurement on them and the system B', using the outcome of the measurement to give an estimate $\widehat{m}$ of the message m sent to him by Alice.

POVM elements as acting on n systems instead of just one. In particular, the state at the end of the protocol is

$$\omega^p_{M\widehat{M}} = (\mathcal{D}_{B^nB'\to\widehat{M}} \circ \mathcal{N}^{\otimes n}_{A\to B} \circ \mathcal{E}_{M'A'\to A^n})(\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'}), \tag{11.2.1}$$

where p is the prior probability distribution over the set of messages $\mathcal{M}$, the encoding channel $\mathcal{E}_{M'A'\to A^n}$ defines a set $\{\mathcal{E}^m_{M'\to A^n}\}_{m\in\mathcal{M}}$ of channels so that

$$\mathcal{E}_{M'A'\to A^n}(\overline{\Phi}^p_{MM'} \otimes \Psi_{A'B'}) = \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \mathcal{E}^m_{A'\to A^n}(\Psi_{A'B'}), \tag{11.2.2}$$

and the decoding channel $\mathcal{D}_{B^nB'\to\widehat{M}}$, with associated POVM $\{\Lambda^m_{B^nB'}\}_{m\in\mathcal{M}}$, is defined as

$$\mathcal{D}_{B^nB'\to\widehat{M}}(\tau_{B^nB'}) = \sum_{m\in\mathcal{M}} \mathrm{Tr}[\Lambda^m_{B^nB'}\tau_{B^nB'}]|m\rangle\langle m|_{\widehat{M}}. \tag{11.2.3}$$

Then, for a given code specified by the encoding and decoding channels, the definitions of the message error probability of the code, the average error probability of the code, and the maximal error probability of the code all follow analogously from their definitions in (11.1.13), (11.1.14), and (11.1.15), respectively, from the one-shot setting.

Definition 11.10 **$(n, |\mathcal{M}|, \varepsilon)$ Entanglement-Assisted Classical Communication Protocol**

Let $(\mathcal{M}, \Psi_{A'B'}, \mathcal{E}_{M'A' \to A^n}, \mathcal{D}_{B^n B' \to \widehat{M}})$ be the elements of an entanglement-assisted classical communication protocol over n uses of the channel $\mathcal{N}_{A \to B}$. The protocol is called an $(n, |\mathcal{M}|, \varepsilon)$ *protocol*, with $\varepsilon \in [0, 1]$, if $p^*_{\text{err}}(\mathcal{P}; \mathcal{N}^{\otimes n}) \leq \varepsilon$.

Note that if there exists an $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol, then there exists an $(n, |\mathcal{M}'|, \varepsilon)$ entanglement-assisted classical communication protocol for all $\mathcal{M}'$ satisfying $|\mathcal{M}'| \leq |\mathcal{M}|$. Indeed, simply take a subset $\mathcal{M}' \subseteq \mathcal{M}$ of size $|\mathcal{M}'|$ and define the encoding and decoding channels $\mathcal{E}'$ and $\mathcal{D}'$ as the restrictions of the original channels $\mathcal{E}, \mathcal{D}$ to the set $\mathcal{M}'$. Then, using the shorthand $\mathcal{P}' \equiv (\Psi, \mathcal{E}', \mathcal{D}')$,

$$p^*_{\text{err}}(\mathcal{P}'; \mathcal{N}) = \max_{m' \in \mathcal{M}'} p_{\text{err}}(m', \mathcal{P}'; \mathcal{N}) \tag{11.2.4}$$

$$\leq \max_{m \in \mathcal{M}} p_{\text{err}}(m, \mathcal{P}; \mathcal{N}) \tag{11.2.5}$$

$$= p^*_{\text{err}}(\mathcal{P}; \mathcal{N}) \tag{11.2.6}$$

$$\leq \varepsilon, \tag{11.2.7}$$

where the first inequality holds because $\mathcal{M}'$ is a subset of $\mathcal{M}$. So we have an $(n, |\mathcal{M}'|, \varepsilon)$ protocol. Similarly, if there does not exist an $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol, then there does not exist an $(n, |\mathcal{M}'|, \varepsilon)$ entanglement-assisted classical communication protocol for all $\mathcal{M}'$ satisfying $|\mathcal{M}'| \geq |\mathcal{M}|$.

The *rate* of an entanglement-assisted classical communication protocol over n uses of a channel is equal to the number of bits that can be transmitted per channel use, i.e.,

$$R(n, |\mathcal{M}|) := \frac{1}{n} \log_2 |\mathcal{M}|. \tag{11.2.8}$$

Observe that the rate depends only on the number of messages in the set and on the number of uses of the channel. In particular, it does not directly depend on the communication channel nor on the encoding and decoding channels. Given a channel $\mathcal{N}_{A \to B}$ and $\varepsilon \in (0, 1)$, the maximum rate of entanglement-assisted classical communication over $\mathcal{N}$ among all $(n, |\mathcal{M}|, \varepsilon)$ protocols is

$$C^{n,\varepsilon}_{\text{EA}}(\mathcal{N}) := \frac{1}{n} C^{\varepsilon}_{\text{EA}}(\mathcal{N}^{\otimes n}) \tag{11.2.9}$$

$$= \sup_{(\mathcal{M},\Psi,\mathcal{E},\mathcal{D})} \left\{ \frac{1}{n} \log_2 |\mathcal{M}| : p^*_{\text{err}}((\Psi, \mathcal{E}, \mathcal{D}); \mathcal{N}^{\otimes n}) \leq \varepsilon \right\}, \tag{11.2.10}$$

where the optimization is over all entanglement-assisted classical communication protocols $(\mathcal{M}, \Psi_{A'B'}, \mathcal{E}_{M'A' \to A^n}, \mathcal{D}_{B^n B' \to \widehat{M}})$ over $\mathcal{N}^{\otimes n}$, with $d_{M'} = d_{\widehat{M}} = |\mathcal{M}|$.

The goal of an entanglement-assisted classical communication protocol in the asymptotic setting is to maximize the rate while at the same time keeping the maximal error probability low, using the number n of channel uses as a tunable parameter. Ideally, we would want the error probability to vanish, and since we want to determine the highest possible rate, we are not necessarily concerned about the practical question regarding how many channel uses might be required, at least in the asymptotic setting. In particular, as indicated by definitions, it might take an arbitrarily large number of channel uses to obtain the highest rate with a vanishing error probability.

Definition 11.11 Achievable Rate for Entanglement-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for entanglement-assisted classical communication over* $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol.

As we prove in Appendix A,

$$R \text{ achievable rate} \iff \lim_{n\to\infty} \varepsilon^*_{\text{EA}}(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) = 0 \quad \forall\, \delta > 0. \tag{11.2.11}$$

In other words, a rate R is achievable if the optimal error probability for a sequence of protocols with rate $R - \delta$, $\delta > 0$, vanishes as the number n of uses of $\mathcal{N}$ increases.

Definition 11.12 Entanglement-Assisted Classical Capacity of a Quantum Channel

The *entanglement-assisted classical capacity* of a quantum channel $\mathcal{N}$, denoted by $C_{\text{EA}}(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$C_{\text{EA}}(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{11.2.12}$$

The entanglement-assisted classical capacity can also be written as

$$C_{\mathrm{EA}}(\mathcal{N}) = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} C_{\mathrm{EA}}^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{11.2.13}$$

See Appendix A for a proof.

Definition 11.13 Weak Converse Rate for Entanglement-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *weak converse rate for entanglement-assisted classical communication over* $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

We show in Appendix A in that

$$R \text{ weak converse rate} \iff \lim_{n\to\infty} \varepsilon_{\mathrm{EA}}^{*}(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) > 0 \quad \forall\, \delta > 0. \tag{11.2.14}$$

In other words, a weak converse rate is a rate above which the optimal error probability cannot be made to vanish in the limit of a large number of channel uses.

Definition 11.14 Strong Converse Rate for Entanglement-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for entanglement-assisted classical communication over* $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}$.

We show in Appendix A that

$$R \text{ strong converse rate} \iff \lim_{n\to\infty} \varepsilon_{\mathrm{EA}}^{*}(2^{n(R+\delta)}; \mathcal{N}^{\otimes n}) = 1 \quad \forall\, \delta > 0. \tag{11.2.15}$$

In other words, unlike the weak converse, in which the optimal error is required to simply be bounded away from zero as the number n of channel uses increases, in order to have a strong converse rate the optimal error has to converge to one as n increases. By comparing (11.2.14) and (11.2.15), it is clear that every strong converse rate is a weak converse rate.

Definition 11.15 Strong Converse Entanglement-Assisted Classical Capacity of a Quantum Channel

The *strong converse entanglement-assisted classical capacity* of a quantum channel $\mathcal{N}$, denoted by $\widetilde{C}_{\text{EA}}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{C}_{\text{EA}}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{11.2.16}$$

As shown in general in Appendix A, we have that

$$C_{\text{EA}}(\mathcal{N}) \leq \widetilde{C}_{\text{EA}}(\mathcal{N}) \tag{11.2.17}$$

for every quantum channel $\mathcal{N}$. We can also write the strong converse entanglement-assisted classical capacity as

$$\widetilde{C}_{\text{EA}}(\mathcal{N}) = \sup_{\varepsilon \in [0,1)} \limsup_{n \to \infty} \frac{1}{n} C_{\text{EA}}^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{11.2.18}$$

See Appendix A for a proof.

Having defined the entanglement-assisted classical capacity of a quantum channel, as well as the strong converse capacity, we now state the main theorem of this chapter, which gives an expression for the entanglement-assisted classical capacity of a quantum channel.

Theorem 11.16 Entanglement-Assisted Classical Capacity

For every quantum channel $\mathcal{N}$, its entanglement-assisted classical capacity $C_{\text{EA}}(\mathcal{N})$ and its strong converse entanglement-assisted classical capacity $\widetilde{C}_{\text{EA}}(\mathcal{N})$ are both equal to the mutual information $I(\mathcal{N})$, i.e.,

$$C_{\text{EA}}(\mathcal{N}) = \widetilde{C}_{\text{EA}}(\mathcal{N}) = I(\mathcal{N}), \tag{11.2.19}$$

where $I(\mathcal{N})$ is defined in (7.11.102).

There are two ingredients to proving Theorem 11.16:

1. *Achievability*: We show that $I(\mathcal{N})$ is an achievable rate. In general, to show that $R \in \mathbb{R}^+$ is achievable, we define the shared entangled state $\Psi_{A'B'}$ and construct

encoding and decoding channels such that for all $\varepsilon \in (0, 1]$ and sufficiently large n, the encoding and decoding channels correspond to $(n, 2^{nr}, \varepsilon)$ protocols, as per Definition 11.10, with rates $r < R$. Thus, if R is an achievable rate, then, for every error probability ε, it is possible to find an n large enough, along with encoding and decoding channels, such that the resulting protocol has rate arbitrarily close to R and maximal error probability bounded from above by ε.

The achievability part of the proof establishes that $C_{\text{EA}}(\mathcal{N}) \geq I(\mathcal{N})$.

2. *Strong Converse*: We show that $I(\mathcal{N})$ is a strong converse rate, from which it follows that $\widetilde{C}_{\text{EA}}(\mathcal{N}) \leq I(\mathcal{N})$. In general, to show that $R \in \mathbb{R}^+$ is a strong converse rate, we show that, given any shared entangled state $\Psi_{A'B'}$ and any encoding and decoding channels, for every rate $r > R$, $\varepsilon \in [0, 1)$, and sufficiently large n, the communication protocol defined by the encoding and decoding channels is not an $(n, 2^{nr}, \varepsilon)$ protocol.

After showing the achievability and strong converse parts, we can use the inequality in (11.2.17) to conclude that

$$I(\mathcal{N}) \leq C_{\text{EA}}(\mathcal{N}) \leq \widetilde{C}_{\text{EA}}(\mathcal{N}) \leq I(\mathcal{N}), \tag{11.2.20}$$

which immediately implies that $C_{\text{EA}}(\mathcal{N}) = \widetilde{C}_{\text{EA}}(\mathcal{N}) = I(\mathcal{N})$.

We first establish in Section 11.2.1 that the rate $I(\mathcal{N})$ is achievable for entanglement-assisted classical communication over $\mathcal{N}$. We then address the additivity of the mutual information of a channel, in particular of the sandwiched Rényi mutual information of a channel, in Section 11.2.2. Finally, we prove that $I(\mathcal{N})$ is a strong converse rate in Section 11.2.3. This implies that $I(\mathcal{N})$ is a weak converse rate; however, in Section 11.2.4, we provide an independent proof of this fact, as the technique used in the proof is useful for alternate communication scenarios (besides entanglement-assisted communication) for which a strong converse theorem is not known to hold.

11.2.1 Proof of Achievability

In this section, we prove that $I(\mathcal{N})$ is an achievable rate for entanglement-assisted classical communication over $\mathcal{N}$.

First, recall from Theorem 11.9 that for all $\varepsilon \in (0, 1)$, $\eta \in (0, \varepsilon)$, and $\alpha \in (0, 1)$, there exists an $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}$ such that

$$\log_2 |\mathcal{M}| \geq \overline{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{\varepsilon - \eta}\right) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right), \tag{11.2.21}$$

where $\overline{I}_\alpha(\mathcal{N})$ is defined in (11.1.135). We obtained this result through a position-based coding strategy along with sequential decoding. A simple corollary of this result is the following.

Corollary 11.17 Lower Bound for Entanglement-Assisted Classical Communication in Asymptotic Setting

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in (0, 1]$, $n \in \mathbb{N}$, and $\alpha \in (0, 1)$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol over n uses of $\mathcal{N}$ such that

$$\frac{1}{n} \log_2 |\mathcal{M}| \geq \overline{I}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)} \log_2\left(\frac{2}{\varepsilon}\right) - \frac{3}{n}. \tag{11.2.22}$$

Proof: Fix $\varepsilon \in (0, 1]$. The inequality (11.2.21) holds for every channel $\mathcal{N}$, which means that it holds for $\mathcal{N}^{\otimes n}$. Applying the inequality in (11.2.21) to $\mathcal{N}^{\otimes n}$ and dividing both sides by n, we obtain

$$\frac{\log_2 |\mathcal{M}|}{n} \geq \frac{1}{n}\overline{I}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{\varepsilon - \eta}\right) - \frac{1}{n} \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{11.2.23}$$

for all $\alpha \in (0, 1)$. By restricting the optimization in the definition of $\overline{I}_\alpha(\mathcal{N}^{\otimes n})$ to tensor-power states, we conclude that $\overline{I}_\alpha(\mathcal{N}^{\otimes n}) \geq n\overline{I}_\alpha(\mathcal{N})$. This follows from the additivity of the Petz–Rényi relative entropy under tensor-product states (see Proposition 7.23). So we obtain

$$\frac{\log_2 |\mathcal{M}|}{n} \geq \overline{I}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{\varepsilon - \eta}\right) - \frac{1}{n} \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{11.2.24}$$

for all $\alpha \in (0, 1)$. Letting $\eta = \frac{\varepsilon}{2}$, and using the fact that $\alpha - 1$ is negative for $\alpha \in (0, 1)$, this inequality becomes

$$\frac{\log_2 |\mathcal{M}|}{n} \geq \overline{I}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)} \log_2\left(\frac{2}{\varepsilon}\right) - \frac{3}{n} \tag{11.2.25}$$

for all $\alpha \in (0, 1)$. Since ε is arbitrary, we find that for all $\varepsilon \in (0, 1]$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ protocol such that (11.2.22) is satisfied, as required. ■

The inequality in (11.2.22) gives us, for every $\varepsilon \in (0, 1]$ and $n \in \mathbb{N}$, a lower bound on the size $|\mathcal{M}|$ of the message set that we can take for a corresponding $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol defined by position-based coding and sequential decoding. If instead we fix a particular communication rate R by letting $|\mathcal{M}| = 2^{nR}$, then we can rearrange the inequality in (11.2.22) to obtain an upper bound on the maximal error probability of the corresponding $(n, 2^{nR}, \varepsilon)$ entanglement-assisted classical communication protocol. Specifically, we conclude that

$$\varepsilon \leq 2 \cdot 2^{-n(1-\alpha)\left(\overline{I}_\alpha(\mathcal{N}) - R - \frac{3}{n}\right)} \tag{11.2.26}$$

for all $\alpha \in (0, 1)$.

The inequality in (11.2.22) implies that

$$C_{\text{EA}}^{n,\varepsilon}(\mathcal{N}) \geq \overline{I}_\alpha(\mathcal{N}) - \frac{1}{n(\alpha - 1)} \log_2\left(\frac{2}{\varepsilon}\right) - \frac{3}{n} \tag{11.2.27}$$

for all $\varepsilon \in (0, 1]$ and $\alpha \in (0, 1)$.

We can now use (11.2.22) to prove that the mutual information $I(\mathcal{N})$ is an achievable rate for entanglement-assisted classical communication over $\mathcal{N}$.

Proof of the Achievability Part of Theorem 11.16

Fix $\varepsilon \in (0, 1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta = \delta_1 + \delta_2. \tag{11.2.28}$$

Set $\alpha \in (0, 1)$ such that

$$\delta_1 \geq I(\mathcal{N}) - \overline{I}_\alpha(\mathcal{N}), \tag{11.2.29}$$

which is possible since $\overline{I}_\alpha(\mathcal{N})$ is monotonically increasing in α (this follows from Proposition 7.23), and since $\lim_{\alpha \to 1^-} \overline{I}_\alpha(\mathcal{N}) = I(\mathcal{N})$ (see Appendix 11.B for a proof). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{1}{n(1 - \alpha)} \log_2\left(\frac{2}{\varepsilon}\right) + \frac{3}{n}. \tag{11.2.30}$$

Now, making use of the inequality in (11.2.22) of Corollary 11.17, there exists an $(n, |\mathcal{M}|, \varepsilon)$ protocol, with n and ε chosen as above, such that

$$\frac{\log_2 |\mathcal{M}|}{n} \geq \overline{I}_\alpha(\mathcal{N}) - \frac{1}{n(1 - \alpha)} \log_2\left(\frac{2}{\varepsilon}\right) - \frac{3}{n}. \tag{11.2.31}$$

Rearranging the right-hand side of this inequality, and using (11.2.28)–(11.2.30), we find that

$$\frac{\log_2 |\mathcal{M}|}{n} \geq I(\mathcal{N}) - \left(I(\mathcal{N}) - \overline{I}_\alpha(\mathcal{N}) + \frac{1}{n(1-\alpha)} \log_2\left(\frac{2}{\varepsilon}\right) + \frac{3}{n} \right) \tag{11.2.32}$$

$$\geq I(\mathcal{N}) - (\delta_1 + \delta_2) \tag{11.2.33}$$

$$= I(\mathcal{N}) - \delta. \tag{11.2.34}$$

We thus have $I(\mathcal{N}) - \delta \leq \frac{1}{n} \log_2 |\mathcal{M}|$. By the fact stated immediately after Definition 11.10, we conclude that there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol with $R = I(\mathcal{N})$ for all sufficiently large n such that (11.2.30) holds. Since ε and δ are arbitrary, we have that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(I(\mathcal{N})-\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol. This means that $I(\mathcal{N})$ is an achievable rate, and thus that $C_{\text{EA}}(\mathcal{N}) \geq I(\mathcal{N})$. See Appendix 11.C for a discussion of a different way of seeing the achievability proof.

11.2.2 Additivity of the Sandwiched Rényi Mutual Information of a Channel

We now turn our attention to establishing converse bounds for entanglement-assisted classical communication in the asymptotic setting. Recall from Theorem 11.6 that, for every quantum channel $\mathcal{N}$, $\varepsilon \in [0, 1)$, and all $(|\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols over $\mathcal{N}$,

$$\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon}[I(\mathcal{N}) + h_2(\varepsilon)], \tag{11.2.35}$$

$$\log_2 |\mathcal{M}| \leq \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{11.2.36}$$

To obtain these inequalities, we considered an entanglement-assisted classical communication protocol over a useless channel and used the hypothesis testing relative entropy to compare this protocol with the actual protocol over the channel $\mathcal{N}$. The useless channel in the asymptotic setting is analogous to the one in Figure 11.2 and is shown in Figure 11.5. A simple corollary of Theorem 11.6, which is relevant for the asymptotic setting, is the following:

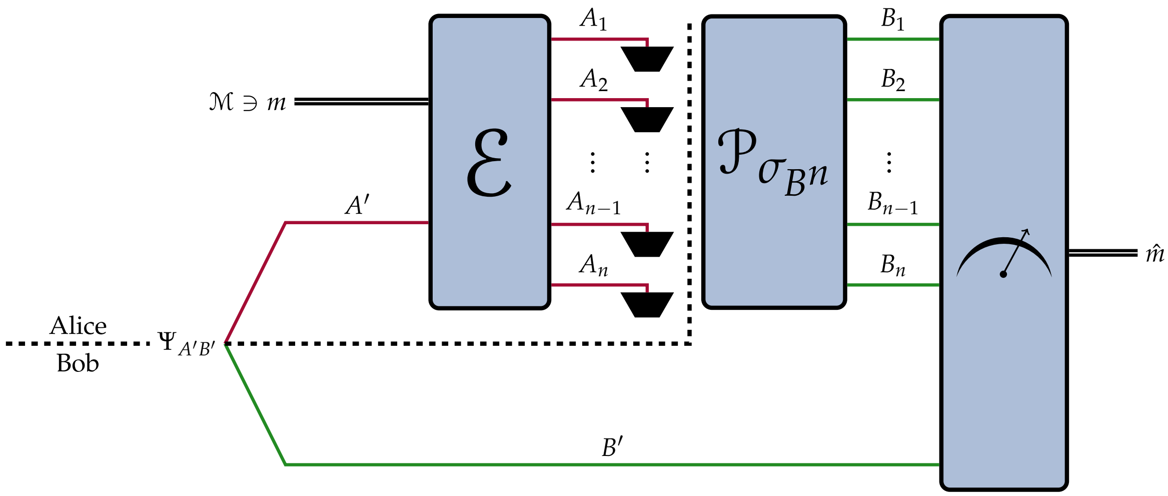


FIGURE 11.5: Depiction of a protocol that is useless for entanglement-assisted classical communication in the asymptotic setting. The state encoding the message m via $\mathcal{E}$ is discarded and replaced by an arbitrary (but fixed) state σ_{B^n}.

Corollary 11.18 Upper Bounds for Entanglement-Assisted Classical Communication in Asymptotic Setting

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in [0,1)$, $n \in \mathbb{N}$, and $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols over n uses of $\mathcal{N}$, the rate of transmitted bits is bounded from above as follows:

$$\frac{1}{n}\log_2|\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(\frac{1}{n}I(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{11.2.37}$$

$$\frac{1}{n}\log_2|\mathcal{M}| \leq \frac{1}{n}\widetilde{I}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{11.2.38}$$

PROOF: Since the inequalities (11.2.35) and (11.2.36) of Theorem 11.6 hold for every channel $\mathcal{N}$, they hold for the channel $\mathcal{N}^{\otimes n}$. Therefore, applying (11.2.35) and (11.2.36) to $\mathcal{N}^{\otimes n}$ and dividing both sides by n, we immediately obtain the desired result. ■

The inequalities in the corollary above give us, for all $\varepsilon \in [0,1)$ and $n \in \mathbb{N}$, an upper bound on the size $|\mathcal{M}|$ of the message set we can take for every corresponding $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol. If instead we fix a particular communication rate R by letting $|\mathcal{M}| = 2^{nR}$, then we can obtain a lower bound on the maximal error probability of the corresponding $(n, 2^{nR}, \varepsilon)$ entanglement-assisted classical communication protocol. Specifically, using (11.2.38), we find that

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{n}\widetilde{I}_\alpha(\mathcal{N}^{\otimes n})\right)} \tag{11.2.39}$$

for all $\alpha > 1$.

The inequality in (11.2.37) can be simplified by using the fact that the mutual information of a channel is additive, as stated in the following theorem:

Theorem 11.19 Additivity of Mutual Information of a Quantum Channel

For every two quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$, the mutual information of $\mathcal{N}_1 \otimes \mathcal{N}_2$ is equal to the sum of mutual informations of $\mathcal{N}_1$ and $\mathcal{N}_2$, i.e.,

$$I(\mathcal{N}_1 \otimes \mathcal{N}_2) = I(\mathcal{N}_1) + I(\mathcal{N}_2). \tag{11.2.40}$$

Proof: We first recall from (7.11.102) that the mutual information $I(\mathcal{N})$ of the channel $\mathcal{N}$ is defined as

$$\begin{aligned} I(\mathcal{N}) &= \sup_{\psi_{RA}} I(R; B)_\omega \\ &= \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA}) \| \psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)), \end{aligned} \tag{11.2.41}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$, ψ_{RA} is a pure state, and R is a system with the same dimension as A. Now that, as shown in Section 7.11 in the context of generalized divergences, optimizing over all states ρ_{RA} is not required, since

$$\sup_{\rho_{RA}} I(R; B)_\omega = \sup_{\psi_{RA}} I(R; B)_\omega. \tag{11.2.42}$$

We also recall that the mutual information $I(A; B)_\rho$ of a bipartite state ρ_{AB} is a special case of conditional mutual information $I(A; B|C)$ with trivial conditioning system C. By applying the additivity of conditional mutual information (see (7.2.133)), we thus conclude additivity of mutual information for product states $\tau_{A_1B_1} \otimes \omega_{A_2B_2}$:

$$I(A_1A_2; B_1B_2)_{\tau\otimes\omega} = I(A_1; B_1)_\tau + I(A_2; B_2)_\omega. \tag{11.2.43}$$

Using these facts, the inequality

$$I(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq I(\mathcal{N}_1) + I(\mathcal{N}_2), \tag{11.2.44}$$

is straightforward to establish. Letting

$$\rho_{R_1R_2B_1B_2} \coloneqq ((\mathcal{N}_1)_{A_1\to B_1} \otimes (\mathcal{N}_2)_{A_2\to B_2})(\psi_{R_1R_2A_1A_2}), \tag{11.2.45}$$

$$\tau_{R_1B_1} \coloneqq (\mathcal{N}_1)_{A_1\to B_1}(\phi_{R_1A_1}), \tag{11.2.46}$$

$$\omega_{R_2B_2} \coloneqq (\mathcal{N}_2)_{A_2\to B_2}(\varphi_{R_2A_2}), \tag{11.2.47}$$

and restricting the optimization in the mutual information of $\mathcal{N}$ to pure product states $\phi \otimes \varphi$, we get that

$$\begin{aligned} I(\mathcal{N}_1 \otimes \mathcal{N}_2) &= \sup_{\psi} I(R_1R_2; B_1B_2)_\rho && (11.2.48)\\ &\geq \sup_{\phi\otimes\varphi} I(R_1R_2; B_1B_2)_{\tau\otimes\omega} && (11.2.49)\\ &= \sup_{\phi\otimes\varphi} \{I(R_1; B_1)_\tau + I(R_2; B_2)_\omega\} && (11.2.50)\\ &= \sup_{\phi} I(R_1; B_1)_\tau + \sup_{\varphi} I(R_2; B_2)_\omega && (11.2.51)\\ &= I(\mathcal{N}_1) + I(\mathcal{N}_2). && (11.2.52) \end{aligned}$$

To prove the reverse inequality, let $\rho_{RB_1B_1} \coloneqq ((\mathcal{N}_1)_{A_1\to B_1} \otimes (\mathcal{N}_2)_{A_2\to B_2})(\psi_{RA_1A_2})$. Then, using the formula in (7.1.8) for the mutual information in terms of the quantum entropy, it is straightforward to verify that

$$I(R; B_1B_2)_\rho = I(R; B_1)_\rho + I(RB_1; B_2)_\rho - I(B_1; B_2)_\rho. \tag{11.2.53}$$

Now, Klein's inequality in Proposition 7.3, implies that the mutual information is non-negative. Using this fact on the last term in (11.2.53), we find that

$$I(R; B_1B_2)_\rho \leq I(R; B_1)_\rho + I(RB_1; B_2)_\rho. \tag{11.2.54}$$

Since $\mathcal{N}_2$ is trace preserving, we have that

$$\rho_{RB_1} = \mathrm{Tr}_{B_2}[\rho_{RB_1B_2}] = (\mathcal{N}_1)_{A_1\to B_1}(\psi_{RA_1}). \tag{11.2.55}$$

Therefore,

$$I(R; B_1)_\rho \leq \sup_{\rho_{RA_1}} I(R; B_1)_\tau = I(\mathcal{N}_1), \tag{11.2.56}$$

where the equality follows from (11.2.42). Similarly, by writing $\rho_{RB_1B_2}$ as

$$\rho_{RB_1B_2} = (\mathcal{N}_2)_{A_2\to B_2}(\omega_{RB_1A_2}), \quad \omega_{RB_1A_2} \coloneqq (\mathcal{N}_1)_{A_1\to B_1}(\psi_{RA_1A_2}), \tag{11.2.57}$$

we get that

$$I(RB_1; B_2)_\rho \leq I(\mathcal{N}_2). \tag{11.2.58}$$

Therefore,

$$I(R; B_1B_1)_\rho \leq I(\mathcal{N}_1) + I(\mathcal{N}_2). \tag{11.2.59}$$

Since the state $\psi_{RA_1A_2}$ that we started with is arbitrary, we obtain

$$I(\mathcal{N}_1 \otimes \mathcal{N}_2) = \sup_{\psi_{RA_1A_2}} I(R; B_1B_2)_\rho \leq I(\mathcal{N}_1) + I(\mathcal{N}_2), \tag{11.2.60}$$

as required. Combining this inequality with that in (11.2.44), we have the required equality, $I(\mathcal{N}_1 \otimes \mathcal{N}_2) = I(\mathcal{N}_1) + I(\mathcal{N}_2)$. ■

Using the additivity of the mutual information of a channel, the inequality in (11.2.37) can be rewritten as

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \frac{1}{1-\varepsilon}\left(I(\mathcal{N}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{11.2.61}$$

which implies that

$$C_{\text{EA}}^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{1-\varepsilon}\left(I(\mathcal{N}) + \frac{1}{n}h_2(\varepsilon)\right) \tag{11.2.62}$$

for all $n \geq 1$ and $\varepsilon \in (0, 1)$. Using this inequality, it is straightforward to conclude that $I(\mathcal{N})$ is a weak converse rate for entanglement-assisted classical communication, and the interested reader can jump ahead to Section 11.2.4 to see this.

We now show that the sandwiched Rényi mutual information $\widetilde{I}_\alpha(\mathcal{N})$ of a channel $\mathcal{N}$ is also additive, i.e., that

$$\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) = \widetilde{I}_\alpha(\mathcal{N}_1) + \widetilde{I}_\alpha(\mathcal{N}_2) \tag{11.2.63}$$

for all $\alpha > 1$. One ingredient of the proof is the additivity of the sandwiched Rényi mutual information of bipartite states with respect to tensor-product states, i.e.,

$$\widetilde{I}_\alpha(A_1A_2; B_1B_2)_{\xi\otimes\omega} = \widetilde{I}_\alpha(A_1; B_1)_\xi + \widetilde{I}_\alpha(A_2; B_2)_\omega, \tag{11.2.64}$$

where $\xi_{A_1B_1}$ and $\omega_{A_2B_2}$ are states. To show this, let us first recall the definition of the sandwiched Rényi mutual information of a bipartite state ρ_{AB} from (7.11.92):

$$\widetilde{I}_\alpha(A; B)_\rho = \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB} \| \rho_A \otimes \sigma_B), \tag{11.2.65}$$

where the optimization is over states σ_B. This quantity, as well as the sandwiched Rényi mutual information of a channel, can be written in an alternate way, as we show in the following lemma, the proof of which can be found in Appendix 11.D.

Lemma 11.20

For every bipartite state ρ_{AB} and $\alpha > 1$, the sandwiched Rényi mutual information $\widetilde{I}_\alpha(A;B)_\rho$ can be written as

$$\widetilde{I}_\alpha(A;B)_\rho = \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_C} \left\| \operatorname{Tr}_{AC}\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}} \otimes \tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\langle\psi|_{ABC}\right]\right\|_{\frac{\alpha}{2\alpha-1}}, \tag{11.2.66}$$

where $|\psi\rangle_{ABC}$ is a purification of ρ_{AB} and τ_C is a state. The sandwiched Rényi mutual information $I(\mathcal{N})$ of a channel $\mathcal{N}$ can be written as

$$\widetilde{I}_\alpha(\mathcal{N}) = \frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2\left\|\mathcal{S}_{\sigma_B}^{(\alpha)} \circ \mathcal{N}\right\|_{\text{CB},1\to\alpha}, \tag{11.2.67}$$

where $\mathcal{S}_{\sigma_B}^{(\alpha)}(\cdot) := \sigma_B^{\frac{1-\alpha}{2\alpha}}(\cdot)\sigma_B^{\frac{1-\alpha}{2\alpha}}$ and

$$\|\mathcal{M}\|_{\text{CB},1\to\alpha} := \sup_{\substack{Y_R>0,\\ \operatorname{Tr}[Y_R]\leq 1}} \left\|\mathcal{M}_{A\to B}\left(Y_R^{\frac{1}{2\alpha}}|\Gamma\rangle\langle\Gamma|_{RA}Y_R^{\frac{1}{2\alpha}}\right)\right\|_\alpha, \tag{11.2.68}$$

with $\mathcal{M}$ a completely positive map. (See Appendix 11.E for an alternate expression for $\|\cdot\|_{\text{CB},1\to\alpha}$.)

Proof: See Appendix 11.D. ■

Using the alternate expression in (11.2.66), we establish the additivity statement in (11.2.64) of the sandwiched Rényi mutual information of a bipartite state.

Proposition 11.21 Additivity of Sandwiched Rényi Mutual Information of Bipartite States

For every product state $\xi_{A_1B_1} \otimes \omega_{A_2B_2}$ and $\alpha > 1$, the sandwiched Rényi mutual information $\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega}$ is additive, i.e.,

$$\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega} = \widetilde{I}_\alpha(A_1;B_1)_\xi + \widetilde{I}_\alpha(A_2;B_2)_\omega. \tag{11.2.69}$$

PROOF: By definition, we have that

$$\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega} = \inf_{\sigma_{B_1B_2}} \widetilde{D}_\alpha(\xi_{A_1B_1}\otimes\omega_{A_2B_2}\|\xi_{A_1}\otimes\omega_{A_2}\otimes\sigma_{B_1B_2}) \quad (11.2.70)$$

If we restrict the optimization to product states $\sigma^1_{B_1}\otimes\sigma^2_{B_2}$, then we find that

$$\begin{aligned}
&\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega} \\
&\quad\leq \inf_{\sigma^1\otimes\sigma^2} \widetilde{D}_\alpha(\xi_{A_1B_1}\otimes\omega_{A_2B_2}\|\xi_{A_1}\otimes\omega_{A_2}\otimes\sigma^1_{B_1}\otimes\sigma^2_{B_2}) && (11.2.71)\\
&\quad= \inf_{\sigma^1,\sigma^2}\left\{\widetilde{D}_\alpha(\xi_{A_1B_1}\|\xi_{A_1}\otimes\sigma^1_{B_1})+\widetilde{D}_\alpha(\omega_{A_2B_2}\|\omega_{A_2}\otimes\sigma^2_{B_2})\right\} && (11.2.72)\\
&\quad= \widetilde{I}_\alpha(A_1;B_1)_\xi+\widetilde{I}_\alpha(A_2;B_2)_\omega. && (11.2.73)
\end{aligned}$$

So

$$\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega}\leq\widetilde{I}_\alpha(A_1;B_1)_\xi+\widetilde{I}_\alpha(A_2;B_2)_\omega. \quad (11.2.74)$$

To show the reverse inequality, we use the alternate expression (11.2.66) in Lemma 11.20 for the sandwiched Rényi mutual information of a bipartite state. In this expression, if we take a product purification $|\psi_1\rangle_{A_1B_1C_1}\otimes|\psi_2\rangle_{A_2B_2C_2}$ of $\xi_{A_1B_1}\otimes\omega_{A_2B_2}$ and restrict the optimization to product states, we obtain

$$\begin{aligned}
&\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\xi\otimes\omega}\\
&= \frac{\alpha}{\alpha-1}\log_2\sup_{\tau_{C_1C_2}}\Bigg\|\mathrm{Tr}_{A_1A_2C_1C_2}\Bigg[\left((\xi_{A_1}\otimes\omega_{A_2})^{\frac{1-\alpha}{\alpha}}\otimes\tau_{C_1C_2}^{\frac{\alpha-1}{\alpha}}\right)\\
&\qquad\qquad\times|\psi_1\rangle\langle\psi_1|_{A_1B_1C_1}\otimes|\psi_2\rangle\langle\psi_2|_{A_2B_2C_2}\Bigg]\Bigg\|_{\frac{\alpha}{2\alpha-1}} && (11.2.75)\\
&\geq \frac{\alpha}{\alpha-1}\log_2\sup_{\tau_{C_1}\otimes\tau_{C_2}}\Bigg\|\mathrm{Tr}_{A_1A_2C_1C_2}\Bigg[\left(\xi_{A_1}^{\frac{1-\alpha}{\alpha}}\otimes\omega_{A_2}^{\frac{1-\alpha}{\alpha}}\otimes\tau_{C_1}^{\frac{\alpha-1}{\alpha}}\otimes\tau_{C_2}^{\frac{\alpha-1}{\alpha}}\right)\\
&\qquad\qquad\times|\psi_1\rangle\langle\psi_1|_{A_1B_1C_1}\otimes|\psi_2\rangle\langle\psi_2|_{A_2B_2C_2}\Bigg]\Bigg\|_{\frac{\alpha}{2\alpha-1}} && (11.2.76)\\
&= \frac{\alpha}{\alpha-1}\log_2\sup_{\tau_{C_1},\tau_{C_2}}\Bigg\|\mathrm{Tr}_{A_1C_1}\left[\left(\xi_{A_1}^{\frac{1-\alpha}{\alpha}}\otimes\tau_{C_1}^{\frac{\alpha-1}{\alpha}}\right)|\psi_1\rangle\langle\psi_1|_{A_1B_1C_1}\right]\\
&\qquad\qquad\otimes\mathrm{Tr}_{A_2C_2}\left[\left(\omega_{A_2}^{\frac{1-\alpha}{\alpha}}\otimes\tau_{C_2}^{\frac{\alpha-1}{\alpha}}\right)|\psi_2\rangle\langle\psi_2|_{A_2B_2C_2}\right]\Bigg\|_{\frac{\alpha}{2\alpha-1}} && (11.2.77)\\
&= \frac{\alpha}{\alpha-1}\log_2\sup_{\tau_{C_1},\tau_{C_2}}\Bigg\{\left\|\mathrm{Tr}_{A_1C_1}\left[\left(\xi_{A_1}^{\frac{1-\alpha}{\alpha}}\otimes\tau_{C_1}^{\frac{\alpha-1}{\alpha}}\right)|\psi_1\rangle\langle\psi_1|_{A_1B_1C_1}\right]\right\|_{\frac{\alpha}{2\alpha-1}}
\end{aligned}$$

$$\times \left\| \operatorname{Tr}_{A_2C_2}\left[\left(\omega_{A_2}^{\frac{1-\alpha}{\alpha}} \otimes \tau_{C_2}^{\frac{\alpha-1}{\alpha}}\right) |\psi_2\rangle\langle\psi_2|_{A_2B_2C_2}\right]\right\|_{\frac{\alpha}{2\alpha-1}} \Bigg\} \tag{11.2.78}$$

$$= \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_{C_1}} \left\| \operatorname{Tr}_{A_1C_1}\left[\left(\xi_{A_1}^{\frac{1-\alpha}{\alpha}} \otimes \tau_{C_1}^{\frac{\alpha-1}{\alpha}}\right) |\psi_1\rangle\langle\psi_1|_{A_1B_1C_1}\right]\right\|_{\frac{\alpha}{2\alpha-1}}$$

$$+ \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_{C_2}} \left\| \operatorname{Tr}_{A_2C_2}\left[\left(\omega_{A_2}^{\frac{1-\alpha}{\alpha}} \otimes \tau_{C_2}^{\frac{\alpha-1}{\alpha}}\right) |\psi_2\rangle\langle\psi_2|_{A_2B_2C_2}\right]\right\|_{\frac{\alpha}{2\alpha-1}} \tag{11.2.79}$$

$$= \widetilde{I}_\alpha(A_1;B_1)_\xi + \widetilde{I}_\alpha(A_2;B_2)_\omega. \tag{11.2.80}$$

We have thus shown (11.2.69), as required. ■

Theorem 11.22 Additivity of Sandwiched Rényi Mutual Information of a Channel

For every two channels $\mathcal{N}_1$ and $\mathcal{N}_2$, and for all $\alpha > 1$,

$$\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) = \widetilde{I}_\alpha(\mathcal{N}_1) + \widetilde{I}_\alpha(\mathcal{N}_2). \tag{11.2.81}$$

PROOF: Recall that

$$\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} \widetilde{I}_\alpha(R;B)_\omega, \tag{11.2.82}$$

where $\omega_{RB} := \mathcal{N}_{A\to B}(\psi_{RA})$, and the supremum is taken over every pure state ψ_{RA}, with R having the same dimension as A. The superadditivity of the sandwiched Rényi mutual information of a channel, namely,

$$\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \widetilde{I}_\alpha(\mathcal{N}_1) + \widetilde{I}_\alpha(\mathcal{N}_2) \tag{11.2.83}$$

follows immediately by restricting the optimization in the definition (11.2.82) to product states and using the additivity of the sandwiched Rényi mutual information of bipartite states, as proven in Proposition 11.21. An explicit proof of this statement goes along the same lines as (11.2.48)–(11.2.52), with I replaced by $\widetilde{I}_\alpha$.

To prove the reverse inequality, namely, the subadditivity of the sandwiched Rényi mutual information of a channel, we use the expression in (11.2.67). By restricting the infimum in (11.2.67) to product states, we find that

$$\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma_{B_1B_2}} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_{B_1B_2}} \circ (\mathcal{N}_1 \otimes \mathcal{N}_2)\right\|_{\text{CB},1\to\alpha} \tag{11.2.84}$$

$$\leq \frac{\alpha}{\alpha-1} \inf_{\sigma^1_{B_1}\otimes\sigma^2_{B_2}} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}\otimes\sigma^2_{B_2}} \circ (\mathcal{N}_1 \otimes \mathcal{N}_2) \right\|_{\text{CB},\,1\to\alpha} \tag{11.2.85}$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma^1_{B_1},\sigma^2_{B_2}} \log_2 \left\| \left(\mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}} \circ \mathcal{N}_1 \right) \otimes \left(\mathcal{S}^{(\alpha)}_{\sigma^2_{B_2}} \circ \mathcal{N}_2 \right) \right\|_{\text{CB},\,1\to\alpha}, \tag{11.2.86}$$

where the last equality follows because

$$\mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}\otimes\sigma^2_{B_2}} = \left(\mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}} \otimes \text{id}_{B_2} \right) \circ \left(\text{id}_{B_1} \otimes \mathcal{S}^{(\alpha)}_{\sigma^2_{B_2}} \right). \tag{11.2.87}$$

Now, consider that the norm $\|\cdot\|_{\text{CB},\,1\to\alpha}$ is multiplicative with respect to tensor products of completely positive maps, i.e.,

$$\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\text{CB},\,1\to\alpha} = \|\mathcal{M}_1\|_{\text{CB},\,1\to\alpha} \, \|\mathcal{M}_2\|_{\text{CB},\,1\to\alpha} \tag{11.2.88}$$

for every two completely positive maps $\mathcal{M}_1, \mathcal{M}_2$ and all $\alpha > 1$ (see Appendix 11.F for a proof). Using this, we find that

$$\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$$

$$\leq \frac{\alpha}{\alpha-1} \inf_{\sigma^1_{B_1},\sigma^2_{B_2}} \log_2 \left\| \left(\mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}} \circ \mathcal{N}_1 \right) \otimes \left(\mathcal{S}^{(\alpha)}_{\sigma^2_{B_2}} \circ \mathcal{N}_2 \right) \right\|_{\text{CB},\,1\to\alpha} \tag{11.2.89}$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma^1_{B_1}} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma^1_{B_1}} \circ \mathcal{N}_1 \right\|_{\text{CB},\,1\to\alpha} + \frac{\alpha}{\alpha-1} \inf_{\sigma^2_{B_2}} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma^2_{B_2}} \circ \mathcal{N}_2 \right\|_{\text{CB},\,1\to\alpha} \tag{11.2.90}$$

$$= \widetilde{I}_\alpha(\mathcal{N}_1) + \widetilde{I}_\alpha(\mathcal{N}_2). \tag{11.2.91}$$

We have thus shown that $\widetilde{I}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq \widetilde{I}_\alpha(\mathcal{N}_1) + \widetilde{I}_\alpha(\mathcal{N}_2)$, and by combining this with (11.2.83), we conclude (11.2.81). ■

Note that the additivity of the mutual information of a channel, i.e., Theorem 11.19, follows straightforwardly from the theorem above by taking the limit $\alpha \to 1$ (see Appendix 11.B for a proof).

Using the additivity of the sandwiched Rényi mutual information of a channel (Theorem 11.21), the inequality in (11.2.38) can be written as

$$\frac{1}{n} \log_2 |\mathcal{M}| \leq \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon} \right) \tag{11.2.92}$$

for all $\alpha > 1$. This implies that

$$C^{n,\varepsilon}_{\text{EA}}(\mathcal{N}) \leq \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right) \tag{11.2.93}$$

for all $n \geq 1$, $\varepsilon \in (0, 1)$, and $\alpha > 1$.

11.2.3 Proof of the Strong Converse

With the inequality in (11.2.93) in hand, we can now prove that the mutual information $I(\mathcal{N})$ is a strong converse rate for entanglement-assisted classical communication over $\mathcal{N}$ and establish that $\widetilde{C}_{\text{EA}}(\mathcal{N}) = I(\mathcal{N})$.

Proof of the Strong Converse Part of Theorem 11.16

Fix $\varepsilon \in [0, 1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \tag{11.2.94}$$

Set $\alpha \in (1, \infty)$ such that

$$\delta_1 \geq \widetilde{I}_\alpha(\mathcal{N}) - I(\mathcal{N}), \tag{11.2.95}$$

which is possible since $\widetilde{I}_\alpha(\mathcal{N})$ is monotonically increasing with α (following from Proposition 7.31), and since $\lim_{\alpha\to 1^+} \widetilde{I}_\alpha(\mathcal{N}) = I(\mathcal{N})$ (see Appendix 11.B for a proof). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{11.2.96}$$

Now, with the values of n and ε as above, every $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocol satisfies (11.2.92). Rearranging the right-hand side of this inequality, and using (11.2.94)–(11.2.96), we obtain

$$\begin{aligned}
\frac{\log_2 |\mathcal{M}|}{n} &\leq I(\mathcal{N}) + \widetilde{I}_\alpha(\mathcal{N}) - I(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right) && (11.2.97)\\
&\leq I(\mathcal{N}) + \delta_1 + \delta_2 && (11.2.98)\\
&= I(\mathcal{N}) + \delta' && (11.2.99)\\
&< I(\mathcal{N}) + \delta. && (11.2.100)
\end{aligned}$$

So we have that $I(\mathcal{N})+\delta > \frac{\log_2|\mathcal{M}|}{n}$ for all $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols and sufficiently large n. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(I(\mathcal{N})+\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol for all sufficiently large n such that (11.2.96) holds, for if it did there would exist a set $\mathcal{M}$ such that $|\mathcal{M}| = 2^{n(I(\mathcal{N})+\delta)}$, which we have just seen is not possible. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(I(\mathcal{N})+\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol. This means that $I(\mathcal{N})$ is a strong converse rate, and thus that $\widetilde{C}_{\text{EA}}(\mathcal{N}) \leq I(\mathcal{N})$. See Appendix 11.G for a different way of understanding the strong converse.

11.2.4 Proof of the Weak Converse

We now conclude Section 11.2 by providing an independent proof of the fact that the mutual information $I(\mathcal{N})$ of a channel $\mathcal{N}$ is a weak converse rate.[3]

Theorem 11.23 Weak Converse for Entanglement-Assisted Classical Communication

For every quantum channel $\mathcal{N}$, the mutual information $I(\mathcal{N})$ is a weak converse rate for entanglement-assisted classical communication over $\mathcal{N}$.

PROOF: Suppose that R is an achievable rate. Then, by definition, for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement-assisted classical communication protocol over $\mathcal{N}$. For all such protocols, the inequality (11.2.61) holds by Corollary 11.18 and the additivity of the mutual information of a channel, i.e.,

$$R - \delta \leq \frac{1}{1-\varepsilon}\left(I(\mathcal{N}) + \frac{1}{n}h_2(\varepsilon)\right). \tag{11.2.101}$$

Since this bound holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$R \leq \frac{1}{1-\varepsilon}I(\mathcal{N}) + \delta. \tag{11.2.102}$$

[3]Recall that any strong converse rate is also a weak converse rate, so that by the proof of the strong converse part of Theorem 11.16 we can immediately conclude that $I(\mathcal{N})$ is a weak converse rate.

Then, since this inequality holds for all $\varepsilon \in (0, 1]$ and $\delta > 0$, we obtain

$$R \leq \lim_{\varepsilon,\delta\to 0} \left\{ \frac{1}{1-\varepsilon} I(\mathcal{N}) + \delta \right\} = I(\mathcal{N}). \tag{11.2.103}$$

We have thus shown that if R is an achievable rate, then $R \leq I(\mathcal{N})$. The contrapositive of this statement is that if $R > I(\mathcal{N})$, then R is not an achievable rate. By definition, therefore, $I(\mathcal{N})$ is a weak converse rate. ■

11.3 Examples

In this section, we determine the entanglement-assisted classical capacity of some of the channels that we introduced in Chapter 4. Recall that Theorem 11.16 states that the entanglement-assisted classical capacity $C_{\mathrm{EA}}(\mathcal{N})$ is given by the mutual information of the channel $\mathcal{N}$, i.e.,

$$C_{\mathrm{EA}}(\mathcal{N}) = I(\mathcal{N}) = \sup_{\psi_{RA}} I(R; B)_{\omega}, \tag{11.3.1}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$ and the optimization is over every pure state ψ_{RA}, with the dimension of R the same as the dimension of the input system A of the channel. The mutual information $I(R; B)_{\omega}$ can be calculated using either the quantum relative entropy or the quantum entropy via

$$I(R; B)_{\omega} = D(\mathcal{N}_{A\to B}(\psi_{RA}) \| \psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)) \tag{11.3.2}$$

$$= H(R)_{\omega} + H(B)_{\omega} - H(AB)_{\omega}. \tag{11.3.3}$$

In what follows, we consider the entanglement-assisted classical capacity of channels that are covariant with respect to a group G, and then we provide an explicit expression for the entanglement-assisted classical capacity of the depolarizing, erasure, and generalized amplitude damping channels (see Section 4.5). See Figure 11.6 for a plot of these capacities.

11.3.1 Covariant Channels

Let us start with covariant channels. Recall from Definition 4.18 that a channel $\mathcal{N}$ is covariant with respect to a group G if there exist projective unitary representations

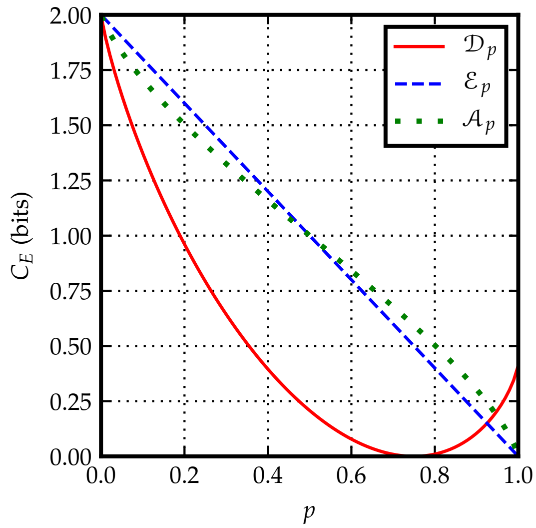


FIGURE 11.6: The entanglement-assisted classical capacity C_{EA} of the depolarizing channel $\mathcal{D}_p$ (expressed in (11.3.19)), the erasure channel $\mathcal{E}_p$ (expressed in (11.3.35)), and the amplitude damping channel $\mathcal{A}_p$ (expressed in (11.3.48)), all of which are defined for the parameter $p \in [0, 1]$.

$\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$ such that

$$\mathcal{N}(U_A^g \rho (U_A^g)^\dagger) = V_B^g \mathcal{N}(\rho)(V_B^g)^\dagger \tag{11.3.4}$$

for every state ρ and all $g \in G$.

Suppose that a channel $\mathcal{N}$ is covariant with respect to a group G such that the representation $\{U_A^g\}_{g\in G}$ of G acting on the input space of $\mathcal{N}$ satisfies

$$\frac{1}{|G|}\sum_{g\in G} U_A^g \rho_A (U_A^g)^\dagger = \text{Tr}[\rho_A]\frac{\mathbb{1}}{d}, \tag{11.3.5}$$

for all ρ_A, where d is the dimension of the input space of the channel $\mathcal{N}$. Such channels are called *irreducibly covariant*.

Let us now recall Proposition 7.86, which tells us that the generalized mutual information for every covariant channel is given as follows:

$$\boldsymbol{I}(\mathcal{N}) = \sup_{\phi_{RA}}\{\boldsymbol{I}(R;B)_\omega : \phi_A = \mathcal{T}_G(\phi_A)\}, \tag{11.3.6}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})$ and $\mathcal{T}_G(\cdot) := \frac{1}{|G|}\sum_{g\in G} U^g(\cdot)U^{g\dagger}$. In other words, for covariant channels, it suffices to optimize over pure states ϕ_{RA} for which the reduced state ϕ_A is invariant under the channel $\mathcal{T}_G$. For irreducibly covariant channels, the expression in (11.3.6) simplifies to

$$\boldsymbol{I}(\mathcal{N}) = \boldsymbol{I}(A;B)_{\rho^{\mathcal{N}}}, \tag{11.3.7}$$

where $\rho^{\mathcal{N}}_{AB} = \mathcal{N}_{A'\to B}(\Phi_{AA'})$ is the Choi state of $\mathcal{N}$.

Using (11.3.6), we immediately obtain the following theorem.

Theorem 11.24 Entanglement-Assisted Classical Capacity of Covariant Channels

If a channel $\mathcal{N}_{A\to B}$ is covariant with respect to a group G as in (11.3.4), then its entanglement-assisted classical capacity is given by

$$C_{\text{EA}}(\mathcal{N}) = \sup_{\phi_{RA}}\{I(R;B)_\omega : \phi_A = \mathcal{T}_G(\phi_A)\}, \tag{11.3.8}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})$. If the representation $\{U^g_A\}_{g\in G}$ is irreducible, then the entanglement-assisted classical capacity of $\mathcal{N}$ is given by the mutual information of its Choi state $\rho^{\mathcal{N}}_{AB}$, i.e.,

$$C_{\text{EA}}(\mathcal{N}) = I(A;B)_{\rho^{\mathcal{N}}}. \tag{11.3.9}$$

11.3.1.1 Depolarizing Channel

In Section 4.5, specifically in (4.5.31), we defined the depolarizing channel on qubits as

$$\mathcal{D}_p(\rho) := (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z) \tag{11.3.10}$$

From this, it follows that the channel is covariant with respect to the Pauli operators, meaning that

$$\mathcal{D}_p(X\rho X) = X\mathcal{D}_p(\rho)X, \tag{11.3.11}$$

$$\mathcal{D}_p(Y\rho Y) = Y\mathcal{D}_p(\rho)Y, \tag{11.3.12}$$

$$\mathcal{D}_p(Z\rho Z) = Z\mathcal{D}_p(\rho)Z. \tag{11.3.13}$$

Furthermore, we have the identity in (4.5.32), which asserts that for every state ρ,

$$\frac{1}{4}\rho + \frac{1}{4}X\rho X + \frac{1}{4}Y\rho Y + \frac{1}{4}Z\rho Z = \frac{\mathbb{1}}{2}. \tag{11.3.14}$$

This means that the operators $\{\mathbb{1}, X, Y, Z\}$ satisfy the property in (11.3.5).[4] By Theorem 11.24, we thus have that the entanglement-assisted classical capacity of the depolarizing channel is simply the mutual information of its Choi state.

It is straightforward to see that the Choi state $\rho_{AB}^{\mathcal{D}_p}$ of the depolarizing channel is

$$\rho_{AB}^{\mathcal{D}_p} = (1-p)|\Phi^+\rangle\langle\Phi^+|_{AB} + \frac{p}{3}\left(|\Psi^+\rangle\langle\Psi^+|_{AB} + |\Psi^-\rangle\langle\Psi^-|_{AB} + |\Phi^-\rangle\langle\Phi^-|_{AB}\right). \quad (11.3.15)$$

Since $H(A)_{\rho^{\mathcal{D}_p}} = H(B)_{\rho^{\mathcal{D}_p}} = \log_2(2) = 1$ and

$$H(AB)_{\rho^{\mathcal{D}_p}} = -(1-p)\log_2(1-p) - p\log_2\left(\frac{p}{3}\right), \quad (11.3.16)$$

we find that

$$C_{\text{EA}}(\mathcal{D}_p) = I(A;B)_{\rho^{\mathcal{D}_p}} = H(A)_{\rho^{\mathcal{D}_p}} + H(B)_{\rho^{\mathcal{D}_p}} - H(AB)_{\rho^{\mathcal{D}_p}} \quad (11.3.17)$$

$$= 2 + (1-p)\log_2(1-p) + p\log_2\left(\frac{p}{3}\right) \quad (11.3.18)$$

$$= 2 - h_2(p) - p\log_2(3) \quad (11.3.19)$$

for all $p \in [0,1]$. See Figure 11.6 above for a plot of the capacity.

Let us also briefly analyze the lower and upper bounds obtained in Corollaries 11.17 and 11.18, respectively. Specifically, let us consider the following bounds on the maximal error probability that results from these bounds, i.e.,

$$\varepsilon \leq 2 \cdot 2^{-n(1-\alpha)\left(\overline{I}_\alpha(\mathcal{D}_p) - R - \frac{3}{n}\right)}, \quad (11.3.20)$$

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \widetilde{I}_\alpha(\mathcal{D}_p)\right)}, \quad (11.3.21)$$

which are rearrangements of (11.2.22) and (11.2.38) and are discussed further in Appendices 11.C and 11.G, respectively. Now, by (11.3.7), we have

$$\widetilde{I}_\alpha(\mathcal{D}_p) = \widetilde{I}_\alpha(R;B)_{\rho^{\mathcal{D}_p}} \leq \overline{\overline{\widetilde{I}}}_\alpha(R;B)_{\rho^{\mathcal{D}_p}}, \quad (11.3.22)$$

where

$$\overline{\overline{\widetilde{I}}}_\alpha(A;B)_\rho \coloneqq \widetilde{D}_\alpha(\rho_{AB}\|\rho_A \otimes \rho_B). \quad (11.3.23)$$

[4] The operators $\{\mathbb{1}, X, Y, Z\}$ form a projective unitary representation of the group $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0),(0,1),(1,0),(1,1)\}$, where $\mathbb{Z}_2$ is the group consisting of the set $\{0,1\}$ with addition modulo two. Specficially, we have $U_{(0,0)} = \mathbb{1}$, $U_{(0,1)} = X$, $U_{(1,0)} = Z$, and $U_{(1,1)} = Y$.

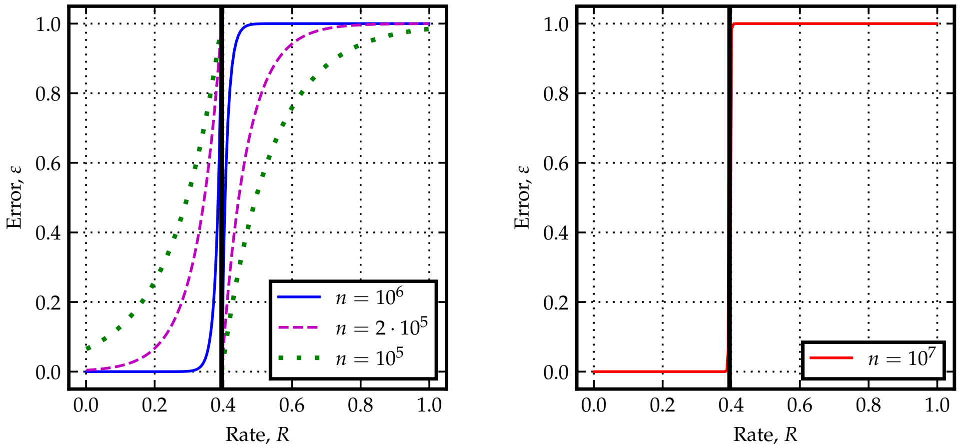


FIGURE 11.7: Plot of the error bounds in (11.3.24) and (11.3.25) for the depolarizing channel $\mathcal{D}_p$ with $p = 0.4$. By increasing the number n of channel uses, it is possible to communicate at rates closer to the capacity (indicated by the black vertical line) with vanishing error probability. Furthermore, for every rate above the capacity, as n increases, the error probability approaches one at an exponential rate, consistent with the fact that the mutual information $I(\mathcal{D}_p)$ is a strong converse rate.

For simplicity, let us use the quantity in (11.3.22), which does not involve an optimization over states σ_B, to place a lower bound on ε, so that

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\overline{\widetilde{I}}_\alpha(R;B)_\omega\right)}, \tag{11.3.24}$$

where $\omega_{RB} = \rho_{RB}^{\mathcal{D}_p}$ is the Choi state of $\mathcal{D}_p$. Similarly, for simplicity, let us take the quantity $\overline{I}_\alpha(\mathcal{N})$, which by definition involves an optimization over all pure states ψ_{RA}, and let ψ_{RA} be the maximally entangled state Φ_{RA}^+. So we take the upper bound on the error probability to be

$$\varepsilon \leq 2 \cdot 2^{-n(1-\alpha)\left(\overline{I}_\alpha(R;B)_\omega - R - \frac{3}{n}\right)}, \tag{11.3.25}$$

where $\omega_{RB} = \rho_{RB}^{\mathcal{D}_p}$ is the Choi state of $\mathcal{D}_p$. Then, for $p = 0.4$, we plot in Figure 11.7 the bounds in (11.3.24) and (11.3.25) (with $\alpha = 1.0001$ and $\alpha = 0.9999$, respectively) to obtain plots that are analogous to the generic plot in Figure 11.8 in Appendix 11.G. As portrayed in Figure 11.8, we indeed see that, as the number n of channel uses increases, the capacity $C_{\text{EA}}(\mathcal{D}_p)$ becomes a clearer dividing point between reliable communication—with nearly-vanishing error probability—and unreliable communication—with error probability approaching one at an exponential rate.

Let us consider the qudit depolarizing channel, as defined in (4.5.36) in terms of the Heisenberg–Weyl operators $\{W_{z,x}\}_{z,x}$ from (3.2.47). From the definition in (4.5.36) and the properties stated in (3.2.50)–(3.2.52), it follows that the qudit depolarizing channel is covariant with respect to the Heisenberg–Weyl operators. Furthermore, the Heisenberg–Weyl operators form an irreducible projective unitary representation of the group $\mathbb{Z}_d \times \mathbb{Z}_d$, and thus satisfy

$$\frac{1}{d^2}\sum_{z,x=0}^{d-1} W_{z,x}\rho W_{z,x}^\dagger = \mathrm{Tr}[\rho]\frac{\mathbb{1}}{d} \tag{11.3.26}$$

for all ρ. Therefore, by Theorem 11.24, we have that

$$C_{\mathrm{EA}}(\mathcal{D}_p^{(p)}) = I(A;B)_{\rho^{\mathcal{D}_p^{(d)}}}. \tag{11.3.27}$$

By calculations analogous to those above, we obtain

$$C_{\mathrm{EA}}(\mathcal{D}_p^{(d)}) = 2\log_2 d - h_2(p) - p\log_2(d^2-1). \tag{11.3.28}$$

11.3.1.2 Erasure Channel

In Section 4.5, specifically in (4.5.18), we defined the qubit erasure channel as

$$\mathcal{E}_p(\rho) \coloneqq (1-p)\rho + p\mathrm{Tr}[\rho]|e\rangle\langle e| \tag{11.3.29}$$

for $p \in [0,1]$, where $|e\rangle$ is a state that is orthogonal to all states in the input qubit system. Recall that if we let the output space simply be a qutrit system with the orthonormal basis $\{|0\rangle, |1\rangle, |2\rangle\}$, then the input qubit space can be naturally embedded into the subspace of the qutrit system spanned by $|0\rangle$ and $|1\rangle$, so that we can let the erasure state simply be $|2\rangle$. This means that we can write the action of the erasure channel as

$$\mathcal{E}_p(\rho) = (1-p)\rho + p|2\rangle\langle 2|. \tag{11.3.30}$$

Observe that, like the depolarizing channel, the erasure channel is covariant with respect to the group $\mathbb{Z}_2 \times \mathbb{Z}_2$, with the representation $\{\mathbb{1}, X, Y, Z\}$ on the input qubit space and the representation $\{\mathbb{1} \oplus |2\rangle\langle 2|, X \oplus |2\rangle\langle 2|, Y \oplus |2\rangle\langle 2|, Z \oplus |2\rangle\langle 2|\}$ on the output space. Then, by Theorem 11.24, the entanglement-assisted classical capacity of the erasure channel is equal to the mutual information of its Choi state.

The Choi state $\rho_{AB}^{\mathcal{E}_p}$ of the erasure channel is

$$\rho_{AB}^{\mathcal{E}_p} = (1-p)|\Phi^+\rangle\langle\Phi^+|_{AB} + p\frac{\mathbb{1}_A}{2} \otimes |2\rangle\langle 2|. \tag{11.3.31}$$

It is straightforward to verify that

$$H(A)_{\rho^{\mathcal{E}_p}} = 1, \tag{11.3.32}$$

$$H(B)_{\rho^{\mathcal{E}_p}} = -(1-p)\log_2\left(\frac{1-p}{2}\right) - p\log_2 p, \tag{11.3.33}$$

$$H(AB)_{\rho^{\mathcal{E}_p}} = -(1-p)\log_2(1-p) - p\log_2\left(\frac{p}{2}\right), \tag{11.3.34}$$

so that

$$C_{\text{EA}}(\mathcal{E}_p) = I(A;B)_{\rho^{\mathcal{E}_p}} = 2(1-p). \tag{11.3.35}$$

In general, as discussed in Section 4.5.2, we can consider an erasure channel $\mathcal{E}_p^{(d)}$ acting on a qudit system with dimension d and orthonormal basis $\{|1\rangle, \dots, |d\rangle\}$, so that the output of the erasure channel is a state on a $(d+1)$-dimensional system, i.e.,

$$\mathcal{E}_p^{(d)}(\rho) = (1-p)\rho + p\text{Tr}[\rho]|d+1\rangle\langle d+1|. \tag{11.3.36}$$

Then, it is straightforward to see that the qudit erasure channel is irreducibly covariant with respect to the group $\mathbb{Z}_d \times \mathbb{Z}_d$, with the corresponding representation on the input space being $\{W_{z,x} : 0 \le z, x \le d-1\}$ and the representation on the output space being $\{W_{z,x} \oplus |d+1\rangle\langle d+1| : 0 \le z, x \le d-1\}$. Therefore, by reasoning analogous to the above, we obtain

$$C_{\text{EA}}(\mathcal{E}_p^{(d)}) = 2(1-p)\log_2 d. \tag{11.3.37}$$

11.3.2 Generalized Amplitude Damping Channel

In (4.5.10), we defined the generalized amplitide damping channel $\mathcal{A}_{\gamma,N}$ as the channel with the four Kraus operators in (4.5.11) and (4.5.12), i.e.,

$$\mathcal{A}_{\gamma,N}(\rho) = A_1\rho A_1^\dagger + A_2\rho A_2^\dagger + A_3\rho A_3^\dagger + A_4\rho A_4^\dagger, \tag{11.3.38}$$

where

$$A_1 = \sqrt{1-N}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad A_2 = \sqrt{1-N}\begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{11.3.39}$$

$$A_3 = \sqrt{N}\begin{pmatrix}\sqrt{1-\gamma} & 0\\ 0 & 1\end{pmatrix}, \quad A_4 = \sqrt{N}\begin{pmatrix}0 & 0\\ \sqrt{\gamma} & 0\end{pmatrix}. \tag{11.3.40}$$

Now, it is straightforward to verify that $ZA_1 = A_1Z$, $ZA_2 = -A_2Z$, $ZA_3 = A_3Z$, and $ZA_4 = -ZA_4$. Therefore,

$$\mathcal{A}_{\gamma,N}(Z\rho Z) = Z\mathcal{A}_{\gamma,N}(\rho)Z \tag{11.3.41}$$

for every state ρ. The generalized amplitude damping channel is thus covariant with respect to $\{\mathbb{1}, Z\}$, which can be viewed as a representation of the group $G = \mathbb{Z}_2$ on both the input and output spaces of the channel. Note that this representation does not satisfy the property in (11.3.5).

Nevertheless, we can use the expression in (11.3.8) to determine the entanglement-assisted classical capacity of the generalized amplitude damping channel. First, observe that the channel $\rho \mapsto \frac{1}{|G|}\sum_{g\in G} U_g\rho U_g^\dagger$ is same as the completely dephasing channel defined in (4.5.28). Since the output of the completely dephasing channel is always diagonal in the basis $\{|0\rangle, |1\rangle\}$, by (11.3.8) it suffices to optimize over pure states ψ_{RA} such that the reduced state ψ_A is diagonal in the basis $\{|0\rangle, |1\rangle\}$. Thus, up to an (irrelevant) unitary on the system R, the pure states ψ_{RA} in (11.3.8) can be taken to have the form

$$|\psi\rangle_{RA} = \sqrt{1-p}|0,0\rangle_{RA} + \sqrt{p}|1,1\rangle_{RA} \tag{11.3.42}$$

for $p \in [0,1]$. For every such state, it is straightforward to show that the corresponding output state $\omega_{RB} = (\mathcal{A}_{\gamma,N})_{A\to B}(\psi_{RA})$ has four eigenvalues: $(1-p)\,\gamma N$, $(1-N)\,p\gamma$, and

$$\lambda_\pm \coloneqq \frac{1}{2}\left(1 - \gamma(N + p - 2Np) \pm f(N,p,\gamma)\right), \tag{11.3.43}$$

$$f(N,p,\gamma) \coloneqq \sqrt{\gamma^2(N-p)^2 - 2\gamma(N+p-2Np) + 1}, \tag{11.3.44}$$

which means that

$$\begin{aligned} H(RB)_{\rho^{\mathcal{A}_{\gamma,N}}} = -(1-p)\gamma N \log_2((1-p)\gamma N) \\ - (1-N)p\gamma \log_2((1-N)p\gamma) \\ - \lambda_+ \log_2 \lambda_+ - \lambda_- \log_2 \lambda_-. \end{aligned} \tag{11.3.45}$$

Since $\omega_R = \psi_R = (1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|$, we have that $H(R)_{\rho^{\mathcal{A}_{\gamma,N}}} = h_2(p)$. Finally, we have

$$H(B)_{\rho^{\mathcal{A}_{\gamma,N}}} = h_2(\gamma(p-N) + 1 - p)). \tag{11.3.46}$$

Therefore,

$$I(\mathcal{A}_{\gamma,N}) = \max_{p\in[0,1]} \{h_2(p) + h_2(\gamma(p-N)+1-p)) + (1-p)\gamma N \log_2((1-p)\gamma N) + (1-N)p\gamma \log_2((1-N)p\gamma) + \lambda_+ \log_2 \lambda_+ + \lambda_- \log_2 \lambda_-\}. \quad (11.3.47)$$

In the case $N = 0$, the channel $\mathcal{A}_{\gamma,0}$ is the amplitude damping channel $\mathcal{A}_\gamma$ (see (4.5.1)). Using (11.3.47), we find that

$$I(\mathcal{A}_\gamma) = \max_{p\in[0,1]} \{h_2(p(1-\gamma)) + h_2(p) - h_2(\gamma p)\} \quad (11.3.48)$$

for all $\gamma \in [0,1]$.

11.4 Summary

In this chapter, we formally defined and studied entanglement-assisted classical communication. We began with the fundamental one-shot setting, in which a quantum channel is used just once for entanglement-assisted classical communication, and we defined the one-shot entanglement-assisted classical capacity in (11.1.43). We then derived upper and lower bounds on the one-shot capacity (Propositions 11.5 and 11.8), making a fundamental link between communication and hypothesis testing for both bounds. To derive the upper bound, the main conceptual point was to compare an actual protocol for entanglement-assisted communication with a useless one. This approach led to the hypothesis testing mutual information as an upper bound. To derive the lower bound, we employed the combined approach of sequential decoding and position-based coding, which at its core, is about how well a correlated state can be distinguished from a product state. Stepping back a bit, this is conceptually similar to the idea behind the converse upper bound, which ultimately features a comparison between a correlated state and a product state. We can consider the one-shot setting to contain the fundamental information theoretic argument for entanglement-assisted communication.

With the one-shot setting in hand, we moved on the asymptotic setting, in which the channel is allowed to be used multiple times (as a model of how communication channels would be used in practice). We defined various notions of communication rates, including achievable rates, capacity, weak converse rates, strong converse rates, and strong converse capacity. With the fundamental one-shot bounds in place, we then substituted one use of the channel $\mathcal{N}$ with n uses (the tensor-product channel $\mathcal{N}^{\otimes n}$) and applied various technical arguments to prove that the mutual information of a channel is equal to both its capacity and strong converse capacity for entanglement-assisted communication. As a main step to establish the capacity, we proved that the mutual information of a channel is additive, and as a main step to establish the strong converse capacity, we proved that the sandwiched Rényi mutual information of a channel is additive.

Finally, we calculated the entanglement-assisted classical for several key channels, including the depolarizing, erasure, and generalized amplitude damping channels, in order to illustrate the theory on some concrete examples.

As it turns out, the strongest results known in quantum information theory are for the entanglement-assisted capacity. The results stated above hold for all quantum channels, and thus can be viewed from the physics perspective as universal physical laws delineating the ultimate limits of entanglement-assisted classical communication for any physical process (i.e., described by a quantum channel). In this sense, shared entanglement simplifies quantum information theory immensely.

Going forward from here, the same concepts such as capacity, achievable rate, etc. can be defined for different communication tasks (i.e., unassisted classical communication, quantum communication, private communication, etc.). What changes is that the known results are not as strong as they are for the entanglement-assisted setting. We know the capacity of these other communication tasks only for certain subclasses of channels. This might be considered unfortunate, but a different perspective is that it is exciting, because rather exotic phenomena such as superadditivity and superactivation can occur.

11.5 Bibliographic Notes

Entanglement-assisted classical communication was devised by Bennett et al. (1999b), as an information-theoretic extension of super-dense coding. The entanglement-assisted classical capacity theorem was proven by Bennett et al. (2002), and Holevo (2002a) gave a different proof for this theorem.

Entanglement-assisted classical communication in the one-shot setting was considered by Datta and Hsieh (2013); Matthews and Wehner (2014); Datta et al. (2016); Anshu et al. (2019); Qi et al. (2018b); Anshu et al. (2019). Proposition 11.5 is due to Matthews and Wehner (2014), and the proof given here was found independently by Qi et al. (2018b) and Anshu et al. (2019).

The position-based coding method was introduced by Anshu et al. (2019). It can be understood as a quantum generalization of pulse position modulation (Verdu, 1990; Cariolaro and Erseghe, 2003). Sequential decoding was considered by Giovannetti et al. (2012); Sen (2012); Wilde (2013); Gao (2015); Oskouei et al. (2019), and Theorem 11.7 is due to Oskouei et al. (2019). Proposition 11.8 is due to Qi et al. (2018b), and the proof given here was found by Oskouei et al. (2019).

The strong converse for entanglement-assisted classical capacity was established by Bennett et al. (2014) and Gupta and Wilde (2015), with the latter paper employing the Rényi entropic method used in this chapter. Eq. (11.2.66) and the additivity of sandwiched Rényi mutual information of bipartite states (Proposition 11.21) were established by Beigi (2013). Eq. (11.2.67) was established by Gupta and Wilde (2015), and the completely-bound $1 \to \alpha$ norm was studied in depth by Devetak et al. (2006). Theorem 11.22 was proven by Gupta and Wilde (2015), by employing the multiplicativity of completely bounded norms (Eq. (11.2.88)) found by Devetak et al. (2006).

The entanglement-assisted classical capacity of the depolarizing and erasure channels was evaluated by Bennett et al. (1999b), the same capacity for the amplitude damping channel was evaluated by Giovannetti and Fazio (2005), and the same capacity for the generalized amplitude damping channel was evaluated by Li-Zhen and Mao-Fa (2007a).

The proofs in Appendix 11.B are due to Cooney et al. (2016), and the proofs in Appendices 11.E and 11.F are due to Jencova (2006) (with the proofs in this book containing some slight variations). The Lieb concavity theorem (Theorem 11.30) is due to Lieb (1973).

Appendix 11.A Proof of Theorem 11.7

Theorem 11.7 is a consequence of the following more general result:

Theorem 11.25

Let $\{P_i\}_{i=1}^N$ be a finite set of projectors. For every vector $|\psi\rangle$ and $c > 0$,

$$\||\psi\rangle\|_2^2 - \|P_N P_{N-1} \cdots P_1 |\psi\rangle\|_2^2 \leq (1+c)\,\|(\mathbb{1} - P_N)|\psi\rangle\|_2^2 + (2 + c + c^{-1}) \sum_{i=2}^{N-1} \|(\mathbb{1} - P_i)|\psi\rangle\|_2^2 + (2 + c^{-1})\,\|(\mathbb{1} - P_1)|\psi\rangle\|_2^2 . \tag{11.A.1}$$

Indeed, recall from Theorem 2.4 that every density operator ρ has a spectral decomposition of the following form:

$$\rho = \sum_{j \in \mathcal{J}} p(j) |\psi_j\rangle\langle\psi_j|, \tag{11.A.2}$$

where $p : \mathcal{J} \to [0, 1]$ is a probability distribution, and $\{|\psi_j\rangle\}_{j \in \mathcal{J}}$ is an orthonormal set of eigenvectors. Applying Theorem 11.25, we find that

$$\begin{aligned}
&1 - \mathrm{Tr}[P_N P_{N-1} \cdots P_1 |\psi_j\rangle\langle\psi_j| P_1 \cdots P_{N-1}] \\
&= \left\||\psi_j\rangle\right\|_2^2 - \left\|P_N P_{N-1} \cdots P_1 |\psi_j\rangle\right\|_2^2 &(11.A.3)\\
&\leq (1+c) \left\|(\mathbb{1} - P_N)\,|\psi_j\rangle\right\|_2^2 + (2 + c + c^{-1}) \sum_{i=2}^{N-1} \left\|(\mathbb{1} - P_i)\,|\psi_j\rangle\right\|_2^2 \\
&\qquad + (2 + c^{-1}) \left\|(\mathbb{1} - P_1)\,|\psi_j\rangle\right\|_2^2 &(11.A.4)\\
&= (1+c)\,\mathrm{Tr}[(\mathbb{1} - P_N)\,|\psi_j\rangle\langle\psi_j|] + (2 + c + c^{-1}) \sum_{i=2}^{N-1} \mathrm{Tr}[(\mathbb{1} - P_i)|\psi_j\rangle\langle\psi_j|] \\
&\qquad + (2 + c^{-1})\mathrm{Tr}[(\mathbb{1} - P_1)\,|\psi_j\rangle\langle\psi_j|]. &(11.A.5)
\end{aligned}$$

The reduction from Theorem 11.25 to Theorem 11.7 follows by averaging the above inequality over the probability distribution $p : \mathcal{J} \to [0, 1]$ and from the fact that the right-hand side of the resulting inequality can be bounded from above by

$$(1+c)\mathrm{Tr}[(\mathbb{1} - P_N)\rho] + (2 + c + c^{-1}) \sum_{i=1}^{N-1} \mathrm{Tr}[(\mathbb{1} - P_i)\rho], \tag{11.A.6}$$

so that

$$\begin{aligned} &1 - \mathrm{Tr}[P_N P_{N-1} \cdots P_1 \rho P_1 \cdots P_{N-1} P_N] \\ &\qquad \leq (1+c)\mathrm{Tr}[(\mathbb{1} - P_N)\rho] + (2 + c + c^{-1}) \sum_{i=1}^{N-1} \mathrm{Tr}[(\mathbb{1} - P_i)\rho]. \end{aligned} \tag{11.A.7}$$

We therefore shift our focus to proving Theorem 11.25, and we do so with the aid of several lemmas. To simplify the notation, we hereafter employ the following shorthands:

$$\|\cdots\|_\psi \equiv \|\cdots|\psi\rangle\|_2 \,, \tag{11.A.8}$$

$$\langle\cdots\rangle_\psi \equiv \langle\psi|\cdots|\psi\rangle, \tag{11.A.9}$$

$$\widehat{P}_i \equiv \mathbb{1} - P_i, \tag{11.A.10}$$

where for a given operator A the expression $\langle A\rangle_\psi = \langle\psi|A|\psi\rangle$ is assumed to mean $\langle\psi|(A|\psi\rangle)$. Furthermore, we also assume without loss of generality that the vector $|\psi\rangle$ in Theorem 11.25 is a unit vector. This assumption can be dropped by scaling the resulting inequality by an arbitrary positive number corresponding to the norm of $|\psi\rangle$.

First recall that, due to the fact that $P^2 = P$ for every projector P, we have the following identities holding for all $i \in \{1, 2, \ldots, N\}$:

$$\begin{aligned} \langle\widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi &= \langle\widehat{P}_i \widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi, \\ \langle P_1 \cdots P_i\rangle_\psi &= \langle P_1 \cdots P_i P_i\rangle_\psi, \end{aligned} \tag{11.A.11}$$

under the convention that $P_{i-1} \cdots P_1 = P_1 \cdots P_{i-1} = \mathbb{1}$ for $i = 1$.

Lemma 11.26

For a set $\{P_i\}_{i=1}^N$, a unit vector $|\psi\rangle$, and employing the shorthand in (11.A.8)–(11.A.10), we have the following identities and inequality:

$$\sum_{i=1}^N \langle\widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi = 1 - \langle P_N \cdots P_1\rangle_\psi, \tag{11.A.12}$$

$$\sum_{i=1}^N \langle P_1 \cdots P_{i-1} \widehat{P}_i\rangle_\psi = 1 - \langle P_1 \cdots P_N\rangle_\psi, \tag{11.A.13}$$

$$\sum_{i=1}^{N}\langle P_1\cdots P_{i-1}\widehat{P}_iP_{i-1}\cdots P_1\rangle_\psi = 1-\langle P_1\cdots P_N\cdots P_1\rangle_\psi, \quad (11.A.14)$$

$$1-\sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1\cdots P_N\cdots P_1\rangle_\psi} \leq \sum_{i=1}^{N}\sqrt{\langle \widehat{P}_i\rangle_\psi}\sqrt{\langle P_1\cdots P_{i-1}\widehat{P}_iP_{i-1}\cdots P_1\rangle_\psi}, \quad (11.A.15)$$

under the convention that $P_{i-1}\cdots P_1 = P_1\cdots P_{i-1} = \mathbb{1}$ for $i=1$.

Proof: The following identities are straightforward to verify:

$$1 = \langle \widehat{P}_1\rangle_\psi + \langle \widehat{P}_2P_1\rangle_\psi + \cdots + \langle \widehat{P}_{N-1}P_{N-2}\cdots P_1\rangle_\psi + \langle \widehat{P}_NP_{N-1}\cdots P_1\rangle_\psi + \langle P_NP_{N-1}\cdots P_1\rangle_\psi, \quad (11.A.16)$$

$$1 = \langle \widehat{P}_1\rangle_\psi + \langle P_1\widehat{P}_2\rangle_\psi + \cdots + \langle P_1\cdots P_{N-2}\widehat{P}_{N-1}\rangle_\psi + \langle P_1\cdots P_{N-1}\widehat{P}_N\rangle_\psi + \langle P_1\cdots P_{N-1}P_N\rangle_\psi, \quad (11.A.17)$$

$$1 = \langle \widehat{P}_1\rangle_\psi + \langle P_1\widehat{P}_2P_1\rangle_\psi + \cdots + \langle P_1\cdots P_{N-2}\widehat{P}_{N-1}P_{N-2}\cdots P_1\rangle_\psi + \langle P_1\cdots P_{N-1}\widehat{P}_NP_{N-1}\cdots P_1\rangle_\psi + \langle P_1\cdots P_{N-1}P_NP_{N-1}\cdots P_1\rangle_\psi. \quad (11.A.18)$$

Consequently, from the equalities in (11.A.16), (11.A.17), and (11.A.18), we obtain (11.A.12), (11.A.13), and (11.A.14), respectively. The following equality is a direct consequence of (11.A.16) and (11.A.11):

$$1 = \langle \widehat{P}_1\rangle_\psi + \langle \widehat{P}_2\widehat{P}_2P_1\rangle_\psi + \cdots + \langle \widehat{P}_{N-1}\widehat{P}_{N-1}P_{N-2}\cdots P_1\rangle_\psi + \langle \widehat{P}_N\widehat{P}_NP_{N-1}\cdots P_1\rangle_\psi + \langle P_NP_NP_{N-1}\cdots P_1\rangle_\psi. \quad (11.A.19)$$

By applying the Cauchy–Schwarz inequality from (2.2.32) to (11.A.19), we find that

$$1 \leq \langle \widehat{P}_1\rangle_\psi + \sqrt{\langle \widehat{P}_2\rangle_\psi}\sqrt{\langle P_1\widehat{P}_2P_1\rangle_\psi} + \cdots + \sqrt{\langle \widehat{P}_N\rangle_\psi}\sqrt{\langle P_1\cdots P_{N-1}\widehat{P}_NP_{N-1}\cdots P_1\rangle_\psi} + \sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1\cdots P_{N-1}P_NP_{N-1}\cdots P_1\rangle_\psi}, \quad (11.A.20)$$

from which (11.A.15) immediately follows. ■

Lemma 11.27

For a set $\{P_i\}_{i=1}^N$ of projectors, a unit vector $|\psi\rangle$, and employing the shorthand in (11.A.8)–(11.A.10), the following inequality holds for $N \geq 2$:

$$\sum_{i=1}^{N} \|\widehat{P}_i(\mathbb{1} - P_{i-1} \cdots P_1)\|_\psi^2 \leq \sum_{i=1}^{N-1} \|\widehat{P}_i\|_\psi^2, \tag{11.A.21}$$

under the convention that $P_{i-1} \cdots P_1 = P_1 \cdots P_{i-1} = \mathbb{1}$ for $i = 1$. Equivalently,

$$\sum_{i=2}^{N} \|\widehat{P}_i(\mathbb{1} - P_{i-1} \cdots P_1)\|_\psi^2 \leq \sum_{i=1}^{N-1} \|\widehat{P}_i\|_\psi^2, \tag{11.A.22}$$

due to the aforementioned convention.

Proof: Consider the following chain of equalities:

$$\sum_{i=1}^{N} \|\widehat{P}_i(\mathbb{1} - P_{i-1} \cdots P_1)\|_\psi^2$$

$$= \sum_{i=1}^{N} \|\widehat{P}_i - \widehat{P}_i P_{i-1} \cdots P_1\|_\psi^2 \tag{11.A.23}$$

$$= \sum_{i=1}^{N} \Big(\|\widehat{P}_i\|_\psi^2 - \langle \widehat{P}_i P_{i-1} \cdots P_1 \rangle_\psi - \langle P_1 \cdots P_{i-1} \widehat{P}_i \rangle_\psi + \langle P_1 \cdots P_{i-1} \widehat{P}_i P_{i-1} \cdots P_1 \rangle_\psi \Big) \tag{11.A.24}$$

$$= \left(\sum_{i=1}^{N} \|\widehat{P}_i\|_\psi^2 \right) - 1 + \langle P_N \cdots P_1 \rangle_\psi - 1 + \langle P_1 \cdots P_N \rangle_\psi + 1 - \langle P_1 \cdots P_N \cdots P_1 \rangle_\psi \tag{11.A.25}$$

$$= \left(\sum_{i=1}^{N} \|\widehat{P}_i\|_\psi^2 \right) - 1 + \langle P_N P_N P_{N-1} \cdots P_1 \rangle_\psi + \langle P_1 \cdots P_{N-1} P_N P_N \rangle_\psi - \langle P_1 \cdots P_N \cdots P_1 \rangle_\psi . \tag{11.A.26}$$

To obtain (11.A.24), we used the identities in (11.A.11). Next, to get (11.A.25), the identities in (11.A.12), (11.A.13), and (11.A.14) of Lemma 11.26 were used.

Continuing, we have that

$$\text{Eq. (11.A.26)} \leq \left(\sum_{i=1}^{N} \|\widehat{P}_i\|_\psi^2\right) - 1 - \langle P_1 \cdots P_N \cdots P_1\rangle_\psi + 2\sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi} \quad (11.A.27)$$

$$= \left(\sum_{i=1}^{N} \|\widehat{P}_i\|_\psi^2\right) - 1 + \langle P_N\rangle_\psi - \left(\sqrt{\langle P_N\rangle_\psi} - \sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right)^2 \quad (11.A.28)$$

$$\leq \left(\sum_{i=1}^{N} \|\widehat{P}_i\|_\psi^2\right) - \|\widehat{P}_N\|_\psi^2 = \sum_{i=1}^{N-1} \|\widehat{P}_i\|_\psi^2. \quad (11.A.29)$$

To obtain (11.A.27), the Cauchy-Schwarz inequality was employed. ■

We are now in a position to prove Theorem 11.25.

Proof of Theorem 11.25

Consider that

$$1 - \|P_N \cdots P_1\|_\psi^2 = 1 - \langle P_1 \cdots P_N \cdots P_1\rangle_\psi + 2\left(1 - \sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right) - 2\left(1 - \sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right) \quad (11.A.30)$$

$$= 2\left(1 - \sqrt{\langle P_N\rangle_\psi}\sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right) - \left(\sqrt{\langle P_N\rangle_\psi} - \sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right)^2 - 1 + \langle P_N\rangle_\psi. \quad (11.A.31)$$

Continuing, we have that

Eq. (11.A.31)

$$\leq -\|\widehat{P}_N\|_\psi^2 + 2\left(1 - \sqrt{P_N}\sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right) \tag{11.A.32}$$

$$\leq -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^N \sqrt{\langle \widehat{P}_i\rangle_\psi}\sqrt{\langle P_1 \cdots P_{i-1}\widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi} \tag{11.A.33}$$

$$\leq -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^N \sqrt{\langle \widehat{P}_i\rangle_\psi}\left(\|\widehat{P}_i\|_\psi + \|\widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi\right). \tag{11.A.34}$$

First, (11.A.32) is obtained by observing that

$$-\left(\sqrt{\langle P_N\rangle_\psi} - \sqrt{\langle P_1 \cdots P_N \cdots P_1\rangle_\psi}\right)^2 - 1 + \langle P_N\rangle_\psi \leq -1 + \langle P_N\rangle_\psi$$

$$= -\|\widehat{P}_N\|_\psi^2. \tag{11.A.35}$$

Next, (11.A.33) follows from (11.A.15) of Lemma 11.26. Then, (11.A.34) is a consequence of the triangle inequality:

$$\sqrt{\langle P_1 \cdots P_{i-1}\widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi} = \sqrt{\langle P_1 \cdots P_{i-1}\widehat{P}_i\widehat{P}_i P_{i-1} \cdots P_1\rangle_\psi} \tag{11.A.36}$$

$$= \|\widehat{P}_i P_{i-1}\cdots P_1\|_\psi \tag{11.A.37}$$

$$= \|\widehat{P}_i(-\mathbb{1} + \mathbb{1} - P_{i-1}\cdots P_1)\|_\psi \tag{11.A.38}$$

$$= \|-\widehat{P}_i + \widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi \tag{11.A.39}$$

$$\leq \|\widehat{P}_i\|_\psi + \|\widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi. \tag{11.A.40}$$

Continuing, we have that

Eq. (11.A.34)

$$= -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^N \|\widehat{P}_i\|_\psi^2 + 2\sum_{i=1}^N\left(\|\widehat{P}_i\|_\psi \|\widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi\right) \tag{11.A.41}$$

$$= -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^N \|\widehat{P}_i\|_\psi^2 + 2\sum_{i=2}^N\left(\|\widehat{P}_i\|_\psi \|\widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi\right) \tag{11.A.42}$$

$$\leq -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^N \|\widehat{P}_i\|_\psi^2$$

$$+ \sum_{i=2}^N\left(c\|\widehat{P}_i\|_\psi^2 + c^{-1}\|\widehat{P}_i(\mathbb{1} - P_{i-1}\cdots P_1)\|_\psi^2\right) \tag{11.A.43}$$

$$\leq -\|\widehat{P}_N\|_\psi^2 + 2\sum_{i=1}^{N}\|\widehat{P}_i\|_\psi^2 + c\sum_{i=2}^{N}\|\widehat{P}_i\|_\psi^2 + c^{-1}\sum_{i=1}^{N-1}\|\widehat{P}_i\|_\psi^2 \tag{11.A.44}$$

$$\leq (1+c)\|\widehat{P}_N\|_\psi^2 + (2+c^{-1})\|\widehat{P}_1\|_\psi^2 + (2+c+c^{-1})\sum_{i=2}^{N-1}\|\widehat{P}_i\|_\psi^2. \tag{11.A.45}$$

Eq. (11.A.42) follows from the convention that $P_{i-1}\cdots P_1 = \mathbb{1}$ for $i = 1$. Eq. (11.A.43) is a consequence of the inequality $2xy \leq cx^2 + c^{-1}y^2$, holding for $x, y \in \mathbb{R}$ and $c > 0$. Finally, (11.A.44) is obtained by using Lemma 11.27.

Appendix 11.B The $\alpha \to 1$ Limit of the Sandwiched Rényi Mutual Information of a Channel

In this appendix, we show that

$$\lim_{\alpha\to 1^-} \overline{I}_\alpha(\mathcal{N}) = \lim_{\alpha\to 1^+} \widetilde{I}_\alpha(\mathcal{N}) = I(\mathcal{N}), \tag{11.B.1}$$

where we recall that

$$\overline{I}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} D_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)), \tag{11.B.2}$$

$$\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_B} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B). \tag{11.B.3}$$

In these definitions, ψ_{RA} is a pure state, with the dimension of R equal to the dimension of A, and in the definition of $\widetilde{I}_\alpha(\mathcal{N})$ the infimum is over states σ_B.

All of the arguments presented here are similar to those in Appendix 10.A, which we refer to for additional details.

As a consequence of the fact that $\overline{I}_\alpha(\mathcal{N})$ increases monotonically with α (see Proposition 7.23), as well as the fact that $\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ (see Proposition 7.22), we find that

$$\lim_{\alpha\to 1^-} \overline{I}_\alpha(\mathcal{N}) = \sup_{\alpha\in(0,1)} \sup_{\psi_{RA}} D_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)) \tag{11.B.4}$$

$$= \sup_{\psi_{RA}} \sup_{\alpha\in(0,1)} D_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)) \tag{11.B.5}$$

$$= \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)) \tag{11.B.6}$$

$$= I(\mathcal{N}), \tag{11.B.7}$$

as required.

Similarly, for the sandwiched Rényi mutual information, we use the fact that it increases monotonically with α (see Proposition 7.31), along with the fact that $\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ (see Proposition 7.30), to obtain

$$\lim_{\alpha\to 1^+} \widetilde{I}_\alpha(\mathcal{N}) = \inf_{\alpha\in(1,\infty)} \sup_{\psi_{RA}} \inf_{\sigma_B} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B) \tag{11.B.8}$$

$$= \sup_{\psi_{RA}} \inf_{\alpha\in(1,\infty)} \inf_{\sigma_B} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B) \tag{11.B.9}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_B} \inf_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B) \tag{11.B.10}$$

$$= \sup_{\psi_{RA}} \inf_{\sigma_B} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B) \tag{11.B.11}$$

$$= \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \mathcal{N}_{A\to B}(\psi_A)) \tag{11.B.12}$$

$$= I(\mathcal{N}). \tag{11.B.13}$$

To obtain the second equality, we made use of the minimax theorem in Theorem 2.25 to exchange $\inf_{\alpha\in(1,\infty)}$ and $\sup_{\psi_{RA}}$. Specifically, we applied that theorem to the function

$$(\alpha, \psi_{RA}) \mapsto \inf_{\sigma_B} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R \otimes \sigma_B), \tag{11.B.14}$$

which is monotonically increasing in the first argument and continuous in the second argument.

Appendix 11.C Achievability from a Different Point of View

Here we show that the mutual information $I(\mathcal{N})$ is an achievable rate based on the alternate definition given in Appendix A. According to that definition, a rate $R \in \mathbb{R}^+$ is an achievable rate for entanglement-asssisted classical communication over $\mathcal{N}$ if there exists a sequence $\{(n, |\mathcal{M}_n|, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols over n uses of $\mathcal{N}$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log_2 |\mathcal{M}_n| \geq R \quad \text{and} \quad \lim_{n\to\infty} \varepsilon_n = 0. \tag{11.C.1}$$

To start, let us recall Corollary 11.17, which states that for all $\varepsilon \in (0,1]$, $n \in \mathbb{N}$, and $\alpha \in (0,1)$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ protocol satisfying

$$\frac{1}{n} \log_2 |\mathcal{M}| \geq \overline{I}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)} \log_2\left(\frac{2}{\varepsilon}\right) - \frac{3}{n}. \tag{11.C.2}$$

Fix constants δ_1, δ_2 satisfying $0 < \delta_2 < \delta_1 < 1$. Pick $\alpha_n \in (0,1)$ and $\varepsilon_n \in (0,1]$ as follows:

$$\alpha_n \coloneqq 1 - n^{-(1-\delta_1)}, \qquad \varepsilon_n \coloneqq 2^{-n^{\delta_2}}. \tag{11.C.3}$$

Plugging in to (11.C.2), we find that there exists a sequence of $\{(n, |\mathcal{M}_n|, \varepsilon_n)\}_{n\in\mathbb{N}}$ protocols satisfying

$$\frac{1}{n} \log_2 |\mathcal{M}_n| \geq \overline{I}_{\alpha_n}(\mathcal{N}) - \frac{1}{n(1-\alpha_n)} \log_2\left(\frac{2}{\varepsilon_n}\right) - \frac{3}{n} \tag{11.C.4}$$

$$= \overline{I}_{\alpha_n}(\mathcal{N}) - \frac{1 + n^{\delta_2}}{n^{\delta_1}} - \frac{3}{n} \tag{11.C.5}$$

$$= \overline{I}_{\alpha_n}(\mathcal{N}) - \frac{1}{n^{\delta_1}} - \frac{1}{n^{\delta_1-\delta_2}} - \frac{3}{n}. \tag{11.C.6}$$

Now taking the limit $n \to \infty$, we find that

$$\liminf_{n\to\infty} \frac{1}{n} \log_2 |\mathcal{M}_n| \geq \liminf_{n\to\infty} \left[\overline{I}_{\alpha_n}(\mathcal{N}) - \frac{1}{n^{\delta_1}} - \frac{1}{n^{\delta_1-\delta_2}} - \frac{3}{n}\right] \tag{11.C.7}$$

$$= I(\mathcal{N}), \tag{11.C.8}$$

and $\lim_{n\to\infty} \varepsilon_n = 0$. The equality above follows because $\lim_{n\to\infty} \alpha_n = 1$ and $\lim_{\alpha\to 1} \overline{I}_\alpha(\mathcal{N}) = I(\mathcal{N})$ (see Appendix 11.B for a proof). Thus, it follows that the mutual information rate $I(\mathcal{N})$ is achievable according to the alternate definition given in Appendix A.

In the approach detailed above, the error probability decays subexponentially to zero (i.e., slower than an exponential decay) and the rate increases to $I(\mathcal{N})$ with increasing n. If we would like to have exponential decay of the error probability, then we can instead fix the rate R to be a constant satisfying $R < I(\mathcal{N})$ and reconsider the analysis. Rearranging the inequality in (11.2.22) in order to get a bound on

ε_n, we find that for all $\alpha \in (0,1)$, there exists a sequence of $\{(n, |\mathcal{M}_n|, \varepsilon_n)\}_{n\in\mathbb{N}}$ protocols satisfying

$$\varepsilon_n \le 2 \cdot 2^{-n(1-\alpha)\left(\overline{I}_\alpha(\mathcal{N})-R-\frac{3}{n}\right)}. \tag{11.C.9}$$

Since $R < I(\mathcal{N})$, $\lim_{\alpha\to 1} \overline{I}_\alpha(\mathcal{N}) = I(\mathcal{N})$, and since $\overline{I}_\alpha(\mathcal{N})$ is monotonically increasing in α (this follows from Proposition 7.31), there exists an $\alpha^* < 1$ such that $\overline{I}_{\alpha^*}(\mathcal{N}) > R$. Applying the bound in (11.C.9) to this value of α, we find that

$$\varepsilon_n \le 2 \cdot 2^{-n(1-\alpha^*)\left(\overline{I}_{\alpha^*}(\mathcal{N})-R-\frac{3}{n}\right)}. \tag{11.C.10}$$

Then, taking the limit $n \to \infty$ on both sides of this inequality, we conclude that $\lim_{n\to\infty} \varepsilon_n = 0$ exponentially fast. Thus, by choosing R as a constant satisfying $R < I(\mathcal{N})$ it follows that there exists a sequence of $\{(n, 2^{nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ protocols such that the error probability ε_n decays exponentially fast to zero.

Appendix 11.D Proof of Lemma 11.20

We start by writing the definition of $\widetilde{I}_\alpha(A;B)_\rho$ as

$$\widetilde{I}_\alpha(A;B)_\rho = \inf_{\sigma_B} \frac{\alpha}{\alpha-1} \log_2 \left\| \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) \rho_{AB} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) \right\|_\alpha \tag{11.D.1}$$

$$= \frac{\alpha}{\alpha-1} \log_2 \inf_{\sigma_B} \left\| \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) \rho_{AB} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) \right\|_\alpha , \tag{11.D.2}$$

where $\alpha > 1$ and the optimization is over states σ_B. Then, for every purification $|\psi\rangle_{ABC}$ of ρ_{AB}, we have

$$\left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) \rho_{AB} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)$$
$$= \mathrm{Tr}_C\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) |\psi\rangle\!\langle\psi|_{ABC} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]. \tag{11.D.3}$$

Now, the operator inside Tr_C on the last line in the equation above is rank one, which means that

$$\mathrm{Tr}_C\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) |\psi\rangle\!\langle\psi|_{ABC} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right] \quad \text{and}$$
$$\mathrm{Tr}_{AB}\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right) |\psi\rangle\!\langle\psi|_{ABC} \left(\rho_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]$$

have the same non-zero eigenvalues. This means that their Schatten norms are equal, so that

$$\begin{aligned}
&\widetilde{I}_\alpha(A;B)_\rho \\
&= \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\left\|\mathrm{Tr}_C\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]\right\|_\alpha &\text{(11.D.4)}\\
&= \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\left\|\mathrm{Tr}_{AB}\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]\right\|_\alpha &\text{(11.D.5)}
\end{aligned}$$

Now, we use the variational characterization of the Schatten norm in (2.2.97), which states that for every operator X,

$$\|X\|_p = \sup_{\|Y\|_{p'}=1}\left|\mathrm{Tr}[Y^\dagger X]\right| \tag{11.D.6}$$

for all $1 \le p \le \infty$, where p' satisfies $\frac{1}{p}+\frac{1}{p'}=1$. Note that if X is positive semi-definite, then we can restrict the optimization to positive semi-definite operators Y. Using (11.D.6), with $p=\alpha$ and $p'=\frac{\alpha}{\alpha-1}$, on the expression in (11.D.5), and since the argument of the norm in that expression is positive semi-definite, we can optimize over positive semi-definite operators τ_C to obtain

$$\begin{aligned}
&\widetilde{I}_\alpha(A;B)_\rho \\
&= \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\sup_{\tau_C}\mathrm{Tr}\left[\tau_C^{\frac{\alpha-1}{\alpha}}\mathrm{Tr}_{AB}\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\times\right.\right.\\
&\qquad\qquad\left.\left.|\psi\rangle\!\langle\psi|_{ABC}\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]\right] &\text{(11.D.7)}\\
&= \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\sup_{\tau_C}\mathrm{Tr}\left[\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{\frac{\alpha-1}{2\alpha}}\right)\right.\\
&\qquad\qquad\left.\times|\psi\rangle\!\langle\psi|_{ABC}\left(\rho_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{\frac{\alpha-1}{2\alpha}}\right)\right] &\text{(11.D.8)}\\
&= \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\sup_{\tau_C}\mathrm{Tr}\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\otimes\tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right] &\text{(11.D.9)}\\
&= \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_C}\inf_{\sigma_B}\mathrm{Tr}\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\otimes\tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right], &\text{(11.D.10)}
\end{aligned}$$

where the last line follows by applying Sion's minimax theorem (Theorem 2.24) to the function

$$(\tau_C,\sigma_B)\mapsto \mathrm{Tr}\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\otimes\tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right], \tag{11.D.11}$$

which is convex in the first argument because $\sigma_B \mapsto \sigma_B^{\frac{1-\alpha}{\alpha}}$ is operator convex and concave in the second argument because $\tau_C \mapsto \tau_C^{\frac{\alpha-1}{\alpha}}$ is operator concave.

Finally, we use Proposition 2.8, which is that

$$\|X\|_p = \inf_{\substack{Y\geq 0,\\ \|Y\|_{p'}=1}} \operatorname{Tr}[XY] \tag{11.D.12}$$

for all $0 < p < 1$, where $\frac{1}{p} + \frac{1}{p'} = 1$. Applying this to (11.D.10) with $p' = \frac{\alpha}{1-\alpha}$, so that $p = \frac{\alpha}{2\alpha-1}$, we conclude that

$$\begin{aligned}
&\widetilde{I}_\alpha(A;B)_\rho \\
&= \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_C}\inf_{\sigma_B} \operatorname{Tr}\!\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{\alpha}} \otimes \tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right] &(11.D.13)\\
&= \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_C}\inf_{\sigma_B} \operatorname{Tr}\!\left[\sigma_B^{\frac{1-\alpha}{\alpha}} \operatorname{Tr}_{AC}\!\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}} \otimes \tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right]\right] &(11.D.14)\\
&= \frac{\alpha}{\alpha-1}\log_2 \sup_{\tau_C}\left\|\operatorname{Tr}_{AC}\!\left[\left(\rho_A^{\frac{1-\alpha}{\alpha}} \otimes \tau_C^{\frac{\alpha-1}{\alpha}}\right)|\psi\rangle\!\langle\psi|_{ABC}\right]\right\|_{\frac{\alpha}{2\alpha-1}}, &(11.D.15)
\end{aligned}$$

the last line of which is (11.2.66), as required.

To prove (11.2.67), we use the fact that the definition of the sandwiched Rényi mutual information of a bipartite state can be written as in (11.D.2), i.e.,

$$\widetilde{I}_\alpha(R;B)_\rho = \frac{\alpha}{\alpha-1}\log_2 \inf_{\sigma_B}\left\|\left(\rho_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\rho_{RB}\left(\rho_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha, \tag{11.D.16}$$

which means that the definition in (11.2.82) can be written as

$$\begin{aligned}
&\widetilde{I}_\alpha(\mathcal{N}) \\
&= \frac{\alpha}{\alpha-1}\sup_{\psi_{RA}}\log_2 \inf_{\sigma_B}\left\|\left(\psi_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\mathcal{N}_{A\to B}(\psi_{RA})\left(\psi_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha.
\end{aligned} \tag{11.D.17}$$

Now, we use the fact mentioned in (2.2.38), which is that for every pure state ψ_{RA}, with the systems R and A having the same dimensions, there exists an operator X_R such that

$$|\psi\rangle_{RA} = (X_R \otimes \mathbb{1}_A)|\Gamma\rangle_{RA}, \tag{11.D.18}$$

and $\mathrm{Tr}[X_R^\dagger X_R] = 1$, with this latter equality following from (2.2.43). By taking a polar decomposition of X_R as $X_R = U_R\sqrt{\tau}_R$ for a unitary U_R and a state τ_R (see Theorem 2.3), we can then write

$$|\psi\rangle_{RA} = (U_R\sqrt{\tau_R} \otimes \mathbb{1}_A)|\Gamma\rangle_{RA}. \tag{11.D.19}$$

This implies that

$$\begin{aligned}\psi_R &= \mathrm{Tr}_B[\mathcal{N}_{A\to B}(\psi_{RA})] && (11.D.20)\\ &= \mathrm{Tr}_B[(U_R\sqrt{\tau_R} \otimes \mathbb{1}_B)\mathcal{N}_{A\to B}(\Gamma_{RA})(\sqrt{\tau_R}U_R^\dagger \otimes \mathbb{1}_B)] && (11.D.21)\\ &= U_R\tau_R U_R^\dagger, && (11.D.22)\end{aligned}$$

where the last equality follows because $\mathcal{N}$ is trace preserving and $\mathrm{Tr}_A[|\Gamma\rangle\langle\Gamma|_{RA}] = \mathbb{1}_R$. Using this, we find that

$$\begin{aligned}&\left(\psi_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\mathcal{N}_{A\to B}(\psi_{RA})\left(\psi_R^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\\ &= \left(U_R\tau_R^{\frac{1-\alpha}{2\alpha}}U_R^\dagger \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)U_R\sqrt{\tau_R}\mathcal{N}_{A\to B}(\Gamma_{RA})\sqrt{\tau_R}U_R^\dagger\left(U_R\tau_R^{\frac{1-\alpha}{2\alpha}}U_R^\dagger \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\\ &= \left(U_R\tau_R^{\frac{1}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\mathcal{N}_{A\to B}(\Gamma_{RA})\left(\tau_R^{\frac{1}{2\alpha}}U_R^\dagger \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right). && (11.D.23)\end{aligned}$$

Therefore, by exploiting unitary invariance of the α-Schatten norm, we can write $\widetilde{I}_\alpha(\mathcal{N})$ as

$$\begin{aligned}&\widetilde{I}_\alpha(\mathcal{N})\\ &= \frac{\alpha}{\alpha-1}\sup_{\rho_R}\log_2\inf_{\sigma_B}\left\|\left(\rho_R^{\frac{1}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\mathcal{N}_{A\to B}(\Gamma_{RA})\left(\rho_R^{\frac{1}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha && (11.D.24)\\ &= \frac{\alpha}{\alpha-1}\sup_{\rho_R}\log_2\inf_{\sigma_B}\left\|\mathcal{N}_{A\to B}(\Gamma_{RA})^{\frac{1}{2}}\left(\rho_R^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{\alpha}}\right)\mathcal{N}_{A\to B}(\Gamma_{RA})^{\frac{1}{2}}\right\|_\alpha. && (11.D.25)\end{aligned}$$

Now, the function

$$(\rho_R, \sigma_B) \mapsto \left\|\mathcal{N}_{A\to B}(\Gamma_{RA})^{\frac{1}{2}}\left(\rho_R^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{\alpha}}\right)\mathcal{N}_{A\to B}(\Gamma_{RA})^{\frac{1}{2}}\right\|_\alpha \tag{11.D.26}$$

is concave in the first argument (this follows from Lemma 11.29 in Appendix 11.F below) and convex in the second argument (this follows from the operator convexity of $\sigma_B \mapsto \sigma_B^{\frac{1-\alpha}{\alpha}}$ for $\alpha > 1$ and convexity of the Schatten norm). Thus, by the Sion minimax theorem (Theorem 2.24), we can exchange $\sup_{\rho_R}$ and $\inf_{\sigma_B}$. Also, we define the completely positive map $\mathcal{S}^{(\alpha)}_{\sigma_B}$ by $\mathcal{S}^{(\alpha)}_{\sigma_B}(\cdot) \coloneqq \sigma_B^{\frac{1-\alpha}{2\alpha}}(\cdot)\sigma_B^{\frac{1-\alpha}{2\alpha}}$. We can then further rewrite $\widetilde{I}_\alpha(\mathcal{N})$ as

$$\begin{aligned}
&\widetilde{I}_\alpha(\mathcal{N}) \\
&= \frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2\sup_{\rho_R}\left\|(\mathcal{S}^{(\alpha)}_{\sigma_B}\circ\mathcal{N}_{A\to B})\left(\rho_R^{\frac{1}{2\alpha}}|\Gamma\rangle\langle\Gamma|_{RA}\rho_R^{\frac{1}{2\alpha}}\right)\right\|_\alpha && (11.D.27)\\
&= \frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2\left\|\mathcal{S}^{(\alpha)}_{\sigma_B}\circ\mathcal{N}\right\|_{\text{CB},\,1\to\alpha}, && (11.D.28)
\end{aligned}$$

where, to arrive at the last line, we used the definition in (11.2.68). Also, consider that the optimum in (11.2.68) is achieved when $\operatorname{Tr}[Y_R] = 1$. Therefore,

$$\widetilde{I}_\alpha(\mathcal{N}) = \frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2\left\|\mathcal{S}^{(\alpha)}_{\sigma_B}\circ\mathcal{N}\right\|_{\text{CB},\,1\to\alpha}, \tag{11.D.29}$$

as required.

Appendix 11.E Alternate Expression for the $1 \to \alpha$ CB Norm

In this section, we show that

$$\|\mathcal{M}\|_{\text{CB},1\to\alpha} = \sup_{Y_{RA}>0}\frac{\|\mathcal{M}_{A\to B}(Y_{RA})\|_\alpha}{\|\operatorname{Tr}_A[Y_{RA}]\|_\alpha} \tag{11.E.1}$$

for every completely positive map $\mathcal{M}$. We start with the expression in (11.2.68) and write it alternatively as follows:

$$\begin{aligned}
\|\mathcal{M}\|_{\text{CB},1\to\alpha} &= \sup_{\substack{Y_R>0,\\ \operatorname{Tr}[Y_R]\le 1}}\left\|\mathcal{M}_{A\to B}\left(Y_R^{\frac{1}{2\alpha}}|\Gamma\rangle\langle\Gamma|_{RA}Y_R^{\frac{1}{2\alpha}}\right)\right\|_\alpha && (11.E.2)\\
&= \sup_{\substack{Y_R>0,\\ \|Y_R\|_\alpha\le 1}}\left\|\mathcal{M}_{A\to B}\left(Y_R^{\frac{1}{2}}|\Gamma\rangle\langle\Gamma|_{RA}Y_R^{\frac{1}{2}}\right)\right\|_\alpha && (11.E.3)
\end{aligned}$$

$$= \sup_{Y_R>0} \frac{\left\|\mathcal{M}_{A\to B}\left(Y_R^{\frac{1}{2}}|\Gamma\rangle\langle\Gamma|_{RA}Y_R^{\frac{1}{2}}\right)\right\|_\alpha}{\|Y_R\|_\alpha}. \tag{11.E.4}$$

Now, we use the fact that there is a one-to-one correspondence between the operators Y_R and the vectors

$$|\Gamma^Y\rangle_{RA} \coloneqq (Y_R^{\frac{1}{2}} \otimes \mathbb{1}_A)|\Gamma\rangle_{RA}. \tag{11.E.5}$$

This allows us to rewrite the optimization in (11.E.4) in terms of such vectors. Then, by employing isometric invariance of the norms with respect to an isometry acting on the reference system R, we can restrict the optimization to arbitrary vectors $|\psi\rangle_{RA}$. Therefore, we have that

$$\|\mathcal{M}\|_{\mathrm{CB},1\to\alpha} = \sup_{\psi_{RA}} \frac{\|\mathcal{M}_{A\to B}(\psi_{RA})\|_\alpha}{\|\mathrm{Tr}_A[\psi_{RA}]\|_\alpha}, \tag{11.E.6}$$

where $\psi_{RA} \equiv |\psi\rangle\langle\psi|_{RA}$. Since the optimization in (11.E.6) is over a subset of the positive semi-definite operators Y_{RA} (and by approximation), we conclude the inequality

$$\|\mathcal{M}\|_{\mathrm{CB},1\to\alpha} \leq \sup_{Y_{RA}>0} \frac{\|\mathcal{M}_{A\to B}(Y_{RA})\|_\alpha}{\|\mathrm{Tr}_A[Y_{RA}]\|_\alpha}. \tag{11.E.7}$$

It remains to show the opposite inequality. Consider a vector $|\phi\rangle_{SRA}$ that purifies $Y_{RA} > 0$, in the sense that $\mathrm{Tr}_S[\phi_{SRA}] = Y_{RA}$. Then we have that

$$\frac{\|\mathcal{M}_{A\to B}(Y_{RA})\|_\alpha}{\|\mathrm{Tr}_A[Y_{RA}]\|_\alpha} = \frac{\|(\mathcal{M}_{A\to B} \otimes \mathrm{Tr}_S)(\phi_{SRA})\|_\alpha}{\|\mathrm{Tr}_{SA}[\phi_{SRA}]\|_\alpha} \tag{11.E.8}$$

$$\leq \sup_{|\phi\rangle_{SRA}} \frac{\|(\mathcal{M}_{A\to B} \otimes \mathrm{Tr}_S)(\phi_{SRA})\|_\alpha}{\|\mathrm{Tr}_{SA}[\phi_{SRA}]\|_\alpha} \tag{11.E.9}$$

$$= \|\mathcal{M} \otimes \mathrm{Tr}\|_{\mathrm{CB},1\to\alpha} \tag{11.E.10}$$

$$= \|\mathcal{M}\|_{\mathrm{CB},1\to\alpha} \|\mathrm{Tr}\|_{\mathrm{CB},1\to\alpha} \tag{11.E.11}$$

$$= \|\mathcal{M}\|_{\mathrm{CB},1\to\alpha} \tag{11.E.12}$$

The third-to-last equality follows from (11.E.6). The second-to-last equality follows from (11.2.88), as shown in Appendix 11.F. The final inequality follows because $\|\mathrm{Tr}\|_{\mathrm{CB},1\to\alpha} = 1$, as can be readily verified.

Appendix 11.F Proof of the Multiplicativity of the $1 \to \alpha$ CB Norm

In this appendix, we prove the statement in (11.2.88), which is that

$$\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\text{CB},1\to\alpha} = \|\mathcal{M}_1\|_{\text{CB},1\to\alpha} \|\mathcal{M}_2\|_{\text{CB},1\to\alpha} \tag{11.F.1}$$

for every two completely positive maps $\mathcal{M}_1$ and $\mathcal{M}_2$ and all $\alpha > 1$. Recall from (11.2.68) that

$$\|\mathcal{M}\|_{\text{CB},1\to\alpha} = \sup_{\substack{Y_R>0,\\ \text{Tr}[Y_R]\le 1}} \left\| Y_R^{\frac{1}{2\alpha}} \Gamma_{RB}^{\mathcal{M}} Y_R^{\frac{1}{2\alpha}} \right\|_\alpha , \tag{11.F.2}$$

where $\Gamma_{RB}^{\mathcal{M}} := \mathcal{M}_{A\to B}(\Gamma_{RA})$ is the Choi representation of $\mathcal{M}$, and the dimension of R is the same as the dimension of A. For a completely positive map $\mathcal{P}_{C\to D}$, let us also define

$$\|\mathcal{P}\|_{\alpha\to\alpha} := \sup_{Z_C>0} \frac{\|\mathcal{P}_{C\to D}(Z_C)\|_\alpha}{\|Z_C\|_\alpha}. \tag{11.F.3}$$

Note that

$$\|\mathcal{P}\|_{\alpha\to\alpha} = \sup_{Z_C>0} \frac{\|\mathcal{P}_{C\to D}(Z_C)\|_\alpha}{\|Z_C\|_\alpha} \tag{11.F.4}$$

$$= \sup_{\substack{Z_C>0,\\ \|Z_C\|_\alpha\le 1}} \|\mathcal{P}_{C\to D}(Z_C)\|_\alpha \tag{11.F.5}$$

$$= \sup_{\substack{Y_C>0,\\ \text{Tr}[Y_C]\le 1}} \left\|\mathcal{P}_{C\to D}(Y_C^{\frac{1}{\alpha}})\right\|_\alpha , \tag{11.F.6}$$

where the last equality follows from the substitution $Y_C = Z_C^\alpha$ so that $\text{Tr}[Y_C] = \text{Tr}[Z_C^\alpha] = \|Z_C\|_\alpha^\alpha$.

Now, it immediately follows that

$$\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\text{CB},1\to\alpha} \ge \|\mathcal{M}_1\|_{\text{CB},1\to\alpha} \|\mathcal{M}_2\|_{\text{CB},1\to\alpha} . \tag{11.F.7}$$

Indeed, due to the fact that the Choi representation of $\mathcal{M}_1 \otimes \mathcal{M}_2$ has a tensor-product form (see (4.2.17)), we can restrict the optimization in the definition of the norm $\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\text{CB},1\to\alpha}$ to tensor-product operators $Y_{R_1} \otimes Y_{R_2}$ to obtain

$$\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\text{CB},1\to\alpha}$$

$$= \sup_{\substack{Y_{R_1R_2}>0,\\ \mathrm{Tr}[Y_{R_1R_2}]\leq 1}} \left\| Y_{R_1R_2}^{\frac{1}{2\alpha}} (\Gamma^{\mathcal{M}_1}_{R_1B_1} \otimes \Gamma^{\mathcal{M}_2}_{R_2B_2}) Y_{R_1R_2}^{\frac{1}{2\alpha}} \right\|_\alpha \tag{11.F.8}$$

$$\geq \sup_{\substack{Y_{R_1}>0, Y_{R_2}>0,\\ \mathrm{Tr}[Y_{R_1}]\leq 1, \mathrm{Tr}[Y_{R_2}]\leq 1}} \left\| (Y_{R_1}^{\frac{1}{2\alpha}} \otimes Y_{R_2}^{\frac{1}{2\alpha}})(\Gamma^{\mathcal{M}_1}_{R_1B_1} \otimes \Gamma^{\mathcal{M}_2}_{R_2B_2})(Y_{R_1}^{\frac{1}{2\alpha}} \otimes Y_{R_2}^{\frac{1}{2\alpha}}) \right\|_\alpha \tag{11.F.9}$$

$$= \sup_{\substack{Y_{R_1}>0, Y_{R_2}>0,\\ \mathrm{Tr}[Y_{R_1}]\leq 1, \mathrm{Tr}[Y_{R_2}]\leq 1}} \left\| Y_{R_1}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_1}_{R_1B_1} Y_{R_1}^{\frac{1}{2\alpha}} \otimes Y_{R_2}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_2}_{R_2B_2} Y_{R_2}^{\frac{1}{2\alpha}} \right\|_\alpha \tag{11.F.10}$$

$$= \sup_{\substack{Y_{R_1}>0, Y_{R_2}>0,\\ \mathrm{Tr}[Y_{R_1}]\leq 1, \mathrm{Tr}[Y_{R_2}]\leq 1}} \left\| Y_{R_1}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_1}_{R_1B_1} Y_{R_1}^{\frac{1}{2\alpha}} \right\|_\alpha \left\| Y_{R_2}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_2}_{R_2B_2} Y_{R_2}^{\frac{1}{2\alpha}} \right\|_\alpha \tag{11.F.11}$$

$$= \sup_{\substack{Y_{R_1}>0,\\ \mathrm{Tr}[Y_{R_1}]\leq 1}} \left\| Y_{R_1}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_1}_{R_1B_1} Y_{R_1}^{\frac{1}{2\alpha}} \right\|_\alpha \sup_{\substack{Y_2>0,\\ \mathrm{Tr}[Y_{R_2}]\leq 1}} \left\| Y_{R_2}^{\frac{1}{2\alpha}} \Gamma^{\mathcal{M}_2}_{R_2B_2} Y_{R_2}^{\frac{1}{2\alpha}} \right\|_\alpha \tag{11.F.12}$$

$$= \|\mathcal{M}_1\|_{\mathrm{CB},1\to\alpha} \|\mathcal{M}_2\|_{\mathrm{CB},1\to\alpha} . \tag{11.F.13}$$

Now we establish the opposite inequality. Let $\mathcal{U}^{\mathcal{M}}_{A\to BE}$ be a linear map that extends $\mathcal{M}_{A\to B}$, in the sense that there is a linear operator $U^{\mathcal{M}}_{A\to BE}$ such that

$$\mathcal{U}^{\mathcal{M}}_{A\to BE}(Y_A) = U^{\mathcal{M}}_{A\to BE} Y_A (U^{\mathcal{M}}_{A\to BE})^\dagger, \tag{11.F.14}$$

$$\mathrm{Tr}_E[\mathcal{U}^{\mathcal{M}}_{A\to BE}(Y_A)] = \mathcal{M}_{A\to B}(Y_A). \tag{11.F.15}$$

Due to the fact that $Y_R^{\frac{1}{2\alpha}} \mathcal{U}^{\mathcal{M}}_{A\to BE}(\Gamma_{RA}) Y_R^{\frac{1}{2\alpha}}$ is a rank-one operator, and from an application of a generalization of the Schmidt decomposition (Theorem 2.2), the following operators have the same non-zero eigenvalues:

$$\mathrm{Tr}_E[Y_R^{\frac{1}{2\alpha}} \mathcal{U}^{\mathcal{M}}_{A\to BE}(\Gamma_{RA}) Y_R^{\frac{1}{2\alpha}}] = Y_R^{\frac{1}{2\alpha}} \mathcal{M}_{A\to B}(\Gamma_{RA}) Y_R^{\frac{1}{2\alpha}} \tag{11.F.16}$$

and

$$\mathrm{Tr}_{RB}[Y_R^{\frac{1}{2\alpha}} \mathcal{U}^{\mathcal{M}}_{A\to BE}(\Gamma_{RA}) Y_R^{\frac{1}{2\alpha}}] = \mathrm{Tr}_R[Y_R^{\frac{1}{2\alpha}} \mathcal{M}^c_{A\to B}(\Gamma_{RA}) Y_R^{\frac{1}{2\alpha}}] \tag{11.F.17}$$

$$= \mathcal{M}^c_{A\to E}((Y_A^{\mathsf{T}})^{\frac{1}{\alpha}}), \tag{11.F.18}$$

where $\mathcal{M}^c_{A\to E} = \mathrm{Tr}_B \circ \mathcal{U}^{\mathcal{M}}_{A\to BE}$ denotes the complementary map and the last equality follows by applying the transpose trick in (2.2.42) and the fact that $\mathrm{Tr}_R[\Gamma_{RA}] = \mathbb{1}_A$.

Then we find that

$$\|\mathcal{M}\|_{\text{CB},1\to\alpha} = \sup_{\substack{Y_A>0,\\ \text{Tr}[Y_A]\leq 1}} \left\|\mathcal{M}^c_{A\to E}((Y_A^{\mathsf{T}})^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.19}$$

$$= \sup_{\substack{Y_A>0,\\ \text{Tr}[Y_A]\leq 1}} \left\|\mathcal{M}^c_{A\to E}(Y_A^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.20}$$

$$= \sup_{Y_A>0} \frac{\left\|\mathcal{M}^c_{A\to E}(Y_A)\right\|_\alpha}{\|Y_A\|_\alpha} \tag{11.F.21}$$

$$= \|\mathcal{M}^c\|_{\alpha\to\alpha}, \tag{11.F.22}$$

where the second equality follows from (11.F.4)–(11.F.6). So we have that

$$\|\mathcal{M}_{A\to B}\|_{\text{CB},1\to\alpha} = \left\|\mathcal{M}^c_{A\to E}\right\|_{\alpha\to\alpha}. \tag{11.F.23}$$

Finally, for an arbitrary operator $Y_{A_1A_2}$ satisfying $Y_{A_1A_2} > 0$, and setting $X_{A_1E_2} \coloneqq \left(\mathcal{M}_2^c\right)_{A_2\to E_2}(Y_{A_1A_2})$, we can write

$$\left((\mathcal{M}_1^c)_{A_1\to E_1} \otimes (\mathcal{M}_2^c)_{A_2\to E_2}\right)(Y_{A_1A_2})$$
$$= (\mathcal{M}_1^c)_{A_1\to E_1}\left((\mathcal{M}_2^c)_{A_2\to E_2}(Y_{A_1A_2})\right) \tag{11.F.24}$$
$$= (\mathcal{M}_1^c)_{A_1\to E_1}(X_{A_1E_2}). \tag{11.F.25}$$

Then, multiplying and dividing by $\left\|X_{A_1E_2}\right\|_\alpha = \left\|(\mathcal{M}_2^c)_{A_2\to E_2}(Y_{A_1A_2})\right\|_\alpha$ gives

$$\frac{\left\|\left(\left(\mathcal{M}_1^c\right)_{A_1\to E_1} \otimes \left(\mathcal{M}_2^c\right)_{A_2\to E_2}\right)(Y_{A_1A_2})\right\|_\alpha}{\left\|Y_{A_1A_2}\right\|_\alpha}$$
$$= \frac{\left\|\left(\mathcal{M}_1^c\right)_{A_1\to E_1}(X_{A_1E_2})\right\|_\alpha}{\left\|X_{A_1E_2}\right\|_\alpha} \frac{\left\|\left(\mathcal{M}_2^c\right)_{A_2\to E_2}(Y_{A_1A_2})\right\|_\alpha}{\left\|Y_{A_1A_2}\right\|_\alpha} \tag{11.F.26}$$
$$\leq \left(\sup_{X_{A_1E_2}>0} \frac{\left\|\left(\mathcal{M}_1^c\right)_{A_1\to E_1}(X_{A_1E_2})\right\|_\alpha}{\left\|X_{A_1E_2}\right\|_\alpha}\right)$$
$$\times \left(\sup_{Y_{A_1A_2}>0} \frac{\left\|\left(\mathcal{M}_2^c\right)_{A_2\to E_2}(Y_{A_1A_2})\right\|_\alpha}{\left\|Y_{A_1A_2}\right\|_\alpha}\right) \tag{11.F.27}$$

$$= \left\| \left(\mathcal{M}_1^c\right)_{A_1 \to E_1} \otimes \mathrm{id}_{E_2} \right\|_{\alpha \to \alpha} \left\| \mathrm{id}_{A_1} \otimes \left(\mathcal{M}_2^c\right)_{A_2 \to E_2} \right\|_{\alpha \to \alpha} \tag{11.F.28}$$

$$= \left\| \left(\mathcal{M}_1^c\right)_{A_1 \to E_1} \right\|_{\alpha \to \alpha} \left\| \left(\mathcal{M}_2^c\right)_{A_2 \to E_2} \right\|_{\alpha \to \alpha} \tag{11.F.29}$$

$$= \|\mathcal{M}_1\|_{\mathrm{CB},1 \to \alpha} \|\mathcal{M}_2\|_{\mathrm{CB},1 \to \alpha} . \tag{11.F.30}$$

The third equality follows from Lemma 11.28 below. The final equality holds by (11.F.23). Since $Y_{A_1 A_2}$ is arbitrary, we find that

$$\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\mathrm{CB},1 \to \alpha} = \left\| \mathcal{M}_1^c \otimes \mathcal{M}_2^c \right\|_{\alpha \to \alpha} \tag{11.F.31}$$

$$= \sup_{Y_{A_1 A_2} > 0} \frac{\left\| (\mathcal{M}_1^c \otimes \mathcal{M}_2^c)(Y_{A_1 A_2}) \right\|_\alpha}{\left\| Y_{A_1 A_2} \right\|_\alpha} \tag{11.F.32}$$

$$\leq \|\mathcal{M}_1\|_{\mathrm{CB},1 \to \alpha} \|\mathcal{M}_2\|_{\mathrm{CB},1 \to \alpha} . \tag{11.F.33}$$

So we have that $\|\mathcal{M}_1 \otimes \mathcal{M}_2\|_{\mathrm{CB},1 \to \alpha} = \|\mathcal{M}_1\|_{\mathrm{CB},1 \to \alpha} \|\mathcal{M}_2\|_{\mathrm{CB},1 \to \alpha}$.

Lemma 11.28

Let $\mathcal{M}$ be a completely positive map. Then, for id an arbitrary identity map, the following equality holds,

$$\|\mathrm{id} \otimes \mathcal{M}\|_{\alpha \to \alpha} = \|\mathcal{M}\|_{\alpha \to \alpha} . \tag{11.F.34}$$

Proof: The inequality $\|\mathrm{id} \otimes \mathcal{M}\|_{\alpha \to \alpha} \geq \|\mathcal{M}\|_{\alpha \to \alpha}$ immediately follows by restricting the optimization on the left-hand side of the inequality. So we now establish the non-trivial inequality $\|\mathrm{id} \otimes \mathcal{M}\|_{\alpha \to \alpha} \leq \|\mathcal{M}\|_{\alpha \to \alpha}$. Letting the identity map act on a reference system R, consider from (11.F.6) that

$$\|\mathrm{id} \otimes \mathcal{M}\|_{\alpha \to \alpha} = \sup_{\substack{Y_{RA} > 0, \\ \mathrm{Tr}[Y_{RA}] \leq 1}} \left\| \mathcal{M}_{A \to B}\left(Y_{RA}^{\frac{1}{\alpha}}\right) \right\|_\alpha . \tag{11.F.35}$$

Let $\{V^i\}_{i=1}^{d_R^2}$ denote a set of Heisenberg–Weyl operators acting on the reference system R (see (3.2.47)), so that

$$\frac{1}{d_R^2} \sum_{i=1}^{d_R^2} V_R^i (\cdot) (V_R^i)^\dagger = \mathrm{Tr}[\cdot] \frac{\mathbb{1}}{d_R}. \tag{11.F.36}$$

Then, for an arbitrary $Y_{RA} > 0$ satisfying $\text{Tr}[Y_{RA}] \leq 1$, we use the unitary invariance of the Schatten norm to obtain

$$\left\|\mathcal{M}_{A\to B}(Y_{RA}^{\frac{1}{\alpha}})\right\|_\alpha = \frac{1}{d_R^2}\sum_{i=1}^{d_R^2}\left\|V_R^i\mathcal{M}_{A\to B}(Y_{RA}^{\frac{1}{\alpha}})(V_R^i)^\dagger\right\|_\alpha \tag{11.F.37}$$

$$= \frac{1}{d_R^2}\sum_{i=1}^{d_R^2}\left\|\mathcal{M}_{A\to B}((V_R^i Y_{RA}(V_R^i)^\dagger)^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.38}$$

$$\leq \left\|\mathcal{M}_{A\to B}\left(\left[\frac{1}{d_R^2}\sum_{i=1}^{d_R^2}V_R^i Y_{RA}(V_R^i)^\dagger\right]^{\frac{1}{\alpha}}\right)\right\|_\alpha \tag{11.F.39}$$

$$= \left\|\mathcal{M}_{A\to B}\left([\pi_R \otimes Y_A]^{\frac{1}{\alpha}}\right)\right\|_\alpha, \tag{11.F.40}$$

where the inequality follows from Lemma 11.29 below, which states that the function $X \mapsto \left\|\mathcal{M}(X^{\frac{1}{\alpha}})\right\|_\alpha$ is concave for all $\alpha > 1$. The last equality follows from (3.2.97), with $\pi_R = \frac{\mathbb{1}_R}{|R|}$ the maximally mixed state and $Y_A = \text{Tr}_R[Y_{RA}]$. Continuing, we find that

$$\left\|\mathcal{M}_{A\to B}\left([\pi_R \otimes Y_A]^{\frac{1}{\alpha}}\right)\right\|_\alpha = \left\|\mathcal{M}_{A\to B}\left(\pi_R^{\frac{1}{\alpha}} \otimes Y_A^{\frac{1}{\alpha}}\right)\right\|_\alpha \tag{11.F.41}$$

$$= \left\|\pi_R^{\frac{1}{\alpha}} \otimes \mathcal{M}_{A\to B}(Y_A^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.42}$$

$$= \left\|\pi_R^{\frac{1}{\alpha}}\right\|_\alpha \left\|\mathcal{M}_{A\to B}(Y_A^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.43}$$

$$= \left\|\mathcal{M}_{A\to B}(Y_A^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.44}$$

$$\leq \sup_{\substack{Y_A>0,\\ \text{Tr}[Y_A]\leq 1}} \left\|\mathcal{M}_{A\to B}(Y_A^{\frac{1}{\alpha}})\right\|_\alpha \tag{11.F.45}$$

$$= \|\mathcal{M}\|_{\alpha\to\alpha}. \tag{11.F.46}$$

Since the inequality holds for arbitrary $Y_{RA} > 0$ satisfying $\text{Tr}[Y_{RA}] \leq 1$, we find that

$$\|\text{id} \otimes \mathcal{M}\|_{\alpha\to\alpha} \leq \|\mathcal{M}\|_{\alpha\to\alpha}, \tag{11.F.47}$$

concluding the proof. ■

Lemma 11.29

Let X be a positive semi-definite operator, and let $\mathcal{M}$ be a completely positive map. For $\alpha > 1$, the following function is concave:

$$X \mapsto \left\| \mathcal{M}(X^{\frac{1}{\alpha}}) \right\|_\alpha . \tag{11.F.48}$$

PROOF: Since $\mathcal{M}$ is completely positive, it has a Kraus representation as

$$\mathcal{M}(Z) = \sum_i M_i Z M_i^\dagger . \tag{11.F.49}$$

From Proposition 2.8, consider that

$$\left\| \mathcal{M}(X^{\frac{1}{\alpha}}) \right\|_\alpha = \sup_{\substack{Y>0, \\ \|Y\|_{\frac{\alpha}{\alpha-1}} \leq 1}} \operatorname{Tr}\left[\mathcal{M}(X^{\frac{1}{\alpha}}) Y \right] \tag{11.F.50}$$

$$= \sup_{\substack{Y>0, \\ \operatorname{Tr}[Y] \leq 1}} \operatorname{Tr}\left[\mathcal{M}(X^{\frac{1}{\alpha}}) Y^{\frac{\alpha-1}{\alpha}} \right] \tag{11.F.51}$$

$$= \sup_{\substack{Y>0, \\ \operatorname{Tr}[Y] \leq 1}} \sum_i \operatorname{Tr}\left[M_i X^{\frac{1}{\alpha}} M_i^\dagger Y^{\frac{\alpha-1}{\alpha}} \right] . \tag{11.F.52}$$

The Lieb concavity theorem (see Theorem 11.30 below) is the statement that the following function is jointly concave with respect to positive semi-definite R and S for arbitrary $t \in (0, 1)$ and an arbitrary operator K:

$$(R, S) \mapsto \operatorname{Tr}[K R^t K^\dagger S^{1-t}]. \tag{11.F.53}$$

Let $X_0, X_1 \geq 0$ and let $Y_0, Y_1 > 0$ be such that $\operatorname{Tr}[Y_0], \operatorname{Tr}[Y_1] \leq 1$. Then for $\lambda \in (0, 1]$, and defining

$$X_\lambda \coloneqq \lambda X_0 + (1 - \lambda) X_1, \qquad Y_\lambda \coloneqq \lambda Y_0 + (1 - \lambda) Y_1, \tag{11.F.54}$$

we find that

$$\begin{aligned} &\lambda \operatorname{Tr}\left[\mathcal{M}(X_0^{\frac{1}{\alpha}}) Y_0^{\frac{\alpha-1}{\alpha}} \right] + (1 - \lambda) \operatorname{Tr}\left[\mathcal{M}(X_1^{\frac{1}{\alpha}}) Y_1^{\frac{\alpha-1}{\alpha}} \right] \\ &= \sum_i \lambda \operatorname{Tr}\left[M_i X_0^{\frac{1}{\alpha}} M_i^\dagger Y_0^{\frac{\alpha-1}{\alpha}} \right] + \sum_i (1 - \lambda) \operatorname{Tr}\left[M_i X_1^{\frac{1}{\alpha}} M_i^\dagger Y_1^{\frac{\alpha-1}{\alpha}} \right] \end{aligned} \tag{11.F.55}$$

$$= \sum_i \lambda \operatorname{Tr}\left[M_i X_0^{\frac{1}{\alpha}} M_i^\dagger Y_0^{\frac{\alpha-1}{\alpha}} \right] + (1-\lambda) \operatorname{Tr}\left[M_i X_1^{\frac{1}{\alpha}} M_i^\dagger Y_1^{\frac{\alpha-1}{\alpha}} \right] \tag{11.F.56}$$

$$\leq \sum_i \left(\operatorname{Tr}\left[M_i X_\lambda^{\frac{1}{\alpha}} M_i^\dagger Y_\lambda^{\frac{\alpha-1}{\alpha}} \right] \right) \tag{11.F.57}$$

$$= \operatorname{Tr}\left[\mathcal{M}(X_\lambda^{\frac{1}{\alpha}}) Y_\lambda^{\frac{\alpha-1}{\alpha}} \right] \tag{11.F.58}$$

$$\leq \sup_{\substack{Y_A > 0, \\ \operatorname{Tr}[Y_A] \leq 1}} \operatorname{Tr}\left[\mathcal{M}(X_\lambda^{\frac{1}{\alpha}}) Y_A^{\frac{\alpha-1}{\alpha}} \right] = \left\| \mathcal{M}(X_\lambda^{\frac{1}{\alpha}}) \right\|_\alpha , \tag{11.F.59}$$

where the first inequality follows from an application of the Lieb concavity theorem, and the second inequality follows from applying (11.F.52) and because Y_λ is a particular operator satisfying $Y_\lambda > 0$ and $\operatorname{Tr}[Y_\lambda] \leq 1$. Since the chain of inequalities holds for arbitary $Y_0, Y_1 > 0$ such that $\operatorname{Tr}[Y_0], \operatorname{Tr}[Y_1] \leq 1$, we conclude that

$$\lambda \left\| \mathcal{M}(X_0^{\frac{1}{\alpha}}) \right\|_\alpha + (1-\lambda) \left\| \mathcal{M}(X_1^{\frac{1}{\alpha}}) \right\|_\alpha \leq \left\| \mathcal{M}(X_\lambda^{\frac{1}{\alpha}}) \right\|_\alpha , \tag{11.F.60}$$

which concludes the proof. ■

Theorem 11.30 Lieb Concavity

The following function is jointly concave with respect to positive semi-definite operators R and S for arbitrary $t \in (0,1)$ and an arbitrary operator K:

$$(R, S) \mapsto \operatorname{Tr}[K R^t K^\dagger S^{1-t}]. \tag{11.F.61}$$

Proof: We begin by restricting the first argument of the function in (11.F.61) to positive definite operators. Defining $|K\rangle_{RA} = K_A^\dagger |\Gamma\rangle_{RA}$, consider that

$$\operatorname{Tr}[K R^t K^\dagger S^{1-t}] = \langle\Gamma|_{RA} \mathbb{1}_R \otimes \left(K R^t K^\dagger S^{1-t} \right)_A |\Gamma\rangle_{RA} \tag{11.F.62}$$

$$= \langle\Gamma|_{RA} (S_R^\mathsf{T})^{1-t} \otimes K_A R_A^t K_A^\dagger |\Gamma\rangle_{RA} \tag{11.F.63}$$

$$= \langle\Gamma|_{RA} K_A \left[(S_R^\mathsf{T})^{1-t} \otimes R_A^t \right] K_A^\dagger |\Gamma\rangle_{RA} \tag{11.F.64}$$

$$= \langle K|_{RA} R_A^{\frac{1}{2}} \left[(S_R^\mathsf{T})^{1-t} \otimes R_A^{t-1} \right] R_A^{\frac{1}{2}} |K\rangle_{RA} \tag{11.F.65}$$

$$= \langle K|_{RA} R_A^{\frac{1}{2}} \left[S_R^\mathsf{T} \otimes R_A^{-1} \right]^{1-t} R_A^{\frac{1}{2}} |K\rangle_{RA} \tag{11.F.66}$$

$$= \langle K|_{RA} R_A^{\frac{1}{2}} g(S_R^{\mathsf{T}} \otimes R_A^{-1}) R_A^{\frac{1}{2}} |K\rangle_{RA}, \tag{11.F.67}$$

where the fourth equality holds by the positive definiteness of R, and where $g(x) := x^{1-t}$ is an operator concave function. For $\lambda \in [0, 1]$, let

$$R_\lambda := \lambda R_0 + (1-\lambda) R_1, \qquad S_\lambda := \lambda S_0 + (1-\lambda) S_1, \tag{11.F.68}$$

where R_0 and R_1 are positive definite and S_0 and S_1 are positive semi-definite. Also, let

$$G_0 := \mathbb{1} \otimes \sqrt{\lambda R_0}\, (R_\lambda)^{-\frac{1}{2}}, \tag{11.F.69}$$

$$G_1 := \mathbb{1} \otimes \sqrt{(1-\lambda) R_1}\, (R_\lambda)^{-\frac{1}{2}}. \tag{11.F.70}$$

Then

$$G_0^\dagger G_0 + G_1^\dagger G_1 = \mathbb{1} \otimes (R_\lambda)^{-\frac{1}{2}} \lambda R_0 (R_\lambda)^{-\frac{1}{2}} + \mathbb{1} \otimes (R_\lambda)^{-\frac{1}{2}} (1-\lambda) R_1 (R_\lambda)^{-\frac{1}{2}} \tag{11.F.71}$$

$$= \mathbb{1} \otimes (R_\lambda)^{-\frac{1}{2}} R_\lambda (R_\lambda)^{-\frac{1}{2}} \tag{11.F.72}$$

$$= \mathbb{1} \otimes \mathbb{1}. \tag{11.F.73}$$

A variation of the operator Jensen inequality (Theorem 2.16) is that the following inequality holds for an operator concave function f, a finite set $\{X_i\}_i$ of Hermitian operators, and a finite set $\{A_i\}_i$ of operators satisfying $\sum_i A_i^\dagger A_i = \mathbb{1}$:

$$\sum_i A_i^\dagger f(X_i) A_i \leq f\left(\sum_i A_i^\dagger X_i A_i\right). \tag{11.F.74}$$

Then from the operator Jensen inequality and (11.F.67), we conclude that

$$\lambda \mathrm{Tr}[K R_0^t K^\dagger S_0^{1-t}] + (1-\lambda) \mathrm{Tr}[K R_1^t K^\dagger S_1^{1-t}]$$

$$= \lambda \langle K|_{RA} (R_0)_A^{\frac{1}{2}} g(S_0^{\mathsf{T}} \otimes R_0^{-1}) (R_0)_A^{\frac{1}{2}} |K\rangle_{RA} + (1-\lambda) \langle K|_{RA} (R_1)_A^{\frac{1}{2}} g(S_1^{\mathsf{T}} \otimes R_1^{-1}) (R_1)_A^{\frac{1}{2}} |K\rangle_{RA} \tag{11.F.75}$$

$$= \lambda \langle K|_{RA} (R_0)_A^{\frac{1}{2}} g(\lambda S_0^{\mathsf{T}} \otimes (\lambda R_0)^{-1}) (R_0)_A^{\frac{1}{2}} |K\rangle_{RA} + (1-\lambda) \langle K|_{RA} (R_1)_A^{\frac{1}{2}} g((1-\lambda) S_1^{\mathsf{T}} \otimes ((1-\lambda) R_1)^{-1}) (R_1)_A^{1/2} |K\rangle_{RA} \tag{11.F.76}$$

$$= \langle K|_{RA} (R_\lambda)_A^{\frac{1}{2}} G_0^\dagger g(\lambda S_0^\top \otimes (\lambda R_0)^{-1}) G_0 (R_\lambda)_A^{\frac{1}{2}} |K\rangle_{RA}$$
$$+ \langle K|_{RA} (R_\lambda)_A^{\frac{1}{2}} G_1^\dagger g((1-\lambda) S_1^\top \otimes ((1-\lambda) R_1)^{-1}) G_1 (R_\lambda)_A^{\frac{1}{2}} |K\rangle_{RA} \quad (11.F.77)$$
$$\leq \langle K|_{RA} (R_\lambda)_A^{\frac{1}{2}} g(L) (R_\lambda)_A^{\frac{1}{2}} |K\rangle_{RA}, \quad (11.F.78)$$

where the third equality follows because $\mathbb{1}_R \otimes \sqrt{\lambda} R_0^{\frac{1}{2}} = \mathbb{1}_R \otimes (R_\lambda)^{\frac{1}{2}} G_0^\dagger$. In the last line, we have let

$$L := G_0^\dagger \left(\lambda S_0^\top \otimes (\lambda R_0)^{-1}\right) G_0 + G_1^\dagger \left((1-\lambda) S_1^\top \otimes ((1-\lambda) R_1)^{-1}\right) G_1. \quad (11.F.79)$$

Consider that

$$L = G_0^\dagger \left(\lambda S_0^\top \otimes (\lambda R_0)^{-1}\right) G_0 + G_1^\dagger \left((1-\lambda) S_1^\top \otimes ((1-\lambda) R_1)^{-1}\right) G_1$$
$$= \left(\mathbb{1} \otimes (R_\lambda)^{-\frac{1}{2}} \sqrt{\lambda R_0}\right) \left(\lambda S_0^\top \otimes (\lambda R_0)^{-1}\right) \left(\mathbb{1} \otimes \sqrt{\lambda R_0} (R_\lambda)^{-\frac{1}{2}}\right)$$
$$+ \left(\mathbb{1} \otimes (R^\lambda)^{-\frac{1}{2}} \sqrt{(1-\lambda) R_1}\right)$$
$$\times \left((1-\lambda) S_1^\top \otimes ((1-\lambda) R_1)^{-1}\right) \left(\mathbb{1} \otimes \sqrt{(1-\lambda) R_1} (R_\lambda)^{-\frac{1}{2}}\right) \quad (11.F.80)$$
$$= \lambda S_0^\top \otimes (R_\lambda)^{-1} + (1-\lambda) S_1^\top \otimes (R_\lambda)^{-1} \quad (11.F.81)$$
$$= S_\lambda^\top \otimes (R_\lambda)^{-1}. \quad (11.F.82)$$

Continuing, we find that

$$\lambda \mathrm{Tr}[K R_0^t K^\dagger S_0^{1-t}] + (1-\lambda) \mathrm{Tr}[K R_1^t K^\dagger S_1^{1-t}]$$
$$\leq \langle K|_{RA} (R_\lambda)_A^{\frac{1}{2}} g(L) (R_\lambda)_A^{\frac{1}{2}} |K\rangle_{RA} \quad (11.F.83)$$
$$= \langle K|_{RA} (R_\lambda)_A^{\frac{1}{2}} g(S_\lambda^\top \otimes (R_\lambda)^{-1}) (R_\lambda)_A^{\frac{1}{2}} |K\rangle_{RA} \quad (11.F.84)$$
$$= \mathrm{Tr}[K R_\lambda^t K^\dagger S_\lambda^{1-t}]. \quad (11.F.85)$$

So the function $(R, S) \mapsto \mathrm{Tr}[K R^t K^\dagger S^{1-t}]$ is jointly concave when the first argument is restricted to be a positive definite operator. The more general case of positive semi-definite operators in the first argument can be established by adding $\varepsilon\mathbb{1}$ to any positive semi-definite operator to ensure that it is positive definite, applying the above inequality, and then taking the limit $\varepsilon \to 0$ at the end. This concludes the proof. ■

Appendix 11.G The Strong Converse from a Different Point of View

Here we show that the mutual information $I(\mathcal{N})$ is a strong converse rate based on the the alternate definition given in Appendix A. According to that definition, a rate $R \in \mathbb{R}^+$ is a strong converse rate for entanglement-assisted classical communication over a channel $\mathcal{N}$ if for every sequence $\{(n, |\mathcal{M}_n|, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols over n uses of $\mathcal{N}$, we have that $\liminf_{n\to\infty} \frac{1}{n}\log_2|\mathcal{M}_n| > R \Rightarrow \lim_{n\to\infty}\varepsilon_n = 1$.

Let us show that the mutual information $I(\mathcal{N})$ of the channel $\mathcal{N}$ is a strong converse rate under this alternate definition. Let $\{(n, |\mathcal{M}_n|, \varepsilon_n)\}_{n\in\mathbb{N}}$ be a sequence of protocols satisfying $\liminf_{n\to\infty} \frac{1}{n}\log_2|\mathcal{M}_n| > I(\mathcal{N})$. Due to this strict inequality, the fact that $\lim_{\alpha\to 1} \widetilde{I}_\alpha(\mathcal{N}) = I(\mathcal{N})$, and since the sandwiched Rényi mutual information $\widetilde{I}_\alpha(\mathcal{N})$ is monotonically increasing in α (this follows from Proposition 7.31), there exists a value $\alpha^* > 1$ such that

$$\liminf_{n\to\infty} \frac{1}{n}\log_2|\mathcal{M}_n| > \widetilde{I}_{\alpha^*}(\mathcal{N}). \tag{11.G.1}$$

Now recall the following bound from (11.2.92), which holds for all $\alpha > 1$ and for every $(n, |\mathcal{M}|, \varepsilon)$ protocol:

$$\frac{1}{n}\log_2|\mathcal{M}| \le \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{11.G.2}$$

We can apply it in our case to conclude that

$$\frac{1}{n}\log_2|\mathcal{M}_n| \le \widetilde{I}_{\alpha^*}(\mathcal{N}) + \frac{\alpha^*}{n(\alpha^*-1)}\log_2\!\left(\frac{1}{1-\varepsilon_n}\right). \tag{11.G.3}$$

Now suppose that

$$\liminf_{n\to\infty} \varepsilon_n = c \in [0,1). \tag{11.G.4}$$

Then it follows that

$$\liminf_{n\to\infty} \frac{1}{n}\log_2|\mathcal{M}_n| \le \liminf_{n\to\infty}\left[\widetilde{I}_{\alpha^*}(\mathcal{N}) + \frac{\alpha^*}{n(\alpha^*-1)}\log_2\!\left(\frac{1}{1-\varepsilon_n}\right)\right] \tag{11.G.5}$$

$$= \widetilde{I}_{\alpha^*}(\mathcal{N}) + \liminf_{n\to\infty}\left[\frac{\alpha^*}{n(\alpha^*-1)}\log_2\!\left(\frac{1}{1-\varepsilon_n}\right)\right] \tag{11.G.6}$$

$$= \widetilde{I}_{\alpha^*}(\mathcal{N}), \tag{11.G.7}$$

where the last equality follows because $\alpha^* > 1$ is a constant and the sequence $\{\varepsilon_n\}_{n\in N}$ converges to a constant $c \in [0, 1)$. However, this contradicts (11.G.1). Thus, (11.G.4) cannot hold, and so we conclude that $\liminf_{n\to\infty} \varepsilon_n = 1$.

The argument given above makes no statement about how fast the error probability converges to one in the large n limit. If we fix the rate R of communication to be a constant satisfying $R > I(\mathcal{N})$, then we can argue that the error probability converges exponentially fast to one. To this end, consider a sequence $\{(n, 2^{nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ protocols, with each element of the sequence having an arbitrary (but fixed) rate $R > I(\mathcal{N})$. For each element of the sequence, the inequality in (11.2.92) holds, which means that

$$R \leq \widetilde{I}_{\alpha}(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\!\left(\frac{1}{1-\varepsilon_n}\right) \tag{11.G.8}$$

for all $\alpha > 1$. Rearranging this inequality leads to the following lower bound on the error probabilities ε_n:

$$\varepsilon_n \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\widetilde{I}_{\alpha}(\mathcal{N})\right)} \tag{11.G.9}$$

for all $\alpha > 1$. Now, since $R > I(\mathcal{N})$, $\lim_{\alpha\to1} \widetilde{I}_{\alpha}(\mathcal{N}) = I(\mathcal{N})$, and since the sandwiched Rényi mutual information $\widetilde{I}_{\alpha}(\mathcal{N})$ is monotonically increasing in α (this follows from Proposition 7.31), there exists an $\alpha^* > 1$ such that $R > \widetilde{I}_{\alpha^*}(\mathcal{N})$. Applying the inequality in (11.G.9) to this value of α, we find that

$$\varepsilon_n \geq 1 - 2^{-n\left(\frac{\alpha^*-1}{\alpha^*}\right)\left(R-\widetilde{I}_{\alpha^*}(\mathcal{N})\right)}. \tag{11.G.10}$$

Then, taking the limit $n \to \infty$ on both sides of this inequality, we conclude that $\lim_{n\to\infty} \varepsilon_n = 1$ and the convergence to one is exponentially fast.

From the arguments above, we find not only that $I(\mathcal{N})$ is a strong converse rate according to the alternate definition provided in Appendix A, but also that the maximal error probability of every sequence of $(n, |\mathcal{M}|, \varepsilon)$ entanglement-assisted classical communication protocols with fixed rate strictly above the mutual information $I(\mathcal{N})$ approaches one at an exponential rate.

In Section 11.C, we showed that the error probability vanishes in the limit $n \to \infty$ for every fixed rate $R < I(\mathcal{N})$. We thus see that, as $n \to \infty$, the mutual information $I(\mathcal{N})$ is a sharp dividing point between reliable, error-free communication and communication with error probability approaching one exponentially fast. This situation is depicted in Figure 11.8.

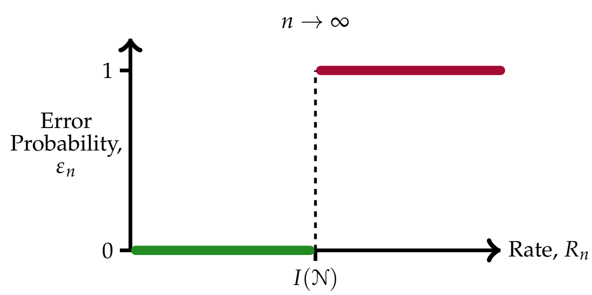


FIGURE 11.8: The error probability ε_n as a function of the rate R_n for entanglement-assisted classical communication over a quantum channel $\mathcal{N}$. As $n \to \infty$, for every rate below the mutual information $I(\mathcal{N})$, there exists a sequence of protocols with error probability converging to zero. For every rate above the mutual information $I(\mathcal{N})$, the error probability converges to one for all possible protocols.

Chapter 12

Classical Communication

We now move on to classical communication over quantum channels. Unlike the previous chapter, here we suppose that Alice and Bob do not have access to shared entanglement prior to communication. Thus, the scenario considered in this chapter is more practical than the entanglement-assisted setting—in the previous chapter, we made the simplifying assumption that shared entanglement is available for free to the sender and receiver. However, without widespread entanglement-sharing networks available, this assumption is not really practical, and so the entanglement-assisted capacity is mostly of academic interest at the moment.

Without shared entanglement available to the sender and receiver, is it still advantageous to use a quantum strategy to send classical information over a quantum channel? At first glance, it may seem that, without prior shared entanglement, there might not be any point in using a quantum strategy to send classical information over a quantum channel. However, when using a channel multiple times, there is still the possibility of encoding a message into a state at the encoder that is entangled across multiple channel uses and then performing a collective measurement at the decoder. For many examples of channels, it is known that collective measurements can enhance communication capacity, and it is known that in principle, there exists a channel for which entangled states at the encoder provides a further enhancement to communication capacity.

Although it may seem that determining the maximum amount of classical information that can be communicated using a given quantum channel, i.e., determining the classical capacity of a quantum channel, might be easier than its entanglement-assisted counterpart, this problem turns out to be one of the most challenging problems in quantum Shannon theory. This is due to the fact that, as we discuss in this chapter, the relevant quantity in calculating the classical capacity of a quantum channel $\mathcal{N}$ is related to its Holevo information $\chi(\mathcal{N})$, and this quantity is not known to be additive for all quantum channels. This means that the best we can say for a given channel is that its Holevo information is an achievable rate for classical communication—we cannot necessarily say that it is the highest possible achievable rate. This is in stark contrast to the case of entanglement-assisted classical communication, for which we know that the mutual information $I(\mathcal{N})$ of a channel is additive for all channels and thus is equal to the entanglement-assisted classical capacity of any quantum channel.

We start in the next brief section by considering some simple motivating examples of communication. Then we consider the one-shot setting. This setting for classical communication is similar to the one-shot setting of entanglement-assisted classical communication from the previous chapter. The only difference is that, in this case, there is no entanglement assistance. We then move on to the asymptotic setting, for which we prove Theorem 12.13, which states that the classical capacity of a quantum channel is equal to the *regularized* Holevo information of the channel. This is a quantity that in general requires computing the Holevo information for an arbitrarily large number of uses of the given channel, and it is therefore intractable unless the Holevo information happens to be additive for the channel. From here, we consider various classes of channels for which the Holevo information is additive or for which we can establish the strong converse. We also consider various methods for bounding the classical capacity from above. Finally, we calculate the classical capacity for some examples of channels.

Simple Example of Classical Communication Over a Quantum Channel

At the beginning of the previous chapter, we stated that super-dense coding is a simple example of an entanglement-assisted classical communication protocol over a noiseless quantum channel. The essense of that protocol is the encoding of d^2 messages into d^2 mutually orthogonal pure states (the maximally entangled states defined in (3.2.57)). The d^2 messages contain $\log_2 d^2 = 2\log_2 d$ bits of classical information, which can be communicated without error from Alice to Bob, with just one use of a noiseless qudit quantum channel.

Now, without the assistance of prior shared entanglement, one result of this chapter is that the maximum amount of classical information that can be communicated over a noiseless quantum channel without error is $\log_2 d$, where d is the dimension of the channel. Let us describe a simple protocol that achieves this number of communicated bits. Consider a discrete set $\mathcal{M}$ of messages, and suppose that Alice encodes each message $m \in \mathcal{M}$ into a quantum state $|m\rangle$, such that the set $\{|m\rangle\}_{m\in\mathcal{M}}$, is orthonormal, i.e., $\langle m|m'\rangle = \delta_{m,m'}$ for all $m, m' \in \mathcal{M}$. Bob, knowing Alice's encoding of the messages, devises a measurement to extract the message described by the POVM $\{|m\rangle\!\langle m|\}_{m\in\mathcal{M}}$. His strategy is to guess that the message sent was "m" if the outcome of his measurement is $m \in \mathcal{M}$. If Alice sends the state $|m\rangle\!\langle m|$ through a noiseless quantum channel, then Bob is guaranteed to receive the state $|m\rangle\!\langle m|$ unaltered, so that his guess will always be correct. Alice can thus send $\log_2 |\mathcal{M}|$ bits of classical information to Bob without error.

Now, if the channel is noisy, the initially orthogonal states in general become non-orthogonal, so that if Alice sends the state $|m\rangle\!\langle m|$ through the channel then Bob generally receive a mixed state ρ^m instead. As a consequence of using a noisy quantum channel, Bob's decoding strategy will not always succeed, meaning that there will be errors. In order to mitigate the effects of noise, Alice can choose a more clever encoding of the message, and similarly Bob can devise a more clever decoding strategy.[1] Alice and Bob can also use the channel multiple times, which can decrease the error in general, while also allowing for the messages to be encoded into higher-dimensional entangled states.

Observe that the task of classical communication over a quantum channel is closely related to the task of state discrimination (see Section 5.3.1). Recall that the goal of state discrimination is to minimize the error probability for a given set $\{\rho^m\}_{m\in\mathcal{M}}$ of states corresponding to the message set $\mathcal{M}$ and a particular decoding POVM $\{\Lambda_B^m\}_{m\in\mathcal{M}}$ indexed by the messages. In classical communication, we focus primarily on maximizing the rate $\frac{1}{n}\log_2 |\mathcal{M}|$ of communication for a given error probability ε, and we are interested in determining the maximum rate R for which ε vanishes as the number n of channel uses increases.

[1] We assume, as in all communication tasks considered in this book, that Alice and Bob know the channel connecting them, so that they can use this knowledge to develop their encoding and decoding.

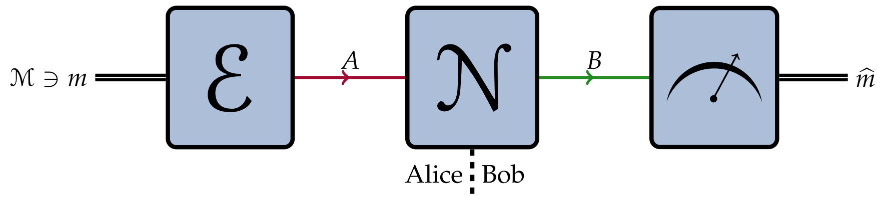


FIGURE 12.1: Depiction of a protocol for classical communication over one use of the quantum channel $\mathcal{N}$. Alice, who wishes to send a message m chosen from a set $\mathcal{M}$ of messages, first encodes the message into a quantum state on a quantum system A, using a classical–quantum encoding channel $\mathcal{E}$. She then sends the quantum system A through the channel $\mathcal{N}_{A\to B}$. After Bob receives the system B, he performs a measurement on it, using the outcome of the measurement to give an estimate $\widehat{m}$ of the message sent by Alice.

12.1 One-Shot Setting

In the one-shot setting, we start by considering a classical communication protocol over a quantum channel $\mathcal{N}$, as depicted in Figure 12.1. The protocol is defined by the triple $(\mathcal{M}, \mathcal{E}_{M\to A}, \mathcal{D}_{B\to\widehat{M}})$, consisting of a message set $\mathcal{M}$, an *encoding channel* $\mathcal{E}_{M\to A}$, and a *decoding channel* $\mathcal{D}_{B\to\widehat{M}}$. The pair $(\mathcal{E}, \mathcal{D})$ of encoding and decoding channels is often called a *code* and denoted by $\mathcal{C} = (\mathcal{E}, \mathcal{D})$. The encoding channel is a classical–quantum channel (see Definition 4.9), and the decoding channel is a quantum–classical or measurement channel (see Definition 4.10).

Now, given that there are $|\mathcal{M}|$ messages in the message set, it follows that each message can be uniquely associated with a bit string of size at least $\log_2 |\mathcal{M}|$. The quantity $\log_2 |\mathcal{M}|$ thus represents the number of bits communicated in the protocol. One of the goals of this section is to obtain upper and lower bounds on maximum number of $\log_2 |\mathcal{M}|$ of bits that can be communicated in any classical communication protocol.

The protocol proceeds as follows: let $p : \mathcal{M} \to [0, 1]$ be a probability distribution over the message set. With probability $p(m)$, Alice picks a message $m \in \mathcal{M}$ and makes a local copy of it. Letting $\{|m\rangle\}_{m\in\mathcal{M}}$ be an orthonormal basis indexed by the messages, her initial state is described by the following classically correlated state:

$$\overline{\Phi}^{p}_{MM'} := \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'}. \qquad (12.1.1)$$

Note that if Alice wishes to send a particular message m deterministically, then she can choose the distribution p to be the degenerate distribution, equal to one for m and zero for all other messages.

She then uses an encoding channel $\mathcal{E}_{M\to A}$ to map the message to a quantum state ρ_A^m. We can explicitly define the encoding channel $\mathcal{E}_{M'\to A}$ as

$$\mathcal{E}_{M'\to A}(|m\rangle\langle m'|_{M'}) = \delta_{m,m'}\rho_A^m \quad \forall\, m, m' \in \mathcal{M}. \tag{12.1.2}$$

Note that this channel has the form of a classical–quantum channel (recall Definition 4.9). The action of the encoding channel on the initial state in (12.1.1) is as follows:

$$\begin{aligned}
\mathcal{E}_{M'\to A}(\overline{\Phi}^p_{MM'}) &= \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \mathcal{E}_{M'\to A}(|m\rangle\langle m|) &(12.1.3)\\
&= \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \rho_A^m &(12.1.4)\\
&=: \rho^p_{MA}. &(12.1.5)
\end{aligned}$$

Alice then sends the system A through the channel $\mathcal{N}_{A\to B}$, resulting in the state

$$\mathcal{N}_{A\to B}(\rho_{MA}) = \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \mathcal{N}_{A\to B}(\rho_A^m). \tag{12.1.6}$$

Bob, whose task is to determine which message Alice sent, performs a decoding measurement on his received system B, which has the corresponding POVM $\{\Lambda_B^m\}_{m\in\mathcal{M}}$. The measurement is associated with the decoding channel $\mathcal{D}_{B\to\widehat{M}}$, which is simply a quantum–classical channel as given in Definition 4.10, i.e.,

$$\mathcal{D}_{B\to\widehat{M}}(\tau_B) := \sum_{m\in\mathcal{M}} \text{Tr}[\Lambda_B^m \tau_B]|m\rangle\langle m|_{\widehat{M}} \tag{12.1.7}$$

for every state τ_B. So the final state of the protocol is

$$\begin{aligned}
\omega^p_{M\widehat{M}} &:= (\mathcal{D}_{B\to\widehat{M}} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}^p_{MM'}) &(12.1.8)\\
&= \sum_{m,\widehat{m}\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \text{Tr}[\Lambda_B^{\widehat{m}}\mathcal{N}_{A\to B}(\rho_A^m)]|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}. &(12.1.9)
\end{aligned}$$

The measurement by Bob induces the conditional probability distribution $q : \mathcal{M}\times\mathcal{M} \to [0,1]$ defined by

$$q(\widehat{m}|m) := \Pr[\widehat{M} = \widehat{m}|M = m] = \text{Tr}[\Lambda_B^{\widehat{m}}\mathcal{N}_{A\to B}(\rho_A^m)]. \tag{12.1.10}$$

Bob's strategy is such that if the outcome $\widehat{m}$ occurs from his measurement, then he guesses that the message sent was $\widehat{m}$. The probability that Bob correctly identifies a given message m is then equal to $q(m|m)$. The *message error probability of the code* is given by

$$\begin{aligned} p_{\text{err}}(m,(\mathcal{E},\mathcal{D});\mathcal{N}) &:= 1-q(m|m) \\ &= \operatorname{Tr}[(\mathbb{1}_B-\Lambda_B^m)\mathcal{N}_{A\to B}(\rho_A^m)] \\ &= \sum_{\widehat{m}\in\mathcal{M}\setminus\{m\}} q(\widehat{m}|m). \end{aligned} \tag{12.1.11}$$

The *average error probability of the code* is

$$\overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) := \sum_{m\in\mathcal{M}} p(m)p_{\text{err}}(m,(\mathcal{E},\mathcal{D});\mathcal{N}) \tag{12.1.12}$$

$$= \sum_{m\in\mathcal{M}} p(m)(1-q(m|m)) \tag{12.1.13}$$

$$= \sum_{m\in\mathcal{M}} \sum_{\widehat{m}\in\mathcal{M}\setminus\{m\}} p(m)q(\widehat{m}|m). \tag{12.1.14}$$

The *maximal error probability of the code* is

$$p_{\text{err}}^*(\mathcal{E},\mathcal{D};\mathcal{N}) := \max_{m\in\mathcal{M}} p_{\text{err}}(m,(\mathcal{E},\mathcal{D});\mathcal{N}). \tag{12.1.15}$$

Just as in the case of entanglement-assisted classical communication in Chapter 11, each of these three error probabilities can be used to assess the reliability of the protocol, i.e., how well the encoding and decoding allows Alice to transmit her message to Bob.

Definition 12.1 ($|\mathcal{M}|,\varepsilon$) Classical Communication Protocol

A classical communication protocol $(\mathcal{M},\mathcal{E}_{M\to A},\mathcal{D}_{B\to\widehat{M}})$ over the channel $\mathcal{N}_{A\to B}$ is called an $(|\mathcal{M}|,\varepsilon)$ *protocol*, with $\varepsilon\in[0,1]$, if $p_{\text{err}}^*(\mathcal{E},\mathcal{D};\mathcal{N})\le\varepsilon$.

As with entanglement-assisted classical communication, the error criterion $p_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N})^*\le\varepsilon$ is equivalent to

$$\max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}_{MM'}^p-\omega_{M\widehat{M}}^p\right\|_1\le\varepsilon, \tag{12.1.16}$$

and the steps to show this are the same as those shown in the proof of Lemma 11.2. In particular, the following equality holds

$$\frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1 = \overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p), \tag{12.1.17}$$

which leads to

$$\begin{aligned} p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) &= \max_{p:\mathcal{M}\to[0,1]} \overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) \\ &= \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1 . \end{aligned} \tag{12.1.18}$$

Also, as in Chapter 11, another way to define the error criterion of the protocol is through a comparator test. Recall that the comparator test is a measurement defined by the two-element POVM $\{\Pi_{M\widehat{M}}, \mathbb{1}_{M\widehat{M}} - \Pi_{M\widehat{M}}\}$, where $\Pi_{M\widehat{M}}$ is the projection defined as

$$\Pi_{M\widehat{M}} := \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\widehat{M}}. \tag{12.1.19}$$

Note that $\text{Tr}[\Pi_{M\widehat{M}}\omega^p_{M\widehat{M}}]$ is simply the probability that the classical registers M and $\widehat{M}$ in the state $\omega^p_{M\widehat{M}}$ have the same values. In particular, following the same steps as in (11.1.38)–(11.1.40), we have

$$\text{Tr}\left[\Pi_{M\widehat{M}}\omega^p_{M\widehat{M}}\right] = 1 - \overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) =: \overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N}), \tag{12.1.20}$$

where we have acknowledged that the expression on the left-hand side can be interpreted as the average success probability of the code $(\mathcal{E},\mathcal{D})$ and denoted it by $\overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N})$.

As mentioned at the beginning of this chapter, our goal is to bound (from above and below) the maximum number $\log_2|\mathcal{M}|$ of transmitted bits for every classical communication protocol over $\mathcal{N}$. Given an error probability threshold of ε, we call the maximum number of transmitted bits the *one-shot classical capacity* of $\mathcal{N}$.

Definition 12.2 One-Shot Classical Capacity of a Quantum Channel

Given a quantum channel $\mathcal{N}_{A\to B}$ and $\varepsilon\in[0,1]$, the *one-shot ε-error classical capacity of* $\mathcal{N}$, denoted by $C^\varepsilon(\mathcal{N})$, is defined to be the maximum number $\log_2|\mathcal{M}|$ of transmitted bits among all $(|\mathcal{M}|,\varepsilon)$ classical communication protocols over

$\mathcal{N}$. In other words,

$$C^{\varepsilon}(\mathcal{N}) := \sup_{(\mathcal{M},\mathcal{E},\mathcal{D})} \{\log_2 |\mathcal{M}| : p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon\}, \tag{12.1.21}$$

where the optimization is over all protocols $(\mathcal{M}, \mathcal{E}_{M' \to A}, \mathcal{D}_{B \to \widehat{M}})$ satisfying $d_{M'} = d_{\widehat{M}} = |\mathcal{M}|$.

In addition to finding, for a given $\varepsilon \in [0, 1]$, the maximum number of transmitted bits among all $(|\mathcal{M}|, \varepsilon)$ classical communication protocols over $\mathcal{N}_{A\to B}$, we can consider the following complementary problem: for a given number of messages $|\mathcal{M}|$, find the smallest possible error probability among all $(|\mathcal{M}|, \varepsilon)$ classical communication protocols, which we denote by $\varepsilon^*_C(|\mathcal{M}|; \mathcal{N})$. In other words, to problem is to determine

$$\varepsilon^*_C(|\mathcal{M}|; \mathcal{N}) := \inf_{\mathcal{E},\mathcal{D}} \{p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) : d_{M'} = d_{\widehat{M}} = |\mathcal{M}|\}, \tag{12.1.22}$$

where the optimization is over encoding channels $\mathcal{E}$ with input space dimension $|\mathcal{M}|$ and decoding channels $\mathcal{D}$ with output space dimension $|\mathcal{M}|$. In this book, we focus primarily on the problem of optimizing the number of transmitted bits rather than the error probability, and so our primary quantity of interest is the one-shot capacity $C^{\varepsilon}(\mathcal{N})$.

12.1.1 Protocol Over a Useless Channel

We now turn to establishing an upper bound on the one-shot classical capacity, and our approach is similar to the approach outlined in Section 11.1.1. With this goal in mind, along with the actual classical communication protocol, we also consider the same protocol but performed over a *useless* channel as depicted in Figure 12.2. This useless channel discards the state encoded with the message and replaces it with some arbitrary (but fixed) state σ_B. In other words,

$$\rho^m_A \mapsto \sigma_B =: (\mathcal{P}_{\sigma_B} \circ \text{Tr})(\rho^m_A) \quad \forall\, m \in \mathcal{M}, \tag{12.1.23}$$

where the encoded states ρ^m_A are defined in (12.1.2). This channel is useless because the state σ_B does not contain any information about the message. As with entanglement-assisted classical communication, comparing this protocol over the

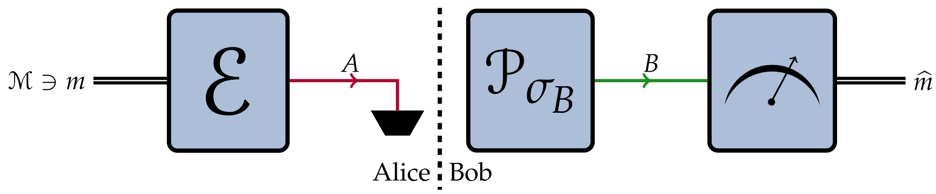


FIGURE 12.2: Depiction of a protocol that is useless for classical communication. The state encoding the message m via $\mathcal{E}$ is discarded and replaced with an arbitrary (but fixed) state σ_B.

useless channel with the actual protocol allows us to obtain an upper bound on the quantity $\log_2 |\mathcal{M}|$, which we recall represents the number of bits that are transmitted over the channel.

The state at the end of the protocol over the useless channel is the following tensor-product state:

$$\tau^p_{M\widehat{M}} := \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \sum_{\widehat{m}\in\mathcal{M}} \operatorname{Tr}[\Lambda^{\widehat{m}}_B \sigma_B]|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}, \tag{12.1.24}$$

which indicates that the decoded message system $\widehat{M}$ is independent of the message system M in this case. Now, recall from (12.1.9) that the state $\omega^p_{M\widehat{M}}$ at the end of the actual protocol over the channel $\mathcal{N}$ is given by

$$\omega^p_{M\widehat{M}} = \sum_{m,\widehat{m}\in\mathcal{M}} p(m)\operatorname{Tr}[\Lambda^{\widehat{m}}_B \mathcal{N}_{A\to B}(\rho^m_A)]|m\rangle\langle m|_M \otimes |\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}. \tag{12.1.25}$$

Similar to the notation from Chapter 11, we let

$$\omega_{M\widehat{M}} := \frac{1}{|\mathcal{M}|} \sum_{m,\widehat{m}\in\mathcal{M}} \operatorname{Tr}[\Lambda^{\widehat{m}}_B \mathcal{N}_{A\to B}(\rho^m_A)]|m\rangle\langle m|_M \otimes |\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}, \tag{12.1.26}$$

be the state $\omega^p_{M\widehat{M}}$ with the probability distribution p over the message set equal to the uniform distribution, i.e., $p(m) = \frac{1}{|\mathcal{M}|}$. We also let

$$\overline{\Phi}_{MM'} := \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'} \tag{12.1.27}$$

be the state in (12.1.1) with p being the uniform distribution over $\mathcal{M}$.

Now, observe that $\operatorname{Tr}_{\widehat{M}}[\omega_{M\widehat{M}}] = \pi_M$. Also, for every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol, the condition $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \le \varepsilon$ holds. By following the same steps as in (11.1.63)–(11.1.66), this condition implies that $\overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) \leq \varepsilon$ for the uniform distribution p. Then, by (12.1.20), we find that

$$\text{Tr}[\Pi_{M\widehat{M}}\omega_{M\widehat{M}}] \geq 1-\varepsilon. \tag{12.1.28}$$

We can therefore use Lemma 11.4 to conclude that

$$\log_2|\mathcal{M}| \leq I_H^{\varepsilon}(M;\widehat{M})_{\omega} \tag{12.1.29}$$

for every $(|\mathcal{M}|,\varepsilon)$ classical communication protocol. This means that, given a particular choice of the encoding and decoding channels, if $p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) \leq \varepsilon$, then the upper bound in (12.1.29) is the maximum number of bits that can be transmitted over the channel $\mathcal{N}$. The optimal value of this upper bound is realized by finding the state $\sigma_{\widehat{M}}$ defining the useless channel that optimizes the quantity $I_H^{\varepsilon}(M;\widehat{M})_{\omega}$ in addition to the measurement that achieves the ε-hypothesis testing relative entropy in (11.1.61). Importantly, a different choice of encoding and decoding produces a different value for this upper bound. We would thus like to find an upper bound that applies regardless of which specific protocol is chosen. In other words, we would like an upper bound that is a function of the channel $\mathcal{N}$ only.

12.1.2 Upper Bound on the Number of Transmitted Bits

We now give a general upper bound on the number of transmitted bits that can be communicated in any classical communication protocol. This result is stated in Theorem 12.4, and the upper bound obtained therein holds independently of the encoding and decoding channels used in the protocol and depends only on the given communication channel $\mathcal{N}$.

Let us start with an arbitrary $(|\mathcal{M}|,\varepsilon)$ classical communication protocol over the channel $\mathcal{N}$, corresponding to, as described at the beginning of this chapter, a message set $\mathcal{M}$, an encoding channel $\mathcal{E}$, and a decoding channel $\mathcal{D}$. The error criterion $p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) \leq \varepsilon$ holds by definition of an $(|\mathcal{M}|,\varepsilon)$ protocol, which implies the upper bound in (12.1.29) for the number $\log_2|\mathcal{M}|$ of transmitted bits in any $(|\mathcal{M}|,\varepsilon)$ classical communication protocol. Using this upper bound, we obtain the following:

Proposition 12.3 Upper Bound on One-Shot Classical Capacity

Let $\mathcal{N}$ be a quantum channel. For every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$, with $\varepsilon \in [0,1]$, the number of bits transmitted over $\mathcal{N}$ is bounded from above by the ε-hypothesis testing Holevo information of $\mathcal{N}$, as defined in (7.11.93), i.e.,

$$\log_2 |\mathcal{M}| \leq \chi_H^\varepsilon(\mathcal{N}). \tag{12.1.30}$$

Therefore,

$$C^\varepsilon(\mathcal{N}) \leq \chi_H^\varepsilon(\mathcal{N}). \tag{12.1.31}$$

Proof: We start with the upper bound in (12.1.29), i.e.,

$$\log_2 |\mathcal{M}| \leq I_H^\varepsilon(M; \widehat{M})_\omega, \tag{12.1.32}$$

where the state $\omega_{M\widehat{M}}$ is defined in (12.1.26). Recall that this bound follows from Lemma 11.4. Note that the state $\omega_{M\widehat{M}}$ can be written as

$$\omega_{M\widehat{M}} = \mathcal{D}_{B\to\widehat{M}}(\theta_{MB}), \tag{12.1.33}$$

where

$$\theta_{MB} := \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes \mathcal{N}_{A\to B}(\rho_A^m). \tag{12.1.34}$$

Now, from the data-processing inequality for the hypothesis testing relative entropy under the action of the decoding channel $\mathcal{D}_{B\to\widehat{M}}$, we find that

$$I_H^\varepsilon(M; \widehat{M})_\omega = \inf_{\sigma'_{\widehat{M}}} D_H^\varepsilon(\omega_{M\widehat{M}} \| \omega_M \otimes \sigma'_{\widehat{M}}) \leq I_H^\varepsilon(M; B)_\theta, \tag{12.1.35}$$

where we have used the fact that $\theta_M = \pi_M = \omega_M$. Note that

$$\theta_{MB} = \mathcal{N}_{A\to B}(\rho_{MA}), \tag{12.1.36}$$

where ρ_{MA} is the classical–quantum state ρ_{MA}^p defined in (12.1.5) with p equal to the uniform probability distribution. Optimizing over all classical–quantum states ξ_{MA} then leads to

$$I_H^\varepsilon(M; B)_\theta \leq \sup_{\xi_{MA}} I_H^\varepsilon(M; B)_\zeta = \chi_H^\varepsilon(\mathcal{N}), \tag{12.1.37}$$

where $\zeta_{MB} = \mathcal{N}_{A\to B}(\xi_{MA})$ and we have used the definition in (7.11.93) for the ε-hypothesis testing Holevo information of a channel. Note that this optimization over all classical–quantum states is effectively an optimization over all possible encoding channels $\mathcal{E}_{M'\to A}$ that define the $(|\mathcal{M}|, \varepsilon)$ protocol. Putting everything together, we obtain

$$\log_2 |\mathcal{M}| \leq I_H^\varepsilon(M; \widehat{M})_\omega \leq I_H^\varepsilon(M; B)_\theta \leq \chi_H^\varepsilon(\mathcal{N}), \tag{12.1.38}$$

as required. ■

The result of Proposition 12.3 can be written explicitly as

$$\log_2 |\mathcal{M}| \leq \sup_{\rho_{MA}} \inf_{\sigma_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{MA}) \| \rho_M \otimes \sigma_B) \tag{12.1.39}$$

$$= \sup_{\rho_{MA}} \inf_{\sigma_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{MA}) \| \mathcal{R}_{A\to B}^{\sigma_B}(\rho_{MA})), \tag{12.1.40}$$

where ρ_{MA} is a classical–quantum state. By doing so, we explictly see here the comparison, via the hypothesis testing relative entropy, between the actual classical communication protocol and the protocol over the useless channels $\mathcal{R}_{A\to B}^{\sigma_B}$, labeled by the states σ_B. The state ρ_{MA} corresponds to the state after the encoding channel, and optimizing over these states is effectively an optimization over all encoding channels.

As an immediate consequence of Propositions 12.3, 7.70, and 7.71, we have the following two bounds:

Theorem 12.4 One-Shot Upper Bounds for Classical Communication

Let $\mathcal{N}$ be a quantum channel, let $\varepsilon \in [0, 1)$, and let $\alpha > 1$. For every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$, the following bounds hold

$$\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon} \left(\chi(\mathcal{N}) + h_2(\varepsilon)\right), \tag{12.1.41}$$

$$\log_2 |\mathcal{M}| \leq \widetilde{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha - 1} \log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{12.1.42}$$

where $\chi(\mathcal{N})$ is the Holevo information of $\mathcal{N}$, as defined in (7.11.106), and $\widetilde{\chi}_\alpha(\mathcal{N})$ is the sandwiched Rényi Holevo information of $\mathcal{N}$, as defined in (7.11.95).

Since the bounds in (12.1.41) and (12.1.42) hold for every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$, we have that

$$C^{\varepsilon}(\mathcal{N}) \leq \frac{1}{1-\varepsilon}(\chi(\mathcal{N}) + h_2(\varepsilon)), \tag{12.1.43}$$

$$C^{\varepsilon}(\mathcal{N}) \leq \widetilde{\chi}_{\alpha}(\mathcal{N}) + \frac{\alpha}{\alpha - 1}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{12.1.44}$$

for all $\varepsilon \in [0, 1)$.

Let us recap the steps that we took to arrive at the bounds in (12.1.41) and (12.1.42).

1. We first compared the classical communication protocol over the channel $\mathcal{N}$ with a protocol over a useless channel, by using the ε-hypothesis testing relative entropy. This led us to the upper bound in (12.1.29).

2. We then used the data-processing inequality for the hypothesis testing relative entropy to obtain a quantity that is independent of the decoding channel, and also optimized over all useless protocols. This is done in (12.1.35) in the proof of Proposition 12.3.

3. Finally, to obtain a bound that is a function solely of the channel $\mathcal{N}$ and the error probability, we optimized over all encoding channels to obtain Proposition 12.3.

4. Using Propositions 7.70 and 7.71, which relate the hypothesis testing relative entropy to the quantum relative entropy and the sandwiched-Rényi relative entropy, respectively, we arrived at Theorem 12.4.

The bounds in (12.1.41) and (12.1.42) are fundamental upper bounds on the number of transmitted bits for *every* classical communication protocol. A natural question to ask is whether the upper bounds in (12.1.41) and (12.1.42) can be achieved. In other words, is it possible to devise protocols such that the number of transmitted bits is equal to the right-hand side of either (12.1.41) or (12.1.42)? We do not know how to, especially if we demand that we exactly attain the right-hand side of either (12.1.41) or (12.1.42). However, when given many uses of a channel (in the asymptotic setting), we can come close to achieving these upper bounds. This motivates finding lower bounds on the number of transmitted bits.

12.1.3 Lower Bound on the Number of Transmitted Bits

Having obtained upper bounds on the number transmitted bits in the previous section, let us now determine lower bounds. The key result of this section is Proposition 12.5, resulting in Theorem 12.6, which contains a lower bound on the number of transmitted bits for every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol.

As we saw in the previous chapter on entanglement-assisted classical communication, in order to obtain a lower bound on the number of transmitted bits, we should devise an explicit classical communication protocol $(\mathcal{M}, \mathcal{E}, \mathcal{D})$ such that the maximal error probability satisfies $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ for $\varepsilon \in [0, 1]$. Recall from (12.1.15) that the maximal error probability is defined as

$$p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) = \max_{m \in \mathcal{M}} p_{\text{err}}(m, (\mathcal{E}, \mathcal{D}); \mathcal{N}), \tag{12.1.45}$$

where, for all $m \in \mathcal{M}$, the message error probability $p_{\text{err}}(m; (\mathcal{E}, \mathcal{D}))$ is defined in (12.1.11) as

$$p_{\text{err}}(m, (\mathcal{E}, \mathcal{D}); \mathcal{N}) = 1 - q(m|m), \tag{12.1.46}$$

with $q(\widehat{m}|m)$ being the probability of identifying the message sent as $\widehat{m}$ given that the message m was sent.

The classical communication protocol discussed here is related to the entanglement-assisted classical communication protocol in Section 11.1.3. We suppose at first that Alice and Bob have some shared randomness prior to communication. This shared randomness is strictly speaking not part of the classical communication protocol as outlined at the beginning of Section 12.1, but the advantage of using it is that we can directly employ all of the developments for the position-based coding and sequential decoding strategy from Section 11.1.3. We then perform what is called *derandomization* and *expurgation* (both of which we outline below) ultimately to remove this shared randomness from the protocol and thus obtain the desired lower bound on the number of transmitted bits for the true unassisted classical communication protocol.

Proposition 12.5 Lower Bound on One-Shot Classical Capacity

Let $\mathcal{N}_{A \to B}$ be a quantum channel. For all $\varepsilon \in (0, 1)$ and $\eta \in \left(0, \frac{\varepsilon}{2}\right)$, there exists an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}_{A\to B}$ such that

$$\log_2 |\mathcal{M}| = \overline{\chi}_H^{\frac{\varepsilon}{2}-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.47}$$

Consequently, for all $\varepsilon \in (0,1)$ and $\eta \in \left(0, \frac{\varepsilon}{2}\right)$,

$$C^{\varepsilon}(\mathcal{N}) \geq \overline{\chi}_H^{\frac{\varepsilon}{2}-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.48}$$

Here,

$$\overline{\chi}_H^{\varepsilon}(\mathcal{N}) \coloneqq \sup_{\rho_{XA}} \overline{I}_H^{\varepsilon}(X;B)_\omega, \tag{12.1.49}$$

where $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$, the state ρ_{XA} is a classical–quantum state, and

$$\overline{I}_H^{\varepsilon}(X;B)_\omega = D_H^{\varepsilon}(\omega_{XB}\|\omega_X \otimes \omega_B). \tag{12.1.50}$$

Remark: The quantity $\overline{\chi}_H^{\varepsilon}(\mathcal{N})$ defined in the statement of Proposition 12.5 above is similar to the quantity $\chi_H^{\varepsilon}(\mathcal{N})$ defined in (7.11.93), except that it is defined with respect to the mutual information $\overline{I}_H^{\varepsilon}(X;B)_\rho$ that we encountered in Proposition 11.8, which does not involve an optimization over states σ_B.

Proof: As described before the statement of the proposition, we start with a protocol based on randomness-assisted classical communication, in which Alice and Bob have shared randomness prior to communication via the state $\rho_{XB'}^{\otimes|\mathcal{M}'|}$, where $\mathcal{M}'$ is a message set and $\rho_{XB'}$ is the following classically correlated state:

$$\rho_{XB'} \coloneqq \sum_{x\in\mathcal{X}} r(x)|x\rangle\langle x|_X \otimes |x\rangle\langle x|_{B'}. \tag{12.1.51}$$

Here, $\mathcal{X}$ is a finite alphabet and $r : \mathcal{X} \to [0,1]$ is a probability distribution on $\mathcal{X}$. The system X is held by Alice and the system B' is held by Bob. We denote the encoding and decoding channels for this protocol by $\mathcal{E}'$ and $\mathcal{D}'$, respectively, and they correspond to the position-based coding and sequential decoding strategies developed in Section 11.1.3. The goal is to use this protocol to determine the existence of encoding and decoding channels $\mathcal{E}$ and $\mathcal{D}$ for an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol.

As a first step, Alice processes each of her X systems with a classical–quantum channel $x \mapsto \rho_A^x$ (see Definition 4.9), so that the state shared by them becomes

$\rho_{A'B'}^{\otimes|\mathcal{M}'|}$, where

$$\rho_{A'B'} = \sum_{x\in\mathcal{X}} r(x)\rho_{A'}^{x} \otimes |x\rangle\langle x|_{B'}. \tag{12.1.52}$$

Just as in Section 11.1.3, the rest of the encoding channel $\mathcal{E}'$ is defined such that if Alice wishes to send the message $m \in \mathcal{M}'$, then she sends the mth A system through the channel. Thus, the state shared by Alice and Bob becomes

$$\rho_{A_1'B_1'} \otimes \cdots \otimes \mathcal{N}_{A\to B}(\rho_{AB_m'}) \otimes \cdots \otimes \rho_{A'_{|\mathcal{M}'|}B'_{|\mathcal{M}'|}} \tag{12.1.53}$$

for all $m \in \mathcal{M}'$, where

$$\rho_{A_i'B_i'} := \sum_{x\in\mathcal{X}} r(x)\rho_{A_i'}^{x} \otimes |x\rangle\langle x|_{B_i'}. \tag{12.1.54}$$

The reduced state on Bob's systems is then

$$\tau^{m}_{B_1'\cdots B_m'\cdots B'_{|\mathcal{M}'|}B} = \rho_{B_1'} \otimes \cdots \otimes \mathcal{N}_{A\to B}(\rho_{AB_m'}) \otimes \cdots \otimes \rho_{B'_{|\mathcal{M}'|}}. \tag{12.1.55}$$

In particular, we have that

$$\tau^{m}_{B'_{\widehat{m}}B} = \begin{cases} \mathcal{N}_{A'\to B}(\rho_{A'B'}) & \text{if } \widehat{m} = m, \\ \rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'}) & \text{if } \widehat{m} \neq m, \end{cases} \tag{12.1.56}$$

for all $m \in \mathcal{M}'$. Bob's task is to perform a test to guess which of the two states $\mathcal{N}_{A'\to B}(\rho_{A'B'})$ and $\rho_{B'} \otimes \mathcal{N}_{A'\to B}(\rho_{A'})$ he has on B and the $|\mathcal{M}'|$ systems $B_1'\cdots B'_{|\mathcal{M}'|}$. Since the shared state in the proof of Proposition 11.8 is arbitrary, we can apply all the arguments in that proof with the state $\rho_{A'B'}$ defined in (12.1.54) above to conclude immediately via (11.1.126) that

$$p_{\text{err}}(m, (\mathcal{E}', \mathcal{D}'); \mathcal{N}) \leq \varepsilon \tag{12.1.57}$$

for all $m \in \mathcal{M}'$, provided that

$$\log_2 |\mathcal{M}'| = \overline{I}_H^{\varepsilon-\eta}(B'; B)_\xi - \log_2\left(\frac{4\varepsilon}{\eta^2}\right), \tag{12.1.58}$$

where $\xi_{B'B} := \mathcal{N}_{A'\to B}(\rho_{A'B'})$. Note that $\xi_{B'B}$ is a classical–quantum state, which means that $\overline{I}_H^{\varepsilon-\eta}(B'; B)_\xi = \overline{\chi}_H^{\varepsilon-\eta}(B'; B)_\xi$. Furthermore, by taking a supremum over every input ensemble $\{(r(x), \rho_A^x)\}_{x\in\mathcal{X}}$, we find that

$$\log_2 |\mathcal{M}'| = \overline{\chi}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.59}$$

Combining (12.1.57) and (12.1.59), we can already conclude that, when shared randomness is available for free, a lower bound on the number of transmitted bits is given by (12.1.59). The condition in (12.1.57) on the message error probability implies that the average error probability $\overline{p}_{\text{err}}((\mathcal{E}', \mathcal{D}'); p)$, with p the uniform distribution over $\mathcal{M}'$, satisfies

$$\overline{p}_{\text{err}}((\mathcal{E}', \mathcal{D}'); p, \mathcal{N}) = \frac{1}{|\mathcal{M}'|} \sum_{m \in \mathcal{M}'} p_{\text{err}}(m, (\mathcal{E}', \mathcal{D}'), \mathcal{N}) \leq \varepsilon. \tag{12.1.60}$$

Let us now use the expression in (11.1.104) to derive an exact expression for the average error probability. The expression in (11.1.104) is

$$\begin{aligned} &p_{\text{err}}(m, (\mathcal{E}, \mathcal{D}); \mathcal{N}) \\ &\qquad = 1 - \text{Tr}[P_m \widehat{P}_{m-1} \cdots \widehat{P}_1 \omega^m_{B'_1 \cdots B'_{|\mathcal{M}'|} B R_1 \cdots R_{|\mathcal{M}'|}} \widehat{P}_1 \cdots \widehat{P}_{m-1} P_m] \end{aligned} \tag{12.1.61}$$

for all $m \in \mathcal{M}'$. First, observe that both $\mathcal{N}_{A' \to B}(\rho_{A'B'})$ and $\rho_{B'} \otimes \mathcal{N}_{A' \to B}(\rho_{A'})$ are classical–quantum states. This means that, for every measurement operator $\Lambda_{BB'}$, we find that

$$\begin{aligned} \text{Tr}[\Lambda_{BB'} \mathcal{N}_{A' \to B}(\rho_{B'A'})] &= \sum_{x \in \mathcal{X}} r(x) \text{Tr}[\Lambda_{B'B}(|x\rangle\langle x|_{B'} \otimes \rho^x_B)] && (12.1.62) \\ &= \sum_{x \in \mathcal{X}} r(x) \text{Tr}[(\langle x|_{B'} \otimes \mathbb{1}_B) \Lambda_{B'B}(|x\rangle_{B'} \otimes \mathbb{1}_B) \rho^x_B] && (12.1.63) \\ &= \sum_{x \in \mathcal{X}} r(x) \text{Tr}[M^x_B \rho^x_B], && (12.1.64) \end{aligned}$$

where $\rho^x_B = \mathcal{N}_{A' \to B}(\rho^x_{A'})$, and we have defined the operators

$$M^x_B \coloneqq \text{Tr}_{B'}[(|x\rangle\langle x|_{B'} \otimes \mathbb{1}_{B'}) \Lambda_{B'B}]. \tag{12.1.65}$$

Similarly, letting $\overline{\rho}_B \coloneqq \sum_{x \in \mathcal{X}} r(x) \rho^x_B$, we find that

$$\begin{aligned} \text{Tr}[\Lambda_{B'B}(\rho_{B'} \otimes \mathcal{N}_{A' \to B}(\rho_{A'}))] &= \sum_{x \in \mathcal{X}} r(x) \text{Tr}[\Lambda_{B'B}(|x\rangle\langle x|_{B'} \otimes \overline{\rho}_B)] && (12.1.66) \\ &= \sum_{x \in \mathcal{X}} r(x) \text{Tr}[M^x_B \overline{\rho}_B]. && (12.1.67) \end{aligned}$$

This implies that the measurement operator $\Lambda^*_{B'B}$ that achieves the optimal value for the quantity $D_H^{\varepsilon-\eta}(\mathcal{N}_{A'\to B}(\rho_{A'B'})\|\rho_{B'}\otimes\mathcal{N}_{A'\to B}(\rho_{A'}))$ can be taken to have the form

$$\Lambda^*_{B'B} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_{B'}\otimes M^x_B. \tag{12.1.68}$$

Now using the fact that

$$\sqrt{\Lambda^*_{B'B}} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_{B'}\otimes\sqrt{M^x_B}, \tag{12.1.69}$$

the projectors $\Pi_{B'BR}$ defined in (11.1.116) have the following form:

$$\Pi_{B'BR} = \sum_{x\in\mathcal{X}} |x\rangle\langle x|_{B'}\otimes\Pi^x_{BR}, \tag{12.1.70}$$

where R is a reference system held by Bob to help with the decoding, and the projector Π^x_{BR} is given by

$$\Pi^x_{BR} \coloneqq (U^x_{BR})^\dagger\,(\mathbb{1}_B\otimes|1\rangle\langle 1|_R)\,U^x_{BR}, \tag{12.1.71}$$

$$\begin{aligned} U^x_{BR} \coloneqq{}& \sqrt{\mathbb{1}_B - M^x_B}\otimes(|0\rangle\langle 0|_R + |1\rangle\langle 1|_R) \\ &+ \sqrt{M^x_B}\otimes(|1\rangle\langle 0|_R - |0\rangle\langle 1|_R)\,. \end{aligned} \tag{12.1.72}$$

This in turn implies that the measurement operators P_i, which are used for the sequential decoding and are defined in (11.1.100), have the form

$$P_i = \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} |\underline{x}\rangle\langle\underline{x}|_{B'_1\cdots B'_{|\mathcal{M}'|}}\otimes P^{x_i}_i, \tag{12.1.73}$$

where $|\underline{x}\rangle \equiv |x_1,\ldots,x_{|\mathcal{M}'|}\rangle$ and

$$P^{x_i}_i \coloneqq \mathbb{1}_{R_1}\otimes\cdots\otimes\mathbb{1}_{R_{i-1}}\otimes\Pi^{x_i}_{BR_i}\otimes\mathbb{1}_{R_{i+1}}\otimes\cdots\otimes\mathbb{1}_{R_{|\mathcal{M}'|}}. \tag{12.1.74}$$

Finally, since we can write the state $\tau^m_{B'_1\cdots B'_{|\mathcal{M}'|}B}$ in (12.1.55) as

$$\tau^m_{B'_1\cdots B'_{|\mathcal{M}'|}B} = \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} r(x_1)\cdots r(x_{|\mathcal{M}'|})|\underline{x}\rangle\langle\underline{x}|_{B'_1\cdots B'_{|\mathcal{M}'|}}\otimes\rho^{x_m}_B, \tag{12.1.75}$$

we find that

$$\omega^m_{B'_1\cdots B'_{|\mathcal{M}'|}BR_1\cdots R_{|\mathcal{M}'|}}$$

$$:= \tau^m_{B'_1\cdots B'_{|\mathcal{M}'|}B} \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1\cdots R_{|\mathcal{M}'|}} \tag{12.1.76}$$

$$= \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} r(x_1)\cdots r(x_{|\mathcal{M}'|})|\underline{x}\rangle\langle\underline{x}|_{B'_1\cdots B'_{|\mathcal{M}'|}} \otimes (\rho^{x_m}_B \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1\cdots R_{|\mathcal{M}'|}}), \tag{12.1.77}$$

where $|\underline{0}\rangle \equiv |0,\ldots,0\rangle$. Therefore, by definition,

$$\begin{aligned} &p_{\text{err}}(m;(\mathcal{E}',\mathcal{D}')) \\ &\quad = 1 - \text{Tr}[P_m \widehat{P}_{m-1}\cdots\widehat{P}_1 \omega_{B'_1\cdots B'_{|\mathcal{M}'|}BR_1\cdots R_{|\mathcal{M}'|}} \widehat{P}_1 \cdots \widehat{P}_{m-1} P_m] && (12.1.78) \\ &\quad = \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} \Big[r(x_1)\cdots r(x_{|\mathcal{M}'|}) \\ &\qquad\qquad \times \Big(1 - \text{Tr}[\Omega^{x_m}_m(\rho^{x_m}_B \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1\cdots R_{|\mathcal{M}'|}})]\Big)\Big], && (12.1.79) \end{aligned}$$

for all $m \in \mathcal{M}'$, where

$$\Omega^{x_m}_m := \widehat{P}^{x_1}_1 \cdots \widehat{P}^{x_{m-1}}_{m-1} P^{x_m}_m \widehat{P}^{x_{m-1}}_{m-1} \cdots \widehat{P}^{x_1}_1. \tag{12.1.80}$$

Therefore, the average error probability is bounded as

$$\begin{aligned} &\overline{p}_{\text{err}}((\mathcal{E}',\mathcal{D}');p) \\ &= \sum_{m\in\mathcal{M}'} \frac{1}{|\mathcal{M}'|} \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} \Big[r(x_1)\cdots r(x_{|\mathcal{M}'|}) \\ &\qquad\qquad \times \Big(1 - \text{Tr}[\Omega^{x_m}_m(\rho^{x_m}_B \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1\cdots R_{|\mathcal{M}'|}})]\Big)\Big] && (12.1.81) \\ &\le \sum_{m\in\mathcal{M}'} \frac{1}{|\mathcal{M}'|} \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} \Big[r(x_1)\cdots r(x_{|\mathcal{M}'|}) \\ &\qquad \times \Big(\gamma_{\text{I}}\text{Tr}[(I_B - M^{x_m}_B)\rho^{x_m}_B] + \gamma_{\text{II}} \sum_{i=1}^{m-1} \text{Tr}[M^{x_i}_B \rho^{x_m}_B]\Big)\Big] && (12.1.82) \\ &\le \varepsilon, && (12.1.83) \end{aligned}$$

where $\gamma_{\text{I}} := 1 + c$ and $\gamma_{\text{II}} := 2 + c + c^{-1}$, with $c = \frac{\eta}{2\varepsilon-\eta}$, and the inequality on the last line holds due to (12.1.60). Exchanging the sum over $\mathcal{M}'$ with the sum over the elements $x_1,\ldots,x_{|\mathcal{M}'|}$ of $\mathcal{X}$ (in the spirit of the famous trick of Shannon), we find that

$$\sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} r(x_1)\cdots r(x_{|\mathcal{M}'|})\overline{p}_{\text{err}}(\mathcal{C};p)$$
$$\leq \sum_{x_1,\ldots,x_{|\mathcal{M}'|}\in\mathcal{X}} r(x_1)\cdots r(x_{|\mathcal{M}'|})\overline{u}_{\text{err}}(\mathcal{C};p) \leq \varepsilon, \quad (12.1.84)$$

where

$$\overline{p}_{\text{err}}(\mathcal{C};p,\mathcal{N}) := \sum_{m\in\mathcal{M}'} \frac{1}{|\mathcal{M}'|}\left(1 - \text{Tr}[\Omega_m^{x_m}(\rho_B^{x_m} \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1\cdots R_{|\mathcal{M}'|}})]\right) \quad (12.1.85)$$

is the average error probability under a code $\mathcal{C}$ in which each message m is encoded as $m \mapsto x_m \mapsto \rho_A^{x_m}$ and

$$\overline{u}_{\text{err}}(\mathcal{C};p,\mathcal{N}) := \sum_{m\in\mathcal{M}'} \frac{1}{|\mathcal{M}'|}\left(\gamma_{\text{I}}\,\text{Tr}[(I_B - M_B^{x_m})\rho_B^{x_m}] + \gamma_{\text{II}}\sum_{i=1}^{m-1}\text{Tr}[M_B^{x_i}\rho_B^{x_m}]\right) \quad (12.1.86)$$

is an upper bound on the average error probability $\overline{p}_{\text{err}}(\mathcal{C};p,\mathcal{N})$. The decoding is defined by the measurement operators $\{\Omega_m^{x_m}\}_{m\in\mathcal{M}'}$. Note that the code $\mathcal{C}$ is a random variable, in the sense that the string $x_1,\ldots,x_{|\mathcal{M}'|}$ of length $|\mathcal{M}'|$ is used for the encoding and decoding with probability $r(x_1)\cdots r(x_{|\mathcal{M}'|})$.

Since the minimum does not exceed the average, the inequality in (12.1.84) implies that there exists a code $\mathcal{C}^*$, with corresponding string $x_1^*,\ldots,x_{|\mathcal{M}'|}^*$, such that

$$\overline{u}_{\text{err}}(\mathcal{C}^*;p,\mathcal{N}) \leq \varepsilon, \quad (12.1.87)$$

and in turn, via Theorem 11.7, that

$$\overline{p}_{\text{err}}(\mathcal{C}^*;p,\mathcal{N}) \leq \overline{u}_{\text{err}}(\mathcal{C}^*;p,\mathcal{N}) \leq \varepsilon. \quad (12.1.88)$$

By choosing this particular code, we can now follow through the entire argument above *without* the shared randomness (in the form of the state $\rho_{A'B'}$) in order to conclude that with the code $\mathcal{C}^*$, the number of transmitted bits is given by (12.1.59), and the average error probability of the code is bounded from above by ε. This completes the *derandomization* part of the proof.

Finally, we are interested in a code, call it $(\mathcal{E},\mathcal{D})$, satisfying the maximal error probability criterion $p_{\text{err}}^*(\mathcal{E},\mathcal{D};\mathcal{N}) \leq \varepsilon$ instead of the average error probability criterion. To find such a code, we can apply *expurgation* to the code $\mathcal{C}^*$ defined above. Formally, this means the following: since we have a code satisfying

(12.1.87), by Markov's inequality (see (2.3.20)), half of the codewords in $\mathcal{C}^*$ (call them $c_1, \dots, c_{\frac{|\mathcal{M}'|}{2}}$) satisfy

$$u_{\text{err}}(m; \mathcal{C}^*) \leq 2\varepsilon, \tag{12.1.89}$$

for all $m \in \mathcal{M}'$ corresponding to the codewords $c_1, \dots, c_{\frac{|\mathcal{M}'|}{2}}$, where

$$u_{\text{err}}(m, \mathcal{C}^*; \mathcal{N}) \coloneqq \gamma_{\text{I}} \text{Tr}[(I_B - M_B^{x_m})\rho_B^{x_m}] + \gamma_{\text{II}} \sum_{i=1}^{m-1} \text{Tr}[M_B^{x_i} \rho_B^{x_m}]. \tag{12.1.90}$$

We thus define a new message set $\mathcal{M} \subset \mathcal{M}'$, with $|\mathcal{M}| = \frac{|\mathcal{M}'|}{2}$, by removing all but those messages in $\mathcal{M}'$ whose encodings are given by $c_1, \dots, c_{\frac{|\mathcal{M}'|}{2}}$. Let $\mathcal{C}$ denote the expurgated code. Due to the fact that all of the terms in $u_{\text{err}}(m, \mathcal{C}^*; \mathcal{N})$ are non-negative, we find for all $m \in \mathcal{M}$ that

$$u_{\text{err}}(m, \mathcal{C}; \mathcal{N}) \leq u_{\text{err}}(m, \mathcal{C}^*; \mathcal{N}), \tag{12.1.91}$$

where

$$u_{\text{err}}(m, \mathcal{C}; \mathcal{N}) \coloneqq \gamma_{\text{I}} \text{Tr}[(I_B - M_B^{c_m})\rho_B^{c_m}] + \gamma_{\text{II}} \sum_{i=1}^{m-1} \text{Tr}[M_B^{c_i} \rho_B^{c_m}]. \tag{12.1.92}$$

Again applying the quantum union bound (Theorem 11.7), we then find that

$$p_{\text{err}}(m, \mathcal{C}; \mathcal{N}) \leq u_{\text{err}}(m, \mathcal{C}; \mathcal{N}), \tag{12.1.93}$$

where

$$\begin{aligned} p_{\text{err}}(m, \mathcal{C}; \mathcal{N}) \coloneqq 1 - \\ \text{Tr}[P_m^{c_m} \widehat{P}_{m-1}^{c_{m-1}} \cdots \widehat{P}_1^{c_1} (\rho_B^{c_m} \otimes |\underline{0}\rangle\langle\underline{0}|_{R_1 \cdots R_{|\mathcal{M}'|}}) \widehat{P}_1^{c_1} \cdots \widehat{P}_{m-1}^{c_{m-1}} P_m^{c_m}]. \end{aligned} \tag{12.1.94}$$

We thus have a code $(\mathcal{E}, \mathcal{D})$ satisfying

$$p_{\text{err}}^*(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq 2\varepsilon. \tag{12.1.95}$$

Specifically, the encoding is given by $m \mapsto c_m$ for all $m \in \mathcal{M}$, and the decoding is given by the sequential decoding procedure consisting of sequentially applying the binary measurements $\{P_m^{c_m}, \widehat{P}_m^{c_m}\}$ for all $m \in \mathcal{M}$ and decoding as message m as soon as the outcome $P_m^{c_m}$ occurs.

Therefore, we can use (12.1.59) to obtain the following for the number $\log_2 |\mathcal{M}|$ of transmitted bits with the reduced message set:

$$\log_2 |\mathcal{M}| = \log_2\left(\frac{|\mathcal{M}'|}{2}\right) = \log_2 |\mathcal{M}'| - \log_2(2) \tag{12.1.96}$$

$$= \overline{\chi}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) - \log_2(2) \tag{12.1.97}$$

$$= \overline{\chi}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{8\varepsilon}{\eta^2}\right). \tag{12.1.98}$$

Since ε and η are arbitrary, we have shown that for all $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$, there exists an $(|\mathcal{M}|, 2\varepsilon)$ classical communication protocol satisfying $\log_2 |\mathcal{M}| = \overline{\chi}_H^{\varepsilon-\eta}(\mathcal{N}) - \log_2\left(\frac{8\varepsilon}{\eta^2}\right)$. By the substitution $2\varepsilon \to \varepsilon$, we can finally say that for all $\varepsilon \in (0,1)$ and $\eta \in \left(0, \frac{\varepsilon}{2}\right)$, there exists an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol satisfying

$$\log_2 |\mathcal{M}| = \overline{\chi}_H^{\frac{\varepsilon}{2}-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.99}$$

This concludes the proof. ■

An immediate consequence of Propositions 12.5 and 7.72 is the following theorem.

Theorem 12.6 One-Shot Lower Bounds for Classical Communication

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For all $\varepsilon \in (0,1)$, $\eta \in \left(0, \frac{\varepsilon}{2}\right)$, and $\alpha \in (0,1)$, there exists an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}_{A\to B}$ such that

$$\log_2 |\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\frac{\varepsilon}{2}-\eta}\right) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.100}$$

Here,

$$\overline{\chi}_\alpha(\mathcal{N}) \coloneqq \sup_{\rho_{XA}} \overline{I}_\alpha(X;B)_\omega, \tag{12.1.101}$$

where $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$, the state ρ_{XA} is a classical–quantum state, and

$$\overline{I}_\alpha(X;B)_\omega \coloneqq D_\alpha(\omega_{XB} \| \omega_X \otimes \omega_B). \tag{12.1.102}$$

REMARK: The quantity $\overline{\chi}_\alpha(\mathcal{N})$ defined in the statement of Theorem 12.6 above is similar to the quantity $\chi_\alpha(\mathcal{N})$ defined in (7.11.94), except that it is defined with respect to the mutual information $\overline{I}_\alpha(X;B)_\omega$ that we encountered in Theorem 11.9, which does not involve an optimization over states σ_B.

PROOF: From Proposition 12.5, we know that for all $\varepsilon \in (0,1)$ and $\eta \in \left(0, \frac{\varepsilon}{2}\right)$, there exists an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol such that

$$\log_2 |\mathcal{M}| = \overline{\chi}_H^{\frac{\varepsilon}{2}-\eta}(\mathcal{N}) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right). \tag{12.1.103}$$

Proposition 7.72 relates the hypothesis testing relative entropy to the Petz–Rényi relative entropy according to

$$D_H^\varepsilon(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right) \tag{12.1.104}$$

for all $\alpha \in (0,1)$, which implies that

$$\overline{\chi}_H^\varepsilon(\mathcal{N}) \geq \overline{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\varepsilon}\right). \tag{12.1.105}$$

Combining this inequality with (12.1.103), we immediately get the desired result. ■

Since the inequality in (12.1.100) holds for every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol, we have that

$$C^\varepsilon(\mathcal{N}) \geq \overline{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\frac{\varepsilon}{2}-\eta}\right) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{12.1.106}$$

for all $\alpha \in (0,1)$, $\varepsilon \in (0,1)$, and $\eta \in \left(0, \frac{\varepsilon}{2}\right)$.

12.2 Classical Capacity of a Quantum Channel

Let us now consider the asymptotic setting of classical communication, as depicted in Figure 12.3. Similar to entanglement-assisted classical communication, instead of encoding the message into one quantum system and consequently using the

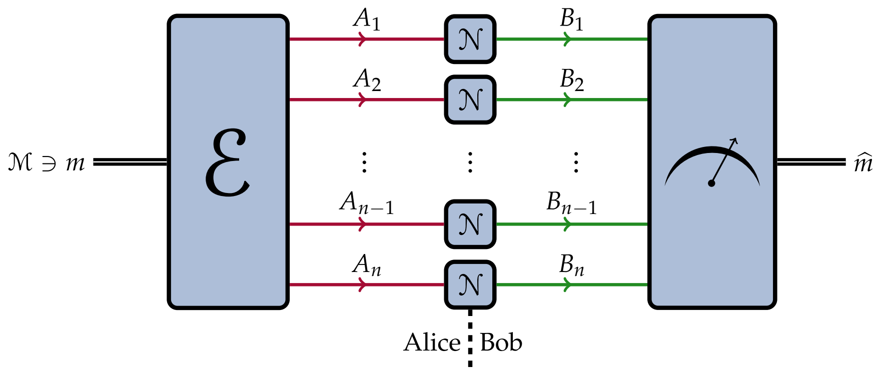


FIGURE 12.3: The most general classical communication protocol over a multiple number $n \geq 1$ uses of a quantum channel $\mathcal{N}$. Alice, who wishes to send a message m selected from a set $\mathcal{M}$, first encodes the message into a quantum state on n quantum systems using a classical–quantum encoding channel $\mathcal{E}$. She then sends each quantum system through the channel $\mathcal{N}$. After Bob receives the systems, he performs a collective measurement on them, using the outcome of the measurement to give an estimate $\widehat{m}$ of the message m sent to him by Alice.

channel $\mathcal{N}$ only once, Alice encodes the message into $n \geq 1$ quantum systems $A_1, \ldots, A_n$, all with the same dimension as A, and sends each one of these through the channel $\mathcal{N}$. We call this the asymptotic setting because the number n of channel uses can be arbitrarily large.

Recall that in the case of entanglement-assisted classical communication, we showed that encoding channels that entangle the n systems $A_1, \ldots, A_n$ do not help to achieve higher rates in the asymptotic setting. This is due to the additivity of the mutual information and the additivity of the sandwiched Rényi mutual information of a channel for all channels and $\alpha > 1$. In the case of classical communication that we consider in this chapter, it turns out that, so far, such a statement is known to be generally false for the Holevo information of a quantum channel (please consult the Bibliographic Notes in Section 12.5). That is, in principle there exists a channel for which the Holevo information is not additive. Therefore, unlike entanglement-assisted classical communication, concrete expressions for the classical capacity exist only for specific classes of channels.

The analysis of the classical communication protocol in the asymptotic setting is almost exactly the same as in the one-shot setting. This is due to the fact that n independent uses of the channel $\mathcal{N}$ can be regarded as a single use of the channel $\mathcal{N}^{\otimes n}$. So the only change that needs to be made is to replace $\mathcal{N}$ with $\mathcal{N}^{\otimes n}$ and to define the states and POVM elements as acting on n systems instead of just one. In particular, the state at the end of the protocol presented in (12.1.8)–(12.1.9) at the beginning of Section 12.1 is

$$\omega^p_{M\widehat{M}} = (\mathcal{D}_{B^n\to\widehat{M}} \circ \mathcal{N}^{\otimes n}_{A\to B} \circ \mathcal{E}_{M'\to A^n})(\overline{\Phi}^p_{MM'}), \tag{12.2.1}$$

where p is the prior probability distribution over the message set $\mathcal{M}$, the encoding channel $\mathcal{E}_{M'\to A^n}$ is defined as

$$\mathcal{E}_{M'\to A^n}(|m\rangle\langle m|_{M'}) = \rho^m_{A^n} \quad \forall\, m \in \mathcal{M}, \tag{12.2.2}$$

and the decoding channel $\mathcal{D}_{B^n\to\widehat{M}}$, with associated POVM $\{\Lambda^m_{B^n}\}_{m\in\mathcal{M}}$, is defined as

$$\mathcal{D}_{B^n\to\widehat{M}}(\tau_{B^n}) = \sum_{m\in\mathcal{M}} \mathrm{Tr}[\Lambda^m_{B^n}\tau_{B^n}]|m\rangle\langle m|_{\widehat{M}}. \tag{12.2.3}$$

Then, for every given code specified by the encoding and decoding channels, the definitions of the message error probability of the code, the average error probability of the code, and the maximal error probability of the code all follow analogously from their definitions in (12.1.11), (12.1.13), and (12.1.15), respectively, in the one-shot setting.

Definition 12.7 $(n, |\mathcal{M}|, \varepsilon)$ Classical Communication Protocol

A classical communication protocol $(\mathcal{M}, \mathcal{E}_{M\to A^n}, \mathcal{D}_{B^n\to\widehat{M}})$ over n uses of the channel $\mathcal{N}_{A\to B}$ is called an $(n, |\mathcal{M}|, \varepsilon)$ *protocol*, with $\varepsilon \in [0, 1]$, if $p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}) \leq \varepsilon$.

Just as in the case of entanglement-assisted classical communication, the *rate* of a classical communication protocol over n uses of a channel is simply the number of bits that can transmitted per channel use, i.e.,

$$R(n, |\mathcal{M}|) := \frac{1}{n}\log_2 |\mathcal{M}|. \tag{12.2.4}$$

Given a channel $\mathcal{N}_{A\to B}$ and $\varepsilon \in [0, 1]$, the maximum rate of classical communication over $\mathcal{N}$ among all $(n, |\mathcal{M}|, \varepsilon)$ protocols is

$$C^{n,\varepsilon}(\mathcal{N}) := \frac{1}{n}C^{\varepsilon}(\mathcal{N}^{\otimes n}) = \sup_{(\mathcal{M},\mathcal{E},\mathcal{D})} \left\{\frac{1}{n}\log_2 |\mathcal{M}| : p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) \leq \varepsilon\right\}, \tag{12.2.5}$$

where the optimization is with respect to every classical communication protocol $(\mathcal{M}, \mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\widehat{M}})$ over $\mathcal{N}^{\otimes n}$, with $d_{M'} = d_{\widehat{M}} = |\mathcal{M}|$.

As with entanglement-assisted classical communication, the goal of a classical communication protocol is to maximize the rate while at the same time keeping the maximal error probability low. Ideally, we would want the error probability to vanish, and since we want to determine the highest possible rate, we are not concerned about the practical question regarding how many channel uses might be required, at least in the asymptotic setting. In particular, as we will see below, it might take an arbitrarily large number of channel uses to obtain the highest rate with a vanishing error probability.

Definition 12.8 Achievable Rate for Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for classical communication over* $\mathcal{N}$ if for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ classical communication protocol.

As we prove in Appendix A,

$$R \text{ acheivable rate} \iff \lim_{n\to\infty} \varepsilon^*_C(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) = 0 \quad \forall\, \delta > 0. \tag{12.2.6}$$

In other words, a rate R is achievable if the optimal error probability for a sequence of protocols with rate $R-\delta$, $\delta > 0$, vanishes as the number n of uses of $\mathcal{N}$ increases.

Definition 12.9 Classical Capacity of a Quantum Channel

The *classical capacity* of a quantum channel $\mathcal{N}$, denoted by $C(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$C(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{12.2.7}$$

The classical capacity can also be written as

$$C(\mathcal{N}) = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} C^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{12.2.8}$$

See Appendix A for a proof.

Definition 12.10 Weak Converse Rate for Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *weak converse rate for classical communication over* $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

We show in Appendix A that

$$R \text{ weak converse rate} \iff \lim_{n\to\infty} \varepsilon_C^*(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) > 0 \quad \forall\, \delta > 0. \tag{12.2.9}$$

In other words, a weak converse rate is a rate above which the optimal error probability cannot be made to vanish in the limit of a large number of channel uses.

Definition 12.11 Strong Converse Rate for Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for classical communication over* $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ classical communication protocol over $\mathcal{N}$.

We show in Appendix A that

$$R \text{ strong converse rate} \iff \lim_{n\to\infty} \varepsilon_C^*(2^{n(R+\delta)}; \mathcal{N}^{\otimes n}) = 1 \quad \forall\, \delta > 0. \tag{12.2.10}$$

In other words, unlike the weak converse, in which the optimal error probability is required to simply be bounded away from zero as the number n of channel uses increases, in order to have a strong converse rate the optimal error has to converge to one as n increases. By comparing (12.2.9) and (12.2.10), it is clear that every strong converse rate is a weak converse rate.

Definition 12.12 Strong Converse Classical Capacity of a Quantum Channel

The *strong converse classical capacity* of a quantum channel $\mathcal{N}$, denoted by $\widetilde{C}(\mathcal{N})$, is defined as the infimum of all strong converse rates,i.e.,

$$\widetilde{C}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{12.2.11}$$

As shown in general in Appendix A, the following inequality holds

$$C(\mathcal{N}) \leq \widetilde{C}(\mathcal{N}) \tag{12.2.12}$$

for every quantum channel $\mathcal{N}$. We can also write the strong converse classical capacity as

$$\widetilde{C}(\mathcal{N}) = \sup_{\varepsilon\in[0,1)} \limsup_{n\to\infty} \frac{1}{n} C^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{12.2.13}$$

See Appendix A for a proof.

Having defined the classical capacity of a quantum channel, as well as the strong converse capacity, we now state one of the main theorems of this chapter, which gives us a formal expression for the classical capacity of every quantum channel.

Theorem 12.13 Classical Capacity of a Quantum Channel

The classical capacity of a quantum channel $\mathcal{N}$ is equal to its *regularized Holevo information* $\chi_{\text{reg}}(\mathcal{N})$ of $\mathcal{N}$, i.e.,

$$C(\mathcal{N}) = \chi_{\text{reg}}(\mathcal{N}) \coloneqq \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{12.2.14}$$

REMARK: The quantity $\chi_{\text{reg}}(\mathcal{N}) \coloneqq \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n})$ is called the *regularization* of the Holevo information. It can be shown that the limit in the definition of $\chi_{\text{reg}}(\mathcal{N})$ does indeed exist (please consult the Bibliographic Notes in Section 12.5).

Note that, unlike the case of entanglement-assisted classical communication in Chapter 11, the right-hand side of (12.2.14) does not depend on only a single use of the channel $\mathcal{N}$. Rather, the capacity formula involves a limit over an arbitrarily large number of uses of the channel and is essentially impossible to compute in general, firstly because of the difficulty of computing the Holevo information of a channel and secondly due to the limit over an arbitrarily large number of uses of the channel. The issue is essentially that the Holevo information of a channel is not known to be additive for all channels, while the mutual information of a channel is known to have this property (we show this in Theorem 11.19). However, as we show below in Section 12.2.3, the Holevo information is *superadditive*, meaning that $\chi(\mathcal{N}^{\otimes n}) \geq n\chi(\mathcal{N})$. This implies that the Holevo information is always a lower bound on the quantum capacity of every channel $\mathcal{N}$:

$$C(\mathcal{N}) \geq \chi(\mathcal{N}) \text{ for all channels } \mathcal{N}. \tag{12.2.15}$$

Channels for which the Holevo information is known to be additive include the following:

1. All entanglement-breaking channels. (See Definition 4.12.)

2. All Hadamard channels. (See Definition 4.16.)

3. The depolarizing channel. (See (4.5.31).)

4. The erasure channel. (See (4.5.18).)

For all of these channels, we thus have that $C(\mathcal{N}) = \chi(\mathcal{N})$.

Also notice that, unlike Theorem 11.16 for entanglement-assisted classical communication, Theorem 12.13 only makes a statement about the classical capacity $C(\mathcal{N})$ of all channels, not about the strong converse classical capacity $\widetilde{C}(\mathcal{N})$. In the case of entanglement-assisted classical communication, proving that the mutual information of a channel is a strong converse rate involved proving that the sandwiched Rényi mutual information is additive for all channels, which we established in Theorem 11.22. Similarly, in the case of classical communication and attempting to follow a similar approach, the relevant quantity is the sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha$, defined in (7.11.95). Unlike the sandwiched Rényi mutual information, the sandwiched Rényi Holevo information is not known to be additive for all channels. However, as we show in Section 12.2.3.1, it is additive for all entanglement-breaking channels. It is also additive for Hadamard and depolarizing channels (please consult the Bibliographic Notes in Section 12.5). The best we can do, at the moment, is to say that the regularized Holevo information is a *weak* converse rate for all channels.

There are two ingredients to the proof of Theorem 12.13:

1. *Achievability*: We show that $\chi_{\text{reg}}(\mathcal{N})$ is an achievable rate. In general, to show that $R \in \mathbb{R}^+$ is achievable, we define encoding and decoding channels such that for all $\varepsilon \in (0, 1]$ and sufficiently large n, the encoding and decoding channels correspond to $(n, 2^{nr}, \varepsilon)$ protocols with rates $r < R$, as per Definition 12.8. Thus, if R is an achievable rate, then, given an error probability ε, it is possible to find an n large enough, along with encoding and decoding channels, such that the resulting protocol has rate arbitrarily close to R and maximal error probability bounded from above by ε.

   The achievability part of the proof establishes that $C(\mathcal{N}) \geq \chi_{\text{reg}}(\mathcal{N})$.

2. *Weak Converse*: We show that $\chi_{\text{reg}}(\mathcal{N})$ is a weak converse rate, from which it follows that $C(\mathcal{N}) \leq \chi_{\text{reg}}(\mathcal{N})$. To show that $\chi_{\text{reg}}(\mathcal{N})$ is a weak converse rate, we show that every achievable rate r satisfies $r \leq \chi_{\text{reg}}(\mathcal{N})$.

The achievability and weak converse proofs establish that the classical capacity is equal to the regularized Holevo information: $C(\mathcal{N}) = \chi_{\text{reg}}(\mathcal{N})$. Theorem 12.13 and the inequality in (12.2.12) allow us to conclude that

$$\widetilde{C}(\mathcal{N}) \geq \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}). \tag{12.2.16}$$

We first establish in Section 12.2.1 that the rate $\chi_{\text{reg}}(\mathcal{N})$ is achievable for classical communication over $\mathcal{N}$. Then, in Section 12.2.2, we prove that $\chi_{\text{reg}}(\mathcal{N})$ is a weak converse rate. We prove that the sandwiched Rényi Holevo information of a entanglement-breaking channel is additive in Section 12.2.3. With this additivity result, we prove in Section 12.2.4 that $C(\mathcal{N}) = \widetilde{C}(\mathcal{N}) = \chi(\mathcal{N})$ for all entanglement-breaking channels.

12.2.1 Proof of Achievability

In this section, we prove that $\chi_{\text{reg}}(\mathcal{N})$ is an achievable rate for classical communication over $\mathcal{N}$.

First, recall from Theorem 12.6 that for all $\varepsilon \in (0,1)$ and $\eta \in (0, \frac{\varepsilon}{2})$, there exists an $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$ such that

$$\log_2 |\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{\frac{\varepsilon}{2}-\eta}\right) - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) \tag{12.2.17}$$

for all $\alpha \in (0,1)$, where we recall from (12.1.101) that

$$\overline{\chi}_\alpha(\mathcal{N}) = \sup_{\rho_{XA}} \overline{I}_\alpha(X;B)_\omega \tag{12.2.18}$$

$$= \sup_{\rho_{XA}} D_\alpha(\mathcal{N}_{A\to B}(\rho_{XA})\|\rho_X \otimes \mathcal{N}_{A\to B}(\rho_A)), \tag{12.2.19}$$

where $\omega_{XB} \coloneqq \mathcal{N}_{A\to B}(\rho_{XA})$ and the optimization is over all classical–quantum states ρ_{XA}. A simple corollary of this result is the following.

Corollary 12.14 Lower Bound for Classical Communication in Asymptotic Setting

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in (0,1]$, $n \in \mathbb{N}$, and $\alpha \in (0,1)$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol over n uses of $\mathcal{N}$ such that

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)}\log_2\!\left(\frac{4}{\varepsilon}\right) - \frac{4}{n}. \tag{12.2.20}$$

PROOF: The inequality (12.2.17) holds for every channel $\mathcal{N}$, which means that it holds for $\mathcal{N}^{\otimes n}$. Applying the inequality in (12.2.17) to $\mathcal{N}^{\otimes n}$ and dividing both sides by n, we obtain

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \frac{1}{n}\overline{\chi}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{\frac{\varepsilon}{2}-\eta}\right) - \frac{1}{n}\log_2\!\left(\frac{4\varepsilon}{\eta^2}\right) \tag{12.2.21}$$

for all $\alpha \in (0,1)$. By restricting the optimization in the definition of $\overline{\chi}_\alpha(\mathcal{N}^{\otimes n})$ to tensor-power states, we find that $\overline{\chi}_\alpha(\mathcal{N}^{\otimes n}) \geq n\overline{\chi}_\alpha(\mathcal{N})$. This follows from the additivity of the Petz–Rényi relative entropy under tensor-product states (see Proposition 7.23). So we obtain

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{\frac{\varepsilon}{2}-\eta}\right) - \frac{1}{n}\log_2\!\left(\frac{4\varepsilon}{\eta^2}\right) \tag{12.2.22}$$

for all $\alpha \in (0,1)$. Now, letting $\eta = \frac{\varepsilon}{4}$ and using the fact that $\alpha - 1$ is negative for $\alpha \in (0,1)$, the following inequality holds for all $\alpha \in (0,1)$:

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)}\log_2\!\left(\frac{4}{\varepsilon}\right) - \frac{4}{n}. \tag{12.2.23}$$

In other words, for all $\varepsilon \in (0,1]$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol such that (12.2.20) is satisfied. This concludes the proof. ■

The inequality in (12.2.20) gives us, for every $\varepsilon \in (0,1]$ and $n \in \mathbb{N}$, a lower bound on the rate of a corresponding $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol, which is known to exist due to Proposition 12.5. If instead we fix a particular communication rate R by letting $|\mathcal{M}| = 2^{nR}$, then we can rearrange the inequality in (12.2.20) to obtain an exponentially decaying upper bound on the maximal error probability of the corresponding $(n, 2^{nR}, \varepsilon)$ classical communication protocol. Specifically, we find that

$$\varepsilon \leq 4 \cdot 2^{-(1-\alpha)\left(\overline{\chi}_\alpha(\mathcal{N})-R-\frac{4}{n}\right)} \tag{12.2.24}$$

for all $\alpha \in (0, 1)$.

The inequality in (12.2.20) implies that

$$C^{n,\varepsilon}(\mathcal{N}) \geq \overline{\chi}_\alpha(\mathcal{N}) - \frac{1}{n(\alpha-1)} \log_2\left(\frac{4}{\varepsilon}\right) - \frac{4}{n} \tag{12.2.25}$$

for all $n \geq 1$, $\varepsilon \in (0, 1)$, and $\alpha \in (0, 1)$.

We can now use (12.2.20) to prove that $\chi_{\text{reg}}(\mathcal{N})$ is an achievable rate for classical communication over $\mathcal{N}$.

Proof of the Achievability Part of Theorem 12.13

Fix $\varepsilon \in (0, 1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta = \delta_1 + \delta_2. \tag{12.2.26}$$

Set $\alpha \in (0, 1)$ such that

$$\delta_1 \geq \chi(\mathcal{N}) - \overline{\chi}_\alpha(\mathcal{N}), \tag{12.2.27}$$

which is possible because $\overline{\chi}_\alpha(\mathcal{N})$ is monotonically increasing in α (this follows from Proposition 7.23), because $\lim_{\alpha\to 1^-} \overline{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})$ (the proof of this is analogous to the one presented in Appendix 11.B). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{1}{n(1-\alpha)} \log_2\left(\frac{4}{\varepsilon}\right) + \frac{4}{n}. \tag{12.2.28}$$

Now, making use of the inequality in (12.2.20) of Corollary 12.14, there exists an $(n, |\mathcal{M}|, \varepsilon)$ protocol, with n and ε chosen as above, such that

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \overline{\chi}_\alpha(\mathcal{N}) - \frac{1}{n(1-\alpha)} \log_2\left(\frac{4}{\varepsilon}\right) - \frac{4}{n}. \tag{12.2.29}$$

Rearranging the right-hand side of this inequality, and using (12.2.26)–(12.2.28), we find that

$$\frac{1}{n}\log_2|\mathcal{M}| \geq \chi(\mathcal{N}) - \left(\chi(\mathcal{N}) - \overline{\chi}_\alpha(\mathcal{N}) + \frac{1}{n(1-\alpha)} \log_2\left(\frac{4}{\varepsilon}\right) + \frac{4}{n}\right) \tag{12.2.30}$$

$$\geq \chi(\mathcal{N}) - (\delta_1 + \delta_2) \tag{12.2.31}$$
$$= \chi(\mathcal{N}) - \delta. \tag{12.2.32}$$

We thus have $\chi(\mathcal{N}) - \delta \leq \frac{1}{n}\log_2 |\mathcal{M}|$. Recall that if an $(n, |\mathcal{M}|, \varepsilon)$ protocol exists, then an $(n, |\mathcal{M}'|, \varepsilon)$ also exists for all $\mathcal{M}'$ satisfying $|\mathcal{M}'| \leq |\mathcal{M}|$. We thus conclude that there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ classical communication with $R = \chi(\mathcal{N})$ for all sufficiently large n such that (12.2.28) holds. Since ε and δ are arbitrary, we conclude that, for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(\chi(\mathcal{N})-\delta)}, \varepsilon)$ classical communication protocol. This means that $\chi(\mathcal{N})$ is an achievable rate, and thus that $C(\mathcal{N}) \geq \chi(\mathcal{N})$.

Now, we can repeat the arguments above for the tensor-power channel $\mathcal{N}^{\otimes k}$ with $k \geq 1$, and we conclude that $\frac{1}{k}\chi(\mathcal{N}^{\otimes k})$ is an achievable rate. Since this holds for all k, we conclude that $\lim_{k\to\infty} \frac{1}{k}\chi(\mathcal{N}^{\otimes k}) = \chi_{\text{reg}}(\mathcal{N})$ is an achievable rate. Therefore, $C(\mathcal{N}) \geq \chi_{\text{reg}}(\mathcal{N})$.

12.2.1.1 Achievability from a Different Point of View

Using arguments similar to those given in Appendix 11.C, we can make the following statement: there exists a sequence $\{(n, 2^{nR_n}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols over $\mathcal{N}$, such that $\liminf_{n\to\infty} R_n \geq \chi(\mathcal{N})$ and $\lim_{n\to\infty} \varepsilon_n = 0$. If we consider a sequence $\{(n, 2^{nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols, this time keeping the rate at an arbitrary (but fixed) value $R < \chi(\mathcal{N})$ and varying the error probability, we conclude that there exists a sequence of protocols for which the error probabilities ε_n approach zero exponentially fast as $n \to \infty$.

12.2.2 Proof of the Weak Converse

We now show that the regularized Holevo information $\chi_{\text{reg}}(\mathcal{N})$ is a weak converse rate. The result is to establish that $C(\mathcal{N}) \leq \chi_{\text{reg}}(\mathcal{N})$ and therefore that $C(\mathcal{N}) = \chi_{\text{reg}}(\mathcal{N})$, completing the proof of Theorem 12.13.

Let us first recall from Theorem 12.4 that for every quantum channel $\mathcal{N}$ we have the following: for all $\varepsilon \in [0, 1)$ and $(|\mathcal{M}|, \varepsilon)$ classical communication protocols

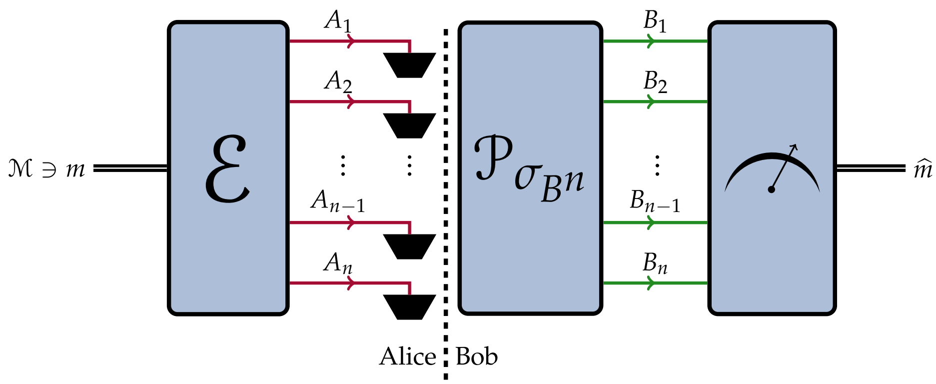


FIGURE 12.4: Depiction of a protocol that is useless for classical communication in the asymptotic setting. The state encoding the message m via $\mathcal{E}$ is discarded and replaced by an arbitrary (but fixed) state σ_{B^n}.

over $\mathcal{N}$,

$$\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(\chi(\mathcal{N}) + h_2(\varepsilon)\right), \tag{12.2.33}$$

$$\log_2 |\mathcal{M}| \leq \widetilde{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right), \quad \forall\, \alpha > 1. \tag{12.2.34}$$

To obtain these inequalities, we considered a classical communication protocol over a useless channel and used the hypothesis testing relative entropy to compare this protocol with the actual protocol over the channel $\mathcal{N}$. The useless channel in the asymptotic setting is analogous to the one in Figure 12.2 and is shown in Figure 12.4. A simple corollary of Theorem 12.4, which is relevant for the asymptotic setting, is the following.

Corollary 12.15 Upper Bounds for Classical Communication in Asymptotic Setting

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in [0,1)$, $n \in \mathbb{N}$, and $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols over n uses of $\mathcal{N}$, the rate of transmitted bits is bounded from above as follows:

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \frac{1}{1-\varepsilon}\left(\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{12.2.35}$$

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \frac{1}{n}\widetilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{12.2.36}$$

Proof: Since the inequalities in (12.2.33) and (12.2.34) of Theorem 12.4 hold for every channel $\mathcal{N}$, they hold for the channel $\mathcal{N}^{\otimes n}$. Therefore, applying (12.2.33) and (12.2.34) to $\mathcal{N}^{\otimes n}$ and dividing both sides by n, we immediately obtain the desired result. ■

The inequalities in the corollary above give us, for every $\varepsilon \in [0, 1)$ and $n \in \mathbb{N}$, an upper bound on the size $|\mathcal{M}|$ of the message set we can take for an arbitrary $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol. If instead we fix a particular communication rate R by letting $|\mathcal{M}| = 2^{nR}$, then we can obtain a lower bound on the maximal error probability of an arbitrary $(n, 2^{nR}, \varepsilon)$ classical communication protocol. Specifically, using (12.2.36), we find that

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{n}\widetilde{\chi}_\alpha(\mathcal{N}^{\otimes n})\right)} \tag{12.2.37}$$

for all $\alpha > 1$.

The inequalities in (12.2.35) and (12.2.36) imply that

$$C^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{1-\varepsilon}\left(\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{12.2.38}$$

$$C^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{n}\widetilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{12.2.39}$$

where $n \geq 1$ and $\varepsilon \in (0, 1)$.

Using (12.2.35), we can now prove the weak converse part of Theorem 12.13.

Proof of the Weak Converse Part of Theorem 12.13

Suppose that R is an achievable rate. Then, by definition, for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ classical communication protocol over $\mathcal{N}$. For all such protocols, the inequality (12.2.35) in Corollary 12.15 holds, so that

$$R - \delta \leq \frac{1}{1-\varepsilon}\left(\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right). \tag{12.2.40}$$

Since this bound holds for all n, it holds in the limit $n \to \infty$, so that

$$R \leq \lim_{n\to\infty} \frac{1}{1-\varepsilon}\left(\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right) + \delta \tag{12.2.41}$$

$$= \frac{1}{1-\varepsilon} \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}) + \delta. \tag{12.2.42}$$

Then, since this inequality holds for all $\varepsilon, \delta > 0$, we then conclude that

$$R \leq \lim_{\varepsilon,\delta\to 0} \left\{ \frac{1}{1-\varepsilon} \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}) + \delta \right\} = \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{12.2.43}$$

We have thus shown that if R is an achievable rate, then $R \leq \chi_{\text{reg}}(\mathcal{N})$. The contrapositive of this statement is that if $R > \chi_{\text{reg}}(\mathcal{N})$, then R is not an achievable rate. By definition, therefore, $\chi_{\text{reg}}(\mathcal{N})$ is a weak converse rate.

Recall that Theorem 12.13 only gives an expression for the capacity $C(\mathcal{N})$, and not for the strong converse capacity $\widetilde{C}(\mathcal{N})$. The sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha(\mathcal{N})$ of a channel $\mathcal{N}$ can be used to obtain the upper bound in (12.2.36), holding for every $(n, |\mathcal{M}|, \varepsilon)$ protocol. This inequality then leads to an expression for the strong converse capacity in the case that $\widetilde{\chi}_\alpha(\mathcal{N})$ happens to be additive for $\mathcal{N}$. We now, therefore, address this question regarding the additivity of the sandwiched Rényi Holevo information.

12.2.3 The Additivity Question

Although we have shown that the classical capacity $C(\mathcal{N})$ of a channel $\mathcal{N}$ is given by the regularized Holevo information $\chi_{\text{reg}}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n})$, as mentioned earlier, without the additivity of $\chi(\mathcal{N})$ this result is not particularly helpful since it is not known how to compute the regularized Holevo information in general.

Note, however, that for all channels $\mathcal{N}_1$ and $\mathcal{N}_2$ we always have the *superadditivity* of the Holevo information, i.e.,

$$\chi(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \chi(\mathcal{N}_1) + \chi(\mathcal{N}_2). \tag{12.2.44}$$

This follows by performing exactly the same steps in (11.2.44)–(11.2.52), but with the systems R_1 and R_2 therein taken to be classical systems. Therefore, to prove the additivity of χ for a channel $\mathcal{N}$, it suffices to show that $\chi(\mathcal{N} \otimes \mathcal{M}) \leq \chi(\mathcal{N}) + \chi(\mathcal{M})$.

Similarly, the sandwiched Rényi Holevo information is superadditive; i.e., for all $\alpha \geq 1$ and all channels $\mathcal{N}_1$ and $\mathcal{N}_2$, it holds that

$$\widetilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \widetilde{\chi}_\alpha(\mathcal{N}_1) + \widetilde{\chi}_\alpha(\mathcal{N}_2). \tag{12.2.45}$$

First, recall that

$$\widetilde{\chi}_\alpha(\mathcal{N}) = \sup_{\rho_{XA}} \widetilde{I}_\alpha(X;B)_\omega, \tag{12.2.46}$$

where $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$, and where we optimize over classical–quantum states $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$, with $\mathcal{X}$ a finite alphabet with associated $|\mathcal{X}|$-dimensional system X and $\{\rho_A^x\}_{x\in\mathcal{X}}$ a set of states. Also, recall that for every state ρ_{AB},

$$\widetilde{I}_\alpha(A;B)_\rho = \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB}\|\rho_A \otimes \sigma_B). \tag{12.2.47}$$

Now, the proof of (12.2.45) proceeds similarly to the proof of the corresponding inequality (12.2.44) for the Holevo information. By restricting the optimization in the definition of $\widetilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2)$ to product states, and letting $\rho'_{X_1X_2B_1B_2}$ be defined as

$$\rho'_{X_1X_2B_1B_2} \coloneqq ((\mathcal{N}_1)_{A_1\to B_1} \otimes (\mathcal{N}_2)_{A_2\to B_2})(\rho_{X_1X_2A_1A_2}), \tag{12.2.48}$$

for a classical–quantum state $\rho_{X_1X_2A_1A_2}$ (X systems classical and A systems quantum), we find that

$$\begin{aligned} \widetilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) &= \sup_{\rho} \widetilde{I}_\alpha(X_1X_2;B_1B_2)_{\rho'} && (12.2.49)\\ &\geq \sup_{\tau\otimes\omega} \widetilde{I}_\alpha(X_1X_2;B_1B_2)_{\xi'\otimes\omega'}, && (12.2.50) \end{aligned}$$

where $\tau'_{X_1B_1} \coloneqq (\mathcal{N}_1)_{A_1\to B_1}(\xi_{X_1A_2})$ and $\omega'_{X_2B_2} \coloneqq (\mathcal{N}_2)_{A_2\to B_2}(\omega_{X_2A_2})$. Proposition 11.21 states that the sandwiched Rényi mutual information $\widetilde{I}_\alpha$ is additive for product states, meaning that

$$\widetilde{I}_\alpha(A_1A_2;B_1B_2)_{\tau\otimes\omega} = \widetilde{I}_\alpha(A_1;B_1)_\tau + \widetilde{I}_\alpha(A_2;B_2)_\omega \tag{12.2.51}$$

for every state $\tau_{A_1B_1} \otimes \omega_{A_2B_2}$. Using this, we find that

$$\begin{aligned} \widetilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) &\geq \sup_{\tau,\omega} \left\{\widetilde{I}_\alpha(X_1;B_1)_{\tau'} + \widetilde{I}_\alpha(X_2;B_2)_{\omega'}\right\} && (12.2.52)\\ &= \sup_{\tau} \widetilde{I}_\alpha(X_1;B_1)_{\tau'} + \sup_{\omega} \widetilde{I}_\alpha(X_2;B_2)_{\omega'} && (12.2.53)\\ &= \widetilde{\chi}_\alpha(\mathcal{N}_1) + \widetilde{\chi}_\alpha(\mathcal{N}_2), && (12.2.54) \end{aligned}$$

i.e.,

$$\widetilde{\chi}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \widetilde{\chi}_\alpha(\mathcal{N}_1) + \widetilde{\chi}_\alpha(\mathcal{N}_2), \tag{12.2.55}$$

as required.

We see that in order to show the additivity of the sandwiched Rényi Holevo information for $\mathcal{N}$, it suffices to show *subadditivity for* $\mathcal{N}$, i.e.,

$$\widetilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) \leq n\widetilde{\chi}_\alpha(\mathcal{N}) \quad \forall\, n \geq 1. \tag{12.2.56}$$

We now show that subadditivity, and thus additivity, of the sandwiched Rényi Holevo information holds for all entanglement-breaking channels.

12.2.3.1 Entanglement-Breaking Channels

In this section, we prove that the sandwiched Rényi Holevo information is additive for all entanglement-breaking channels.

Theorem 12.16 Additivity of $\widetilde{\chi}_\alpha$ for Entanglement-Breaking Channels

For an entanglement-breaking channel $\mathcal{N}$ and an arbitrary channel $\mathcal{M}$, the following equality holds for all $\alpha > 1$:

$$\widetilde{\chi}_\alpha(\mathcal{N} \otimes \mathcal{M}) = \widetilde{\chi}_\alpha(\mathcal{N}) + \widetilde{\chi}_\alpha(\mathcal{M}). \tag{12.2.57}$$

The proof of this theorem relies on two lemmas, the first of which states that the sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha(\mathcal{N})$ of a channel $\mathcal{N}$ is equal to a quantity $\widetilde{K}_\alpha(\mathcal{N})$, called the *sandwiched Rényi information radius of* $\mathcal{N}$.

Lemma 12.17

For every quantum channel $\mathcal{N}$ and $\alpha > 1$, the following equality holds

$$\widetilde{\chi}_\alpha(\mathcal{N}) = \inf_\sigma \sup_\rho \widetilde{D}_\alpha(\mathcal{N}(\rho)\|\sigma) =: \widetilde{K}_\alpha(\mathcal{N}), \tag{12.2.58}$$

where the optimizations are over states ρ and σ. The quantity $\widetilde{K}_\alpha(\mathcal{N})$ is called the *sandwiched Rényi information radius of* $\mathcal{N}$.

PROOF: To prove this lemma, we show that $\widetilde{\chi}_\alpha(\mathcal{N}) \leq \widetilde{K}_\alpha(\mathcal{N})$ and $\widetilde{\chi}_\alpha(\mathcal{N}) \geq \widetilde{K}_\alpha(\mathcal{N})$.

First, using the definition in (7.11.95) of $\widetilde{\chi}_\alpha(\mathcal{N})$, we find that for every state τ_B,

$\widetilde{\chi}_\alpha(\mathcal{N})$

$$= \sup_{\rho_{XA}} \widetilde{I}_\alpha(X;B)_\omega \tag{12.2.59}$$

$$= \sup_{\rho_{XA}} \inf_{\sigma_B} \widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \mathcal{N}(\rho_A^x) \middle\| \sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \sigma_B\right) \tag{12.2.60}$$

$$\leq \sup_{\rho_{XA}} \widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \mathcal{N}(\rho_A^x) \middle\| \sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \tau_B\right), \tag{12.2.61}$$

where $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$ and the supremum is over all classical–quantum states ρ_{XA} of the form $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \rho_A^x$, with $\mathcal{X}$ a finite alphabet with associated $|\mathcal{X}|$-dimensional quantum system X and $\{\rho_A^x\}_{x\in\mathcal{X}}$ is a set of states. Now, recall from (7.5.174) that the sandwiched Rényi relative entropy is jointly quasi-convex for $\alpha > 1$ and invariant under tensoring in the same state $|x\rangle\!\langle x|$, which implies that

$$\widetilde{\chi}_\alpha(\mathcal{N}) \leq \sup_{\rho_{XA}} \widetilde{D}_\alpha\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \mathcal{N}(\rho_A^x) \middle\| \sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \tau_B\right) \tag{12.2.62}$$

$$\leq \sup_{\rho_{XA}} \max_{x\in\mathcal{X}} \widetilde{D}_\alpha(\mathcal{N}(\rho_A^x)\|\tau_B) \tag{12.2.63}$$

$$\leq \sup_{\rho_A} \widetilde{D}_\alpha(\mathcal{N}(\rho_A)\|\tau_B). \tag{12.2.64}$$

The final inequality above holds for every state τ_B, which implies that

$$\widetilde{\chi}_\alpha(\mathcal{N}) \leq \inf_{\tau_B} \sup_{\rho_A} \widetilde{D}_\alpha(\mathcal{N}(\rho_A)\|\tau_B) = \widetilde{K}_\alpha(\mathcal{N}), \tag{12.2.65}$$

i.e., $\widetilde{\chi}_\alpha(\mathcal{N}) \leq \widetilde{K}_\alpha(\mathcal{N})$.

We now show that $\widetilde{K}_\alpha(\mathcal{N}) \leq \widetilde{\chi}_\alpha(\mathcal{N})$. First, consider that

$$\widetilde{K}_\alpha(\mathcal{N}) = \inf_{\sigma_B} \sup_{\rho_A} \widetilde{D}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B) \tag{12.2.66}$$

$$= \frac{1}{\alpha-1} \inf_{\sigma_B} \sup_{\rho_A} \log_2 \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B) \tag{12.2.67}$$

$$= \frac{1}{\alpha-1} \log_2 \inf_{\sigma_B} \sup_{\rho_A} \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B). \tag{12.2.68}$$

Now, by taking a supremum over all probability measures μ on the set of all states ρ_A, we find that

$$\sup_{\rho_A} \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B) \leq \sup_{\mu} \int \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B)\, \mathrm{d}\mu(\rho_A). \tag{12.2.69}$$

So we have that

$$\widetilde{K}_\alpha(\mathcal{N}) \leq \frac{1}{\alpha - 1} \log_2 \inf_{\sigma_B} \sup_\mu \int \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B)\, \mathrm{d}\mu(\rho_A). \tag{12.2.70}$$

We now apply the Sion minimax theorem (Theorem 2.24) to exchange $\inf_{\sigma_B}$ and $\sup_\mu$. This theorem is applicable because the function

$$(\mu, \sigma_B) \mapsto \int \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B)\, \mathrm{d}\mu(\rho_A) \tag{12.2.71}$$

is linear in the measure μ and convex in the states σ_B. The latter is indeed true because

$$\widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B) = \left\|\mathcal{N}(\rho)^{\frac{1}{2}} \sigma_B^{\frac{1-\alpha}{\alpha}} \mathcal{N}(\rho_A)^{\frac{1}{2}}\right\|_\alpha^\alpha, \tag{12.2.72}$$

for all $\alpha > 1$ the function $\sigma_B \mapsto \sigma_B^{\frac{1-\alpha}{\alpha}}$ is operator convex, the Schatten norm $\|\cdot\|_\alpha$ is convex, and the function $x \mapsto x^\alpha$ is convex for all $x \geq 0$. So we find that

$$\widetilde{K}_\alpha(\mathcal{N}) \leq \frac{1}{\alpha - 1} \log_2 \sup_\mu \inf_{\sigma_B} \int \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B)\, \mathrm{d}\mu(\rho_A). \tag{12.2.73}$$

Now, by Carathéodory's theorem (Theorem 2.23), if ρ is a density operator acting on a d-dimensional space, then there exists an alphabet $\mathcal{X}$ of size no more than d^2, a probability distribution $p : \mathcal{X} \to [0,1]$ on $\mathcal{X}$, and an ensemble $\{(p(x), \rho_A^x)\}_{x\in\mathcal{X}}$ of states such that

$$\int \widetilde{Q}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B)\, \mathrm{d}\mu(\rho) = \sum_{x\in\mathcal{X}} p(x) \widetilde{Q}_\alpha(\mathcal{N}(\rho_A^x)\|\sigma_B). \tag{12.2.74}$$

Therefore,

$$\widetilde{K}_\alpha(\mathcal{N}) \leq \frac{1}{\alpha - 1} \log_2 \sup_{\{(p(x),\rho_A^x)\}_x} \inf_{\sigma_B} \sum_{x\in\mathcal{X}} p(x) \widetilde{Q}_\alpha(\mathcal{N}(\rho_A^x)\|\sigma_B) \tag{12.2.75}$$

$$= \frac{1}{\alpha - 1} \log_2 \sup_{\{(p(x),\rho_A^x)\}_x} \inf_{\sigma_B} \widetilde{Q}_\alpha(\mathcal{N}_{A\to B}(\rho_{XB})\|\rho_X \otimes \sigma_B) \tag{12.2.76}$$

$$= \sup_{\rho_{XA}} \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{XA}\|\rho_X \otimes \sigma_B) \tag{12.2.77}$$

$$= \sup_{\rho_{XA}} \widetilde{I}_\alpha(X;B)_\omega \tag{12.2.78}$$

$$= \widetilde{\chi}_\alpha(\mathcal{N}), \tag{12.2.79}$$

where $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$, and to obtain the first equality we used the direct-sum property of $\widetilde{Q}_\alpha$ in (7.5.41). So we have $\widetilde{K}_\alpha(\mathcal{N}) \leq \widetilde{\chi}_\alpha(\mathcal{N})$ in addition to $\widetilde{K}_\alpha(\mathcal{N}) \geq \widetilde{\chi}_\alpha(\mathcal{N})$, which means that $\widetilde{K}_\alpha(\mathcal{N}) = \widetilde{\chi}_\alpha(\mathcal{N})$, as required. ■

Using the fact that $\lim_{\alpha\to 1^+} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})$ (the proof of this analogous to the one presented in Appendix 11.B), we obtain the following alternate formula for the Holevo information of a quantum channel $\mathcal{N}$:

$$\chi(\mathcal{N}) = \inf_\sigma \sup_\rho D(\mathcal{N}(\rho)\|\sigma), \tag{12.2.80}$$

where the optimizations are over states ρ and σ.

Before stating the following lemma, let us define an entanglement-breaking map $\mathcal{N}_{A\to B}$ to be a completely positive map such that $\mathcal{N}_{A\to B}(X_{RA})$ is a separable operator of systems R and B for every positive semi-definite input operator X_{RA}. We can think of it as a generalization of an entanglement-breaking channel (Definition 4.12) in which there is no requirement of trace preservation.

Lemma 12.18

Let $\mathcal{M}_{A\to B}$ be a completely positive map, and let P_{RA} be a positive semi-definite separable operator, i.e., such that it can be written in the following form:

$$P_{RA} = \sum_{x\in\mathcal{X}} C_R^x \otimes D_A^x, \tag{12.2.81}$$

where $\mathcal{X}$ is a finite alphabet and $C_R^x, D_A^x \geq 0$ for all $x \in \mathcal{X}$. Let $P_R = \mathrm{Tr}_A[P_{RA}] = \sum_{x\in\mathcal{X}} \mathrm{Tr}[D_A^x] C_R^x$. For all $\alpha \geq 1$, the following inequality holds

$$\|\mathcal{M}_{A\to B}(P_{RA})\|_\alpha \leq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \|P_R\|_\alpha\,, \tag{12.2.82}$$

where

$$\nu_\alpha(\mathcal{P}) := \sup_\rho \|\mathcal{P}(\rho)\|_\alpha\,, \tag{12.2.83}$$

$\mathcal{P}$ is a completely positive map, and the supremum is taken over every density operator in the domain of $\mathcal{P}$. As a consequence, if $\mathcal{N}_{A'\to B'}$ is an entanglement-breaking map, then the following equality holds for all $\alpha \geq 1$:

$$\nu_\alpha(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'}) = \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \nu_\alpha(\mathcal{N}_{A'\to B'}). \tag{12.2.84}$$

Proof: Without loss of generality, we can suppose that each D_A^x is normalized, in the sense that $\operatorname{Tr}[D_A^x] = 1$. If it is not the case, then we can redefine C_R^x as $C_R^x \operatorname{Tr}[D_A^x]$ and D_A^x as $D_A^x/\operatorname{Tr}[D_A^x]$ without changing the separable operator P_{RA}. Next, we observe that

$$\mathcal{M}_{A\to B}(P_{RA}) = \sum_{x\in\mathcal{X}} C_R^x \otimes \mathcal{M}_{A\to B}(D_A^x) \tag{12.2.85}$$

$$= VTV^\dagger, \tag{12.2.86}$$

where

$$V := \sum_{x\in\mathcal{X}} \langle x| \otimes \sqrt{C_R^x} \otimes \mathbb{1}_B, \tag{12.2.87}$$

$$T := \sum_{x\in\mathcal{X}} |x\rangle\langle x| \otimes \mathbb{1}_R \otimes \mathcal{M}_{A\to B}(D_A^x). \tag{12.2.88}$$

This implies that

$$\operatorname{Tr}[(\mathcal{M}_{A\to B}(P_{RA}))^\alpha] = \operatorname{Tr}[(VTV^\dagger)^\alpha]. \tag{12.2.89}$$

Now let us apply the Araki–Lieb–Thirring inequality (Lemma 2.15), which states that for all positive semi-definite operators X and Y,

$$\operatorname{Tr}\left[\left(Y^{\frac{1}{2}} X Y^{\frac{1}{2}}\right)^{rq}\right] \leq \operatorname{Tr}\left[\left(Y^{\frac{r}{2}} X^r Y^{\frac{r}{2}}\right)^{q}\right] \tag{12.2.90}$$

for all $q \geq 0$ and $r \geq 1$. For $q = 1$, we obtain

$$\operatorname{Tr}\left[\left(Y^{\frac{1}{2}} X Y^{\frac{1}{2}}\right)^{r}\right] \leq \operatorname{Tr}[X^r Y^r]\,. \tag{12.2.91}$$

Now, for every operator Z, note that $ZXZ^\dagger$ has the same non-zero eigenvalues as $\left(Z^\dagger Z\right)^{\frac{1}{2}} X \left(Z^\dagger Z\right)^{\frac{1}{2}}$ (this follows by considering the polar decomposition of Z). In addition, since $Z^\dagger Z$ is positive semi-definite, applying (12.2.91) with $Y = Z^\dagger Z$ gives us

$$\operatorname{Tr}\left[\left(ZXZ^\dagger\right)^r\right] = \operatorname{Tr}\left[\left(\left(Z^\dagger Z\right)^{\frac{1}{2}} X \left(Z^\dagger Z\right)^{\frac{1}{2}}\right)^r\right] \tag{12.2.92}$$

$$\leq \operatorname{Tr}\left[X^r (Z^\dagger Z)^r\right]. \tag{12.2.93}$$

Substituting $r = \alpha$, $Z = V$, and $X = T$ into this inequality gives us

$$\operatorname{Tr}[(\mathcal{M}_{A\to B}(P_{RA}))^\alpha] = \operatorname{Tr}[(VTV^\dagger)^\alpha] \tag{12.2.94}$$

$$\leq \text{Tr}[(V^\dagger V)^\alpha T^\alpha]. \tag{12.2.95}$$

Letting

$$S := \sum_{x\in\mathcal{X}} \langle x| \otimes \sqrt{C_R^x}, \tag{12.2.96}$$

observe that

$$V^\dagger V = S^\dagger S \otimes \mathbb{1}_B, \tag{12.2.97}$$

which implies that

$$(V^\dagger V)^\alpha = (S^\dagger S)^\alpha \otimes \mathbb{1}_B. \tag{12.2.98}$$

Therefore, since T, and thus T^α, is block diagonal, we find that

$$\text{Tr}[(V^\dagger V)^\alpha T^\alpha]$$
$$= \text{Tr}\left[\left((S^\dagger S)^\alpha \otimes \mathbb{1}_B\right)\left(\sum_{x\in\mathcal{X}} |x\rangle\langle x| \otimes \mathbb{1}_R \otimes (\mathcal{M}_{A\to B}(D_A^x))^\alpha\right)\right] \tag{12.2.99}$$
$$= \sum_{x\in\mathcal{X}} \text{Tr}\left[\left((S^\dagger S)^\alpha\right)_x\right] \text{Tr}[(\mathcal{M}_{A\to B}(D_A^x))^\alpha], \tag{12.2.100}$$

where

$$\left((S^\dagger S)^\alpha\right)_x := (\langle x| \otimes \mathbb{1}_R)(S^\dagger S)^\alpha(|x\rangle \otimes \mathbb{1}_R). \tag{12.2.101}$$

Now,

$$\text{Tr}[(\mathcal{M}_{A\to B}(D_A^x))^\alpha]^{\frac{1}{\alpha}} = \left\|\mathcal{M}_{A\to B}(D_A^x)\right\|_\alpha \leq \nu_\alpha(\mathcal{M}_{A\to B}) \tag{12.2.102}$$
$$\Rightarrow \text{Tr}[(\mathcal{M}_{A\to B}(D_A^x))^\alpha] \leq \nu_\alpha(\mathcal{M}_{A\to B})^\alpha. \tag{12.2.103}$$

By taking the partial trace over system A of P_{RA}, we find that

$$P_R = \text{Tr}_A[P_{RA}] = \sum_{x\in\mathcal{X}} C_R^x = SS^\dagger. \tag{12.2.104}$$

Using this, we find that

$$\sum_{x\in\mathcal{X}} \text{Tr}\left[\left((S^\dagger S)^\alpha\right)_x\right] = \text{Tr}[(S^\dagger S)^\alpha] \tag{12.2.105}$$
$$= \text{Tr}[(SS^\dagger)^\alpha] \tag{12.2.106}$$
$$= \text{Tr}[P_R^\alpha] \tag{12.2.107}$$
$$= \|P_R\|_\alpha^\alpha. \tag{12.2.108}$$

Putting everything together, we conclude that

$$\|\mathcal{M}_{A\to B}(P_{RA})\|_\alpha = (\operatorname{Tr}[(\mathcal{M}_{A\to B}(P_{RA}))^\alpha])^{\frac{1}{\alpha}} \tag{12.2.109}$$
$$= \left(\operatorname{Tr}[(VTV^\dagger)^\alpha]\right)^{\frac{1}{\alpha}} \tag{12.2.110}$$
$$\leq \left(\operatorname{Tr}[(V^\dagger V)^\alpha T^\alpha]\right)^{\frac{1}{\alpha}} \tag{12.2.111}$$
$$\leq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \|P_R\|_\alpha . \tag{12.2.112}$$

To see the equality in (12.2.84), we prove it in two steps. First, consider that the following inequality holds for all completely positive maps $\mathcal{M}_{A\to B}$ and $\mathcal{N}_{A'\to B'}$:

$$\nu_\alpha(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'}) \geq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \nu_\alpha(\mathcal{N}_{A'\to B'}). \tag{12.2.113}$$

This follows simply by restricting the optimization in the definition of $\nu_\alpha(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'})$ to tensor-product states. Specifically,

$$\nu_\alpha(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'}) = \sup_{\rho_{AA'}} \|(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'})(\rho_{AA'})\|_\alpha \tag{12.2.114}$$
$$\geq \sup_{\sigma_A,\omega_{A'}} \|(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'})(\sigma_A \otimes \omega_{A'})\|_\alpha \tag{12.2.115}$$
$$= \sup_{\sigma_A,\omega_{A'}} \|(\mathcal{M}_{A\to B}(\sigma_A) \otimes \mathcal{N}_{A'\to B'}(\omega_{A'})\|_\alpha \tag{12.2.116}$$
$$= \sup_{\sigma_A} \|(\mathcal{M}_{A\to B}(\sigma_A)\|_\alpha \cdot \sup_{\omega_{A'}} \|\mathcal{N}_{A'\to B'}(\omega_{A'})\|_\alpha \tag{12.2.117}$$
$$= \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \nu_\alpha(\mathcal{N}_{A'\to B'}). \tag{12.2.118}$$

The following reverse inequality

$$\nu_\alpha(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'}) \leq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \nu_\alpha(\mathcal{N}_{A'\to B'}) \tag{12.2.119}$$

holds when $\mathcal{N}_{A'\to B'}$ is an entanglement-breaking map. Indeed, considering an arbitrary input state $\rho_{AA'}$, the output state $\omega_{AB'} := \mathcal{N}_{A'\to B'}(\rho_{AA'})$ is a separable operator. Applying (12.2.82) to the separable operator $\omega_{AB'}$ and identifying system B' with R in (12.2.82), we conclude that

$$\|(\mathcal{M}_{A\to B} \otimes \mathcal{N}_{A'\to B'})(\rho_{AA'})\|_\alpha = \|\mathcal{M}_{A\to B}(\omega_{AB'})\|_\alpha \tag{12.2.120}$$
$$\leq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \|\omega_{B'}\|_\alpha \tag{12.2.121}$$
$$= \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \|\mathcal{N}_{A'\to B'}(\rho_{A'})\|_\alpha \tag{12.2.122}$$
$$\leq \nu_\alpha(\mathcal{M}_{A\to B}) \cdot \nu_\alpha(\mathcal{N}_{A'\to B'}). \tag{12.2.123}$$

Since the inequality holds for every input state $\rho_{AA'}$, we conclude the inequality in (12.2.119). ■

With Lemmas 12.17 and 12.18 in hand, we can now prove Theorem 12.16.

Proof of Theorem 12.16

We start by using (7.5.3) to write the definition of $\widetilde{K}_\alpha(\mathcal{N})$ as

$$\widetilde{K}_\alpha(\mathcal{N}) = \inf_{\sigma_B} \sup_{\rho_A} \widetilde{D}_\alpha(\mathcal{N}(\rho_A)\|\sigma_B) \tag{12.2.124}$$

$$= \inf_{\sigma_B} \sup_{\rho_A} \frac{\alpha}{\alpha-1} \log_2 \left\| \sigma_B^{\frac{1-\alpha}{2\alpha}} \mathcal{N}(\rho_A) \sigma_B^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha . \tag{12.2.125}$$

Then, for every channel $\mathcal{M}$,

$$\widetilde{K}_\alpha(\mathcal{N}\otimes\mathcal{M})$$

$$= \inf_{\sigma_{A'B'}} \sup_{\rho_{AB}} \widetilde{D}_\alpha((\mathcal{N}\otimes\mathcal{M})(\rho_{AB})\|\sigma_{A'B'}) \tag{12.2.126}$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma_{A'B'}} \sup_{\rho_{AB}} \log_2 \left\| \sigma_{A'B'}^{\frac{1-\alpha}{2\alpha}} ((\mathcal{N}\otimes\mathcal{M})(\rho_{AB}))\, \sigma_{A'B'}^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \tag{12.2.127}$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma_{A'B'}} \log_2 \sup_{\rho_{AB}} \left\| \sigma_{A'B'}^{\frac{1-\alpha}{2\alpha}} ((\mathcal{N}\otimes\mathcal{M})(\rho_{AB}))\, \sigma_{A'B'}^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \tag{12.2.128}$$

$$\leq \frac{\alpha}{\alpha-1} \inf_{\sigma_{A'},\tau_{B'}} \log_2 \sup_{\rho_{AB}} \left\| \left(\sigma_{A'}^{\frac{1-\alpha}{2\alpha}} \otimes \tau_{B'}^{\frac{1-\alpha}{2\alpha}}\right) ((\mathcal{N}\otimes\mathcal{M})(\rho_{AB})) \left(\sigma_{A'}^{\frac{1-\alpha}{2\alpha}} \otimes \tau_{B'}^{\frac{1-\alpha}{2\alpha}}\right) \right\|_\alpha , \tag{12.2.129}$$

where to obtain the inequality we have restricted the infimum to tensor product states. Now, observe that since $\mathcal{N}$ is entanglement-breaking, then sandwiching the output of the channel by the positive semi-definite operator $\sigma_{A'}^{\frac{1-\alpha}{2\alpha}}$ leads to a new map $\mathcal{N}'$ that is a completely positive entanglement-breaking map (though not necessarily trace preserving). Similarly, sandwiching the output of $\mathcal{M}$ by the positive semi-definite operator $\tau_{B'}^{\frac{1-\alpha}{2\alpha}}$ leads to a new completely positive map $\mathcal{M}'$. Therefore, using Lemma 12.18, we obtain

$$\widetilde{K}_\alpha(\mathcal{N}\otimes\mathcal{M})$$

$$\leq \frac{\alpha}{\alpha-1} \inf_{\sigma_{A'},\tau_{B'}} \log_2 \left(\sup_{\rho_{AB}} \|(\mathcal{N}'\otimes\mathcal{M}')(\rho_{AB})\|_\alpha \right) \tag{12.2.130}$$

$$= \frac{\alpha}{\alpha-1} \inf_{\sigma_{A'},\tau_{B'}} \log_2 \nu_\alpha(\mathcal{N}'\otimes\mathcal{M}') \tag{12.2.131}$$

$$= \frac{\alpha}{\alpha - 1} \inf_{\sigma_{A'}, \tau_{B'}} \log_2(\nu_\alpha(\mathcal{N}')\nu_\alpha(\mathcal{M}')) \tag{12.2.132}$$

$$= \frac{\alpha}{\alpha - 1} \inf_{\sigma_{A'}, \tau_{B'}} \left[\log_2 \nu_\alpha(\mathcal{N}') + \log_2 \nu_\alpha(\mathcal{M}')\right] \tag{12.2.133}$$

$$= \inf_{\sigma_{A'}} \frac{\alpha}{\alpha - 1} \log_2 \sup_{\rho_A} \|\mathcal{N}'(\rho_A)\|_\alpha + \inf_{\tau_{B'}} \frac{\alpha}{\alpha - 1} \log_2 \sup_{\omega_B} \|\mathcal{M}'(\omega_B)\|_\alpha \tag{12.2.134}$$

$$= \inf_{\sigma_{A'}} \sup_{\rho_A} \frac{\alpha}{\alpha - 1} \log_2 \|\mathcal{N}'(\rho_A)\|_\alpha + \inf_{\tau_{B'}} \sup_{\omega_B} \frac{\alpha}{\alpha - 1} \log_2 \|\mathcal{M}'(\omega_B)\|_\alpha \tag{12.2.135}$$

$$= \widetilde{K}_\alpha(\mathcal{N}) + \widetilde{K}_\alpha(\mathcal{M}). \tag{12.2.136}$$

So we have that $\widetilde{K}_\alpha(\mathcal{N} \otimes \mathcal{M}) \leq \widetilde{K}_\alpha(\mathcal{N}) + \widetilde{K}_\alpha(\mathcal{M})$ for every channel $\mathcal{M}$. Using Lemma 12.17, we obtain the desired result.

Note that the additivity of the Holevo information of every entanglement-breaking channel follows from the additivity of the sandwiched Rényi Holevo information of such channels by taking the limit $\alpha \to 1^+$ (the proof is analogous to the one presented in Appendix 11.B).

12.2.4 Proof of the Strong Converse for Entanglement-Breaking Channels

Having shown that the sandwiched Rényi Holevo information is additive for all entanglement-breaking channels, we can now proceed further from (12.2.36) to prove a strong converse theorem for all entanglement-breaking channels. Moreover, since the sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha(\mathcal{N})$ satisfies $\lim_{\alpha \to 1^+} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})$ (the proof of this is analogous to the one presented in Appendix 11.B), we can go beyond the statement of Theorem 12.13 and say that $C(\mathcal{N}) = \chi(\mathcal{N})$ for all entanglement-breaking channels $\mathcal{N}$.

Theorem 12.19 Classical Capacity of Entanglement-Breaking Channels

For every entanglement-breaking channel $\mathcal{N}$,

$$C(\mathcal{N}) = \widetilde{C}(\mathcal{N}) = \chi(\mathcal{N}). \tag{12.2.137}$$

REMARK: Note that this theorem holds more generally for every channel $\mathcal{N}$ for which the sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha(\mathcal{N})$ is additive.

PROOF: Since $\lim_{\alpha\to 1^+} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})$ (the proof of this is analogous to the one presented in Appendix 11.B), we find that the Holevo information is additive for all entanglement-breaking channels. The equality $C(\mathcal{N}) = \chi(\mathcal{N})$ then follows from Theorem 12.13.

The remainder of the proof is devoted to establishing that $\chi(\mathcal{N})$ is a strong converse rate for classical communication over $\mathcal{N}$, from which it follows that $\widetilde{C}(\mathcal{N}) \leq \chi(\mathcal{N})$, which in turn implies, via (12.2.12), that $\widetilde{C}(\mathcal{N}) = \chi(\mathcal{N})$.

Fix $\varepsilon \in [0,1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \tag{12.2.138}$$

Set $\alpha \in (1,\infty)$ such that

$$\delta_1 \geq \widetilde{\chi}_\alpha(\mathcal{N}) - \chi(\mathcal{N}), \tag{12.2.139}$$

which is possible since $\widetilde{\chi}_\alpha(\mathcal{N})$ is monotonically increasing with α (this follows from Proposition 7.31), and since $\lim_{\alpha\to 1^+} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N})$. With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.140}$$

Now, with the values of n and ε chosen as above, every $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol satisfies (12.2.36) in Corollary 12.15. In particular, using the additivity of the sandwiched Rényi Holevo information for all $\alpha > 1$, we can write (12.2.36) as

$$\frac{1}{n}\log_2 |\mathcal{M}| \leq \widetilde{\chi}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.141}$$

Rearranging the right-hand side of this inequality, and using the assumptions in (12.2.138)–(12.2.140), we obtain

$$\frac{1}{n}\log_2 |\mathcal{M}| \leq \chi(\mathcal{N}) + \widetilde{\chi}_\alpha(\mathcal{N}) - \chi(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right) \tag{12.2.142}$$

$$\leq \chi(\mathcal{N}) + \delta_1 + \delta_2 \tag{12.2.143}$$

$$= \chi(\mathcal{N}) + \delta' \tag{12.2.144}$$
$$< \chi(\mathcal{N}) + \delta. \tag{12.2.145}$$

So we have that $\chi(\mathcal{N}) + \delta > \frac{1}{n} \log_2 |\mathcal{M}|$ for all $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols with n sufficiently large. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(\chi(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol for all sufficiently large n such that (12.2.140) holds, for if it did there would exist some message set $\mathcal{M}$ such that $\frac{1}{n} \log_2 |\mathcal{M}| = \chi(\mathcal{N}) + \delta$, which we have just seen is not possible. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in [0, 1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(\chi(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol. This means that $\chi(\mathcal{N})$ is a strong converse rate, which completes the proof. ■

12.2.4.1 The Strong Converse from a Different Point of View

Just as we did in the case of entanglement-assisted classical communication in Appendix 11.G, we can use the alternative definitions of classical capacity and strong converse classical capacity (stated in Appendix A) to see that $C(\mathcal{N}) = \widetilde{C}(\mathcal{N}) = \chi(\mathcal{N})$ for all channels $\mathcal{N}$ for which the sandwiched Rényi Holevo information is additive and that, as shown in Figure 12.5, the quantity $\chi(\mathcal{N})$ is a sharp dividing point between reliable, error-free communication and communication with error approaching one exponentially fast. Specifically, by following the arguments in Appendix 11.G, we obtain the following: for every sequence $\{(n, 2^{nR}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of $(n, |\mathcal{M}|, \varepsilon)$ protocols, with each element of the sequence having an arbitrary (but fixed) rate $R > \chi(\mathcal{N})$, the sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ of error probabilities approaches one at an exponential rate.

12.2.5 General Upper Bounds on the Strong Converse Classical Capacity

The difficulty in proving the additivity of the Holevo information for a general channel, and thus obtaining an upper bound on its classical capacity, has motivated the study of other, more tractable upper bounds on the classical capacity of a quantum channel. In this section, we present two upper bounds on the strong converse classical capacity of a quantum channel.

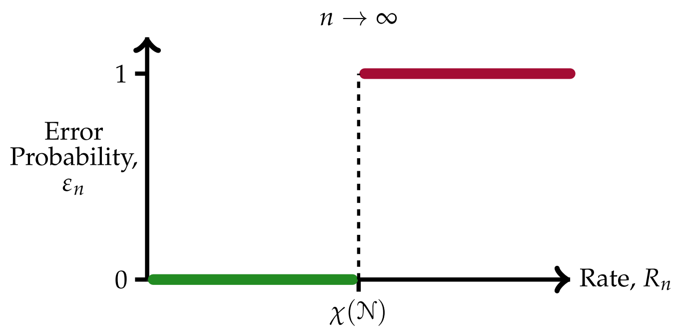


FIGURE 12.5: The error probability ε_n as a function of the rate R_n for classical communication over a quantum channel $\mathcal{N}$ for which the sandwiched Rényi Holevo information $\widetilde{\chi}_\alpha(\mathcal{N})$ is additive. As $n \to \infty$, for every rate below the Holevo information $\chi(\mathcal{N})$, there exists a sequence of protocols with error probability converging to zero. For every rate above the Holevo information $\chi(\mathcal{N})$, the error probability converges to one for all possible protocols.

12.2.5.1 Υ-Information Upper Bound

Recall from Proposition 12.3 that the following upper bound on the number of transmitted bits holds for every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol:

$$\log_2 |\mathcal{M}| \leq \chi_H^\varepsilon(\mathcal{N}), \tag{12.2.146}$$

where the ε-hypothesis testing Holevo information $\chi_H^\varepsilon(\mathcal{N})$ of the quantum channel $\mathcal{N}$ is defined in (7.11.93) as

$$\chi_H^\varepsilon(\mathcal{N}) = \sup_{\rho_{XA}} I_H^\varepsilon(X;B)_\omega \tag{12.2.147}$$

$$= \sup_{\rho_{XA}} \inf_{\sigma_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{XA}) \| \rho_X \otimes \sigma_B). \tag{12.2.148}$$

Here, $\omega_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$, and the optimization is over classical–quantum states ρ_{XA} of the form $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$, where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{(p(x), \rho_A^x)\}_{x\in\mathcal{X}}$ is an ensemble of states.

We derived the inequality in (12.2.146) by comparing the actual classical communication protocol over the channel $\mathcal{N}$ with a classical communication protocol over the replacement channel $\mathcal{R}$ (see (12.1.23) in Section 12.1.1). The replacement channel is useless for classical communication because it discards the state encoded with the message and replaces it with a fixed state. We can make the comparison between the channel $\mathcal{N}$ and the channel $\mathcal{R}$ for classical communication more explicit by writing the quantity $\chi_H^\varepsilon(\mathcal{H})$ as

$$\chi_H^\varepsilon(\mathcal{N}) = \sup_{\rho_{XA}} \inf_{\mathcal{R}_{A\to B}} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{XA})\|\mathcal{R}_{A\to B}(\rho_{XA})), \tag{12.2.149}$$

where $\mathcal{R}_{A\to B}$ has the action $\mathcal{R}_{A\to B}(\rho_{XA}) = \rho_X \otimes \sigma_B$.

Intuitively, an approach for obtaining an alternative upper bound on the strong converse classical capacity is to expand the set of useless channels from the replacement channels to the following set of completely positive trace non-increasing maps:

$$\mathfrak{F} := \{\mathcal{F}_{A\to B} : \exists\, \sigma_B \geq 0,\ \mathrm{Tr}[\sigma_B] \leq 1,\ \mathcal{F}_{A\to B}(\rho_A) \leq \sigma_B\, \forall \rho_A \in \mathrm{D}(\mathcal{H}_A)\}, \tag{12.2.150}$$

This set of maps, even though they contain completely positive, non-trace-preserving maps, can be thought of intuitively as also being useless for classical communication, and it contains the set of replacement channels. Using this set, we define the generalized Υ-information as follows:

Definition 12.20 Generalized Υ-Information

Let $\boldsymbol{D}$ be a generalized divergence (see Definition 7.15). For every quantum channel $\mathcal{N}_{A\to B}$, we define the *generalized Υ-information of* $\mathcal{N}$ as

$$\boldsymbol{\Upsilon}(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.2.151}$$

where the supremum is over pure states ψ_{RA}, with the dimension of R the same as the dimension of A.

Remark: Note that it suffices to optimize over pure states ψ_{RA}, with the dimension of R equal to the dimension of A, when calculating the generalized Υ-information of a channel, i.e., for general states ρ_{RA} (with the dimension of R not necessarily equal to the dimension of A),

$$\sup_{\rho_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{F}_{A\to B}(\rho_{RA})) = \boldsymbol{\Upsilon}(\mathcal{N}). \tag{12.2.152}$$

The proof of this proceeds analogously to the steps in (7.11.4)–(7.11.2) for proving that it suffices to optimize over pure states when calculating the generalized channel divergence.

In this section, we are interested in the following generalized Υ-information channel quantities:

1. The Υ-*information of* $\mathcal{N}$,

$$\Upsilon(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})). \tag{12.2.153}$$

2. The ε-*hypothesis testing* Υ-*information of* $\mathcal{N}$,

$$\Upsilon_H^\varepsilon(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D_H^\varepsilon(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})). \tag{12.2.154}$$

3. The *sandwiched Rényi* Υ-*information of* $\mathcal{N}$,

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.2.155}$$

where $\alpha \in [1/2, 1) \cup (1, \infty)$.

Proposition 12.21 Holevo Information and Υ-Information

The Υ-information $\Upsilon(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is greater than or equal to its Holevo information $\chi(\mathcal{N})$:

$$\Upsilon(\mathcal{N}) \geq \chi(\mathcal{N}). \tag{12.2.156}$$

PROOF: To see this, we apply Proposition 7.83 to obtain

$$\Upsilon(\mathcal{N}) = \sup_{\rho_A} \inf_{\mathcal{F}\in\mathfrak{F}} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{F}}_{AB}\sqrt{\rho_A}) \tag{12.2.157}$$

$$= \inf_{\mathcal{F}\in\mathfrak{F}} \sup_{\rho_A} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{F}}_{AB}\sqrt{\rho_A}). \tag{12.2.158}$$

Then, by employing (12.2.158), we find that

$$\Upsilon(\mathcal{N}) = \inf_{\mathcal{F}\in\mathfrak{F}} \sup_{\rho_A} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{F}}_{AB}\sqrt{\rho_A}) \tag{12.2.159}$$

$$\geq \inf_{\mathcal{F}\in\mathfrak{F}} \sup_{\rho_A} D(\mathcal{N}_{A\to B}(\rho_A)\|\mathcal{F}_{A\to B}(\rho_A)), \tag{12.2.160}$$

where, to obtain the inequality, we used the fact that

$$\sqrt{\rho_A}\Gamma^{\mathcal{N}}_{AB}\sqrt{\rho_A} = (\sqrt{\rho_{A'}} \otimes \mathbb{1}_B)\mathcal{N}_{A\to B}(\Gamma_{A'A})(\sqrt{\rho_{A'}} \otimes \mathbb{1}_B) \tag{12.2.161}$$

$$= \mathcal{N}_{A\to B}\left((\sqrt{\rho_{A'}} \otimes \mathbb{1}_A)\Gamma_{A'A}(\sqrt{\rho_{A'}} \otimes \mathbb{1}_A)\right) \tag{12.2.162}$$

$$= \mathcal{N}_{A\to B}\left((\mathbb{1}_{A'} \otimes \sqrt{\rho_A^{\mathsf{T}}})\Gamma_{AA}(\mathbb{1}_{A'} \otimes \sqrt{\rho_A^{\mathsf{T}}})\right). \tag{12.2.163}$$

In the last line we used the transpose trick (see (2.2.42). Then, we applied the monotonicity of the quantum relative entropy with respect to the partial trace Tr_A. Finally, we used the fact that ρ_A^{T} is a state for every ρ_A, so that the optimization over states remains unchanged.

Continuing, we have that

$$\Upsilon(\mathcal{N}) \geq \inf_{\mathcal{F}\in\mathfrak{F}} \sup_{\rho_A} D(\mathcal{N}_{A\to B}(\rho_A)\|\sigma_{\mathcal{F}}) \tag{12.2.164}$$

$$\geq \inf_{\sigma_B} \sup_{\rho_A} D(\mathcal{N}_{A\to B}(\rho_A)\|\sigma_B) \tag{12.2.165}$$

$$= \chi(\mathcal{N}). \tag{12.2.166}$$

To obtain the first inequality, we first used the fact that for every map $\mathcal{F} \in \mathfrak{F}$ there exists a state, which we call $\sigma_{\mathcal{F}}$, such that $\mathcal{F}_{A\to B}(\rho_A) \leq \sigma_{\mathcal{F}}$ for all input states ρ_A. We then used 2.(d) in Proposition 7.3. To obtain the last inequality, we simply enlarged the set over which the infimum is performed to include all states. Then, to obtain the equality on the last line, we used the expression in (12.2.80) for the Holevo information. ■

We now prove an analogue of Proposition 12.3 involving the ε-hypothesis testing Υ-information, $\Upsilon_H^{\varepsilon}(\mathcal{N})$, in place of the ε-hypothesis testing Holevo information $\chi_H^{\varepsilon}(\mathcal{N})$.

Proposition 12.22

Let $\mathcal{N}$ be a quantum channel. For every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$, the number of bits transmitted over $\mathcal{N}$ is bounded from above by the ε-hypothesis testing Υ-information of $\mathcal{N}$, i.e.,

$$\log_2 |\mathcal{M}| \leq \Upsilon_H^{\varepsilon}(\mathcal{N}). \tag{12.2.167}$$

Proof: For every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol, with encoding and decoding channel given by $\mathcal{E}$ and $\mathcal{D}$, respectively, the maximal error probability criterion $p_{\mathrm{err}}^*(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ holds. This implies $\overline{p}_{\mathrm{err}}((\mathcal{E}, \mathcal{D}); p, \mathcal{N}) \leq \varepsilon$ for the average probability, where $p : \mathcal{M} \to [0, 1]$ is the uniform prior probability distribution over the messages in $\mathcal{M}$. If the encoding channel $\mathcal{E}$ is defined such that we obtain the set $\{\rho_A^m\}_{m\in\mathcal{M}}$ of states associated to each message $m \in \mathcal{M}$ (see (12.1.2)), and the decoding channel $\mathcal{D}$ is defined by the POVM $\{\Lambda_B^{\widehat{m}}\}_{\widehat{m}\in\mathcal{M}}$, then we can write the average success probability $\overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N})$ of the code $(\mathcal{E},\mathcal{D})$ as

$$\overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) = 1 - \overline{p}_{\text{err}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) \tag{12.2.168}$$

$$= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \text{Tr}[\Lambda_B^m \mathcal{N}_{A\to B}(\rho_A^m)], \tag{12.2.169}$$

and we have that $\overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N}) \geq 1-\varepsilon$. Now, recall from (4.2.5) that we can write the action of $\mathcal{N}_{A\to B}$ in terms of its Choi representation $\Gamma_{AB}^{\mathcal{N}}$ as

$$\mathcal{N}_{A\to B}(\rho_A^m) = \text{Tr}_A\left[\left((\rho_A^m)^{\mathsf{T}} \otimes \mathbb{1}_B\right)\Gamma_{AB}^{\mathcal{N}}\right] \tag{12.2.170}$$

for all $m \in \mathcal{M}$. Also, let us define the average state

$$\overline{\rho}_A := \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \rho_A^m, \tag{12.2.171}$$

and a purification of it

$$|\overline{\phi}\rangle_{AA'} := (\mathbb{1}_A \otimes \sqrt{\overline{\rho}_{A'}})|\Gamma\rangle_{AA'} = \left(\sqrt{\overline{\rho}_A^{\mathsf{T}}} \otimes \mathbb{1}_{A'}\right)|\Gamma\rangle_{A'A}, \tag{12.2.172}$$

where we used the transpose trick in (2.2.42) to obtain the last equality. Then, observe that

$$\mathcal{N}_{A'\to B}(\overline{\phi}_{AA'}) = \sqrt{\overline{\rho}_A^{\mathsf{T}}}\mathcal{N}_{A'\to B}(\Gamma_{AA'})\sqrt{\overline{\rho}_A^{\mathsf{T}}} \tag{12.2.173}$$

$$= \sqrt{\overline{\rho}_A^{\mathsf{T}}}\Gamma_{AB}^{\mathcal{N}}\sqrt{\overline{\rho}_A^{\mathsf{T}}}, \tag{12.2.174}$$

which implies that the Choi operator $\Gamma_{AB}^{\mathcal{N}}$ can be written as[2]

$$\Gamma_{AB}^{\mathcal{N}} = (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}}\mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})(\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}}. \tag{12.2.175}$$

Therefore,

$$\overline{p}_{\text{succ}}((\mathcal{E},\mathcal{D});p,\mathcal{N})$$

$$= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \text{Tr}[\Lambda_B^m \mathcal{N}_{A\to B}(\rho_A^m)] \tag{12.2.176}$$

[2]Note that if $\overline{\rho}_A^{\mathsf{T}}$ is not invertible, then the inverse is understood to be on the support of $\overline{\rho}_A^{\mathsf{T}}$.

$$= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}\big[\big((\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\big)\, \Gamma_{AB}^{\mathcal{N}}\big] \tag{12.2.177}$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}\Big[\big((\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\big)\, (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})(\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}}\Big] \tag{12.2.178}$$

$$= \text{Tr}\left[(\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \left(\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} (\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\right) (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})\right] \tag{12.2.179}$$

$$= \text{Tr}[\Omega_{AB} \mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})], \tag{12.2.180}$$

where

$$\Omega_{AB} \coloneqq (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \left(\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} (\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\right) (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}}. \tag{12.2.181}$$

Note that Ω_{AB} is positive semi-definite, i.e., $\Omega_{AB} \geq 0$. Also, observe that since $\Lambda_B^m \leq \mathbb{1}_B$ for all $m \in \mathcal{M}$, we have that

$$\Omega_{AB} \leq (\overline{\rho}_A^{\mathsf{T}}) \left(\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} (\rho_A^m)^{\mathsf{T}} \otimes \mathbb{1}_B\right) (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} = \mathbb{1}_{AB}. \tag{12.2.182}$$

Together with $\Omega_{AB} \geq 0$, this means that Ω_{AB} is a measurement operator. So we have that

$$p_{\text{succ}}((\mathcal{E}, \mathcal{D}); p, \mathcal{N}) = \text{Tr}[\Omega_{AB} \mathcal{N}_{A\to B}(\overline{\phi}_{AA'})] \geq 1 - \varepsilon. \tag{12.2.183}$$

Now, let $\mathcal{F} \in \mathfrak{F}$. This means that there exists a state, call it σ_B, such that $\mathcal{F}(\rho_A) \leq \sigma_B$ for all states ρ_A. We find that

$$\text{Tr}[\Omega_{AB} \mathcal{F}_{A'\to B}(\overline{\phi}_{AA'})]$$

$$= \text{Tr}\left[(\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \left(\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} (\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\right) (\overline{\rho}_A^{\mathsf{T}})^{-\frac{1}{2}} \mathcal{F}_{A'\to B}(\overline{\phi}_{AA'})\right] \tag{12.2.184}$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}\big[\big((\rho_A^m)^{\mathsf{T}} \otimes \Lambda_B^m\big)\, \Gamma_{AB}^{\mathcal{F}}\big] \tag{12.2.185}$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}[\Lambda_B^m \mathcal{F}_{A\to B}(\rho_A^m)] \tag{12.2.186}$$

$$\leq \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}[\Lambda_B^m \sigma_B] \tag{12.2.187}$$

$$\leq \frac{1}{|\mathcal{M}|}, \tag{12.2.188}$$

where we used (12.2.175) to obtain the second equality and we used the fact that $\mathcal{F}(\rho_A) \leq \sigma_B$ for every input state ρ_A to obtain the second-to-last inequality.

Now, by optimizing the quantity $\text{Tr}[\Omega_{AB}\mathcal{F}_{A'\to B}(\overline{\phi}_{AA'})]$ over all measurement operators, subject to the constraint $\text{Tr}[\Omega_{AB}\mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})] \geq 1-\varepsilon$, we get that

$$\log_2 |\mathcal{M}| \leq D_H^{\varepsilon}(\mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})\|\mathcal{F}_{A'\to B}(\overline{\phi}_{AA'})). \tag{12.2.189}$$

Since this holds for every $\mathcal{F} \in \mathfrak{F}$, we have that

$$\log_2 |\mathcal{M}| \leq \inf_{\mathcal{F}\in\mathfrak{F}} D_H^{\varepsilon}(\mathcal{N}_{A'\to B}(\overline{\phi}_{AA'})\|\mathcal{F}_{A'\to B}(\overline{\phi}_{AA'})). \tag{12.2.190}$$

Finally, optimizing over all pure states $\overline{\phi}_{AA'}$, we conclude that

$$\log_2 |\mathcal{M}| \leq \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) = \Upsilon_H^{\varepsilon}(\mathcal{N}), \tag{12.2.191}$$

as required. ■

As an immediate consequence of Propositions 12.22, 7.70, and 7.71, we have the following two bounds:

Proposition 12.23

Let $\mathcal{N}$ be a quantum channel, let $\varepsilon \in [0,1)$, and let $\alpha > 1$. For every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol over $\mathcal{N}$, the following bounds hold

$$\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(\Upsilon(\mathcal{N}) + h_2(\varepsilon)\right), \tag{12.2.192}$$

$$\log_2 |\mathcal{M}| \leq \widetilde{\Upsilon}_{\alpha}(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.193}$$

In the asymptotic setting, the bounds above become the following:

$$\frac{1}{n}\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(\frac{1}{n}\Upsilon(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{12.2.194}$$

$$\frac{1}{n}\log_2 |\mathcal{M}| \leq \frac{1}{n}\widetilde{\Upsilon}_{\alpha}(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{12.2.195}$$

These upper bounds hold for every $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol over a quantum channel $\mathcal{N}$, where $n \in \mathbb{N}$ and $\varepsilon \in [0, 1)$.

Now, as with the Holevo information and the sandwiched Rényi Holevo information, we are faced with the additivity of the Υ-information and the sandwiched Rényi Υ-information. Our primary focus is on the latter, since we would like to make a statement about the strong converse for channels more general than entanglement-breaking channels. It turns out that the sandwiched Rényi Υ-information is additive for irreducibly-covariant channels.

Recall from Definition 4.18 that a channel $\mathcal{N}_{A\to B}$ is covariant with respect to a group G if there exist projective unitary representations $\{U_A^g\}_{g\in G}$ and $\{V_B^g\}_{g\in G}$ such that

$$\mathcal{N}(U_A^g \rho_A (U_A^g)^\dagger) = V_B^g \mathcal{N}(\rho_A)(V_B^g)^\dagger \tag{12.2.196}$$

for all states ρ_A and all $g \in G$. The channel $\mathcal{N}$ is called irreducibly covariant if the representation $\{U_A^g\}_{g\in G}$ acting on the input space of the channel is irreducible, which means that it satisfies

$$\frac{1}{|G|}\sum_{g\in G} U_A^g \rho_A (U_A^g)^\dagger = \frac{\mathbb{1}}{d_A} \tag{12.2.197}$$

for every state ρ_A.

Proposition 12.24 Generalized Υ-Information for Irreducibly-Covariant Channels

Let $\mathcal{N}_{A\to B}$ be an irreducibly-covariant quantum channel. Then the generalized Υ-information of $\mathcal{N}$ can be calculated using the maximally entangled state Φ_{RA}, i.e.,

$$\begin{aligned}\boldsymbol{\Upsilon}(\mathcal{N}) &= \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA})) \\ &= \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\rho_{RB}^{\mathcal{N}}\|\rho_{RB}^{\mathcal{F}}).\end{aligned} \tag{12.2.198}$$

Proof: By simply restricting the optimization over states ψ_{RA} in the definition of the generalized Υ-information to the maximally entangled state Φ_{RA}, we obtain

$$\boldsymbol{\Upsilon}(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) \tag{12.2.199}$$

$$\geq \inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA})). \tag{12.2.200}$$

To prove the reverse inequality, let us recall Proposition 7.84, specifically its proof. Let G be the group with respect to which $\mathcal{N}$ is irreducibly covariant, let $\{U_A^g\}_{g\in G}$ be the irreducible representation of G acting on the input space of $\mathcal{N}_{A\to B}$, and let $\{V_B^g\}_{g\in G}$ be the representation of G acting on the output space of $\mathcal{N}$. Since the maps $\mathcal{F}_{A\to B}$ in $\mathfrak{F}$, in particular the map achieving the infimum in the definition of $\boldsymbol{\Upsilon}(\mathcal{N})$, need not be irreducibly covariant, we cannot use Proposition 7.84 directly. Instead, we consider (7.11.41) in its proof. For every $\mathcal{F}_{A\to B}\in\mathfrak{F}$, by using (7.11.52), the inequality in (7.11.41) becomes

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA}))$$
$$\geq \boldsymbol{D}\left(\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_{R'}\otimes\left((\mathcal{V}_B^g)^\dagger\circ\mathcal{N}_{A\to B}\circ\mathcal{U}_A^g\right)(\psi_{RA})\right\| \tag{12.2.201}$$
$$\left.\frac{1}{|G|}\sum_{g\in G}|g\rangle\langle g|_{R'}\otimes\left((\mathcal{V}_B^g)^\dagger\circ\mathcal{F}_{A\to B}\circ\mathcal{U}_A^g\right)(\psi_{RA})\right) \tag{12.2.202}$$

for every pure state ψ_{RA}, with the dimension of R equal to the dimension of A. Using the data-processing inequality for the generalized divergence with respect to the partial trace Tr_R, and using the fact that $\mathcal{N}$ is covariant, we find that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA}))$$
$$\geq \boldsymbol{D}\left(\mathcal{N}_{A\to B}(\psi_{RA})\middle\|\frac{1}{|G|}\sum_{g\in G}\left((\mathcal{V}_B^g)^\dagger\circ\mathcal{F}_{A\to B}\circ\mathcal{U}_A^g\right)(\psi_{RA})\right). \tag{12.2.203}$$

Now, observe that the map $\mathcal{F}'_{A\to B}:=\frac{1}{|G|}\sum_{g\in G}(\mathcal{V}_B^g)^\dagger\circ\mathcal{F}_{A\to B}\circ\mathcal{U}_A^g$ is in the set $\mathfrak{F}$ since $\mathcal{F}$ is. Therefore, optimizing over all $\mathcal{F}'\in\mathfrak{F}$, we find that

$$\boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA}))$$
$$\geq \inf_{\mathcal{F}'\in\mathfrak{F}}\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}'_{A\to B}(\psi_{RA})). \tag{12.2.204}$$

Since the map $\mathcal{F}\in\mathfrak{F}$ is arbitrary, we conclude that

$$\inf_{\mathcal{F}\in\mathfrak{F}}\boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA}))$$
$$\geq \inf_{\mathcal{F}'\in\mathfrak{F}}\boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}'_{A\to B}(\psi_{RA})). \tag{12.2.205}$$

Finally, since the state ψ_{RA} is arbitrary, we obtain

$$\inf_{\mathcal{F}\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA}))$$
$$\geq \sup_{\psi_{RA}} \inf_{\mathcal{F}'\in\mathfrak{F}} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}'_{A\to B}(\psi_{RA})) \tag{12.2.206}$$
$$= \boldsymbol{\Upsilon}(\mathcal{N}). \tag{12.2.207}$$

We thus conclude (12.2.198), as required. ■

By Proposition 12.24, we have that for irreducibly-covariant channels the sandwiched Rényi Υ-information can be calculated without an optimization over all pure states ψ_{RA}—we simply set $\psi_{RA} = \Phi_{RA}$. Using this fact, we obtain the following:

Proposition 12.25 Subadditivity of Sandwiched Rényi Υ-Information for Irreducibly-Covariant Channels

Let $\mathcal{N}_{A\to B}$ be an irreducibly covariant quantum channel. Then, for all $\alpha \in [1/2, 1) \cup (1, \infty)$, the sandwiched Rényi Υ-information $\widetilde{\Upsilon}_\alpha(\mathcal{N})$ is subadditive, i.e.,

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}) \leq n\widetilde{\Upsilon}_\alpha(\mathcal{N}) \tag{12.2.208}$$

for all $n \geq 1$.

Proof: Since $\mathcal{N}$ is irreducibly covariant, we use Proposition 12.24 to conclude that

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n}) = \inf_{\mathcal{F}\in\mathfrak{F}_n} \widetilde{D}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\Phi_{R^nA^n})\|\mathcal{F}_{A^n\to B^n}(\Phi_{R^nA^n})), \tag{12.2.209}$$

where $\mathfrak{F}_n$ is the set of completely positive maps in (12.2.150) acting on the space of the system A^n. Now, the maximally entangled state $\Phi_{R^nA^n}$ on n identical copies $R_1 \cdots R_n$ and $A_1 \cdots A_n$ of the systems R and A splits into a tensor product in the following way:

$$\Phi_{R^nA^n} = \Phi_{R_1A_1} \otimes \cdots \otimes \Phi_{R_nA_n}. \tag{12.2.210}$$

Furthermore, if we restrict the optimization over maps $\mathcal{F}_{A^n\to B^n} \in \mathfrak{F}_n$ to a tensor product of identical maps $\mathcal{G}$ in the set $\mathfrak{F}$ such that

$$\mathcal{F}_{A^n\to B^n} = \mathcal{G}_{A_1\to B_1} \otimes \cdots \otimes \mathcal{G}_{A_n\to B_n}, \tag{12.2.211}$$

then, we arrive at the following bound:

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}^{\otimes n})$$
$$\leq \inf_{\mathcal{G}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A_1\to B_1}(\Phi_{R_1A_1})\otimes\cdots\otimes\mathcal{N}_{A_n\to B_n}(\Phi_{R_nA_n})\|$$
$$\mathcal{G}_{A_1\to B_1}(\Phi_{R_1A_1})\otimes\cdots\otimes\mathcal{G}_{A_n\to B_n}(\Phi_{R_nA_n})) \quad (12.2.212)$$
$$= \inf_{\mathcal{G}\in\mathfrak{F}_1} n\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\Phi_{RA})\|\mathcal{G}_{A\to B}(\Phi_{RA})) \quad (12.2.213)$$
$$= n\widetilde{\Upsilon}_\alpha(\mathcal{N}), \quad (12.2.214)$$

as required, where the first equality follows from the additivity of the sandwiched Rényi relative entropy for tensor-product states (see (7.5.40) in Proposition 7.31). ■

With the subadditivity of the sandwiched Rényi Υ-information for irreducibly covariant channels, we can now state the following strong converse theorem.

Theorem 12.26 Υ-Information Upper Bound on the Strong Converse Classical Capacity of Irreducibly-Covariant Channels

The Υ-information $\Upsilon(\mathcal{N})$ of an irreducibly-covariant quantum channel $\mathcal{N}$ is a strong converse rate for classical communication over $\mathcal{N}$; i.e.,

$$\widetilde{C}(\mathcal{N}) \leq \Upsilon(\mathcal{N}). \quad (12.2.215)$$

Proof: Fix $\varepsilon \in [0,1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \quad (12.2.216)$$

Set $\alpha \in (1,\infty)$ such that

$$\delta_1 \geq \widetilde{\Upsilon}_\alpha(\mathcal{N}) - \Upsilon(\mathcal{N}), \quad (12.2.217)$$

which is possible because $\widetilde{\Upsilon}_\alpha(\mathcal{N})$ is monotonically increasing with α (this follows from Proposition 7.31) and $\lim_{\alpha\to 1^+}\widetilde{\Upsilon}_\alpha(\mathcal{N}) = \Upsilon(\mathcal{N})$ (see Appendix 12.A for a proof). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \quad (12.2.218)$$

Now, with the values of n and ε chosen as above, every $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol satisfies (12.2.195). In particular, using the subadditivity of the sandwiched Rényi Υ-information for all $\alpha > 1$, we can write (12.2.195) as

$$\frac{1}{n}\log_2|\mathcal{M}| \leq \widetilde{\Upsilon}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.219}$$

Rearranging the right-hand side of this inequality, and using the assumptions in (12.2.216)–(12.2.218), we obtain

$$\frac{1}{n}\log_2|\mathcal{M}| \leq \Upsilon(\mathcal{N}) + \widetilde{\Upsilon}_\alpha(\mathcal{N}) - \Upsilon(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{12.2.220}$$

$$\leq \Upsilon(\mathcal{N}) + \delta_1 + \delta_2 \tag{12.2.221}$$

$$= \Upsilon(\mathcal{N}) + \delta' \tag{12.2.222}$$

$$< \Upsilon(\mathcal{N}) + \delta. \tag{12.2.223}$$

So we have that $\Upsilon(\mathcal{N}) + \delta > \frac{1}{n}\log_2|\mathcal{M}|$ for all $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols with n sufficiently large. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(\Upsilon(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol for all sufficiently large n such that (12.2.218) holds, for if it did there would exist some message set $\mathcal{M}$ such that $\frac{1}{n}\log_2|\mathcal{M}| = \Upsilon(\mathcal{N}) + \delta$, which we have just seen is not possible. Since ε and δ are arbitrary, we have that for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(\Upsilon(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol. This means that $\Upsilon(\mathcal{N})$ is a strong converse rate, which completes the proof. ■

Theorem 12.26 thus gives us an upper bound on the strong converse classical capacity $\widetilde{C}(\mathcal{N})$ of any irreducibly-covariant channel $\mathcal{N}$, namely,

$$\widetilde{C}(\mathcal{N}) \leq \Upsilon(\mathcal{N}), \tag{12.2.224}$$

which in turn implies, via (12.2.12), that

$$C(\mathcal{N}) \leq \Upsilon(\mathcal{N}) \tag{12.2.225}$$

for every irreducibly-covariant channel $\mathcal{N}$. Recall that the Holevo information $\chi(\mathcal{N})$ is an achievable rate for classical communication over any quantum channel, which implies that $C(\mathcal{N}) \geq \chi(\mathcal{N})$. (This in fact gives another way for concluding that $\Upsilon(\mathcal{N}) \geq \chi(\mathcal{N})$.)

It turns out that the Υ-information $\Upsilon(\mathcal{N})$ is *equal* to the Holevo information in the case of the erasure channel $\mathcal{E}_p^{(d)}$. Recall from Section 11.3.1.2 that the erasure channel is irreducibly covariant. This fact, along with other reasoning, allows us to conclude that

$$C(\mathcal{E}_p^{(d)}) = \widetilde{C}(\mathcal{E}_p^{(d)}) = \chi(\mathcal{E}_p^{(d)}) = (1-p)\log_2 d \tag{12.2.226}$$

for all dimensions $d \geq 2$ and all $p \in [0,1]$. We provide a proof of this chain of equalities in Section 12.3.1.2 below.

12.2.5.2 SDP Upper Bound

While the Υ-information gives us an upper bound on the strong converse classical capacity of any irreducibly-covariant channel, computing it is relatively challenging due to the minimization over the set $\mathfrak{F}$. In this section, we define a subset of $\mathfrak{F}$, denoted by $\mathfrak{F}_\beta$, that allows us to obtain a quantity that can be computed using a semi-definite program (SDP). Furthermore, this quantity turns out to be additive for *all* channels, which means that it is an upper bound on the strong converse classical capacity for all channels.

The set $\mathfrak{F}_\beta$ is defined as the following set of completely positive maps:

$$\mathfrak{F}_\beta := \{\mathcal{F} \text{ completely positive} : \beta(\mathcal{F}) \leq 1\}, \tag{12.2.227}$$

where $\beta(\mathcal{F})$ is defined as the solution to the following optimization problem:

$$\beta(\mathcal{F}) := \begin{cases} \text{infimum} & \operatorname{Tr}[S_B] \\ \text{subject to} & -R_{AB} \leq (\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \leq R_{AB}, \\ & -\mathbb{1}_A \otimes S_B \leq R_{AB}^{\mathsf{T}_B} \leq \mathbb{1}_A \otimes S_B \end{cases} \tag{12.2.228}$$

Note that the optimization occurs over the operators S_B and R_{AB}, and that the optimization problem as a whole is an SDP. Indeed, recalling the general form of an SDP from Section 2.4, we can write it as

$$\beta(\mathcal{F}) = \begin{cases} \text{infimum} & \operatorname{Tr}[CX] \\ \text{subject to} & \Phi(X) \geq D, \\ & X \geq 0, \end{cases} \tag{12.2.229}$$

where

$$X = \begin{pmatrix} S_B & 0 \\ 0 & R_{AB} \end{pmatrix}, \quad C = \begin{pmatrix} \mathbb{1}_B & 0 \\ 0 & 0_{AB} \end{pmatrix}, \tag{12.2.230}$$

$$\Phi(X) = \begin{pmatrix} R_{AB} & 0 & 0 & 0 \\ 0 & R_{AB} & 0 & 0 \\ 0 & 0 & \mathbb{1}_R \otimes S_B - R_{AB}^{\mathsf{T}_B} & 0 \\ 0 & 0 & 0 & \mathbb{1}_R \otimes S_B + R_{AB}^{\mathsf{T}_B} \end{pmatrix}, \tag{12.2.231}$$

$$D = \begin{pmatrix} (\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} & 0 & 0 & 0 \\ 0 & -(\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{12.2.232}$$

Note that the constraints in (12.2.228) imply that S_B and R_{AB} are positive semi-definite, since the constraint $-R_{AB} \leq (\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \leq R_{AB}$ implies that

$$R_{AB} - (\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \geq 0, \tag{12.2.233}$$

$$R_{AB} + (\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \geq 0. \tag{12.2.234}$$

Adding the two inequalities leads to $R_{AB} \geq 0$. Similarly, the constraint $-\mathbb{1}_A \otimes S_B \leq R_{AB}^{\mathsf{T}_B} \leq \mathbb{1}_A \otimes S_B$ implies that $S_B \geq 0$.

It is straightforward to show that $\mathfrak{F}_\beta$ is a subset of $\mathfrak{F}$, i.e., $\mathfrak{F}_\beta \subseteq \mathfrak{F}$. Indeed, suppose that for a given completely positive map $\mathcal{F} \in \mathfrak{F}_\beta$, the quantity $\beta(\mathcal{F})$ in (12.2.228) is achieved by the operators (R_{AB}^*, S_B^*). Note that since $\mathcal{F} \in \mathfrak{F}_\beta$, by definition we have that $\mathrm{Tr}[S_B^*] \leq 1$, and we also have that $(\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \leq R_{AB}^*$ and $(R_{AB}^*)^{\mathsf{T}_B} \leq \mathbb{1}_A \otimes S_B^*$. Letting $\sigma_B \equiv S_B^*$, for every state ρ_A we find that

$$\begin{aligned}
\mathcal{F}_{A\to B}(\rho_A) &= \mathrm{Tr}_A\big[(\rho_A^{\mathsf{T}} \otimes \mathbb{1}_B)\Gamma_{AB}^{\mathcal{F}}\big] && (12.2.235)\\
&= \Big(\mathrm{Tr}_A\big[(\rho_A^{\mathsf{T}} \otimes \mathbb{1}_B)(\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B}\big]\Big)^{\mathsf{T}} && (12.2.236)\\
&= \Big(\mathrm{Tr}_A\Big[(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)(\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B}(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)\Big]\Big)^{\mathsf{T}} && (12.2.237)\\
&\leq \Big(\mathrm{Tr}_A\Big[(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)R_{AB}^*(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)\Big]\Big)^{\mathsf{T}} && (12.2.238)\\
&= \mathrm{Tr}_A\Big[(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)(R_{AB}^*)^{\mathsf{T}_B}(\sqrt{\rho_A^{\mathsf{T}}} \otimes \mathbb{1}_B)\Big] && (12.2.239)\\
&\leq \mathrm{Tr}_A\big[(\rho_A^{\mathsf{T}} \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes \sigma_B)\big] && (12.2.240)\\
&= \sigma_B, && (12.2.241)
\end{aligned}$$

where to obtain the first inequality we used $(\Gamma_{AB}^{\mathcal{F}})^{\mathsf{T}_B} \leq R_{AB}^*$ and to obtain the second inequality we used $(R_{AB}^*)^{\mathsf{T}_B} \leq \mathbb{1}_A \otimes \sigma_B$. Therefore, $\mathcal{F}_{A\to B}(\rho_A) \leq \sigma_B$ for all ρ_A, which means that $\mathcal{F}_{A\to B} \in \mathfrak{F}$. Since $\mathcal{F} \in \mathfrak{F}_\beta$ is arbitrary, we conclude that $\mathfrak{F}_\beta \subset \mathfrak{F}$.

By replacing the set $\mathfrak{F}$ in the definition of the generalized Υ-information of a channel $\mathcal{N}$ with the set $\mathfrak{F}_\beta$, we obtain the following quantity:

$$\boldsymbol{\Upsilon}^{\beta}(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}_\beta} \boldsymbol{D}(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.2.242}$$

which we call the Υ^β*-information of* $\mathcal{N}$. When we take the generalized divergence $\boldsymbol{D}$ to be the quantum relative entropy, the hypothesis testing relative entropy, and the sandwiched Rényi relative entropy, we have

$$\Upsilon^{\beta}(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}_\beta} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.2.243}$$

$$\Upsilon^{\beta,\varepsilon}_H(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}_\beta} D^{\varepsilon}_H(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.2.244}$$

$$\widetilde{\Upsilon}^{\beta}_\alpha(\mathcal{N}) := \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}_\beta} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})). \tag{12.2.245}$$

Since the set $\mathfrak{F}_\beta$ is a subset of $\mathfrak{F}$, minimizing over $\mathfrak{F}_\beta$ can never lead to a smaller value compared to minimizing over $\mathfrak{F}$, which means that

$$\boldsymbol{\Upsilon}(\mathcal{N}) \leq \boldsymbol{\Upsilon}^{\beta}(\mathcal{N}). \tag{12.2.246}$$

Therefore, using the Υ^β-information, the bound in Proposition 12.22 on every $(|\mathcal{M}|, \varepsilon)$ classical communication protocol thus becomes

$$\log_2 |\mathcal{M}| \leq \Upsilon^{\beta,\varepsilon}_H(\mathcal{N}). \tag{12.2.247}$$

This bound is looser than the one in Proposition 12.22, but it has the advantange that it can be computed using an SDP. This is due to the fact that the hypothesis testing relative entropy can itself be computed via an SDP.

Although we get an efficiently computable upper bound in the one-shot setting via the Υ^β-information, in the asymptotic setting this bound is not known to be additive, making its evaluation computationally prohibitive as the number n of channel uses increases. Instead, for the purpose of obtaining an efficiently computable upper bound in the asymptotic setting, we define the following quantity for every quantum channel $\mathcal{N}$:

$$C_\beta(\mathcal{N}) = \log_2 \beta(\mathcal{N}), \tag{12.2.248}$$

Since $\beta(\mathcal{N})$ can be computed using an SDP (in particular, via the optimization problem in (12.2.228)), we have that $C_\beta(\mathcal{N})$ can also be computed using an SDP.

A useful fact about the quantity $C_\beta(\mathcal{N})$ is the fact that it is additive, i.e.,

$$C_\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_\beta(\mathcal{N}_1) + C_\beta(\mathcal{N}_2) \tag{12.2.249}$$

for all channels $\mathcal{N}_1$ and $\mathcal{N}_2$, as proved in Appendix 12.B by employing semi-definite programming duality. We use this fact to prove that $C_\beta(\mathcal{N})$ is a strong converse rate for classical communication over a channel $\mathcal{N}$ in Theorem 12.28 below. However, first we establish the following proposition:

Proposition 12.27

For a quantum channel $\mathcal{N}$, the following inequalities hold for all $\alpha > 1$:

$$\Upsilon(\mathcal{N}) \leq C_\beta(\mathcal{N}), \qquad \widetilde{\Upsilon}_\alpha(\mathcal{N}) \leq C_\beta(\mathcal{N}). \tag{12.2.250}$$

Proof: Let $\mathcal{F} = \frac{1}{\beta(\mathcal{N})}\mathcal{N}$. Then, $\beta(\mathcal{F}) = \frac{1}{\beta(\mathcal{N})}\beta(\mathcal{N}) = 1$, which means that $\mathcal{F} \in \mathfrak{F}_\beta$. Then, since $\mathfrak{F}_\beta \subseteq \mathfrak{F}$, we can choose $\mathcal{F}$ as above when performing the infimum in the definition of $\Upsilon(\mathcal{N})$. This leads to

$$\Upsilon(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) \tag{12.2.251}$$

$$\leq \sup_{\psi_{RA}} D\left(\mathcal{N}_{A\to B}(\psi_{RA})\middle\|\frac{1}{\beta(\mathcal{N})}\mathcal{N}_{A\to B}(\psi_{RA})\right) \tag{12.2.252}$$

$$= \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{N}_{A\to B}(\psi_{RA})) + \log_2 \beta(\mathcal{N}) \tag{12.2.253}$$

$$= C_\beta(\mathcal{N}), \tag{12.2.254}$$

as required, where the second equality follows because $D(\rho\|\frac{1}{x}\sigma) = D(\rho\|\sigma)+\log_2 x$ for every state ρ, positive semi-definite operator σ, and $x > 0$ (see (7.2.26)). The last equality follows because $D(\rho\|\rho) = 0$.

Similarly, using the fact that $\widetilde{D}_\alpha(\rho\|\frac{1}{x}\sigma) = \widetilde{D}_\alpha(\rho\|\sigma) + \log_2 x$, and using the same choice for the map $\mathcal{F}$ as above, we find that $\widetilde{\Upsilon}_\alpha(\mathcal{N}) \leq C_\beta(\mathcal{N})$. ■

The additivity of C_β, along with Proposition 12.27, leads us to the following:

Theorem 12.28 SDP Upper Bound on Strong Converse Classical Capacity

For a quantum channel $\mathcal{N}$, the quantity $C_\beta(\mathcal{N})$ is a strong converse rate for classical communication over $\mathcal{N}$.

PROOF: We start by observing that, using Proposition 12.27, the inequality in (12.2.195) can be written as

$$\frac{1}{n}\log_2|\mathcal{M}| \le \frac{1}{n}C_\beta(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \tag{12.2.255}$$

for all $\alpha > 1$. This inequality holds for all $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols. Using the additivity of C_β, we find that

$$\frac{1}{n}\log_2|\mathcal{M}| \le C_\beta(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.256}$$

Now, let us fix $\varepsilon \in [0,1)$ and $\delta > 0$. Let δ' be such that $\delta > \delta'$. Since $C_\beta(\mathcal{N})$ does not depend on α, let us choose α such that the right-hand side of the above inequality is as small as possible, which occurs as $\alpha \to \infty$. With this choice of α, take n large enough so that

$$\delta' \ge \frac{1}{n}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{12.2.257}$$

Then, we obtain

$$\begin{aligned}
\frac{1}{n}\log_2|\mathcal{M}| &\le C_\beta(\mathcal{N}) + \frac{1}{n}\log_2\!\left(\frac{1}{1-\varepsilon}\right) && (12.2.258)\\
&\le C_\beta(\mathcal{N}) + \delta' && (12.2.259)\\
&< C_\beta(\mathcal{N}) + \delta. && (12.2.260)
\end{aligned}$$

So we have that $C_\beta(\mathcal{N}) + \delta > \frac{1}{n}\log_2|\mathcal{M}|$ for all $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocols with n sufficiently large. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(C_\beta(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol for all sufficiently large n such that (12.2.257) holds, for if it did there would exist some message set $\mathcal{M}$ such that $\frac{1}{n}\log_2|\mathcal{M}| = C_\beta(\mathcal{N}) + \delta$, which we have just seen is not possible. Since ε and δ are arbitrary, we have that for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(C_\beta(\mathcal{N})+\delta)}, \varepsilon)$ classical communication protocol. This means that $C_\beta(\mathcal{N})$ is a strong converse rate, which completes the proof. ■

By examining (12.2.258) in the above proof, we see that the following bound holds for an arbitrary $(n, |\mathcal{M}|, \varepsilon)$ classical communication protocol:

$$\frac{1}{n}\log_2|\mathcal{M}| \leq C_\beta(\mathcal{N}) + \frac{1}{n}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \qquad (12.2.261)$$

If we fix the rate $R = \frac{1}{n}\log_2|\mathcal{M}|$, then this bound can be rewritten as follows:

$$1-\varepsilon \leq 2^{-n(R-C_\beta(\mathcal{N}))}, \qquad (12.2.262)$$

which indicates that communicating at a rate $R > C_\beta(\mathcal{N})$ implies the success probability $1-\varepsilon$ of every sequence of such protocols decays exponentially fast to zero.

12.3 Examples

In this section, we present various examples of channels with known formulas for the Holevo information and/or known results on additivity of the Holevo information.

Let us start by making some observations about the Holevo information $\chi(\mathcal{N})$ of a channel $\mathcal{N}$. First, by expanding the definition of the Holevo information using the expression for the mutual information in terms of the relative entropy, we arrive at the following:

Proposition 12.29 Alternate Forms for Channel Holevo Information

For a channel $\mathcal{N}$, the following equalities hold

$$\chi(\mathcal{N}) = \sup_{\{(p(x),\psi_A^x)\}_x}\left[H(\mathcal{N}(\overline{\rho}_A)) - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}(\psi_A^x))\right], \qquad (12.3.1)$$

$$= \sup_{\{(p(x),\psi_A^x)\}_x}\left[H(\mathcal{N}(\overline{\rho}_A)) - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}^c(\psi_A^x))\right]. \qquad (12.3.2)$$

where $\{(p(x),\psi_A^x)\}_{x\in\mathcal{X}}$ is an ensemble of pure states, $\overline{\rho}_A \coloneqq \sum_{x\in\mathcal{X}} p(x)\psi_A^x$, and we recall from (7.1.1) and (7.2.88) that $H(\rho) = -\mathrm{Tr}[\rho\log\rho]$ is the quantum entropy of ρ.

PROOF: We start by recalling the definition of the Holevo information $\chi(\mathcal{N})$ of $\mathcal{N}$ from (7.11.106):

$$\chi(\mathcal{N}) \coloneqq \sup_{\rho_{XA}} I(X;B)_\omega, \tag{12.3.3}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\rho_{XA})$, and the supremum is over all classical-quantum states of the form $\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_A^x$, with $\mathcal{X}$ a finite alphabet with associated $|\mathcal{X}|$-dimensional system X, $\{\rho_A^x\}_{x\in\mathcal{X}}$ a set of states, and $p : \mathcal{X} \to [0,1]$ a probability distribution on $\mathcal{X}$. Recall from Proposition 7.87 that to compute the Holevo information it suffices to take ensembles consisting only of pure states.

Defining the classical–quantum state ρ_{XA} as

$$\rho_{XA} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \psi_A^x, \tag{12.3.4}$$

where $\{\psi_A^x\}_{x\in\mathcal{X}}$ is a set of pure states, and defining $\rho'_{XB} = \mathcal{N}_{A\to B}(\rho_{XA})$ is another classical–quantum state, it follows from Proposition 7.14 that

$$I(X;B)_{\rho'} = H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\psi_A^x)\right) - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}(\psi_A^x)) \tag{12.3.5}$$

for all states ρ_{XA}. Since optimizing over classical–quantum states is equivalent to optimizing over ensembles $\{(p(x), \psi_A^x)\}_{x\in\mathcal{X}}$, we obtain (12.3.1).

To prove (12.3.2), let $V_{A\to BE}$ be an isometric extension of $\mathcal{N}$. Then, for every pure state ψ on the channel input system A, we can write

$$\mathcal{N}(\psi) = \mathrm{Tr}_E[V\psi V^\dagger]. \tag{12.3.6}$$

Since $V|\psi\rangle$ is a pure state, it follows that $\mathrm{Tr}_E[V\psi V^\dagger]$ and $\mathrm{Tr}_B[V\psi V^\dagger]$ have the same (non-zero) eigenvalues. The latter state is equal to $\mathcal{N}^c(\psi_A)$ by definition, which means that

$$H(\mathcal{N}(\psi_A)) = H(\mathcal{N}^c(\psi_A)). \tag{12.3.7}$$

Therefore, (12.3.2) follows. ■

12.3.1 Covariant Channels

For irreducibly-covariant channels (see Definition 4.18), the Holevo information takes a particularly simple form.

Theorem 12.30 Holevo Information of Irreducibly-Covariant Channels

Suppose $\mathcal{N}_{A\to B}$ is a covariant channel with respect to a finite group G, with an irreducible representation $\{U_A^g\}_{g\in G}$ of G acting on the input space of the channel and another representation $\{V_B^g\}_{g\in G}$ of G acting on the output space of the channel. Then,

$$\chi(\mathcal{N}) = H(\mathcal{N}(\pi_A)) - H_{\min}(\mathcal{N}), \tag{12.3.8}$$

where $\pi_A = \frac{\mathbb{1}_A}{d_A}$ and

$$H_{\min}(\mathcal{N}) \coloneqq \min_{\rho_A} H(\mathcal{N}(\rho_A)) \tag{12.3.9}$$

is called the *minimum output entropy* of $\mathcal{N}$. The minimization is with respect to all input states ρ_A in the domain of $\mathcal{N}$.

Proof: We have

$$\chi(\mathcal{N}) = \sup_{\{(p(x),\rho_A^x)\}_x} \left[H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\rho_A^x)\right) - \sum_{x\in\mathcal{X}} p(x) H(\mathcal{N}(\rho_A^x)) \right] \tag{12.3.10}$$

$$\leq \sup_{\{(p(x),\rho_A^x)\}_x} \left\{ H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\rho_A^x)\right)\right\} + \sup_{\{(p(x),\rho_A^x)\}_x} \left\{ -\sum_{x\in\mathcal{X}} p(x) H(\mathcal{N}(\rho_A^x)) \right\} \tag{12.3.11}$$

$$\leq \sup_{\rho_A} H(\mathcal{N}(\rho_A)) + \sup_{\rho_A} \{-H(\mathcal{N}(\rho_A))\} \tag{12.3.12}$$

$$= \sup_{\rho_A} H(\mathcal{N}(\rho_A)) - \inf_{\rho_A} H(\mathcal{N}(\rho_A)). \tag{12.3.13}$$

Now, by the unitary invariance of the quantum entropy, for every state ρ_A we obtain

$$H(\mathcal{N}(\rho_A)) = H(V_B^g \mathcal{N}(\rho_A)(V_B^g)^\dagger) = H(\mathcal{N}(U_A^g \rho_A (U_A^g)^\dagger)) \tag{12.3.14}$$

for all $g \in G$. This implies that

$$H(\mathcal{N}(\rho_A)) = \frac{1}{|G|}\sum_{g\in G} H(\mathcal{N}(U_A^g \rho_A (U_A^g)^\dagger)) \tag{12.3.15}$$

$$\leq H\left(\sum_{g\in G} \frac{1}{|G|} \mathcal{N}(U_A^g \rho_A (U_A^g)^\dagger)\right) \tag{12.3.16}$$

$$= H\left(\mathcal{N}\left(\frac{1}{|G|}\sum_{g\in G} U_A^g \rho_A (U_A^g)^\dagger\right)\right) \tag{12.3.17}$$

$$= H(\mathcal{N}(\pi_A)), \tag{12.3.18}$$

where the inequality follows from concavity of the quantum entropy and the last equality follows because $\{U^g\}_{g\in G}$ is an irreducible representation, which implies that

$$\frac{1}{|G|}\sum_{g\in G} U_A^g \rho_A (U_A^g)^\dagger = \frac{\mathbb{1}_A}{d_A} \tag{12.3.19}$$

for every state ρ_A. Then, since we are optimizing a continuous function over a compact and convex set, the infimum in (12.3.13) can be achieved, meaning that we can replace the infimum in (12.3.13) with a minimum, which means that

$$\chi(\mathcal{N}) \leq H(\mathcal{N}(\pi_A)) - H_{\min}(\mathcal{N}). \tag{12.3.20}$$

To show the reverse inequality, let ρ_A^* be a state for which $H(\mathcal{N}(\rho_A^*)) = H_{\min}(\mathcal{N})$. Then, we consider the ensemble $\left\{\left(\frac{1}{|G|}, U_A^g \rho_A^* (U_A^g)^\dagger\right)\right\}_{g\in G}$ and obtain

$$\chi(\mathcal{N}) \geq H\left(\sum_{g\in G}\frac{1}{|G|}\mathcal{N}(U_A^g \rho_A^* (U_A^g)^\dagger)\right) - \sum_{g\in G}\frac{1}{|G|} H(\mathcal{N}(U_A^g \rho_A^* (U_A^g)^\dagger)) \tag{12.3.21}$$

$$= H(\mathcal{N}(\pi_A)) - \sum_{g\in G}\frac{1}{|G|} H(V_B^g \rho_A^* (V_B^g)^\dagger) \tag{12.3.22}$$

$$= H(\mathcal{N}(\pi_A)) - \sum_{g\in G}\frac{1}{|G|} H(\mathcal{N}(\rho_A^*)) \tag{12.3.23}$$

$$= H(\mathcal{N}(\pi_A)) - H(\mathcal{N}(\rho_A^*)) \tag{12.3.24}$$

$$= H(\mathcal{N}(\pi_A)) - H_{\min}(\mathcal{N}). \tag{12.3.25}$$

Therefore,

$$\chi(\mathcal{N}) \geq H(\mathcal{N}(\pi_A)) - H_{\min}(\mathcal{N}), \tag{12.3.26}$$

and the proof is complete. ■

We note that to compute the minimum output entropy of any channel $\mathcal{N}$, it suffices to optimize over pure states. Indeed, by restricting the optimization to pure states ψ in the definition of $H_{\min}(\mathcal{N})$, we find that

$$H_{\min}(\mathcal{N}) = \min_{\rho} H(\mathcal{N}(\rho)) \leq \min_{\psi} H(\mathcal{N}(\psi)). \tag{12.3.27}$$

On the other hand, since every state ρ can be written as a convex combination of pure states, so that $\rho = \sum_{x\in\mathcal{X}} p(x)\psi^x$, we see that

$$\begin{aligned} H(\mathcal{N}(\rho)) &= H\left(\mathcal{N}\left(\sum_{x\in\mathcal{X}} p(x)\psi^x\right)\right) && (12.3.28)\\ &= H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\psi^x)\right) && (12.3.29)\\ &\geq \sum_{x\in\mathcal{X}} p(x) H(\mathcal{N}(\psi^x)) && (12.3.30)\\ &\geq \min_{x\in\mathcal{X}} H(\mathcal{N}(\psi^x)) && (12.3.31)\\ &\geq \min_{\psi} H(\mathcal{N}(\psi)), && (12.3.32) \end{aligned}$$

where the first inequality follows from concavity of the quantum entropy. So we have

$$H_{\min}(\mathcal{N}) = \min_{\psi} H(\mathcal{N}(\psi)). \tag{12.3.33}$$

We now look at two irreducibly covariant channels, the depolarizing channel and the erasure channel. A plot of the Holevo information for these channels is given in Figure 12.6.

12.3.1.1 Depolarizing Channel

In Section 4.5, specifically in (4.5.31), we defined the qubit depolarizing channel as

$$\mathcal{D}_p(\rho) := (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z) \tag{12.3.34}$$

for all $p \in [0,1]$. From the arguments in Section 11.3.1.1, it follows that this channel is covariant with respect to the Pauli operators on both the input and output spaces. Furthermore, the Pauli operators $\{\mathbb{1}, X, Y, Z\}$ satisfy the property in (12.3.19). We thus conclude that

$$\chi(\mathcal{D}_p) = H\big(\mathcal{D}_p(\pi)\big) - H_{\min}(\mathcal{D}_p) = 1 - H_{\min}(\mathcal{D}_p), \tag{12.3.35}$$

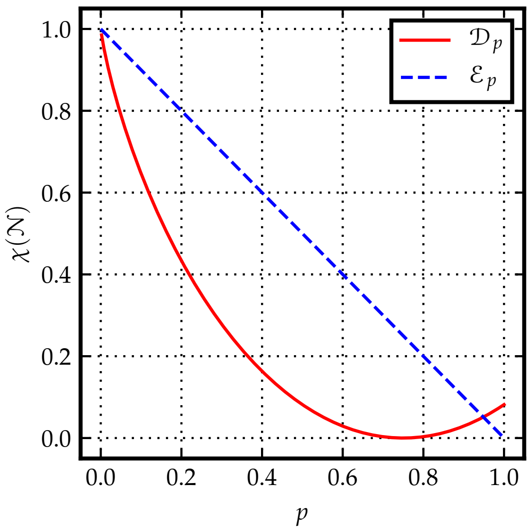


FIGURE 12.6: The Holevo information of the depolarizing channel $\mathcal{D}_p$ (expressed in (12.3.37)) and the erasure channel $\mathcal{E}_p$ (expressed in (12.3.57)), both of which are defined for the parameter $p \in [0, 1]$.

where we used the fact that the depolarizing channel is unital, i.e., $\mathcal{D}_p(\mathbb{1}) = \mathbb{1}$, and that $H(\pi) = \log_2 2 = 1$. To compute the minimum output entropy, we use the fact that it suffices to minimize over pure states. It is straightforward to show that the two eigenvalues of $\mathcal{D}_p(\psi)$ are $(^{2p}/_3, 1 - {}^{2p}/_3)$ for *every* pure state ψ. Therefore,

$$H_{\min}(\mathcal{D}_p) = h_2\left(\frac{2p}{3}\right) \tag{12.3.36}$$

so that

$$\chi(\mathcal{D}_p) = 1 - h_2\left(\frac{2p}{3}\right) \tag{12.3.37}$$

for $p \in [0, 1]$. Since the Holevo information is known to be additive for the depolarizing channel (please consult the Bibliographic Notes in Section 12.5), it follows that $\chi(\mathcal{D}_p)$ is equal to the classical capacity of the depolarizing channel. See Figure 12.6 for a plot of the Holevo information $\chi(\mathcal{D}_p)$ of the depolarizing channel.

For the qudit depolarizing channel $\mathcal{D}_p^{(d)}$, recall from the discussion around (11.3.26) that it is irreducibly covariant. Therefore, by Theorem 12.30, we obtain

$$\chi(\mathcal{D}_p^{(d)}) = \log_2 d - H_{\min}(\mathcal{D}_p^d). \tag{12.3.38}$$

The calculation of the minimum output entropy for the qudit depolarizing channel is analogous to the calculation of the minimum output entropy of the qubit depolarizing channel. In particular, for every pure state ψ, the eigenvalues of $\mathcal{D}_p^{(d)}(\psi)$ are $1-\frac{d}{d+1}p$ (with multiplicity one) and $\frac{d}{d^2-1}p$ (with multiplicity $d-1$). Indeed, by using the parameterization in (4.5.37) with $q = \frac{pd^2}{d^2-1}$, consider that

$$\mathcal{D}_p^{(d)}(\psi) = (1-q)\,\psi + q\frac{I}{d} \tag{12.3.39}$$

$$= \left(1 - q + \frac{q}{d}\right)\psi + \frac{q}{d}\,(I-\psi) \tag{12.3.40}$$

$$= \left(1 - \frac{pd}{d+1}\right)\psi + \frac{pd}{d^2-1}\,(I-\psi)\,. \tag{12.3.41}$$

Therefore,

$$H_{\min}(\mathcal{D}_p^{(d)}) = -\left(1-\frac{dp}{d+1}\right)\log_2\left(1-\frac{dp}{d+1}\right) - \frac{dp}{d+1}\log_2\left(\frac{dp}{d^2-1}\right), \tag{12.3.42}$$

so that

$$\chi(\mathcal{D}_p^{(d)}) = \log_2 d + \left(1-\frac{dp}{d+1}\right)\log_2\left(1-\frac{dp}{d+1}\right) + \frac{dp}{d+1}\log_2\left(\frac{dp}{d^2-1}\right) \tag{12.3.43}$$

for $d \geq 2$ and $p \in [0,1]$.

The Holevo information is also known to be additive for the qudit depolarizing channel, which means that the expression in (12.3.43) is equal to its classical capacity.

Theorem 12.31 Additivity of the Holevo Information for the Depolarizing Channel

For every channel $\mathcal{M}$,

$$\chi(\mathcal{D}_p^{(d)} \otimes \mathcal{M}) = \chi(\mathcal{D}_p^{(d)}) + \chi(\mathcal{M}) \tag{12.3.44}$$

for all $d \geq 2$ and all $p \in [0,1]$. Consequently,

$$C(\mathcal{D}_p^{(d)}) = \chi(\mathcal{D}_p^{(d)}). \tag{12.3.45}$$

Proof: Please consult the Bibliographic Notes in Section 12.5. ■

It also holds that the Holevo information is the strong converse classical capacity of the qudit depolarizing channel, i.e.,

$$\widetilde{C}(\mathcal{D}_p^{(d)}) = \chi(\mathcal{D}_p^{(d)}) \tag{12.3.46}$$

for all $d \geq 2$ and all $p \in [0,1]$. Please consult the Bibliographic Notes in Section 12.5 for a reference to the proof.

12.3.1.2 Erasure Channel

Let us now consider the erasure channel. Recall from (4.5.18) that the erasure channel $\mathcal{E}_p$, with $p \in [0,1]$, is defined as

$$\mathcal{E}_p(\rho) = (1-p)\rho + p\text{Tr}[\rho]|e\rangle\langle e|, \tag{12.3.47}$$

where $|e\rangle$ is called the erasure state and is not in the Hilbert space of the input system A. In other words, the state $|e\rangle\langle e|$ is supported on the space orthogonal to the input space. As argued in Section 11.3.1.2, we can consider the output space of the channel to be a qutrit system with the orthonormal basis $\{|0\rangle, |1\rangle, |2\rangle\}$, and we can let the state $|2\rangle$ be the erasure state. Then,

$$\mathcal{E}_p(\rho) = (1-p)\rho + p|2\rangle\langle 2| \tag{12.3.48}$$

for every state ρ.

We also argued in Section 11.3.1.2 that the erasure channel is irreducibly covariant. Therefore, by Theorem 12.30, we have that

$$\chi(\mathcal{E}_p) = H\big(\mathcal{E}_p(\pi)\big) - H_{\min}(\mathcal{E}_p). \tag{12.3.49}$$

Now, on the input qubit space, we have $\pi = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. Therefore,

$$\mathcal{E}_p(\pi) = \frac{1-p}{2}|0\rangle\langle 0| + \frac{1-p}{2}|1\rangle\langle 1| + p|2\rangle\langle 2|, \tag{12.3.50}$$

which means that

$$H\big(\mathcal{E}_p(\pi)\big) = -(1-p)\log_2\left(\frac{1-p}{2}\right) - p\log_2 p \tag{12.3.51}$$

$$= 1 - p + h_2(p). \tag{12.3.52}$$

In addition, for every pure state ψ, we have

$$H(\mathcal{E}_p(\psi)) = H((1-p)\psi + p|2\rangle\langle 2|) \tag{12.3.53}$$
$$= -(1-p)\log_2(1-p) - p\log_2 p, \tag{12.3.54}$$
$$= h_2(p), \tag{12.3.55}$$

where the second equality follows because the state $|2\rangle\langle 2|$ is orthogonal to ψ. Therefore,

$$H_{\min}(\mathcal{E}_p) = h_2(p), \tag{12.3.56}$$

which means that the Holevo information of the erasure channel is

$$\chi(\mathcal{E}_p) = 1 - p. \tag{12.3.57}$$

This is consistent with what one might expect intuitively because communication over the erasure channel is only possible with probability $1-p$, when no erasure occurs, and conditioned on this outcome, the erasure channel is simply the identity channel.

In general, for the qudit erasure channel $\mathcal{E}_p^{(d)}$, whose action can be defined on the d-dimensional space with orthonormal basis $\{|1\rangle, \dots, |d\rangle\}$ such that the state $|d+1\rangle$ is the erasure state, we have that it is irreducibly covariant (see Section 11.3.1.2). Using this fact, which implies that

$$\chi(\mathcal{E}_p^{(d)}) = H\left(\mathcal{E}_p^{(d)}(\pi)\right) - H_{\min}(\mathcal{E}_p^{(d)}), \tag{12.3.58}$$

along with arguments analogous to those presented above, we obtain

$$\chi(\mathcal{E}_p^{(d)}) = (1-p)\log_2 d. \tag{12.3.59}$$

Proposition 12.32 Υ-Information of the Erasure Channel

The Υ-information of the qudit erasure channel $\mathcal{E}_p^{(d)}$ is given by

$$\Upsilon(\mathcal{E}_p^{(d)}) = \chi(\mathcal{E}_p^{(d)}) = (1-p)\log_2 d. \tag{12.3.60}$$

Proof: By combining (12.3.59) and Proposition 12.21, we conclude that $\Upsilon(\mathcal{E}_p^{(d)}) \geq \chi(\mathcal{E}_p^{(d)}) = (1-p)\log_2 d$. So we establish the opposite inequality.

Since the erasure channel is irreducibly covariant, Theorem 12.24 implies that the optimization over states ψ_{RA} in the definition of the Υ-information is unnecessary, and we have that

$$\Upsilon(\mathcal{E}_p^{(d)}) = \inf_{\mathcal{F}\in\mathfrak{F}} D((\mathcal{E}_p^{(d)})_{A\to B}(\Phi_{RA})\|\mathcal{F}_{A\to B}(\Phi_{RA})), \tag{12.3.61}$$

where Φ_{RA} is the maximally entangled state with Schmidt rank d. The Choi state $\rho_{RB}^{\mathcal{E}_p^{(d)}} = (\mathcal{E}_p^{(d)})_{A\to B}(\Phi_{RA})$ of the erasure channel is

$$\rho_{RB}^{\mathcal{E}_p^{(d)}} = (1-p)\Phi_{RB} + p\frac{\mathbb{1}_R}{d} \otimes |e\rangle\langle e|. \tag{12.3.62}$$

Now, let us make a particular choice of the map $\mathcal{F}$ in the minimization over the completely positive maps in $\mathfrak{F}$. Suppose that $\mathcal{F}_{A\to B}$ is such that

$$\mathcal{F}_{A\to B}(\Phi_{RA}) = \frac{1-p}{d}\Phi_{RB} + p\frac{\mathbb{1}_R}{d} \otimes |e\rangle\langle e| =: \sigma_{RB}^{\mathcal{F}}, \tag{12.3.63}$$

which implies via (4.2.14) that its action on a general input state ρ_A is as follows:

$$\mathcal{F}_{A\to B}(\rho_A) = \frac{1-p}{d}\rho_A + p|e\rangle\langle e| \leq \frac{1-p}{d}\mathbb{1}_A + p|e\rangle\langle e|. \tag{12.3.64}$$

Note that the map $\mathcal{F}_{A\to B}$ defined in this way is indeed in the set $\mathfrak{F}$, as demonstrated by the inequality in (12.3.64) and the fact that the operator on the right-hand side of (12.3.64) is a quantum state. Then,

$$\Upsilon(\mathcal{E}_p^{(d)}) \leq D(\rho_{RB}^{\mathcal{E}_p^{(d)}}\|\sigma_{RB}^{\mathcal{F}}). \tag{12.3.65}$$

It is straightforward to show that

$$\mathrm{Tr}\left[\rho_{RB}^{\mathcal{E}_p^{(d)}} \log_2 \rho_{RB}^{\mathcal{E}_p^{(d)}}\right] = (1-p)\log_2(1-p) + p\log_2\left(\frac{p}{d}\right), \tag{12.3.66}$$

and

$$\mathrm{Tr}\left[\rho_{RB}^{\mathcal{E}_p^{(d)}} \log_2 \sigma_{RB}^{\mathcal{F}}\right] = (1-p)\log_2\left(\frac{1-p}{d}\right) + p\log_2\left(\frac{p}{d}\right), \tag{12.3.67}$$

which means that

$$\Upsilon(\mathcal{E}_p^{(d)}) \leq (1-p)\log_2 d = \chi(\mathcal{E}_p^{(d)}). \tag{12.3.68}$$

This concludes the proof. ■

The classical capacity of the quantum erasure channel and its strong converse now follow as a direct corollary of (12.3.59), (12.2.15), Proposition 12.32, the irreducible covariance of the erasure channel, and Theorem 12.26.

Theorem 12.33 Classical Capacity of the Erasure Channel

For $d \geq 2$ and $p \in [0,1]$, the following equality holds for the classical capacity and strong converse classical capacity of the quantum erasure channel $\mathcal{E}_p^{(d)}$:

$$C(\mathcal{E}_p^{(d)}) = \widetilde{C}(\mathcal{E}_p^{(d)}) = (1-p)\log_2 d. \tag{12.3.69}$$

12.3.2 Amplitude Damping Channel

Recall from (4.5.1) that the amplitude damping channel is defined as

$$\mathcal{A}_\gamma(\rho) = A_1 \rho A_1^\dagger + A_2 \rho A_2^\dagger, \tag{12.3.70}$$

where

$$A_1 = \sqrt{\gamma}|0\rangle\langle 1|, \quad A_2 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|. \tag{12.3.71}$$

It can be shown that

$$\chi(\mathcal{A}_\gamma) = \frac{1}{2}\left(f(r^*) - \log_2(1-q^2) - qf'(q)\right), \tag{12.3.72}$$

where

$$f(x) := (1+x)\log_2(1+x) - (1-x)\log_2(1-x), \tag{12.3.73}$$

$$f'(x) = \log_2\left(\frac{1+x}{1-x}\right), \tag{12.3.74}$$

$$r^* := \sqrt{1-\gamma - \frac{(q-\gamma)^2}{1-\gamma} + q^2}, \tag{12.3.75}$$

and q is determined via

$$(\gamma q - \gamma^2 - \gamma(1-\gamma))f'(r^*) = -r^*(1-\gamma)f'(q). \tag{12.3.76}$$

(Please consult the Bibliographic Notes in Section 12.5.) It is worth noting that neither the additivity of the Holevo information for nor the classical capacity of the amplitude damping channel are not known.

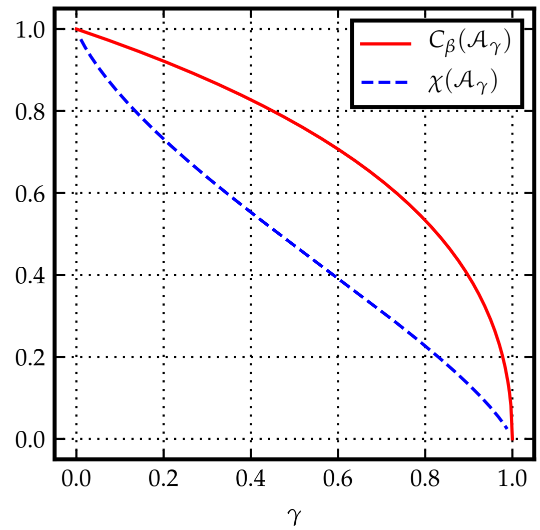


FIGURE 12.7: The Holevo information $\chi(\mathcal{A}_\gamma)$ in (12.3.72) of the amplitude damping channel, which represents a lower bound on its classical capacity. Also shown is the upper bound $C_\beta(\mathcal{A}_\gamma)$ on the classical capacity, which is defined in (12.2.248) via the SDP in (12.2.228), and for the amplitude damping channel it is given by the expression in (12.3.77).

As we have seen, the quantity $C_\beta(\mathcal{A}_\gamma)$ is an upper bound on the classical capacity of the amplitude damping channel, and it can be shown (please consult the Bibliographic Notes in Section 12.5) that

$$C_\beta(\mathcal{A}_\gamma) = \log_2(1 + \sqrt{1-\gamma}). \tag{12.3.77}$$

See Figure 12.7 for a plot of both $\chi(\mathcal{A}_\gamma)$ and $C_\beta(\mathcal{A}_\gamma)$.

12.3.3 Hadamard Channels

In this section, we prove that the Holevo information is additive for all Hadamard channels.

Theorem 12.34 Additivity of Holevo Information for Hadamard Channels

For a Hadamard channel $\mathcal{N}$ and an arbitrary channel $\mathcal{M}$, the following additivity relation holds

$$\chi(\mathcal{N} \otimes \mathcal{M}) = \chi(\mathcal{N}) + \chi(\mathcal{M}). \tag{12.3.78}$$

Proof: Using the expression in (12.3.2), we have that

$$\chi(\mathcal{N} \otimes \mathcal{M}) = \sup_{\{(p(x),\psi^x)\}_{x\in\mathcal{X}}} \left[H\left(\sum_{x\in\mathcal{X}} p(x)(\mathcal{N} \otimes \mathcal{M})(\psi^x_{A_1A_2})\right) - \sum_{x\in\mathcal{X}} p(x) H((\mathcal{N}^c \otimes \mathcal{M}^c)(\psi^x_{A_1A_2})) \right]. \tag{12.3.79}$$

Now, for every bipartite state ρ_{AB}, it follows from strong subadditivity that $H(\rho_{AB}) \leq H(\rho_A) + H(\rho_B)$ a fact known as the subadditivity of the quantum entropy (consider (10.6.88) with system C trivial). Using this for the first term in (12.3.79), we find that

$$H\left(\sum_{x\in\mathcal{X}} p(x)(\mathcal{N} \otimes \mathcal{M})(\psi^x_{A_1A_2})\right) \leq H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\psi^x_{A_1})\right) + H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{M}(\psi^x_{A_2})\right). \tag{12.3.80}$$

We now make use of the following identity, which is straightforward to verify: for every finite alphabet $\mathcal{X}$ and ensemble $\{(p(x), \rho^x_{AB})\}$,

$$\sum_{x\in\mathcal{X}} p(x) D(\rho^x_{AB} \| \rho^x_A \otimes \rho^x_B) = \sum_{x\in\mathcal{X}} p(x) H(\rho^x_A) + \sum_{x\in\mathcal{X}} p(x) H(\rho^x_B) - \sum_{x\in\mathcal{X}} p(x) H(\rho^x_{AB}). \tag{12.3.81}$$

We use this for the second term in (12.3.79) to conclude that

$$\sum_{x\in\mathcal{X}} p(x)H((\mathcal{N}^c\otimes\mathcal{M}^c)(\psi^x_{A_1A_2}))$$
$$= \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}^c(\psi^x_{A_1})) + \sum_{x\in\mathcal{X}} p(x)H(\mathcal{M}^c(\psi^x_{A_2})) - \sum_{x\in\mathcal{X}} p(x)D((\mathcal{N}^c\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})\|\mathcal{N}^c(\psi^x_{A_1})\otimes\mathcal{M}^c(\psi^x_{A_2})). \tag{12.3.82}$$

Now, let us focus on the relative entropy term in the expression above. Since $\mathcal{N}$ is a Hadamard channel, by Proposition 4.17 we know that the complementary channel $\mathcal{N}^c$ is entanglement-breaking. Then, from Theorem 4.15, we know that every entanglement-breaking channel can be written as the composition of a measurement channel followed by a preparation channel. This means that we can write $\mathcal{N}^c$ as $\mathcal{N}^c = \mathcal{P}\circ\mathcal{M}_{\text{qc}}$, where $\mathcal{M}_{\text{qc}}$ is the measurement (or quantum–classical) channel, and $\mathcal{P}$ is the preparation channel. Using the data-processing inequality for the quantum relative entropy, for all $x\in\mathcal{X}$ we obtain

$$D((\mathcal{N}^c\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})\|\mathcal{N}^c(\psi^x_{A_1})\otimes\mathcal{M}(\psi^x_{A_2}))$$
$$= D((\mathcal{P}\circ\mathcal{M}^c_{\text{qc}}\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})\|(\mathcal{P}\circ\mathcal{M}_{\text{qc}})(\psi^x_{A_1})\otimes\mathcal{M}^c(\psi^x_{A_2})) \tag{12.3.83}$$
$$\leq D((\mathcal{M}_{\text{qc}}\otimes\mathcal{M})(\psi^x_{A_1A_2})\|\mathcal{M}^c_{\text{qc}}(\psi^x_{A_1})\otimes\mathcal{M}^c(\psi^x_{A_2})). \tag{12.3.84}$$

Then, using the identity (12.3.81) once again, we obtain

$$\sum_{x\in\mathcal{X}} p(x)D((\mathcal{M}_{\text{qc}}\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})\|\mathcal{M}_{\text{qc}}(\psi^x_{A_1})\otimes\mathcal{M}^c(\psi^x_{A_2}))$$
$$= \sum_{x\in\mathcal{X}} p(x)H(\mathcal{M}_{\text{qc}}(\psi^x_{A_1})) + \sum_{x\in\mathcal{X}} p(x)H(\mathcal{M}^c(\psi^x_{A_2})) - \sum_{x\in\mathcal{X}} p(x)H((\mathcal{M}_{\text{qc}}\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})). \tag{12.3.85}$$

Now, let the measurement channel $\mathcal{M}_{\text{qc}}$ have the associated POVM $\{M^y_{A_1}\}_{y\in\mathcal{Y}}$ for some finite alphabet $\mathcal{Y}$. Then, letting $q(y|x) \coloneqq \text{Tr}[M_y\psi^x_{A_1}]$, we have that

$$H(\mathcal{M}_{\text{qc}}(\psi^x_{A_1})) = H(Y|X=x) \tag{12.3.86}$$

for all $x\in\mathcal{X}$. Also, for every $x\in\mathcal{X}$, we find that

$$(\mathcal{M}_{\text{qc}}\otimes\mathcal{M}^c)(\psi^x_{A_1A_2}) = \sum_{y\in\mathcal{Y}} q(y|x)|y\rangle\langle y|_Y\otimes\rho^{x,y}_{A_2}, \tag{12.3.87}$$

where $\rho^{x,y}_{A_2} := \frac{1}{q(y|x)}\mathrm{Tr}_{A_1}[(M^y_{A_1} \otimes \mathbb{1}_{A_2})\psi^x_{A_1A_2}]$. Note that

$$\begin{aligned}\sum_{y\in\mathcal{Y}} q(y|x)\rho^{x,y}_{A_2} &= \sum_{y\in\mathcal{Y}} \mathrm{Tr}_{A_1}[(M^y_{A_1} \otimes \mathbb{1}_{A_2})\psi^x_{A_1A_2}] \\ &= \mathrm{Tr}_{A_1}[\psi^x_{A_1A_2}] \\ &= \psi^x_{A_2}\end{aligned} \tag{12.3.88}$$

for all $x \in \mathcal{X}$. Therefore, for all $x \in \mathcal{X}$,

$$\begin{aligned}&H((\mathcal{M}_{\mathrm{qc}} \otimes \mathcal{M})(\psi^x_{A_1A_2})) \\ &\quad = H\left(\sum_{y\in\mathcal{Y}} q(y|x)|y\rangle\langle y|_Y \otimes \mathcal{M}^c(\rho^{x,y}_{A_2})\right) &(12.3.89)\\ &\quad = H(Y|X=x) + \sum_{y\in\mathcal{Y}} q(y|x)H(\mathcal{M}^c(\rho^{x,y}_{A_2})), &(12.3.90)\end{aligned}$$

where the last equality follows from the direct-sum property of the quantum entropy. Putting everything together, we obtain

$$\begin{aligned}&H\left(\sum_{x\in\mathcal{X}} p(x)(\mathcal{N}\otimes\mathcal{M})(\psi^x_{A_1A_2})\right) - \sum_{x\in\mathcal{X}} p(x)H((\mathcal{N}^c\otimes\mathcal{M}^c)(\psi^x_{A_1A_2})) \\ &\quad \leq H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\psi^x_{A_1})\right) - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}^c(\psi^x_{A_1})) + H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{M}(\psi^x_{A_2})\right) \\ &\qquad - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{M}^c(\psi^x_{A_2})) + \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \\ &\qquad + \sum_{x\in\mathcal{X}} p(x)H(\mathcal{M}^c(\psi^x_{A_2})) - \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \\ &\qquad - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q(y|x)H(\mathcal{M}^c(\psi^{x,y}_{A_2})) &(12.3.91)\\ &\quad = H\left(\sum_{x\in\mathcal{X}} p(x)\mathcal{N}(\psi^x_{A_1})\right) - \sum_{x\in\mathcal{X}} p(x)H(\mathcal{N}^c(\psi^x_{A_1})) &(12.3.92)\\ &\qquad + H\left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q(y|x)\mathcal{M}(\rho^{x,y}_{A_2})\right) \\ &\qquad - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q(y|x)H(\mathcal{M}^c(\rho^{x,y}_{A_2}))\end{aligned}$$

$$\leq \chi(\mathcal{N}) + \chi(\mathcal{M}), \tag{12.3.93}$$

where we have used (12.3.88) and, to obtain the last inequality, the fact that the first two terms in (12.3.93) are of the form of the objective function in the expression in (12.3.1) for the Holevo information $\chi(\mathcal{N})$, and similarly for the last two terms, in which the ensemble is $\{(p(x)q(y|x), \rho_{A_2}^{x,y}) : x \in \mathcal{X},\ y \in \mathcal{Y}\}$.

Since the ensemble $\{(p(x), \psi_{A_1 A_2}^x)\}_{x \in \mathcal{X}}$ used to obtain (12.3.93) is arbitrary, we conclude that

$$\chi(\mathcal{N} \otimes \mathcal{M}) \leq \chi(\mathcal{M}) + \chi(\mathcal{N}), \tag{12.3.94}$$

which implies, via the superadditivity in (12.2.44) that $\chi(\mathcal{N} \otimes \mathcal{M}) = \chi(\mathcal{N}) + \chi(\mathcal{M})$, as required. ■

Exercise 12.1

Prove that the classical capacity of the d-dimensional dephasing channel, $d \geq 2$, is $\log_2 d$.

12.4 Summary

In this chapter, we developed the theory of classical communication over a quantum channel, adopting a similar structure to that of the previous chapter. We began with the one-shot setting of classical communication, and we defined the one-shot classical capacity of a quantum channel in Definition 12.2. We then derived upper (Proposition 12.3) and lower (Proposition 12.5) bounds on the one-shot classical capacity in terms of the hypothesis testing Holevo information of a quantum channel. The approaches to doing so are conceptually similar to those from the previous chapter. However, there are extra steps involved in deriving the lower bound, called derandomization and expurgation, that establish the existence of a code with maximum error probability no larger than a given threshold and number of bits transmitted roughly equal to the one-shot Holevo information.

With the fundamental information-theoretic arguments established in the one-shot setting, we then moved on to the asymptotic setting of classical communication. One of the main results is that the regularized Holevo information of a channel is equal to its classical capacity (Theorem 12.13). We then considered some special cases: for entanglement-breaking, Hadamard, depolarizing, and erasue channels, the Holevo information is not only equal to the classical capacity but also equal to the strong converse classical capacity (we showed the proofs in full for entanglement-breaking and erasure channels, but deferred to the literature for the others). We discussed general upper bounds on the classical capacity, including the Υ-information and C_β semi-definite programming bound.

Going forward from here, the methods of position-based coding and sequential decoding are useful for the tasks of secret key distillation (Chapter 15) and private communication (Chapter 16), and the concept of derandomization appears again in the context of private communication. The Holevo information will also play a role in achievable rates for these tasks.

12.5 Bibliographic Notes

Classical communication over quantum channels is one of the earliest settings considered in quantum information theory. A key early work on the topic includes Holevo (1973), in which the Holevo upper bound on classical capacity was established. Many years later, after the advent of quantum computing, the Holevo information lower bound on classical capacity was established by Holevo (1998) and Schumacher and Westmoreland (1997). Prior to these works, Hausladen et al. (1996) proved the same lower bound for the special case of a channel that accepts classical inputs and outputs pure quantum states.

Classical communication in the one-shot setting has been considered by a number of authors, including Hayashi and Nagaoka (2003); Hayashi (2007); Mosonyi and Datta (2009); Mosonyi and Hiai (2011); Renes and Renner (2011); Wang and Renner (2012); Matthews and Wehner (2014); Sharma and Warsi (2013); Datta et al. (2013); Tomamichel and Hayashi (2013); Wilde (2013); Anshu et al. (2019); Qi et al. (2018b); Oskouei et al. (2019). Proposition 12.3 is due to Matthews and Wehner (2014). The second part of Theorem 12.4 is due to Wilde et al. (2014). A variation of Theorem 12.5 (for average error probability with uniformly random message) is due to Wang and Renner (2012). The proof given here is due to Oskouei et al. (2019), however with some variations given in this book to account for maximal error probability. At the same time, the proof uses the method of position-based coding (Anshu et al., 2019), with the derandomization argument as given by Qi et al. (2018b).

Additivity of Holevo information for entanglement-breaking channels was established by Shor (2002a), for Hadamard channels by King et al. (2007), for the depolarizing channel by King (2003b), and for the erasure channel by Bennett et al. (1997). The fact that the Holevo information is the strong converse classical capacity of the depolarizing channel was proven by Koenig and Wehner (2009). Additivity of the sandwiched Rényi-Holevo information for entanglement-breaking channels (Theorem 12.16) was established by Wilde et al. (2014), by building upon earlier seminal results of King (2003a) subsequently generalized by Holevo (2006). That is, Lemma 12.18 is due to King (2003a); Holevo (2006). Lemma 12.17 is due to Wilde et al. (2014).

The Υ-information of a quantum channel and its variants were defined by Wang et al. (2019c). The same authors established bounds on classical capacity involving Υ-information. The strong converse for the classical capacity of the quantum erasure channel is due to Wilde and Winter (2014), but here we have followed the approach of Wang et al. (2019c). The semi-definite programming upper bound $C_\beta(\mathcal{N})$ for the classical capacity of a quantum channel $\mathcal{N}$ was established by Wang et al. (2018).

The Holevo information of covariant channels was studied by Holevo (2002b).

A proof of the fact that the limit in the definition of the regularized Holevo information of a channel exists was given by Barnum et al. (1998).

The formula in (12.3.72) for the Holevo information of the amplitude damping channel was derived by Li-Zhen and Mao-Fa (2007b), using the techniques of Cortese (2002) and Berry (2005). The formula in (12.3.77) for the quantity C_β for the same channel was determined by Wang et al. (2018) (see also Khatri et al. (2020)).

Appendix 12.A The $\alpha \to 1$ Limit of the Sandwiched Rényi Υ-Information of a Channel

In this section, we show that

$$\lim_{\alpha\to 1^+} \widetilde{\Upsilon}_\alpha(\mathcal{N}) = \Upsilon(\mathcal{N}), \tag{12.A.1}$$

where we recall that

$$\widetilde{\Upsilon}_\alpha(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.A.2}$$

$$\Upsilon(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})). \tag{12.A.3}$$

Here, ψ_{RA} is a pure state, with the dimension of R equal to the dimension of A, and the infimum is over the set $\mathfrak{F}$ of completely positive maps defined as

$$\mathfrak{F} = \{\mathcal{F}_{A\to B} : \exists\, \sigma_B \geq 0,\ \mathrm{Tr}[\sigma_B] \leq 1,\ \mathcal{F}_{A\to B}(\rho_A) \leq \sigma_B\ \forall\, \rho_A \in \mathrm{D}(\mathcal{H}_A)\}. \tag{12.A.4}$$

Now, since the sandwiched Rényi relative entropy increases monotonically with α (see Proposition 7.31), and since $\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$ (see Proposition 7.30), we obtain

$$\begin{aligned}
\lim_{\alpha\to 1^+} \widetilde{\Upsilon}_\alpha(\mathcal{N}) &= \inf_{\alpha\in(1,\infty)} \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) && (12.A.5)\\
&= \sup_{\psi_{RA}} \inf_{\alpha\in(1,\infty)} \inf_{\mathcal{F}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) && (12.A.6)\\
&= \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} \inf_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) && (12.A.7)\\
&= \sup_{\psi_{RA}} \inf_{\mathcal{F}\in\mathfrak{F}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})) && (12.A.8)\\
&= \Upsilon(\mathcal{N}), && (12.A.9)
\end{aligned}$$

as required, where to obtain the second equality we made use of the minimax theorem in Theorem 2.25 to exchange $\inf_{\alpha\in(1,\infty)}$ and $\sup_{\psi_{RA}}$. Specifically, we applied that theorem to the function

$$(\alpha, \psi_{RA}) \mapsto \inf_{\mathcal{F}\in\mathfrak{F}} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathcal{F}_{A\to B}(\psi_{RA})), \tag{12.A.10}$$

which is monotonically increasing in the first argument and continuous in the second argument.

Appendix 12.B Proof of the Additivity of $C_\beta(\mathcal{N})$

In this section, we prove that

$$C_\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_\beta(\mathcal{N}_1) + C_\beta(\mathcal{N}_2). \tag{12.B.1}$$

Noting that $C_\beta(\mathcal{N}) = \log_2 \beta(\mathcal{N})$, with $\beta(\mathcal{N})$ defined in (12.2.228), the expression above for the additivity of $C_\beta(\mathcal{N})$ is equivalent to the multiplicativity of $\beta(\mathcal{N})$, i.e.,

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) = \beta(\mathcal{N}_1) \cdot \beta(\mathcal{N}_2). \tag{12.B.2}$$

We now prove that this equality holds. We start with the following lemma.

Lemma 12.35

Let A and B be Hermitian operators such that $-A \leq B \leq A$, and let C and D be Hermitian operators such that $-C \leq D \leq C$. Then,

$$-A \otimes C \leq B \otimes D \leq A \otimes C. \tag{12.B.3}$$

PROOF: The condition $-A \leq B \leq A$ is equivalent to $A - B \geq 0$ and $A + B \geq 0$, and the condition $-C \leq D \leq C$ is equivalent to $C - D \geq 0$ and $C + D \geq 0$. These inequalities imply that

$$(A - B) \otimes (C - D) \geq 0, \tag{12.B.4}$$

$$(A + B) \otimes (C + D) \geq 0, \tag{12.B.5}$$

$$(A - B) \otimes (C + D) \geq 0, \tag{12.B.6}$$

$$(A + B) \otimes (C - D) \geq 0. \tag{12.B.7}$$

Expanding the left-hand side of these inequalities gives

$$A \otimes C - B \otimes C - A \otimes D + B \otimes D \geq 0, \tag{12.B.8}$$

$$A \otimes C + B \otimes C + A \otimes D + B \otimes D \geq 0, \tag{12.B.9}$$

$$A \otimes C - B \otimes C + A \otimes D - B \otimes D \geq 0, \tag{12.B.10}$$

$$A \otimes C + B \otimes C - A \otimes D - B \otimes D \geq 0. \tag{12.B.11}$$

Now, adding the first two of these inequalities implies that $A \otimes C + B \otimes D \geq 0$, which is equivalent to the left-hand side of (12.B.3). Adding the last two inequalities implies that $A \otimes C - B \otimes D \geq 0$, which is equivalent to the right-hand side of (12.B.3). ■

An immediate corollary of the lemma above is the following: for all Hermitian operators A, B, C, D such that $0 \leq B \leq A$ and $0 \leq D \leq C$, it holds that

$$0 \leq B \otimes D \leq A \otimes C. \tag{12.B.12}$$

Indeed, the condition $0 \leq B \leq A$ implies that $A \geq 0$, which is equivalent to $-A \leq 0$, which means that $-A \leq B$ holds. Similarly, we get that $-C \leq D$. So we have that $-A \leq B \leq A$ and $-C \leq D \leq C$. The result then follows by applying the lemma above.

Now, let us start the proof of (12.B.2) by showing

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq \beta(\mathcal{N}_1) \cdot \beta(\mathcal{N}_2). \tag{12.B.13}$$

Recall from (12.2.228) that

$$\beta(\mathcal{N}) = \begin{cases} \text{infimum} & \text{Tr}[S_B] \\ \text{subject to} & -R_{AB} \leq \text{T}_B[\Gamma^{\mathcal{N}}_{AB}] \leq R_{AB}, \\ & -\mathbb{1}_A \otimes S_B \leq \text{T}_B[R_{AB}] \leq \mathbb{1}_A \otimes S_B. \end{cases} \tag{12.B.14}$$

Now, let (R^1_{AB}, S^1_B) be a feasible point in the SDP for $\beta(\mathcal{N}_1)$, and let (R^2_{AB}, S^2_B) be a feasible point in the SDP for $\beta(\mathcal{N}_2)$. Each pair thus satisfies the constraints in (12.B.14). Using Lemma 12.35, the first of these constraints implies that

$$-R^1_{A_1B_1} \otimes R^2_{A_2B_2} \leq \text{T}_{B_1}[\Gamma^{\mathcal{N}_1}_{A_1B_1}] \otimes \text{T}_{B_2}[\Gamma^{\mathcal{N}_2}_{A_2B_2}] \leq R^1_{A_1B_1} \otimes R^2_{A_2B_2}. \tag{12.B.15}$$

Furthermore, observe that

$$\text{T}_{B_1}[\Gamma^{\mathcal{N}_1}_{A_1B_1}] \otimes \text{T}_{B_2}[\Gamma^{\mathcal{N}_2}_{A_2B_2}] = \text{T}_{B_1B_2}[\Gamma^{\mathcal{N}_1}_{A_1B_1} \otimes \Gamma^{\mathcal{N}_2}_{A_2B_2}] \tag{12.B.16}$$

$$= \text{T}_{B_1B_2}[\Gamma^{\mathcal{N}_1 \otimes \mathcal{N}_2}_{A_1A_2B_1B_2}]. \tag{12.B.17}$$

Using this, along with Lemma 12.35, the second constraint in (12.B.14) implies that

$$-\mathbb{1}_{A_1A_2} \otimes S^1_{B_1} \otimes S^2_{B_2} \leq \text{T}_{B_1B_2}[R^1_{A_1B_1} \otimes R^2_{A_2B_2}] \leq \mathbb{1}_{A_1A_2} \otimes S^1_{B_1} \otimes S^2_{B_2}. \tag{12.B.18}$$

Now, the inequalities in (12.B.17) and (12.B.18) imply that $(R^1_{A_1B_1} \otimes R^2_{A_2B_2}, S^1_{B_1} \otimes S^2_{B_2})$ is a feasible point in the SDP for $\beta(\mathcal{N}_1 \otimes \mathcal{N}_2)$. This means that

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq \text{Tr}[S^1_{B_1} \otimes S^2_{B_2}] = \text{Tr}[S^1_{B_1}]\text{Tr}[S^2_{B_2}]. \tag{12.B.19}$$

Since $(R^1_{A_1B_1}, S^1_{B_1})$ and $(R^2_{A_2B_2}, S^2_{B_2})$ are arbitrary feasible points in the SDPs for $\beta(\mathcal{N}_1)$ and $\beta(\mathcal{N}_2)$, respectively, the inequality in (12.B.19) holds for the feasible points achieving $\beta(\mathcal{N}_1)$ and $\beta(\mathcal{N}_2)$. This means that

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq \beta(\mathcal{N}_1) \cdot \beta(\mathcal{N}_2), \tag{12.B.20}$$

as required.

The prove the reverse inequality, i.e.,

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \beta(\mathcal{N}_1)\beta(\mathcal{N}_2), \tag{12.B.21}$$

we turn to the SDP dual to the one in (12.B.14).

Lemma 12.36

For every quantum channel $\mathcal{N}$, the SDP dual to the SDP in (12.B.14) for $\beta(\mathcal{N})$ is given by

$$\widehat{\beta}(\mathcal{N}) := \begin{cases} \text{supremum} & \operatorname{Tr}\left[\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}](K_{AB} - M_{AB})\right], \\ \text{subject to} & K_{AB} + M_{AB} \leq \mathrm{T}_B[E_{AB} + F_{AB}], \\ & E_B + F_B \leq \mathbb{1}_B, \\ & K_{AB}, M_{AB}, E_{AB}, F_{AB} \geq 0. \end{cases} \tag{12.B.22}$$

Furthermore, it holds that $\widehat{\beta}(\mathcal{N}) = \beta(\mathcal{N})$.

Proof: Using the formulation of the SDP for $\beta(\mathcal{N})$ as in (12.2.229), the dual to the SDP for $\beta(\mathcal{N})$ is simply

$$\widehat{\beta}(\mathcal{N}) = \begin{cases} \text{supremum} & \operatorname{Tr}[DY] \\ \text{subject to} & \Phi^\dagger(Y) \leq C, \\ & Y \geq 0, \end{cases} \tag{12.B.23}$$

where

$$C = \begin{pmatrix} \mathbb{1}_B & 0 \\ 0 & 0_{AB} \end{pmatrix}, \quad D = \begin{pmatrix} \mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}] & 0 & 0 & 0 \\ 0 & -\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}] & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{12.B.24}$$

and the map Φ is defined as

$$\Phi(X) = \begin{pmatrix} R_{AB} & 0 & 0 & 0 \\ 0 & R_{AB} & 0 & 0 \\ 0 & 0 & \mathbb{1}_R \otimes S_B - \mathrm{T}_B[R_{AB}] & 0 \\ 0 & 0 & 0 & \mathbb{1}_R \otimes S_B + \mathrm{T}_B[R_{AB}] \end{pmatrix}, \tag{12.B.25}$$

$$X = \begin{pmatrix} S_B & 0 \\ 0 & R_{AB} \end{pmatrix}. \tag{12.B.26}$$

To determine the adjoint $\Phi^\dagger$, we first observe that, since the operators C and D are block diagonal, the objective function $\operatorname{Tr}[DY]$ of the dual problem involves only the diagonal blocks of Y. Furthermore, the fact that $\Phi(X)$ and X are block diagonal means that the condition $\operatorname{Tr}[\Phi(X)Y] = \operatorname{Tr}[X\Phi^\dagger(Y)]$ defining the adjoint map $\Phi^\dagger$ involves only the diagonal blocks of Y. Therefore, if the dual problem is feasible, then there is always a feasible point Y that is block diagonal. This means that, without loss of generality, we can let

$$Y = \begin{pmatrix} Y_{AB}^1 & 0 & 0 & 0 \\ 0 & Y_{AB}^2 & 0 & 0 \\ 0 & 0 & Y_{AB}^3 & 0 \\ 0 & 0 & 0 & Y_{AB}^4 \end{pmatrix}, \tag{12.B.27}$$

with $Y_{AB}^1, Y_{AB}^2, Y_{AB}^3, Y_{AB}^4 \geq 0$. Then,

$$\begin{aligned} &\operatorname{Tr}[\Phi(X)Y] \\ &= \operatorname{Tr}\left[\begin{pmatrix} R_{AB} & 0 & 0 & 0 \\ 0 & R_{AB} & 0 & 0 \\ 0 & 0 & \mathbb{1}_R \otimes S_B - \mathrm{T}_B[R_{AB}] & 0 \\ 0 & 0 & 0 & \mathbb{1}_R \otimes S_B + \mathrm{T}_B[R_{AB}] \end{pmatrix}\right. \\ &\qquad\qquad \left.\times \begin{pmatrix} Y_{AB}^1 & 0 & 0 & 0 \\ 0 & Y_{AB}^2 & 0 & 0 \\ 0 & 0 & Y_{AB}^3 & 0 \\ 0 & 0 & 0 & Y_{AB}^4 \end{pmatrix}\right] &(12.B.28) \\ &= \operatorname{Tr}\big[R_{AB}Y_{AB}^1 + R_{AB}Y_{AB}^2 + (\mathbb{1}_R \otimes S_B - \mathrm{T}_B[R_{AB}])Y_{AB}^3 \\ &\qquad + (\mathbb{1}_R \otimes S_B + \mathrm{T}_B[R_{AB}])Y_{AB}^4\big] &(12.B.29) \\ &= \operatorname{Tr}[S_B(Y_B^3 + Y_B^4)] + \operatorname{Tr}[R_{AB}(Y_{AB}^1 + Y_{AB}^2 + \mathrm{T}_B[Y_{AB}^4 - Y_{AB}^3])] &(12.B.30) \\ &= \operatorname{Tr}\left[\begin{pmatrix} S_B & 0 \\ 0 & R_{AB} \end{pmatrix} \begin{pmatrix} Y_B^3 + Y_B^4 & 0 \\ 0 & Y_{AB}^1 + Y_{AB}^2 + \mathrm{T}_B[Y_{AB}^4 - Y_{AB}^3] \end{pmatrix}\right]. &(12.B.31) \end{aligned}$$

For the last line to be equal to $\operatorname{Tr}[X\Phi^\dagger(Y)]$, we must have

$$\Phi^\dagger(Y) = \begin{pmatrix} Y_B^3 + Y_B^4 & 0 \\ 0 & Y_{AB}^1 + Y_{AB}^2 + \mathrm{T}_B[Y_{AB}^4 - Y_{AB}^3] \end{pmatrix}. \tag{12.B.32}$$

Then, the condition $\Phi^\dagger(Y) \leq C$ is given by

$$\begin{pmatrix} Y_B^3 + Y_B^4 & 0 \\ 0 & Y_{AB}^1 + Y_{AB}^2 + \mathrm{T}_B[Y_{AB}^4 - Y_{AB}^3] \end{pmatrix} \leq \begin{pmatrix} \mathbb{1}_B & 0 \\ 0 & 0_{RB} \end{pmatrix}, \tag{12.B.33}$$

which implies that

$$Y_B^3 + Y_B^4 \leq \mathbb{1}_B, \tag{12.B.34}$$

$$Y^1_{AB} + Y^2_{AB} \leq \mathrm{T}_B[Y^3_{AB} - Y^4_{AB}]. \tag{12.B.35}$$

Then,

$$\mathrm{Tr}[DY] = \mathrm{Tr}[\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}]Y^1_{AB}] - \mathrm{Tr}[\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}]Y^2_{AB}] \tag{12.B.36}$$

$$= \mathrm{Tr}[\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}](Y^1_{AB} - Y^2_{AB})]. \tag{12.B.37}$$

Therefore, the dual is given by

$$\widehat{\beta}(\mathcal{N}) = \begin{cases} \text{supremum} & \mathrm{Tr}\big[\mathrm{T}_B[\Gamma^{\mathcal{N}}_{AB}](K_{AB} - M_{AB})\big] \\ \text{subject to} & K_{AB} + M_{AB} \leq \mathrm{T}_B[E_{AB} - F_{AB}], \\ & E_B + F_B \leq \mathbb{1}_B, \\ & K_{AB}, M_{AB}, E_{AB}, F_{AB} \geq 0, \end{cases} \tag{12.B.38}$$

as required.

To show that $\widehat{\beta}(\mathcal{N}) = \beta(\mathcal{N})$, we need to check that Slater's condition holds (Theorem 2.28). We can pick $E_{AB} = \frac{\mathbb{1}_{AB}}{3d_A}$, $F_{AB} = \frac{\mathbb{1}_{AB}}{6d_A}$, and $K_{AB} = M_{AB} = \frac{\mathbb{1}_{AB}}{24d_A}$, where d_A is the dimension of the space of the system A. Then we have strict inequalities for all of the constraints of the dual problem, which means that Slater's condition holds. The primal $\beta(\mathcal{N})$ and dual $\widehat{\beta}(\mathcal{N})$ are thus equal. ■

With the dual problem in hand, we can now prove (12.B.21). Let $(K^1_{A_1B_1}, M^1_{A_1B_1}, E^1_{A_1B_1}, F^1_{A_1B_1})$ be a feasible point for the dual SDP for $\mathcal{N}_1$, and let $(K^2_{A_2B_2}, M^2_{A_2B_2}, E^2_{A_2B_2}, F^2_{A_2B_2})$ be a feasible point for the dual SDP for $\mathcal{N}_2$. Then, pick

$$K_{A_1B_1A_2B_2} = K^1_{A_1B_1} \otimes K^2_{A_2B_2} + M^1_{A_1B_1} \otimes M^2_{A_2B_2}, \tag{12.B.39}$$

$$M_{A_1B_1A_2B_2} = K^1_{A_1B_1} \otimes M^2_{A_2B_2} + M^1_{A_1B_1} \otimes K^2_{A_2B_2}, \tag{12.B.40}$$

$$E_{A_1B_1A_2B_2} = E^1_{A_1B_1} \otimes E^2_{A_2B_2} + F^1_{A_1B_1} \otimes F^2_{A_2B_2}, \tag{12.B.41}$$

$$F_{A_1B_1A_2B_2} = E^1_{A_1B_1} \otimes F^2_{A_2B_2} + F^1_{A_1B_1} \otimes E^2_{A_2B_2}. \tag{12.B.42}$$

Note that $K_{A_1B_1A_2B_2}, M_{A_1B_1A_2B_2}, E_{A_1B_1A_2B_2}, F_{A_1B_1A_2B_2} \geq 0$. Then,

$$K_{A_1B_1A_2B_2} - M_{A_1B_1A_2B_2} = (K^1_{A_1B_1} - M^1_{A_1B_1}) \otimes (K^2_{A_2B_2} - M^2_{A_2B_2}), \tag{12.B.43}$$

$$K_{A_1B_1A_2B_2} + M_{A_1B_1A_2B_2} = (K^1_{A_1B_1} + M^1_{A_1B_1}) \otimes (K^2_{A_2B_2} + M^2_{A_2B_2}), \tag{12.B.44}$$

$$E_{A_1B_1A_2B_2} - F_{A_1B_1A_2B_2} = (E^1_{A_1B_1} - F^1_{A_1B_1}) \otimes (E^2_{A_2B_2} - F^2_{A_2B_2}), \tag{12.B.45}$$

$$E_{A_1B_1A_2B_2} + F_{A_1B_1A_2B_2} = (E^1_{A_1B_1} + F^1_{A_1B_1}) \otimes (E^2_{A_2B_2} + F^2_{A_2B_2}). \tag{12.B.46}$$

Consider that

$$\begin{aligned}
&K_{A_1B_1A_2B_2} + M_{A_1B_1A_2B_2} \\
&\quad = (K^1_{A_1B_1} + M^1_{A_1B_1}) \otimes (K^2_{A_2B_2} + M^2_{A_2B_2}) && (12.B.47)\\
&\quad \leq \mathrm{T}_{B_1}[E^1_{A_1B_1} - F^1_{A_1B_1}] \otimes \mathrm{T}_{B_2}[E^2_{A_2B_2} - F^2_{A_2B_2}] && (12.B.48)\\
&\quad = \mathrm{T}_{B_1B_2}[(E^1_{A_1B_1} - F^1_{A_1B_1}) \otimes (E^2_{A_2B_2} - F^2_{A_2B_2})] && (12.B.49)\\
&\quad = \mathrm{T}_{B_1B_2}[E_{A_1B_1A_2B_2} - F_{A_1B_1A_2B_2}], && (12.B.50)
\end{aligned}$$

where the inequality follows from the constraints $K^i_{A_iB_i}, M^i_{A_iB_i} \geq 0$ and $K^i_{A_iB_i} + M^i_{A_iB_i} \leq \mathrm{T}_{B_i}[E^i_{A_iB_i} - F^i_{A_iB_i}]$ for $i \in \{1, 2\}$ and from an application of (12.B.12). Furthermore, we have that

$$\begin{aligned}
E_{B_1B_2} + F_{B_1B_2} &= (E^1_{B_1} + F^1_{B_1}) \otimes (E^2_{B_2} + F^2_{B_2}) && (12.B.51)\\
&\leq \mathbb{1}_{B_1} \otimes \mathbb{1}_{B_2} && (12.B.52)\\
&= \mathbb{1}_{B_1B_2}, && (12.B.53)
\end{aligned}$$

where the inequality follows from the constraints $E^i_{B_i}, F^i_{B_i} \geq 0$ and $E^i_{B_i} + F^i_{B_i} \leq \mathbb{1}_{B_i}$ for $i \in \{1, 2\}$ and from an application of (12.B.12). The collection

$$(K_{A_1B_1A_2B_2}, M_{A_1B_1A_2B_2}, E_{A_1B_1A_2B_2}, F_{A_1B_1A_2B_2}) \tag{12.B.54}$$

thus constitutes a feasible point for the SDP in (12.B.38). By restricting the optimization in the SDP to this point, we find that

$$\begin{aligned}
&\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) && (12.B.55)\\
&\quad \geq \mathrm{Tr}\Big[\mathrm{T}_{B_1B_2}[\Gamma^{\mathcal{N}_1\otimes\mathcal{N}_2}_{A_1A_2B_1B_2}](K_{A_1B_1A_2B_2} - M_{A_1B_1A_2B_2})\Big] && (12.B.56)\\
&\quad = \mathrm{Tr}\Big[\Big(\mathrm{T}_{B_1}[\Gamma^{\mathcal{N}_1}_{A_1B_1}] \otimes \mathrm{T}_{B_2}[\Gamma^{\mathcal{N}_2}_{A_2B_2}]\Big)(K_{A_1B_1A_2B_2} - M_{A_1B_1A_2B_2})\Big] && (12.B.57)\\
&= \mathrm{Tr}\Big[\Big(\mathrm{T}_{B_1}[\Gamma^{\mathcal{N}_1}_{A_1B_1}] \otimes \mathrm{T}_{B_2}[\Gamma^{\mathcal{N}_2}_{A_2B_2}]\Big) \\
&\qquad\qquad \times\Big((K^1_{A_1B_1} - M^1_{A_1B_1}) \otimes (K^2_{A_2B_2} - M^2_{A_2B_2})\Big)\Big] && (12.B.58)\\
&= \mathrm{Tr}\Big[\mathrm{T}_{B_1}[\Gamma^{\mathcal{N}_1}_{A_1B_1}](K^1_{A_1B_1} - M^1_{A_1B_1})\Big] \\
&\qquad\qquad \times \mathrm{Tr}\Big[\mathrm{T}_{B_2}[\Gamma^{\mathcal{N}_2}_{A_2B_2}](K^2_{A_2B_2} - M^2_{A_2B_2})\Big]. && (12.B.59)
\end{aligned}$$

Now, since $(K^1_{AB}, M^1_{AB}, E^1_{AB}, F^1_{AB})$ $(K^2_{AB}, M^2_{AB}, E^2_{AB}, F^2_{AB})$ were arbitrary feasible points in the SDPs for $\widehat{\beta}(\mathcal{N}_1) = \beta(\mathcal{N}_1)$ and $\widehat{\beta}(\mathcal{N}_2) = \beta(\mathcal{N}_2)$, respectively, the inequality in (12.B.59) holds for the feasible points achieving $\beta(\mathcal{N}_1)$ and $\beta(\mathcal{N}_2)$. Therefore,

$$\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq \beta(\mathcal{N}_1) \cdot \beta(\mathcal{N}_2). \tag{12.B.60}$$

We have thus shown that $\beta(\mathcal{N}_1 \otimes \mathcal{N}_2) = \beta(\mathcal{N}_1) \cdot \beta(\mathcal{N}_2)$.

Chapter 13

Entanglement Distillation

In the last two chapters, we explored classical communication over quantum channels, in which classical information is encoded into a quantum state, transmitted over a quantum channel, and decoded at the receiving end. In this chapter, we begin our exploration of quantum communication. The goal here is to send quantum information between two spatially separated parties. By "quantum information," we mean that a particular quantum state is transmitted, which is carried physically by some quantum system. As was the case in previous chapters, the particular information carrier is unimportant to us when developing the theoretical results; however, the most common physical manifestation is a photonic encoding, which is useful for long-distance quantum communication.

A basic quantum communication protocol is teleportation, which we developed in Section 5.1. In this protocol, the sender, Alice, initially shares a maximally entangled state with the receiver, Bob. This shared entanglement, along with classical communication, can be used to transmit an arbitrary quantum state perfectly from Alice to Bob. Specifically, if Alice and Bob share a maximally entangled state of Schmidt rank $d \geq 2$, then using this entanglement along with $2\log_2 d$ bits of classical communication, Alice can perfectly transmit an arbitrary state of $\log_2 d$ qubits to Bob. Thus, the quantum teleportation protocol realizes a noiseless quantum channel between Alice and Bob without having to physically transport the particles carrying the quantum information. Of course, this achievement comes at the cost of having a pre-shared maximally entangled state.

How do we obtain maximally entangled states in the first place? In practice, due to noise and other device imperfections, physical sources of entanglement

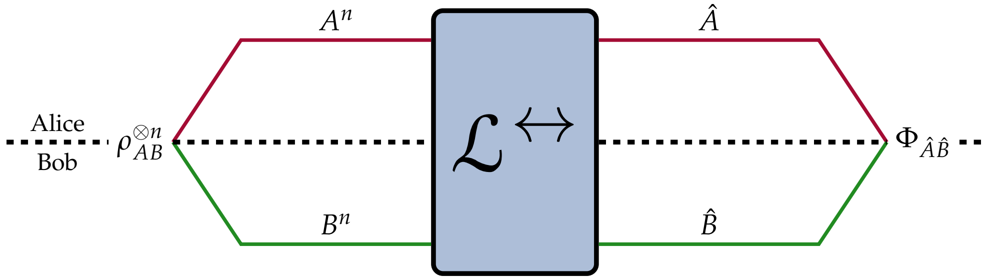


FIGURE 13.1: Given a bipartite state ρ_{AB} shared by Alice and Bob, the task of entanglement distillation is to find the largest d for which a maximally entangled state $|\Phi\rangle_{\hat{A}\hat{B}}$ of Schmidt rank d can be extracted from n copies of ρ_{AB} with the smallest possible error, given a two-way LOCC channel $\mathcal{L}^{\leftrightarrow}_{A^nB^n\to\hat{A}\hat{B}}$ between Alice and Bob.

often only produce mixed entangled states, not the pure, maximally entangled states that are needed for quantum teleportation. The purpose of this chapter is to show that many copies of a mixed entangled state can be used to extract, or *distill*, some smaller number of pure maximally entangled states. These distilled maximally entangled states can then be used for quantum communication via the teleportation protocol. This is a basic strategy for quantum communication that we consider in more detail in Chapter 19, in order to obtain achievable rates for quantum communication over a quantum channel.

Similar to quantum teleportation, in which the allowed resources are local operations by Alice and Bob and one-way classical communication from Alice to Bob, in entanglement distillation we allow Alice and Bob local operations with *two-way* classical communication (that is, communication from Alice to Bob and from Bob to Alice); see Figure 13.1. The goal is to determine, given many copies of a quantum state ρ_{AB}, the maximum rate at which maximally entangled states (i.e., ebits) can be distilled approximately from ρ_{AB}, where the rate is defined as the ratio $\frac{1}{n}\log_2 d$ between the number $\log_2 d$ of approximate ebits extracted and the initial number n of copies of ρ_{AB}. In the asymptotic setting, this maximum rate of entanglement distillation is called the *distillable entanglement of* ρ_{AB}, and we denote it by $E_D(\rho_{AB})$. We often write $E_D(\rho_{AB})$ as $E_D(A;B)_\rho$ in order to explicitly indicate the bipartition between the subsystems.

The shared resource state ρ_{AB} for entanglement distillation has to be entangled to begin with in order for entanglement distillation to be successful. If ρ_{AB} is separable to begin with, then it stays separable after the application of an LOCC channel, and it is not possible to distill high fidelity maximally entangled states from a separable state. This intuitive reasoning becomes formalized in this chapter: some of the entanglement measures from Chapter 9 serve as upper bounds on the distillable entanglement, in both the one-shot (Section 13.1) and asymptotic (Section 13.2) settings. In particular, the Rains relative entropy and squashed entanglement are upper bounds on distillable entanglement. These entanglement measures are currently the best known upper bounds on distillable entanglement, and so we focus exclusively on them for this purpose in this chapter. It is then a trivial consequence of Proposition 9.24, (9.1.149), and Proposition 9.35 that log-negativity, relative entropy of entanglement, and entanglement of formation are upper bounds on distillable entanglement, and so we do not focus on these entanglement measures in this chapter.

We also consider lower bounds on distillable entanglement in this chapter: the lower bound on distillable entanglement in the one-shot setting in Section 13.1.2 is based on the concept of *decoupling*, which is an important concept that we discuss later. This lower bound, when applied in the asymptotic setting, leads to the coherent information lower bound $E_D(\rho_{AB}) \geq I(A\rangle B)_\rho$ on distillable entanglement.

13.1 One-Shot Setting

The one-shot setting for entanglement distillation begins with Alice and Bob sharing the state ρ_{AB}. An *entanglement distillation protocol for* ρ_{AB} is defined by the pair $(d, \mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}})$, where $d \in \mathbb{N}$, $d \geq 1$, and $\mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}$ is an LOCC channel (Definition 4.22), with $d_{\hat{A}} = d_{\hat{B}} = d$. The *distillation error* $p_{\mathrm{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB})$ of the protocol is given by the *infidelity*, defined as

$$p_{\mathrm{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}) \coloneqq 1 - F(\Phi_{\hat{A}\hat{B}}, \mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}(\rho_{AB})) \tag{13.1.1}$$

$$= 1 - \langle\Phi|_{\hat{A}\hat{B}} \mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}(\rho_{AB})|\Phi\rangle_{\hat{A}\hat{B}}, \tag{13.1.2}$$

where $\Phi_{\hat{A}\hat{B}}$ is the maximally entangled state of Schmidt rank d, defined as

$$|\Phi\rangle_{\hat{A}\hat{B}} = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_{\hat{A}} \otimes |i\rangle_{\hat{B}}, \tag{13.1.3}$$

and F is the fidelity (see Section 6.2). To obtain (13.1.2), we used the formula in (6.2.2) for the fidelity between a pure state and a mixed state.

The figure of merit in (13.1.2) is sensible: the error probability $p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB})$ is equal to the probability that the state $\omega_{\hat{A}\hat{B}} := \mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}(\rho_{AB})$ fails an "entanglement test," which is a measurement defined by the POVM

$$\{\Phi_{\hat{A}\hat{B}}, \mathbb{1}_{\hat{A}\hat{B}} - \Phi_{\hat{A}\hat{B}}\}. \tag{13.1.4}$$

Passing the test corresponds to the measurement operator $\Phi_{\hat{A}\hat{B}}$ and failing corresponds to $\mathbb{1}_{\hat{A}\hat{B}} - \Phi_{\hat{A}\hat{B}}$. If $1 - \text{Tr}[\omega_{\hat{A}\hat{B}}\Phi_{\hat{A}\hat{B}}] \leq \varepsilon \in [0,1]$, and $d_{\hat{A}} = d_{\hat{B}} = d \geq 1$, then we say that the final state $\omega_{\hat{A}\hat{B}}$ contains $\log_2 d$ *ε-approximate ebits*.

Definition 13.1 (d,ε) Entanglement Distillation Protocol

An entanglement distillation protocol $(d, \mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}})$ for the state ρ_{AB} is called a (d,ε) *protocol*, with $\varepsilon \in [0,1]$, if $p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB}) \leq \varepsilon$.

Given $\varepsilon \in [0,1]$, the largest number $\log_2 d$ of ε-approximate ebits that can be extracted from a state ρ_{AB} among all (d,ε) entanglement distillation protocols is called the *one-shot ε-distillable entanglement of ρ_{AB}*.

Definition 13.2 One-Shot Distillable Entanglement

Given a bipartite state ρ_{AB} and $\varepsilon \in [0,1]$, the *one-shot ε-distillable entanglement of ρ_{AB}*, denoted by $E_D^{\varepsilon}(\rho_{AB}) \equiv E_D^{\varepsilon}(A;B)_{\rho}$, is defined as

$$E_D^{\varepsilon}(A;B)_{\rho} := \sup_{(d,\mathcal{L}^{\leftrightarrow})} \{\log_2 d : p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB}) \leq \varepsilon\}, \tag{13.1.5}$$

where the optimization is over all $d \geq 1$ and every LOCC channel $\mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}$ with $d_{\hat{A}} = d_{\hat{B}} = d$.

In addition to finding the largest number $\log_2 d$ of ε-approximate ebits that can be extracted from all (d,ε) entanglement distilltion protocols for a given $\varepsilon \in [0,1]$, we can consider the following complementary question: for a given $d \geq 1$, what is the lowest value of ε that can be attained among all (d,ε) entanglement distillation protocols? In other words, what is the value of

$$\varepsilon_D^*(d;\rho_{AB}) := \inf_{\mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}\in\text{LOCC}} \{p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB}) : d_{\hat{A}} = d_{\hat{B}} = d\}, \tag{13.1.6}$$

where the optimization is over every LOCC channel $\mathcal{L}^{\leftrightarrow}_{AB\to\hat{A}\hat{B}}$, with $d_{\hat{A}} = d_{\hat{B}} = d$? In this book, we focus primarily on the problem of optimizing the number of extracted (approximate) ebits rather than the error, and so our primary quantity of interest is the one-shot distillable entanglement $E_D^\varepsilon(\rho_{AB})$.

Calculating the one-shot distillable entanglement is generally a difficult task, because it involves optimizing over every Schmidt rank $d \geq 1$ of the maximally entangled state $\Phi_{\hat{A}\hat{B}}$ and over every LOCC channel $\mathcal{L}_{AB\to\hat{A}\hat{B}}$, with $d_{\hat{A}} = d_{\hat{B}} = d$. We therefore try to estimate the one-shot distillable entanglement by devising upper and lower bounds. We begin in the next section with upper bounds.

13.1.1 Upper Bounds on the Number of Ebits

In this section, we provide three different upper bounds on one-shot distillable entanglement, based on coherent information, Rains relative entropy, and squashed entanglement. Our study of upper bounds on one-shot distillable entanglement begins with coherent information and the following lemma. This can be understood as a fully quantum generalization of Lemma 11.4, in which we are performing the entanglement test in (13.1.4) rather than the comparator test from (11.1.37).

Lemma 13.3

Let A and B be quantum systems with the same dimension $d \geq 1$. Let Φ_{AB} be a maximally entangled state of Schmidt rank d, and let ω_{AB} be an arbitrary bipartite state. If the probability $\text{Tr}[\Phi_{AB}\omega_{AB}]$ that the state ω_{AB} passes the entanglement test defined by the POVM $\{\Phi_{AB}, \mathbb{1}_{AB} - \Phi_{AB}\}$ satisfies

$$\text{Tr}[\Phi_{AB}\omega_{AB}] \geq 1 - \varepsilon \tag{13.1.7}$$

for some $\varepsilon \in [0, 1]$, then

$$\log_2 d \leq I_H^\varepsilon(A\rangle B)_\omega, \tag{13.1.8}$$

where $I_H^\varepsilon(A\rangle B)_\omega$ is the ε-hypothesis testing coherent information (see (7.11.97)).

If, in addition to (13.1.7), we have that $\text{Tr}_B[\omega_{AB}] = \pi_A = \frac{\mathbb{1}_A}{d}$, then

$$2\log_2 d \leq I_H^\varepsilon(A; B)_\omega. \tag{13.1.9}$$

PROOF: By assumption, we have that

$$F(\Phi_{AB}, \rho_{AB}) = \mathrm{Tr}[\Phi_{AB}\omega_{AB}] \geq 1 - \varepsilon. \tag{13.1.10}$$

Now, for every state σ_B, we have that

$$\begin{aligned}
\mathrm{Tr}[\Phi_{AB}(\mathbb{1}_A \otimes \sigma_B)] &= \frac{1}{d}\langle\Gamma|_{AB}(\mathbb{1}_A \otimes \sigma_B)|\Gamma\rangle_{AB} && (13.1.11)\\
&= \frac{1}{d}\mathrm{Tr}[\sigma_B] && (13.1.12)\\
&= \frac{1}{d}, && (13.1.13)
\end{aligned}$$

where we used (2.2.43). Next, recall from (7.11.97) that

$$I_H^\varepsilon(A\rangle B)_\omega = \inf_{\sigma_B} D_H^\varepsilon(\omega_{AB}\|\mathbb{1}_A \otimes \sigma_B), \tag{13.1.14}$$

where

$$D_H^\varepsilon(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) = -\log_2 \inf_{\Lambda_{AB}}\{\mathrm{Tr}[\Lambda_{AB}(\mathbb{1}_A \otimes \sigma_B)] : 0 \leq \Lambda_{AB} \leq \mathbb{1}_{AB},\\ \mathrm{Tr}[\Lambda_{AB}\omega_{AB}] \geq 1 - \varepsilon\}. \tag{13.1.15}$$

Based on (13.1.10), we see that Φ_{AB} is a measurement operator satisfying the constraints for the optimization in the definition of $D_H^\varepsilon(\omega_{AB}\|\mathbb{1}_A \otimes \sigma_B)$. Therefore,

$$\mathrm{Tr}[\Phi_{AB}(\mathbb{1}_A \otimes \sigma_B)] = \frac{1}{d} \geq 2^{-D_H^\varepsilon(\omega_{AB}\|\mathbb{1}_A\otimes\sigma_B)}, \tag{13.1.16}$$

which implies that

$$\log_2 d \leq D_H^\varepsilon(\omega_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{13.1.17}$$

for every state σ_B. Optimizing over σ_B leads to

$$\log_2 d \leq \inf_{\sigma_B} D_H^\varepsilon(\omega_{AB}\|\mathbb{1}_A \otimes \sigma_B) = I_H^\varepsilon(A\rangle B)_\omega, \tag{13.1.18}$$

which is precisely (13.1.8).

Similarly,

$$\mathrm{Tr}[\Phi_{AB}(\pi_A \otimes \sigma_B)] = \frac{1}{d}\mathrm{Tr}[\Phi_{AB}(\mathbb{1}_A \otimes \sigma_B)] \tag{13.1.19}$$

$$= \frac{1}{d^2} \langle \Gamma |_{AB} (\mathbb{1}_A \otimes \sigma_B) | \Gamma \rangle_{AB} \tag{13.1.20}$$

$$= \frac{1}{d^2}, \tag{13.1.21}$$

where the last line follows from the same reasoning for (13.1.11)–(13.1.13). Next, recall that

$$I_H^\varepsilon(A;B)_\omega = \inf_{\sigma_B} D_H^\varepsilon(\omega_{AB} \| \omega_A \otimes \sigma_B). \tag{13.1.22}$$

Therefore, by definition of the hypothesis testing relative entropy,

$$\mathrm{Tr}[\Phi_{AB}(\pi_A \otimes \sigma_B)] = \frac{1}{d^2} \geq 2^{-D_H^\varepsilon(\omega_{AB} \| \pi_A \otimes \sigma_B)}, \tag{13.1.23}$$

which implies that

$$2 \log_2 d \leq D_H^\varepsilon(\omega_{AB} \| \pi_A \otimes \sigma_B). \tag{13.1.24}$$

Since the state σ_B is arbitrary, we obtain

$$2 \log_2 d \leq \inf_{\sigma_B} D_H^\varepsilon(\omega_{AB} \| \pi_A \otimes \sigma_B) = I_H^\varepsilon(A;B)_\omega, \tag{13.1.25}$$

which is precisely (13.1.9). To obtain the last equality, we made use of the assumption $\mathrm{Tr}_B[\omega_{AB}] = \pi_A$. ■

Note that the result of Lemma 13.3 is general and applies to every bipartite state that is close in fidelity to a maximally entangled state. Applying it to the state $\omega_{\hat{A}\hat{B}} = \mathcal{L}_{AB \to \hat{A}\hat{B}}(\rho_{AB})$ at the output of a (d, ε) entanglement distillation protocol for a state ρ_{AB}, we obtain the following result:

Theorem 13.4 Upper Bound on One-Shot Distillable Entanglement

Let ρ_{AB} be a bipartite state. For every (d, ε) entanglement distillation protocol $(d, \mathcal{L}_{AB \to \hat{A}\hat{B}})$ for ρ_{AB}, with $\varepsilon \in (0,1]$ and $d_{\hat{A}} = d_{\hat{B}} = d$, the number of ε-approximate ebits extracted at the end of the protocol is bounded from above by the LOCC-optimized ε-hypothesis testing coherent information of ρ_{AB}, i.e.,

$$\log_2 d \leq \sup_{\mathcal{L}} I_H^\varepsilon(A' \rangle B')_{\mathcal{L}(\rho)}, \tag{13.1.26}$$

where the optimization is over every LOCC channel $\mathcal{L}_{AB \to A'B'}$. Consequently, for the one-shot ε-distillable entanglement, we obtain

$$E_D^{\varepsilon}(A;B)_\rho \leq \sup_{\mathcal{L}} I_H^{\varepsilon}(A'\rangle B')_{\mathcal{L}(\rho)}. \tag{13.1.27}$$

Proof: For a (d,ε) entanglement distillation protocol $(d,\mathcal{L}_{AB\to\hat{A}\hat{B}})$ for ρ_{AB}, by definition the state $\omega_{\hat{A}\hat{B}} = \mathcal{L}_{AB\to\hat{A}\hat{B}}(\rho_{AB})$ satisfies $\mathrm{Tr}[\Phi_{\hat{A}\hat{B}}\omega_{\hat{A}\hat{B}}] \geq 1-\varepsilon$. Therefore, using (13.1.8), we conclude that $\log_2 d \leq I_H^{\varepsilon}(\hat{A}\rangle\hat{B})_{\mathcal{L}(\rho)}$. Since $\mathcal{L}_{AB\to\hat{A}\hat{B}}$ is a particular LOCC channel, we conclude that

$$I_H^{\varepsilon}(\hat{A}\rangle\hat{B})_{\mathcal{L}(\rho)} \leq \sup_{\mathcal{L}} I_H^{\varepsilon}(A'\rangle B')_{\mathcal{L}(\rho)}. \tag{13.1.28}$$

We thus conclude (13.1.26). Now using the definition of $E_D^{\varepsilon}(A;B)_\rho$ in (13.1.5), we obtain the inequality in (13.1.27). ■

We now consider an upper bound based on the Rains relative entropy. In order to place an upper bound on the one-shot distillable entanglement $E_D^{\varepsilon}(\rho_{AB})$ for a given state ρ_{AB} and $\varepsilon \in [0,1]$, we consider states that are useless for entanglement distillation. This is entirely analogous conceptually to what is done for classical and entanglement-assisted classical communication in the previous two chapters, in which we used the set of replacement channels (which are useless for both of these communication tasks) to place an upper bound on the number of transmitted bits in an $(|\mathcal{M}|,\varepsilon)$ protocol, such that the upper bound depends only on the channel $\mathcal{N}$ being used for communication.

What states are useless for entanglement distillation? Note that an intuitive necessary condition for successful entanglement distillation is that the initial state ρ_{AB} should be entangled: if ρ_{AB} is separable, then the output state $\mathcal{L}_{AB\to\hat{A}\hat{B}}(\rho_{AB})$ of an arbitrary entanglement distillation protocol is still a separable state. This suggests that *separable states are useless for entanglement distillation*. To be more precise, separable states are useless for entanglement distillation because they have a very small probability of passing the entanglement test. As we show in Lemma 13.5 below, the following bound holds for every separable state σ_{AB}:

$$\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}, \tag{13.1.29}$$

where d is the Schmidt rank of Φ_{AB}. More generally, operators in the set PPT′, defined as

$$\mathrm{PPT}'(A\!:\!B) = \{\sigma_{AB} : \sigma_{AB} \geq 0,\ \|\mathrm{T}_B(\sigma_{AB})\|_1 \leq 1\}, \tag{13.1.30}$$

are also useless for entanglement distillation, in the sense that a statement analgous to (13.1.29) can be made for them. We now prove this statement.

Lemma 13.5

Let A and B be quantum systems with the same dimension $d \geq 1$. Let Φ_{AB} be a maximally entangled state of Schmidt rank d. If $\sigma_{AB} \in \mathrm{PPT}'(A\!:\!B)$, then $\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}$.

REMARK: Note that Lemma 13.5 implies that $\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}$ for every separable state σ_{AB} because $\mathrm{SEP}(A\!:\!B) \subseteq \mathrm{PPT}'(A\!:\!B)$ (recall Figure 9.2).

PROOF: Using the fact that the partial transpose is self-inverse and self-adjoint, as discussed in (3.2.113) and (3.2.114), respectively, we find that

$$\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] = \mathrm{Tr}[\mathrm{T}_B(\Phi_{AB})\mathrm{T}_B(\sigma_{AB})] \tag{13.1.31}$$

$$= \frac{1}{d}\mathrm{Tr}[U_B F_{AB} U_B^\dagger \mathrm{T}_B(\sigma_{AB})], \tag{13.1.32}$$

where F_{AB} is the unitary swap operator and U_B is a local unitary acting on system B. Here we applied the identity $\mathrm{T}_B(\Phi_{AB}) = \frac{1}{d}U_B F_{AB} U_B^\dagger$ from (3.2.125). Since $U_B F_{AB} U_B^\dagger$ is a unitary operator, by the variational characterization of the trace norm (see (2.2.115)), we obtain

$$\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}\left\|\mathrm{T}_B(\sigma_{AB})\right\|_1 \tag{13.1.33}$$

$$\leq \frac{1}{d}, \tag{13.1.34}$$

where the last line follows from the definition of the set $\mathrm{PPT}'(A\!:\!B)$ in (13.1.30). ■

Due to the fact that SEP ⊆ PPT ⊆ PPT′ (see Figure 9.2), Lemma 13.5 tells us that both separable and PPT states are useless for entanglement distillation. However, due to the fact that separable states are strictly contained in the set of PPT states for all bipartite states except for qubit-qubit and qubit-qutrit states, it follows that there are PPT entangled states that are useless for entanglement distillation. We elaborate upon this point further in Section 13.2.0.1 below, and we show that the distillable entanglement (in the asymptotic setting) vanishes for all PPT states.

The steps followed in the proof of Lemma 13.5 above are completely analogous to the steps in (13.1.11)–(13.1.13) and in (13.1.19)–(13.1.21) of the proof of Lemma 13.3. Therefore, just as Lemma 13.3 was used to establish Proposition 13.4, we can use Lemma 13.5 to place an upper bound on the number $\log_2 d$ of approximate ebits in a bipartite state ω_{AB}.

Proposition 13.6

Fix $\varepsilon \in [0, 1]$, and let A and B be quantum systems with the same dimension $d \geq 1$. Fix a maximally entangled state Φ_{AB} of Schmidt rank d. Let ω_{AB} be an ε-approximate maximally entangled state, in the sense that

$$F(\Phi_{AB}, \omega_{AB}) = \mathrm{Tr}[\Phi_{AB}\omega_{AB}] \geq 1 - \varepsilon. \tag{13.1.35}$$

Then, the number $\log_2 d$ of ε-approximate ebits in ω_{AB} is bounded from above as follows:

$$\log_2 d \leq R_H^{\varepsilon}(A; B)_{\omega}, \tag{13.1.36}$$

where $R_H^{\varepsilon}(A; B)_{\omega}$ is the ε-hypothesis testing Rains relative entropy of ω_{AB} (see (9.3.5)).

Proof: Let σ_{AB} be an arbitrary operator in $\mathrm{PPT}'(A\!:\!B)$. The inequality $\mathrm{Tr}[\Phi_{AB}\omega_{AB}] \geq 1 - \varepsilon$ guarantees that ω_{AB} passes the entanglement test with probability greater than $1 - \varepsilon$. Thus, we conclude that Φ_{AB} is a particular measurement operator satisfying the constraints for $2^{-D_H^{\varepsilon}(\omega_{AB}\|\sigma_{AB})}$. Applying Lemma 13.5 and the definition of $D_H^{\varepsilon}(\omega_{AB}\|\sigma_{AB})$, we conclude that

$$2^{-D_H^{\varepsilon}(\omega_{AB}\|\sigma_{AB})} \leq \mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}. \tag{13.1.37}$$

Rearranging this leads to

$$\log_2 d \leq D_H^{\varepsilon}(\omega_{AB}\|\sigma_{AB}) \tag{13.1.38}$$

Since this inequality holds for every operator $\sigma_{AB} \in \mathrm{PPT}'(A\!:\!B)$, we conclude that

$$\log_2 d \leq \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} D_H^{\varepsilon}(\omega_{AB}\|\sigma_{AB}) = R_H^{\varepsilon}(A; B)_{\omega}, \tag{13.1.39}$$

where we used the definition of $R_H^{\varepsilon}(A; B)_{\omega}$ in (9.3.5) to obtain the last equality. ■

A consequence of Proposition 13.6 is the following upper bound on the one-shot distillable entanglement of ρ_{AB}.

Theorem 13.7 Rains Upper Bound on One-Shot Distillable Entanglement

Let ρ_{AB} be a bipartite state. For every (d, ε) entanglement distillation protocol for ρ_{AB}, with $\varepsilon \in [0, 1]$, we have that

$$\log_2 d \leq R_H^\varepsilon(A; B)_\rho. \tag{13.1.40}$$

Consequently, for the one-shot distillable entanglement, we have

$$E_D^\varepsilon(A; B)_\rho \leq R_H^\varepsilon(A; B)_\rho \tag{13.1.41}$$

for every state ρ_{AB} and all $\varepsilon \in [0, 1]$.

PROOF: Consider a (d, ε) entanglement distillation protocol for ρ_{AB} with the corresponding LOCC channel $\mathcal{L}_{AB \to \hat{A}\hat{B}}$. Then, by definition, we have that

$$p_{\text{err}}(\mathcal{L}; \rho_{AB}) = 1 - \text{Tr}[\Phi_{\hat{A}\hat{B}} \mathcal{L}_{AB \to \hat{A}\hat{B}}(\rho_{AB})] \leq \varepsilon. \tag{13.1.42}$$

Letting $\omega_{\hat{A}\hat{B}} = \mathcal{L}_{AB \to \hat{A}\hat{B}}(\rho_{AB})$, we have that $\text{Tr}[\Phi_{\hat{A}\hat{B}} \omega_{\hat{A}\hat{B}}] \geq 1 - \varepsilon$. The output state $\omega_{\hat{A}\hat{B}}$ of the entanglement distillation protocol therefore satisfies the conditions of Proposition 13.6, which means that

$$\log_2 d \leq R_H^\varepsilon(\hat{A}; \hat{B})_\omega. \tag{13.1.43}$$

Now, it follows from Proposition 9.25 that $R_H^\varepsilon(\hat{A}; \hat{B})$ is an entanglement measure. Thus, it satisfies the data-processing inequality under LOCC channels, which means that $R_H^\varepsilon(\hat{A}; \hat{B})_\omega \leq R_H^\varepsilon(A; B)_\rho$. We thus have $\log_2 d \leq R_H^\varepsilon(A; B)_\rho$. Since this inequality holds for all $d \geq 1$ and every LOCC channel $\mathcal{L}_{AB \to \hat{A}\hat{B}}$, by definition of one-shot ε-distillable entanglement, we obtain $E_D^\varepsilon(A; B)_\rho \leq R_H^\varepsilon(A; B)_\rho$, as required. ■

The main step that allows us to conclude the bound in (13.1.40) in terms of the state ρ_{AB} alone is the fact that R_H^ε is an entanglement measure, meaning that it is monotone non-increasing under LOCC channels. In other words, the set of PPT′ operators is preserved under LOCC channels. This fact is not true for the set $\{\mathbb{1}_A \otimes \sigma_B : \sigma_B \in \text{D}(\mathcal{H})\}$ appearing in the optimization that defines the ε-hypothesis testing coherent information, meaning that the operator $\mathcal{L}_{AB\to\hat{A}\hat{B}}(\mathbb{1}_A \otimes \sigma_B)$ (where $\mathcal{L}_{AB\to\hat{A}\hat{B}}$ is an LOCC channel) is not in general of the form $\mathbb{1}_{\hat{B}} \otimes \tau_{\hat{B}}$ for some state $\tau_{\hat{B}}$. We therefore cannot use the data-processing inequality for the bound in (13.4) in order to reduce it to $\log_2 d \le I_H^\varepsilon(A\rangle B)_\rho$.

Combining Theorems 13.4 and 13.7 with Propositions 7.70 and 7.71 immediately leads to the following upper bounds.

Corollary 13.8

Let ρ_{AB} be a bipartite state, and let $\varepsilon \in [0, 1/2)$. For every (d, ε) entanglement distillation protocol $(d, \mathcal{L}_{AB\to\hat{A}\hat{B}})$, with $d_{\hat{A}} = d_{\hat{B}} = d$, we have that

$$\log_2 d \le \frac{1}{1-2\varepsilon}\left(\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)} + h_2(\varepsilon)\right), \tag{13.1.44}$$

where $I(A'\rangle B')_{\mathcal{L}(\rho)}$ is the coherent information of $\mathcal{L}_{AB\to A'B'}(\rho_{AB})$ (see (7.2.92)). For $\varepsilon \in [0, 1)$,

$$\log_2 d \le \widetilde{R}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{13.1.45}$$

where

$$\widetilde{R}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \tag{13.1.46}$$

is the sandwiched Rényi Rains relative entropy of ρ_{AB} (see (9.3.6)).

Proof: Combining the upper bound in (13.1.26) from Theorem 13.4 with the upper bound in (7.9.52) from Proposition 7.70, we obtain

$$\log_2 d \le \sup_{\mathcal{L}} I_H^\varepsilon(A'\rangle B')_{\mathcal{L}(\rho)} \tag{13.1.47}$$

$$\le \frac{1}{1-\varepsilon}\left(\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)} + h_2(\varepsilon)\right) + \frac{\varepsilon}{1-\varepsilon}\log_2 d. \tag{13.1.48}$$

Rearranging this and simplifying leads to

$$\log_2 d \le \frac{1}{1-2\varepsilon}\left(\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)} + h_2(\varepsilon)\right), \tag{13.1.49}$$

which is the inequality in (13.1.44). The inequality in (13.1.45) follows from Theorem 13.7 and (7.9.59) in Proposition 7.71. ■

Since the upper bounds in (13.1.44) and (13.1.45) hold for all (d,ε) entanglement distillation protocols, we conclude the following upper bounds on distillable entanglement:

$$E_D^{\varepsilon}(A;B)_\rho \leq \frac{1}{1-2\varepsilon}\left(\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)} + h_2(\varepsilon)\right), \tag{13.1.50}$$

$$E_D^{\varepsilon}(A;B)_\rho \leq \widetilde{R}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{13.1.51}$$

We finally turn to squashed entanglement and establish it as an upper bound on one-shot distillable entanglement:

Theorem 13.9 Squashed Entanglement Upper Bound on One-Shot Distillable Entanglement

Let ρ_{AB} be a bipartite state. For every (d,ε) entanglement distillation protocol for ρ_{AB}, with $\varepsilon \in [0,1)$, we have that

$$\log_2 d \leq \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\mathrm{sq}}(A;B)_\rho + g_2(\sqrt{\varepsilon})\right), \tag{13.1.52}$$

where $E_{\mathrm{sq}}(A;B)_\rho$ is the squashed entanglement of ρ_{AB} (see (9.1.162)) and $g_2(\delta) \coloneqq (\delta+1)\log_2(\delta+1) - \delta\log_2\delta$. Consequently, for the one-shot distillable entanglement, we have

$$E_D^{\varepsilon}(A;B)_\rho \leq \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\mathrm{sq}}(A;B)_\rho + g_2(\sqrt{\varepsilon})\right) \tag{13.1.53}$$

for every state ρ_{AB} and $\varepsilon \in [0,1)$.

Proof: Consider a (d,ε) entanglement distillation protocol for ρ_{AB} with the corresponding LOCC channel $\mathcal{L}_{AB\to\hat{A}\hat{B}}$. From the LOCC monotonicity of squashed entanglement (Theorem 9.33), we have that

$$E_{\mathrm{sq}}(\hat{A};\hat{B})_\omega \leq E_{\mathrm{sq}}(A;B)_\rho, \tag{13.1.54}$$

where $\omega_{\hat{A}\hat{B}} = \mathcal{L}_{AB\to\hat{A}\hat{B}}(\rho_{AB})$. Continuing, by definition, the following inequality holds

$$p_{\mathrm{err}}(\mathcal{L};\rho_{AB}) = 1 - \mathrm{Tr}[\Phi_{\hat{A}\hat{B}}\mathcal{L}_{AB\to\hat{A}\hat{B}}(\rho_{AB})] \leq \varepsilon. \tag{13.1.55}$$

It follows that $\mathrm{Tr}[\Phi_{\hat{A}\hat{B}}\omega_{\hat{A}\hat{B}}] \geq 1 - \varepsilon$, which is the same as

$$F(\Phi_{\hat{A}\hat{B}}, \omega_{\hat{A}\hat{B}}) \geq 1 - \varepsilon. \tag{13.1.56}$$

As a consequence of Proposition 9.38, we find that

$$E_{\mathrm{sq}}(\hat{A};\hat{B})_\omega \geq E_{\mathrm{sq}}(\hat{A};\hat{B})_\Phi - \left(\sqrt{\varepsilon}\log_2 \min\left\{\left|\hat{A}\right|, \left|\hat{B}\right|\right\} + g_2(\sqrt{\varepsilon})\right) \tag{13.1.57}$$
$$= \log_2 d - \left(\sqrt{\varepsilon}\log_2 d + g_2(\sqrt{\varepsilon})\right) \tag{13.1.58}$$
$$= (1 - \sqrt{\varepsilon})\log_2 d - g_2(\sqrt{\varepsilon}). \tag{13.1.59}$$

The first equality follows from Proposition 9.36. We can finally rearrange the established inequality $E_{\mathrm{sq}}(A;B)_\rho \geq (1 - \sqrt{\varepsilon})\log_2 d - g_2(\sqrt{\varepsilon})$ to be in the form stated in the theorem. ■

13.1.2 Lower Bound on the Number of Ebits via Decoupling

Having found upper bounds on one-shot distillable entanglement, we now focus on lower bounds. In order to find a lower bound on distillable entanglement, we have to find an explicit entanglement distillation protocol that works for an arbitrary bipartite state ρ_{AB} and an arbitrary error $\varepsilon \in (0,1)$. Recall that the goal of entanglement distillation is for two parties, Alice and Bob, to make use of LOCC to transform their shared bipartite state ρ_{AB} to the maximally entangled state $\Phi_{\hat{A}\hat{B}}$, for some $d_{\hat{A}} = d_{\hat{B}} = d \geq 1$. Now, the initial state ρ_{AB} has some purification $|\psi^\rho\rangle_{ABE}$, with the purifying system E in general correlated with A and B. However, because the maximally entangled state is pure, every purification of it must be of the form $\Phi_{\hat{A}\hat{B}} \otimes \phi_{E'}$, with the system E' in tensor product with systems $\hat{A}$ and $\hat{B}$. Since the goal of entanglement distillation is to distill a maximally entangled state and the maximally entangled state has this property, we can thus think of entanglement distillation as the task of *decoupling* A and B from their environment E; see Figure 13.2. Our lower bound on one-shot distillable entanglement tells us what dimension d of $\hat{A}$ and $\hat{B}$ is sufficient in order to achieve this decoupling up to error ε.

The lower bound on one-shot distillable entanglement that we determine in this section is expressed in terms of an information measure that is derived from a *smoothed* version of the max-relative entropy $D_{\max}$, which we briefly cover in Section 7.8.1. Recall from Definition 7.58 that the max-relative entropy of a state

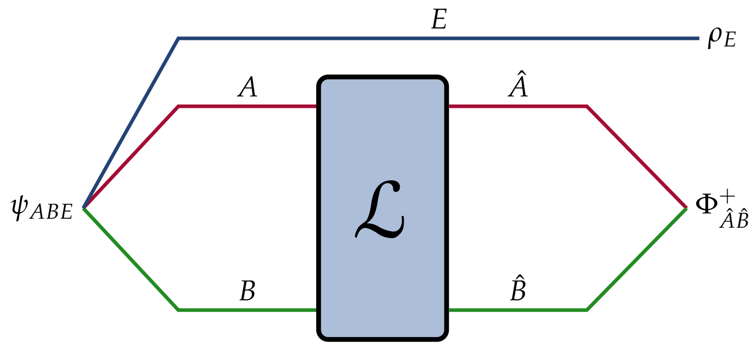


FIGURE 13.2: The task of entanglement distillation can be understood from the perspective of decoupling: given a bipartite state ρ_{AB} with purification ψ_{ABE}, the entanglement distillation protocol given by the LOCC channel $\mathcal{L}$ should result in the pure maximally entangled state $\Phi_{\hat{A}\hat{B}}$, which by definition is in tensor product with the environment, so that the joint state is $\Phi_{\hat{A}\hat{B}} \otimes \rho_E$, with $\rho_E = \text{Tr}_{AB}[\psi_{ABE}]$.

ρ and a positive semi-definite operator σ is defined as

$$D_{\max}(\rho\|\sigma) = \log_2 \left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\infty . \tag{13.1.60}$$

Using this, we define the conditional min-entropy as

$$H_{\min}(A|B)_\rho := -\inf_{\sigma_B} D_{\max}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{13.1.61}$$

where the optimization is with respect to every state σ_B.

From Definition 7.62, the smooth max-relative entropy is defined as

$$D^\varepsilon_{\max}(\rho\|\sigma) = \inf_{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} D_{\max}(\widetilde{\rho}\|\sigma), \tag{13.1.62}$$

where we recall from (7.8.41) that

$$\mathcal{B}^\varepsilon(\rho) = \{\widetilde{\rho} : P(\rho,\widetilde{\rho}) \le \varepsilon\}, \tag{13.1.63}$$

and the sine distance $P(\rho,\widetilde{\rho})$ is given by (see Definition 6.16)

$$P(\rho,\widetilde{\rho}) = \sqrt{1 - F(\rho,\widetilde{\rho})}. \tag{13.1.64}$$

Using the smooth max-relative entropy, we define the smooth conditional min-entropy of ρ_{AB} as

$$H^\varepsilon_{\min}(A|B)_\rho = -\inf_{\sigma_B} D^\varepsilon_{\max}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{13.1.65}$$

$$= \sup_{\widetilde{\rho} \in \mathcal{B}^{\varepsilon}(\rho)} H_{\min}(A|B)_{\widetilde{\rho}} \tag{13.1.66}$$

for all $\varepsilon \in (0,1)$, where the optimization in the first line is with respect to states σ_B.

We also need the *smooth conditional max-entropy* of ρ_{AB}, which is defined as

$$H_{\max}^{\varepsilon}(A|B)_{\rho} := \inf_{\widetilde{\rho} \in \mathcal{B}^{\varepsilon}(\rho)} H_{\max}(A|B)_{\rho} \tag{13.1.67}$$

for all $\varepsilon \in (0,1)$, where

$$H_{\max}(A|B)_{\rho} := \widetilde{H}_{\frac{1}{2}}(A|B)_{\rho} \tag{13.1.68}$$

$$= -\inf_{\sigma_B} \widetilde{D}_{\frac{1}{2}}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) \tag{13.1.69}$$

$$= \sup_{\sigma_B} \log_2 F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B). \tag{13.1.70}$$

The obtain the last equality, we made use of (7.5.8), and the optimization therein is with respect to states σ_B.

For every state ρ_{AB}, the conditional min- and max-entropies are related as follows:

$$H_{\max}(A|B)_{\rho} = -H_{\min}(A|E)_{\psi} \tag{13.1.71}$$

$$H_{\max}^{\varepsilon}(A|B)_{\rho} = -H_{\min}^{\varepsilon}(A|E)_{\psi}, \tag{13.1.72}$$

for all $\varepsilon \in (0,1)$, where ψ_{ABE} is a purification of ρ_{AB}.

Both the conditional min-entropy and the smooth conditional min-entropy can be formulated as semi-definite programs. The same is true for the conditional max-entropy and the smooth conditional max-entropy. Please consult the Bibliographic Notes in Section 13.5 for details.

Finally, we need the quantity

$$\widetilde{H}_2(A|B)_{\rho} := -\inf_{\sigma_B} \widetilde{D}_2(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) \tag{13.1.73}$$

$$= -\inf_{\sigma_B} \log_2 \operatorname{Tr}\left[\left(\sigma_B^{-\frac{1}{4}} \rho_{AB} \sigma_B^{-\frac{1}{4}}\right)^2\right], \tag{13.1.74}$$

which is known as the "conditional collision entropy" of ρ_{AB}, where the optimization is with respect to states σ_B. Due to monotonicity in α of the sandwiched Rényi relative entropy $\widetilde{D}_{\alpha}$ (see Proposition 7.31), we have that

$$H_{\min}(A|B)_{\rho} \leq \widetilde{H}_2(A|B)_{\rho} \tag{13.1.75}$$

for every bipartite state ρ_{AB}.

We are now ready to state a lower bound on one-shot distillable entanglement.

Theorem 13.10 Lower Bound on One-Shot Distillable Entanglement

Let ρ_{AB} be a quantum state. For all $\varepsilon \in (0,1]$ and $\eta \in [0, \sqrt{\varepsilon})$, there exists a (d, ε) one-way entanglement distillation protocol for ρ_{AB} with

$$\log_2 d = -H_{\max}^{\sqrt{\varepsilon}-\eta}(A|B)_\rho + 4\log_2 \eta. \tag{13.1.76}$$

Consequently, for the one-shot distillable entanglement of ρ_{AB}, we have

$$E_D^\varepsilon(A;B)_\rho \geq \sup_{\mathcal{L}} \left(-H_{\max}^{\sqrt{\varepsilon}-\eta}(A'|B')_{\mathcal{L}(\rho)} \right) + 4\log_2 \eta \tag{13.1.77}$$

for all $\varepsilon \in [0,1]$ and $\eta \in [0, \sqrt{\varepsilon})$, where the optimization is over every LOCC channel $\mathcal{L}_{AB\to A'B'}$.

In order to prove Theorem 13.10, we exhibit an entanglement distillation protocol $(d, \mathcal{L}_{AB\to\hat{A}\hat{B}})$, with $d_{\hat{A}} = d_{\hat{B}} = d = 2^{-H_{\max}^{\sqrt{\varepsilon}-\eta}(A|B)_\rho + 4\log_2 \eta}$, such that $p_{\text{err}}(\mathcal{L}; \rho_{AB}) \leq \varepsilon$ for all $\varepsilon \in (0,1]$. To this end, we construct a *one-way* LOCC channel $\mathcal{L}^{\to}_{AB\to\hat{A}\hat{B}}$ of the form

$$\mathcal{L}^{\to}_{AB\to\hat{A}\hat{B}} = \sum_{x\in\mathcal{X}} \mathcal{E}^x_{A\to\hat{A}} \otimes \mathcal{D}^x_{B\to\hat{B}}, \tag{13.1.78}$$

where $\mathcal{X}$ is a finite alphabet, $\{\mathcal{E}^x_{A\to\hat{A}}\}_{x\in\mathcal{X}}$ is a set of completely positive maps such that $\sum_{x\in\mathcal{X}} \mathcal{E}^x_{A\to\hat{A}}$ is trace preserving, and $\{\mathcal{D}^x_{B\to\hat{B}}\}_{x\in\mathcal{X}}$ is a set of channels. Recall from Section 4.6.2 that every one-way Alice-to-Bob LOCC channel can be written as

$$\mathcal{L}^{\to}_{AB\to\hat{A}\hat{B}} = \mathcal{D}_{BX_B\to\hat{B}} \circ \mathcal{C}_{X_A\to X_B} \circ \mathcal{E}_{A\to\hat{A}X_A}, \tag{13.1.79}$$

where $\mathcal{E}_{A\to\hat{A}X_A}$ is a local channel for Alice that corresponds to the quantum instrument given by the maps $\{\mathcal{E}^x_{A\to\hat{A}}\}_{x\in\mathcal{X}}$, i.e., (see also (4.4.53))

$$\mathcal{E}_{A\to\hat{A}X_A}(\rho_A) = \sum_{x\in\mathcal{X}} \mathcal{E}^x_{A\to\hat{A}}(\rho_A) \otimes |x\rangle\!\langle x|_{X_A}. \tag{13.1.80}$$

The map $\mathcal{C}_{X_A\to X_B}$ is a noiseless classical channel that transforms the classical register X_A, held by Alice, to the classical register X_B (which is simply a copy of

X_A), held by Bob. The final channel $\mathcal{D}_{BX_B\to\hat{B}}$ is a local channel for Bob defined as

$$\mathcal{D}_{BX_B\to\hat{B}}(\rho_B \otimes |x\rangle\langle x|_{X_B}) = \mathcal{D}^x_{B\to\hat{B}}(\rho_B) \tag{13.1.81}$$

for all $x \in \mathcal{X}$. In the proof below, we explicitly construct the CP maps $\{\mathcal{E}_{A\to\hat{A}}\}_{x\in\mathcal{X}}$ and the channels $\{\mathcal{D}^x_{B\to\hat{B}}\}_{x\in\mathcal{X}}$.

In addition to providing explicit forms for the channels $\mathcal{E}_{A\to\hat{A}X_A}$ and $\mathcal{D}_{BX_B\to\hat{B}}$ involved in the LOCC channel $\mathcal{L}^{\to}_{AB\to\hat{A}\hat{B}}$ in (13.1.78), we prove that $p_{\text{err}}(\mathcal{L};\rho_{AB}) \leq \varepsilon$. To do this, we make use of the following general decoupling result, which we explain and prove in Appendix 13.A.

Theorem 13.11

Given a subnormalized state ρ_{AE} (i.e., $\text{Tr}[\rho_{AE}] \leq 1$), and a completely positive map $\mathcal{N}_{A\to A'}$, the following bound holds

$$\int_{U_A} \left\| \mathcal{N}_{A\to A'}(U_A\rho_{AE}U_A^\dagger) - \Phi^{\mathcal{N}}_{A'} \otimes \rho_E \right\|_1 \, \mathrm{d}U_A \leq 2^{-\frac{1}{2}\widetilde{H}_2(A|E)_\rho - \frac{1}{2}\widetilde{H}_2(A|A')_{\Phi^{\mathcal{N}}}}, \tag{13.1.82}$$

where $\Phi^{\mathcal{N}}_{A'} := \text{Tr}_A[\Phi^{\mathcal{N}}_{AA'}]$, $\Phi^{\mathcal{N}}_{AA'}$ is given by $\mathcal{N}_{A\to A'}(\Phi_{AA})$, $\rho_E := \text{Tr}_A[\rho_{AE}]$, and the integral is over unitaries U_A acting on system A, taken with respect to the Haar measure.

Proof: See Appendix 13.A. ■

Remark: The integral in (13.1.82) with respect to the Haar measure should be thought of as a uniform average over the continuous set of all unitaries U_A acting on the system A. In other words, the integral is analogous to a uniform average over a discrete set of unitaries. In fact, for every dimension $d \geq 1$, there exists a set $\{U_x\}_{x\in\mathcal{X}}$ of unitaries, called a *unitary one-design*, such that

$$\int_U UXU^\dagger \, \mathrm{d}U = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} U_x X U_x^\dagger = \text{Tr}[X]\frac{\mathbb{1}}{d} \tag{13.1.83}$$

for every operator X. An example of a unitary one-design is the Heisenberg-Weyl operators $\{W_{z,x} : 0 \leq z, x \leq d-1\}$, which are defined in (3.2.47)–(3.2.49). Please consult the Bibliographic Notes in Section 13.5 for more information about integration over unitaries with respect to the Haar measure and about unitary designs. A simple argument for the right-most equality in (13.1.83) goes as follows. First, it follows for a unitary V that

$$V\left(\int_U UXU^\dagger \, \mathrm{d}U\right)V^\dagger = \int_U VUX(VU)^\dagger \, \mathrm{d}U = \int_U UXU^\dagger \, \mathrm{d}U, \tag{13.1.84}$$

where the final equality follows because the Haar measure is a unitarily invariant measure. So it follows that the operator $\int_U UXU^\dagger\, dU$ commutes with all unitaries. The only operator that does so is the identity operator, which implies that $\int_U UXU^\dagger\, dU \propto \mathbb{1}$. The normalization factor of $\text{Tr}[X]/d$ follows by taking a trace of the left-hand side, using its cyclicity, and the fact that dU is a probability measure.

Proof of Theorem 13.10

Fix $\eta \in (0, \sqrt{\varepsilon})$, and let ψ_{ABE} be a purification of ρ_{AB}, with $\rho_{AE} := \text{Tr}_B[\psi_{ABE}]$. Fix d such that (13.1.76) holds. Then, starting from the expression in (13.1.76) and using (13.1.72), we find that

$$\log_2 d = H_{\min}^{\sqrt{\varepsilon}-\eta}(A|E)_\rho + 4\log_2 \eta. \tag{13.1.85}$$

Now, pick a state $\widetilde{\rho}_{AE} \in \mathcal{B}^{\sqrt{\varepsilon}-\eta}(\rho_{AE})$ such that

$$H_{\min}(A|E)_{\widetilde{\rho}} = H_{\min}^{\sqrt{\varepsilon}-\eta}(A|E)_\rho. \tag{13.1.86}$$

Then, using (13.1.75), we find that

$$\log_2 d = H_{\min}(A|E)_{\widetilde{\rho}} + 4\log_2 \eta \tag{13.1.87}$$

$$\leq \widetilde{H}_2(A|E)_{\widetilde{\rho}} + 4\log_2 \eta \tag{13.1.88}$$

$$= \widetilde{H}_2(A|E)_{\widetilde{\rho}} - 2\log_2\left(\frac{1}{\eta^2}\right), \tag{13.1.89}$$

where

$$\widetilde{H}_2(A|E)_{\widetilde{\rho}} = -\inf_{\sigma_E} \log_2 \text{Tr}\left[\left(\sigma_E^{-\frac{1}{4}}\widetilde{\rho}_{AE}\sigma_E^{-\frac{1}{4}}\right)^2\right] \tag{13.1.90}$$

We now define a channel $\mathcal{E}_{A\to\hat{A}X_A}$ as follows:

$$\mathcal{E}_{A\to\hat{A}X_A}(\cdot) := \sum_{x\in\mathcal{X}} V^x_{A\to\hat{A}}\Pi^x_A(\cdot)\Pi^x_A V^{x\dagger}_{A\to\hat{A}} \otimes |x\rangle\langle x|_{X_A}, \tag{13.1.91}$$

where $d_{\hat{A}} = d$, $\mathcal{X}$ is a finite alphabet with[1] $|\mathcal{X}| = d_{X_A} = \frac{d_A}{d}$, $\{\Pi^x_A\}_{x\in\mathcal{X}}$ is a set of

[1]We assume that d divides d_A without loss of generality. If it is not the case, then we can repeat the whole analysis with the system A embedded in a larger Hilbert space that is divided by d. We would also need to start with a state $\widetilde{\rho}_{AE}$ such that (13.1.86) holds with the definition $H_{\min}^{\sqrt{\varepsilon}-\eta}(A|E)_\rho := -\inf_{\sigma_E} D_{\max}^{\sqrt{\varepsilon}-\eta}(\rho_{AE}\|\Pi_A \otimes \sigma_B)$, where Π_A is the projection onto the support of $\text{Tr}_E[\rho_{AE}]$. Then we would repeat the whole analysis with such a $\widetilde{\rho}_{AE}$. We do not go into further details here.

projectors such that $\sum_{x\in\mathcal{X}} \Pi_A^x = \mathbb{1}_A$, and $\{V^x_{A\to\hat{A}}\}_{x\in\mathcal{X}}$ is a set of isometries. So we have that

$$\mathcal{E}^x_{A\to\hat{A}}(\cdot) := V^x_{A\to\hat{A}}\Pi^x_A(\cdot)\Pi^x_A V^{x\dagger}_{A\to\hat{A}} \tag{13.1.92}$$

for all $x \in \mathcal{X}$. Each isometry $V^x_{A\to\hat{A}}$ takes the subspace of $\mathcal{H}_A$ onto which Π^x_A projects and embeds it into the fixed d-dimensional space $\mathcal{H}_{\hat{A}}$, i.e., $\text{im}(V^x_{A\to\hat{A}}) = \mathcal{H}_{\hat{A}}$ for all $x \in \mathcal{X}$. The projectors $\{\Pi^x_A\}_{x\in\mathcal{X}}$ correspond to a measurement of the input state, with $\Pi^x_A(\cdot)\Pi^x_A$ the (unnormalized) post-measurement state, and the isometries $\{V^x_{A\to\hat{A}}\}_{x\in\mathcal{X}}$ can be thought of as encodings of the initial system A into the system $\hat{A}$ on which one share of the desired maximally entangled state $\Phi_{\hat{A}\hat{B}}$ is to be generated. We have

$$\begin{aligned}
\mathcal{E}_{A\to\hat{A}X_A}(\rho_{AB}) &= \sum_{x\in\mathcal{X}} \mathcal{E}_{A\to\hat{A}}(\rho_{AB}) \otimes |x\rangle\langle x|_{X_A} && (13.1.93)\\
&= \sum_{x\in\mathcal{X}} \mathcal{V}^x_{A\to\hat{A}}(\Pi^x_A\rho_{AB}\Pi^x_A) \otimes |x\rangle\langle x|_{X_A} && (13.1.94)\\
&= \sum_{x\in\mathcal{X}} p(x)\omega^x_{\hat{A}B} \otimes |x\rangle\langle x|_{X_A}, && (13.1.95)
\end{aligned}$$

where

$$\begin{aligned}
p(x) &:= \text{Tr}[\Pi_A\rho_A], && (13.1.96)\\
\omega^x_{\hat{A}B} &:= \frac{1}{p(x)}\mathcal{V}^x_{A\to\hat{A}}(\Pi^x_A\rho_{AB}\Pi^x_A). && (13.1.97)
\end{aligned}$$

Now, by Theorem 13.11, the following inequality holds

$$\int_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \Phi^{\mathcal{E}}_{\hat{A}X_A} \otimes \widetilde{\rho}_E \right\|_1 \, \mathrm{d}U_A \\
\leq 2^{-\frac{1}{2}\widetilde{H}_2(A|E)_{\widetilde{\rho}} - \frac{1}{2}\widetilde{H}_2(A|\hat{A}X_A)_{\Phi^{\mathcal{E}}}}, \tag{13.1.98}$$

where $\Phi^{\mathcal{E}}_{A\hat{A}X_A}$ is the Choi state of $\mathcal{E}_{A\to\hat{A}X_A}$. Given that $\log_2 d \leq \widetilde{H}_2(A|E)_{\widetilde{\rho}} - 2\log_2\left(\frac{1}{\eta^2}\right)$, we obtain

$$2^{-\frac{1}{2}\widetilde{H}_2(A|E)_{\widetilde{\rho}}} \leq \frac{1}{\sqrt{d}}\eta^2. \tag{13.1.99}$$

We also have that

$$\widetilde{H}_2(A|\hat{A}X_A)_{\Phi^{\mathcal{E}}} = \sup_{\sigma_{\hat{A}X_A}} \left\{ -\log_2 \text{Tr}\left[\left(\sigma^{-\frac{1}{4}}_{\hat{A}X_A} \Phi^{\mathcal{E}}_{A\hat{A}X_A} \sigma^{-\frac{1}{4}}_{\hat{A}X_A} \right)^2 \right] \right\} \tag{13.1.100}$$

$$\geq -\log_2 d. \tag{13.1.101}$$

Indeed, in the optimization over $\sigma_{\hat{A}X_A}$, take

$$\sigma_{\hat{A}X_A} = \frac{1}{d_{X_A}} \sum_{x\in\mathcal{X}} \pi_{\hat{A}} \otimes |x\rangle\langle x|_{X_A} \tag{13.1.102}$$

$$= \pi_{\hat{A}} \otimes \pi_{X_A} \tag{13.1.103}$$

$$= \frac{1}{d_A} \mathbb{1}_{\hat{A}} \otimes \mathbb{1}_{X_A}. \tag{13.1.104}$$

With this choice of $\sigma_{\hat{A}X_A}$, we find that

$$\mathrm{Tr}\left[\left(\sigma_{\hat{A}X_A}^{-\frac{1}{4}} \Phi^{\mathcal{E}}_{A\hat{A}X_A} \sigma_{\hat{A}X_A}^{-\frac{1}{4}}\right)^2\right]$$

$$= d_A \, \mathrm{Tr}\left[\left(\Phi^{\mathcal{E}}_{A\hat{A}X_A}\right)^2\right] \tag{13.1.105}$$

$$= d_A \, \mathrm{Tr}\left[\left(\sum_{x\in\mathcal{X}} V^x_{A\to\hat{A}} \Pi^x_A \Phi_{AA} \Pi^x_A (V^x_{A\to\hat{A}})^\dagger \otimes |x\rangle\langle x|_{X_A}\right)^2\right] \tag{13.1.106}$$

$$= d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left[\left(V^x_{A\to\hat{A}} \Pi^x_A \Phi_{AA} \Pi^x_A (V^x_{A\to\hat{A}})^\dagger\right)^2\right] \tag{13.1.107}$$

$$= d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left[\left(V^x_{A\to\hat{A}} \Pi^x_A \frac{1}{d_A} \sum_{i,j=0}^{d_A-1} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_A \Pi^x_A (V^x_{A\to\hat{A}})^\dagger\right)^2\right] \tag{13.1.108}$$

$$= d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left[\left(\frac{1}{d_A} \sum_{i,j=0}^{d-1} |i\rangle\langle j|_{\hat{A}} \otimes |i\rangle\langle j|_{\hat{A}}\right)^2\right] \tag{13.1.109}$$

$$= d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left[\left(\frac{d}{d_A} \frac{1}{d} \sum_{i,j=0}^{d-1} |i\rangle\langle j|_{\hat{A}} \otimes |i\rangle\langle j|_{\hat{A}}\right)^2\right] \tag{13.1.110}$$

$$= \frac{d^2}{d_A} \sum_{x\in\mathcal{X}} \mathrm{Tr}\left[\left(\frac{1}{d} \sum_{i,j=0}^{d-1} |i\rangle\langle j|_{\hat{A}} \otimes |i\rangle\langle j|_{\hat{A}}\right)^2\right] \tag{13.1.111}$$

$$= \frac{d^2}{d_A} |\mathcal{X}| \tag{13.1.112}$$

$$= d, \tag{13.1.113}$$

where we recall that $d_{X_A} = |\mathcal{X}| = \frac{d_A}{d}$. We thus have

$$2^{-\frac{1}{2}\widetilde{H}_2(A|\hat{A}X_A)_{\Phi^{\mathcal{E}}}} \leq \sqrt{d}, \tag{13.1.114}$$

which means that

$$\int_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \Phi^{\mathcal{E}}_{\hat{A}X_A} \otimes \widetilde{\rho}_E \right\|_1 \, \mathrm{d}U_A \leq \eta^2. \tag{13.1.115}$$

Note that

$$\begin{aligned}
\Phi^{\mathcal{E}}_{\hat{A}X_A} &= \frac{1}{d_A} \sum_{x\in\mathcal{X}} V^x_{A\to\hat{A}} \Pi^x_A \mathbb{1}_A \Pi^x_A V^{x\dagger}_{A\to\hat{A}} \otimes |x\rangle\langle x|_{X_A} && (13.1.116)\\
&= \frac{1}{d_A} \sum_{x\in\mathcal{X}} V^x_{A\to\hat{A}} \Pi^x_A V^{x\dagger}_{A\to\hat{A}} \otimes |x\rangle\langle x|_{X_A} && (13.1.117)\\
&= \frac{1}{d_A} \mathbb{1}_{\hat{A}} \otimes \mathbb{1}_{X_A} && (13.1.118)\\
&= \pi_{\hat{A}} \otimes \pi_{X_A}, && (13.1.119)
\end{aligned}$$

where the last equality follows because $d_{X_A} = \frac{d_A}{d}$. So we have

$$\int_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E \right\|_1 \, \mathrm{d}U_A \leq \eta^2. \tag{13.1.120}$$

Now, since the average over a set of elements is never less than the minimum over the same set, we have that

$$\begin{aligned}
&\int_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E \right\|_1 \, \mathrm{d}U_A \\
&\qquad \geq \min_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E \right\|_1 \qquad (13.1.121)
\end{aligned}$$

This implies that there exists a unitary U_A (in particular, one that achieves the minimum on the right-hand side of the above inequality) such that

$$\eta^2 \geq \int_{U_A} \left\| \mathcal{E}_{A\to\hat{A}X_A}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E \right\|_1 \, \mathrm{d}U_A$$

$$\geq \left\| \mathcal{E}_{A \to \hat{A} X_A}(U_A \widetilde{\rho}_{AE} U_A^\dagger) - \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E \right\|_1 \quad (13.1.122)$$

Now, let

$$\widetilde{\omega}_{\hat{A} X_A E} = \mathcal{E}_{A \to \hat{A} X_A}(U_A \widetilde{\rho}_{AE} U_A^\dagger), \quad \widetilde{\tau}_{\hat{A} X_A E} = \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \widetilde{\rho}_E, \quad (13.1.123)$$

$$\omega_{\hat{A} X_A E} = \mathcal{E}_{A \to \hat{A} X_A}(U_A \rho_{AE} U_A^\dagger), \quad \tau_{\hat{A} X_A E} = \pi_{\hat{A}} \otimes \pi_{X_A} \otimes \rho_E. \quad (13.1.124)$$

Then, by the Fuchs–van de Graaf inequality (see (6.2.88)), and by the definition of the sine distance (see Definition 6.16), we have that

$$\eta^2 \geq \left\| \widetilde{\omega}_{\hat{A} X_A E} - \widetilde{\tau}_{\hat{A} X_A E} \right\|_1 \quad (13.1.125)$$

$$\geq 2 - 2\sqrt{F(\widetilde{\omega}_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E})} \quad (13.1.126)$$

$$\geq 2 - 2\sqrt{1 - P(\widetilde{\omega}_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E})^2}, \quad (13.1.127)$$

which implies that

$$P(\widetilde{\omega}_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E}) \leq \sqrt{1 - \left(1 - \frac{\eta^2}{2}\right)^2} \quad (13.1.128)$$

$$= \eta \sqrt{1 - \frac{\eta^2}{4}} \quad (13.1.129)$$

$$\leq \eta. \quad (13.1.130)$$

Then, by the triangle inequality for sine distance (Lemma 6.17), we have

$$P(\omega_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E}) \leq P(\omega_{\hat{A} X_A E}, \widetilde{\omega}_{\hat{A} X_A E}) + P(\widetilde{\omega}_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E}) \quad (13.1.131)$$

$$\leq P(\rho_{AE}, \widetilde{\rho}_{AE}) + \eta \quad (13.1.132)$$

$$\leq \sqrt{\varepsilon} - \eta + \eta \quad (13.1.133)$$

$$= \sqrt{\varepsilon}, \quad (13.1.134)$$

where the second inequality follows from the data-processing inequality for the sine distance, unitary invariance of the sine distance, and the inequality in (13.1.130). To obtain the last inequality, we used the definition of the state $\widetilde{\rho}_{AE}$ as one that is $(\sqrt{\varepsilon} - \eta)$-close to ρ_{AE} in sine distance. We can write the inequality in (13.1.134) in terms of fidelity as

$$F(\omega_{\hat{A} X_A E}, \widetilde{\tau}_{\hat{A} X_A E}) \geq 1 - \varepsilon. \quad (13.1.135)$$

Note that both $\omega_{\hat{A}X_AE}$ and $\widetilde{\tau}_{\hat{A}X_AE}$ are classical-quantum states because X_A is a classical register. In particular,

$$\omega_{\hat{A}X_AE} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_A} \otimes \omega^x_{\hat{A}E}, \tag{13.1.136}$$

$$p(x) = \mathrm{Tr}[\Pi^x_A U_A \rho_A U_A^\dagger], \tag{13.1.137}$$

$$\omega^x_{\hat{A}E} = \frac{1}{p(x)}\mathcal{V}^x_{A\to\hat{A}}(\Pi^x_A U_A \rho_{AE} U_A^\dagger \Pi^x_A). \tag{13.1.138}$$

Also,

$$\widetilde{\tau}_{\hat{A}X_AE} = \pi_{\hat{A}} \otimes \frac{1}{d_{X_A}} \sum_{x\in\mathcal{X}} |x\rangle\langle x|_{X_A} \otimes \widetilde{\rho}_E. \tag{13.1.139}$$

Then, using the direct-sum property of the root fidelity (see (6.2.58)), we obtain

$$F(\omega_{\hat{A}X_AE}, \widetilde{\tau}_{\hat{A}X_AE}) = \left(\sqrt{F}(\omega_{\hat{A}X_AE}, \widetilde{\tau}_{\hat{A}X_AE})\right)^2 \tag{13.1.140}$$

$$= \left(\sum_{x\in\mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\omega^x_{\hat{A}E}, \pi_{\hat{A}} \otimes \widetilde{\rho}_E)\right)^2. \tag{13.1.141}$$

Now, let

$$\psi^x_{\hat{A}BE} := \frac{1}{p(x)} \mathcal{V}^x_{A\to\hat{A}}(\Pi^x_A U_A \psi_{ABE} U_A^\dagger \Pi^x_A) \tag{13.1.142}$$

be a purification of $\omega^x_{\hat{A}E}$ for all $x \in \mathcal{X}$, and let $\Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}$ be a purification of $\pi_{\hat{A}} \otimes \widetilde{\rho}_E$. Then, by Uhlmann's theorem (Theorem 6.8), for every $x \in \mathcal{X}$ there exists an isometric channel $\mathcal{W}^x_{B\to\hat{B}B'}$ such that

$$\sqrt{F}(\omega^x_{\hat{A}E}, \pi_{\hat{A}} \otimes \widetilde{\rho}_E) = \sqrt{F}(\mathcal{W}^x_{B\to\hat{B}B'}(\psi^x_{\hat{A}BE}), \Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}) \tag{13.1.143}$$

for all $x \in \mathcal{X}$. Using the set $\{\mathcal{W}^x_{B\to\hat{B}B'}\}_{x\in\mathcal{X}}$, we define the quantum channels $\{\mathcal{D}^x_{B\to\hat{B}}\}_{x\in\mathcal{X}}$ as follows:

$$\mathcal{D}^x_{B\to\hat{B}} := \mathrm{Tr}_{B'} \circ \mathcal{W}^x_{B\to\hat{B}B'}. \tag{13.1.144}$$

By the data-processing inequality for fidelity (see Theorem 6.9) under the partial trace channel $\mathrm{Tr}_{EB'}$, we obtain

$$\sqrt{F}(\omega^x_{\hat{A}E}, \pi_{\hat{A}} \otimes \widetilde{\rho}_E)$$

$$= \sqrt{F}(\mathcal{W}^x_{B\to\hat{B}B'}(\psi^x_{\hat{A}BE}), \Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}) \tag{13.1.145}$$

$$\leq \sqrt{F}(\mathrm{Tr}_{EB'}[\mathcal{W}^x_{B\to\hat{B}B'}(\psi^x_{\hat{A}BE})], \mathrm{Tr}_{EB'}[\Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}]) \tag{13.1.146}$$

$$= \sqrt{F}(\mathcal{D}^x_{B\to\hat{B}}(\omega^x_{\hat{A}B}), \Phi_{\hat{A}\hat{B}}), \tag{13.1.147}$$

for all $x \in \mathcal{X}$, where we recall the definition of $\omega^x_{\hat{A}B}$ from (13.1.97). Since the inequality in (13.1.147) holds for all $x \in \mathcal{X}$, we have that

$$\left(\sum_{x\in\mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\omega^x_{\hat{A}E}, \pi_{\hat{A}} \otimes \widetilde{\rho}_E)\right)^2 \leq \left(\sum_{x\in\mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\mathcal{D}^x_{B\to\hat{B}}(\omega^x_{\hat{A}B}), \Phi_{\hat{A}\hat{B}})\right)^2. \tag{13.1.148}$$

Now, the final state of Alice and Bob after executing the LOCC channel defined by (13.1.79), with $\mathcal{E}_{A\to\hat{A}X_A}$ defined by (13.1.91) and $\mathcal{D}_{BX_B\to\hat{B}}$ defined by (13.1.81) and (13.1.144), is

$$\omega_{\hat{A}\hat{B}} = (\mathcal{D}_{X_BB\to\hat{B}} \circ \mathcal{C}_{X_A\to X_B} \circ \mathcal{E}_{A\to\hat{A}X_A})(\rho_{AB}) \tag{13.1.149}$$

$$= \sum_{x\in\mathcal{X}} \mathcal{D}_{X_BB\to\hat{B}}(\mathcal{E}^x_{A\to\hat{A}}(\rho_{AB}) \otimes |x\rangle\langle x|_{X_B}) \tag{13.1.150}$$

$$= \sum_{x\in\mathcal{X}} (\mathcal{E}^x_{A\to\hat{A}} \otimes \mathcal{D}^x_{B\to\hat{B}})(\rho_{AB}) \tag{13.1.151}$$

$$= \mathrm{Tr}_{X_B}\left[\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_B} \otimes \mathcal{D}^x_{B\to\hat{B}}(\omega^x_{\hat{A}B})\right], \tag{13.1.152}$$

where in the third equality we recognize the required form in (13.1.78) for a one-way Alice-to-Bob LOCC channel, and in the last inequality we made use of (13.1.97). Using the form of $\omega_{\hat{A}\hat{B}}$ in the last equality, along with all of the developments above, we finally obtain

$$F(\omega_{\hat{A}\hat{B}}, \Phi_{\hat{A}\hat{B}})$$

$$= F\left(\mathrm{Tr}_{X_B}\left[\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_B} \otimes \mathcal{D}^x_{B\to\hat{B}}(\omega^x_{\hat{A}B})\right], \mathrm{Tr}_{X_B}[\pi_{X_B} \otimes \Phi_{\hat{A}\hat{B}}]\right) \tag{13.1.153}$$

$$\geq F\left(\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_{X_B} \otimes \mathcal{D}^x_{B\to\hat{B}}(\omega^x_{\hat{A}B}), \frac{1}{d_{X_A}} \sum_{x\in\mathcal{X}} |x\rangle\langle x|_{X_A} \otimes \Phi_{\hat{A}\hat{B}}\right) \tag{13.1.154}$$

$$= \left(\sum_{x \in \mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\mathcal{D}^x_{B \to \hat{B}}(\omega^x_{\hat{A}B}), \Phi_{\hat{A}\hat{B}}) \right)^2 \tag{13.1.155}$$

$$= \left(\sum_{x \in \mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}((\mathrm{Tr}_{EB'} \circ \mathcal{W}^x_{B \to \hat{B}B'})(\psi^x_{\hat{A}BE}), \mathrm{Tr}_{EB'}[\Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}]) \right)^2 \tag{13.1.156}$$

$$\geq \left(\sum_{x \in \mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\mathcal{W}^x_{B \to \hat{B}B'}(\psi^x_{\hat{A}BE}), \Phi_{\hat{A}\hat{B}} \otimes \widetilde{\phi}_{EB'}) \right)^2 \tag{13.1.157}$$

$$= \left(\sum_{x \in \mathcal{X}} \sqrt{\frac{p(x)}{d_{X_A}}} \sqrt{F}(\omega^x_{\hat{A}E}, \pi_{\hat{A}} \otimes \widetilde{\rho}_E) \right)^2 \tag{13.1.158}$$

$$\geq 1 - \varepsilon. \tag{13.1.159}$$

Therefore,

$$p_{\mathrm{err}}(\mathcal{L}; \rho_{AB}) = 1 - F(\omega_{\hat{A}\hat{B}}, \Phi_{\hat{A}\hat{B}}) \leq \varepsilon. \tag{13.1.160}$$

To summarize, we have shown that, given a state ρ_{AB} and $\varepsilon \in (0,1)$, there exists a (d, ε) one-way entanglement distillation protocol $\mathcal{L}_{AB \to \hat{A}\hat{B}}$ if the dimension $d = d_{\hat{A}} = d_{\hat{B}}$ satisfies $\log_2 d = -H^{\sqrt{\varepsilon}-\eta}_{\max}(A|B)_\rho + 4 \log_2 \eta$, where $\eta \in [0, \sqrt{\varepsilon})$. Although we explicitly constructed the encoding channels $\{\mathcal{E}^x_{A \to \hat{A}}\}_{x \in \mathcal{X}}$ on Alice's side, on Bob's side we relied on Uhlmann's theorem to guarantee the existence of a set of decoding channels $\{\mathcal{D}^x_{B \to \hat{B}}\}_{x \in \mathcal{X}}$ such that the overall LOCC channel $\mathcal{L}_{AB \to \hat{A}\hat{B}}$ satisfies $p_{\mathrm{err}}(\mathcal{L}; \rho_{AB}) \leq \varepsilon$. ■

Combining Theorem 13.10 with (7.8.83), and using (13.1.72), leads to the following lower bound on the one-shot distillable entanglement:

Corollary 13.12

Let ρ_{AB} be a bipartite quantum state with purification ψ_{ABE}. For all $\varepsilon \in (0,1)$, $\eta \in [0, \sqrt{\varepsilon})$, and $\alpha > 1$, there exists a (d, ε) one-way entanglement distillation protocol for ρ_{AB} satisfying

$$\log_2 d \geq \widetilde{H}_\alpha(A|E)_\psi - \frac{1}{\alpha - 1} \log_2 \left(\frac{1}{(\sqrt{\varepsilon} - \eta)^2} \right)$$

$$- \log_2\left(\frac{1}{1-(\sqrt{\varepsilon}-\eta)^2}\right) + 4\log_2 \eta. \quad (13.1.161)$$

PROOF: The inequality follows from taking the results of Theorem 13.10, using (13.1.72), and applying the inequality in (7.8.83). ■

Since the inequality in (13.1.161) holds for all (d, ε) entanglement distillation protocols, we obtain the following bound for all $\varepsilon \in (0, 1)$, $\eta \in [0, \sqrt{\varepsilon})$, and $\alpha > 1$:

$$E_D^{\varepsilon}(A; B)_\rho \geq \sup_{\mathcal{L}} \widetilde{H}_\alpha(A'|E')_\phi - \frac{1}{\alpha - 1}\log_2\left(\frac{1}{(\sqrt{\varepsilon}-\eta)^2}\right) - \log_2\left(\frac{1}{1-(\sqrt{\varepsilon}-\eta)^2}\right) + 4\log_2 \eta, \quad (13.1.162)$$

where the optimization is with respect to every LOCC channel $\mathcal{L}_{AB\to A'B'}$, such that $\phi_{A'B'E'}$ is a purification of $\mathcal{L}_{AB\to A'B'}(\rho_{AB})$. This comes about by first applying the LOCC channel $\mathcal{L}_{AB\to A'B'}$ to ρ_{AB} for free, applying Corollary 13.12 to the state $\mathcal{L}_{AB\to A'B'}(\rho_{AB})$, and finally optimizing over every LOCC channel $\mathcal{L}_{AB\to A'B'}$.

13.2 Distillable Entanglement of a Quantum State

Having found upper and lower bounds on the one-shot distillable entanglement $E_D^{\varepsilon}(A; B)_\rho$ of a bipartite quantum state ρ_{AB}, let us now move on to the asymptotic setting. In this setting, we allow Alice and Bob to make use of an arbitrarily large number n of copies of the state ρ_{AB} in order to obtain a maximally entangled state. An *entanglement distillation protocol for n copies of* ρ_{AB} is defined by the triple $(n, d, \mathcal{L}_{A^nB^n\to\hat{A}\hat{B}})$, consisting of the number n of copies of ρ_{AB}, an integer $d \geq 1$, and an LOCC channel $\mathcal{L}_{A^nB^n\to\hat{A}\hat{B}}$ with $d_{\hat{A}} = d_{\hat{B}} = d$. Observe that an entanglement distillation protocol for n copies of ρ_{AB} is equivalent to a (one-shot) entanglement distillation protocol for the state $\rho_{AB}^{\otimes n}$. All of the results of Section 13.1 thus carry over to the asymptotic setting simply by replacing ρ_{AB} with $\rho_{AB}^{\otimes n}$. In particular, the error probability for an entanglement distillation protocol for ρ_{AB} defined by $(n, d, \mathcal{L}_{A^nB^n\to\hat{A}\hat{B}})$ is equal to

$$p_{\text{err}}(\mathcal{L}; \rho_{AB}^{\otimes n}) = 1 - \langle\Phi|_{\hat{A}\hat{B}}\mathcal{L}_{A^nB^n\to\hat{A}\hat{B}}(\rho_{AB}^{\otimes n})|\Phi\rangle_{\hat{A}\hat{B}}. \quad (13.2.1)$$

Definition 13.13 (n, d, ε) Entanglement Distillation Protocol

An entanglement distillation protocol $(n, d, \mathcal{L}_{A^nB^n\to\hat{A}\hat{B}})$ for n copies of ρ_{AB}, with $d_{\hat{A}} = d_{\hat{B}} = d$, is called an (n, d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}(\mathcal{L}; \rho_{AB}^{\otimes n}) \leq \varepsilon$.

Based on the discussion above, we note that an (n, d, ε) entanglement distillation protocol for ρ_{AB} is a (d, ε) entanglement distillation protocol for $\rho_{AB}^{\otimes n}$.

The *rate* $R(n, d)$ of an (n, d, ε) entanglement distillation protocol for n copies of a given state is

$$R(n, d) := \frac{\log_2 d}{n}, \tag{13.2.2}$$

which can be thought of as the number of ε-approximate ebits contained in the final state of the protocol, per copy of the given initial state. Given a state ρ_{AB} and $\varepsilon \in [0, 1]$, the maximum rate of entanglement distillation among all (n, d, ε) entanglement distillation protocols for ρ_{AB} is given by

$$E_D^{n,\varepsilon}(\rho_{AB}) \equiv E_D^{n,\varepsilon}(A; B)_\rho := \frac{1}{n} E_D^{\varepsilon}(\rho_{AB}^{\otimes n}) \tag{13.2.3}$$

$$= \sup_{(d,\mathcal{L})} \left\{ \frac{\log_2 d}{n} : p_{\text{err}}(\mathcal{L}; \rho_{AB}^{\otimes n}) \leq \varepsilon \right\}, \tag{13.2.4}$$

where the optimization is with respect to all $d \geq 1$ and every LOCC channel $\mathcal{L}_{A^nB^n\to\hat{A}\hat{B}}$ with $d_{\hat{A}} = d_{\hat{B}} = d$.

Definition 13.14 Achievable Rate for Entanglement Distillation

Given a bipartite quantum state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for entanglement distillation for* ρ_{AB} if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB}.

As we prove in Appendix A,

$$R \text{ achievable rate } \iff \lim_{n\to\infty} \varepsilon_D(2^{n(R-\delta)}; \rho_{AB}^{\otimes n}) = 0 \quad \forall\, \delta > 0. \tag{13.2.5}$$

In other words, a rate R is achievable if the optimal error probability for a sequence of protocols with rate $R - \delta$, $\delta > 0$, vanishes as the number n of copies of ρ_{AB} increases.

Definition 13.15 Distillable Entanglement of a Quantum State

The *distillable entanglement of a bipartite state* ρ_{AB}, denoted by $E_D(A;B)_\rho$, is defined to be the supremum of all achievable rates for entanglement distillation for ρ_{AB}, i.e.,

$$E_D(A;B)_\rho \coloneqq \sup\{R : R \text{ is an achievable rate for } \rho_{AB}\}. \quad (13.2.6)$$

The distillable entanglement can also be written as

$$E_D(A;B)_\rho = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} E_D^\varepsilon(\rho_{AB}^{\otimes n}). \quad (13.2.7)$$

See Appendix A for a proof.

Definition 13.16 Weak Converse Rate for Entanglement Distillation

Given a bipartite state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called a *weak converse rate for entanglement distillation for* ρ_{AB} if every $R' > R$ is not an achievable rate for ρ_{AB}.

As we show in Appendix A,

$$R \text{ weak converse rate} \iff \lim_{n\to\infty} \varepsilon_D(2^{n(R-\delta)}; \rho_{AB}^{\otimes n}) > 0 \quad \forall\, \delta > 0. \quad (13.2.8)$$

Definition 13.17 Strong Converse Rate for Entanglement Distillation

Given a bipartite state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for entanglement distillation for* ρ_{AB} if for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB}.

We show in Appendix A that

$$R \text{ strong converse rate} \iff \lim_{n\to\infty} \varepsilon_D(2^{n(R+\delta)}; \rho_{AB}) = 1 \quad \forall\, \delta > 0. \quad (13.2.9)$$

Definition 13.18 Strong Converse Distillable Entanglement of a Quantum State

The *strong converse distillable entanglement* of a bipartite state ρ_{AB}, denoted by $\widetilde{E}_D(A;B)_\rho$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{E}_D(A;B)_\rho := \inf\{R : R \text{ is a strong converse rate for } \rho_{AB}\}. \tag{13.2.10}$$

Note that

$$E_D(A;B)_\rho \leq \widetilde{E}_D(A;B)_\rho \tag{13.2.11}$$

for all bipartite states ρ_{AB}. We can also write the strong converse distillable entanglement as

$$\widetilde{E}_D(A;B)_\rho = \sup_{\varepsilon\in[0,1)} \limsup_{n\to\infty} \frac{1}{n} E_D^\varepsilon(\rho_{AB}^{\otimes n}). \tag{13.2.12}$$

See Appendix A for a proof.

We are now ready to present a general expression for the distillable entanglement of a bipartite quantum state, as well as two upper bounds on it.

Theorem 13.19 Distillable Entanglement of a Bipartite State

The distillable entanglement of a bipartite state ρ_{AB} is given by

$$E_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{(n)}} I(A'\rangle B')_{\mathcal{L}^{(n)}(\rho^{\otimes n})}, \tag{13.2.13}$$

where the optimization is with respect to (two-way) LOCC channels $\mathcal{L}^{(n)}_{A^nB^n\to A'B'}$. Furthermore, the Rains relative entropy $R(A;B)_\rho$ from (9.3.4) is a strong converse rate for distillable entanglement, in the sense that

$$\widetilde{E}_D(A;B)_\rho \leq R(A;B)_\rho, \tag{13.2.14}$$

and the squashed entanglement from (9.4.1) is a weak converse rate, in the sense that

$$E_D(A;B)_\rho \leq E_{\text{sq}}(A;B)_\rho. \tag{13.2.15}$$

If we define

$$D^{\leftrightarrow}(\rho_{AB}) \equiv D^{\leftrightarrow}(A;B)_{\rho} := \sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)}, \tag{13.2.16}$$

then we can write (13.2.13) as

$$E_D(A;B)_{\rho} = \lim_{n\to\infty} \frac{1}{n} D^{\leftrightarrow}(\rho_{AB}^{\otimes n}) =: D^{\leftrightarrow}_{\text{reg}}(\rho_{AB}), \tag{13.2.17}$$

so that the distillable entanglement can be viewed as the regularized version of $D^{\leftrightarrow}$, and it is reminiscent of the regularized Holevo information that gives the classical capacity of a quantum channel.

Let us make the following observations about Theorem 13.19.

- The coherent information of a bipartite state ρ_{AB} is an achievable rate for entanglement distillation, i.e.,

  $$E_D(A;B)_{\rho} \geq I(A\rangle B)_{\rho} = H(B)_{\rho} - H(AB)_{\rho}. \tag{13.2.18}$$

  This follows immediately from (13.2.13) by dropping the optimization over two-way LOCC channels and due to the fact that the coherent information is additive for product states, meaning that $I(A^n\rangle B^n)_{\rho^{\otimes n}} = nI(A\rangle B)_{\rho}$. As we show in Section 13.2.1 below, the strategy to attain the coherent information rate is essentially the one-way entanglement distillation protocol considered in Section 13.1.2 for the one-shot lower bound, and it is sometimes called the "hashing protocol." For this reason, the inequality in (13.2.18) is known as the *hashing bound* (please consult the Bibliographic Notes in Section 13.5 for pointers to the research literature).

- In order to obtain a higher entanglement distillation rate than $I(A\rangle B)_{\rho}$, one strategy is to use $n \geq 2$ copies of ρ_{AB} along with a two-way LOCC channel $\mathcal{L}_{A^nB^n\to A'B'}$ in order to obtain a state $\omega_{A'B'} := \mathcal{L}_{A^nB^n\to A'B'}(\rho_{AB}^{\otimes n})$ whose coherent information is potentially larger than that of ρ_{AB}. Then, we can apply the hashing protocol to the state $\omega_{A'B'}$. The overall rate of this strategy (the two-way LOCC channel followed by the hashing protocl) is then $\frac{1}{n}I(A'\rangle B')_{\omega}$, and Theorem 13.19 tells us that such a strategy is optimal in the large n limit. With increasingly more copies of ρ_{AB} to start with, it might be possible to obtain a better rate, which is why we need to regularize in general.

As with the proof of the entanglement-assisted classical capacity and classical capacity theorems in Chapters 11 and 12, respectively, we prove Theorem 13.19 in two steps:

1. *Achievability*: We show that the right-hand side of (13.2.13) is an achievable rate for entanglement distillation for ρ_{AB}. Doing so involves exhibiting an explicit entanglement distillation protocol. The protocol we use is based on the one we used in Section 13.1.2 to obtain a lower bound on the one-shot distillable entanglement.

   The achievability part of the proof establishes that

$$E_D(A;B)_\rho \geq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})}. \tag{13.2.19}$$

2. *Weak converse*: We show that the right-hand side of (13.2.13) is a weak converse rate for entanglement distillation for ρ_{AB}, from which it follows that $E_D(A;B)_\rho \leq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})}$. In order to show this, we use the one-shot upper bounds from Section 13.1.1 to prove that every achievable rate R satisfies $R \leq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})}$.

We go through the achievability part of the proof of Theorem 13.19 in Section 13.2.1. We then proceed with the weak converse part in Section 13.2.2.

The expression in (13.2.13) for the distillable entanglement involves both a limit over an unbounded number of copies of the state ρ_{AB}, as well as an optimization over all two-way LOCC channels. Computing the distillable entanglement is therefore intractable in general. After establishing a proof of (13.2.13), we proceed to establish upper bounds on distillable entanglement that depend only on the given state ρ_{AB}. Specifically, in Section 13.2.3, we use the one-shot results in Section 13.1.1 to show that the Rains relative entropy is a strong converse rate for entanglement distillation. We also show that the squashed entanglement is a weak converse rate for entanglement distillation.

13.2.0.1 Bound Entanglement

The inequality in (13.2.14) implies that $E_D(A;B)_\rho \leq 0$ for every PPT state ρ_{AB}, because, by definition, the Rains relative entropy vanishes for all PPT states. On the other hand, we always have $E_D(A;B)_\rho \geq 0$ for every state ρ_{AB}. Therefore,

$$E_D(A;B)_\rho = 0 \text{ for all PPT states.} \tag{13.2.20}$$

Recall from the discussion after Lemma 13.5 (see also Section 3.2.9) that there exist PPT entangled states in higher dimensional bipartite systems because $\text{SEP}(A:B) \neq$

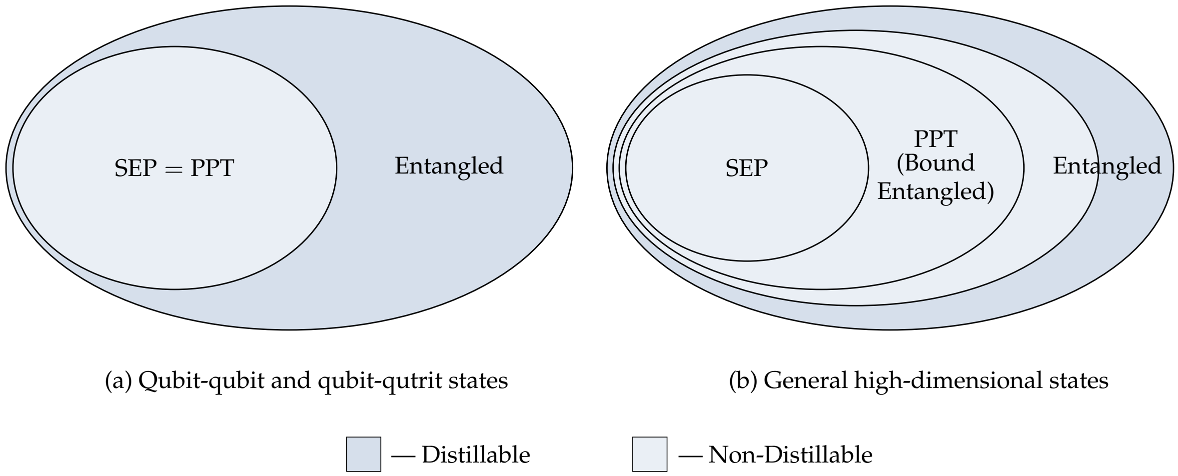


FIGURE 13.3: The set of all bipartite states can be split into distillable and non-distillable sets. (a) For qubit-qubit and qubit-qutrit states entanglement and distillability are in one-to-one correspondence because all PPT states are separable. (b) In higher dimensions, there are entangled states belonging to the set PPT, which we call bound entangled states. Distillability and entanglement are thus not synonymous in general for high-dimensional quantum systems.

$\mathrm{PPT}(A\!:\!B)$ except for when A and B are both qubits or when one is a qubit and the other is a qutrit. All of these entangled states have zero distillable entanglement, and thus we refer to them as *bound entangled*. Remarkably, therefore, except for qubit-qubit and qubit-qutrit states, prior entanglement is only necessary, but not sufficient, for distilling pure maximally entangled states. Please consult the Bibliographic Notes in Section 13.5 for more information about bound entanglement.

As shown in Figure 13.3, we can use entanglement distillation to split up the set of all bipartite states into distillable and non-distillable states. For two-qubit states and qubit-qutrit states, non-distillable states are exactly equal to the set of separable states by the PPT criterion. For higher dimensions, as stated above, this is not the case. Also in higher dimensions, it is in general possible to have states with negative partial transpose (NPT) that are nonetheless non-distillable. These *NPT bound entangled* states are shown in Figure 13.3(b) as the region between the PPT bound entangled states and the distillable entangled states. It is not known whether NPT bound entangled states exist, but since they have not been ruled out, we nevertheless depict them in the figure.

13.2.1 Proof of Achievability

As the first step in proving the achievability part of Theorem 13.19, let us recall Corollary 13.12: given a bipartite state ρ_{AB} with purification ψ_{ABE}, for all $\varepsilon \in (0,1)$, $\eta \in [0, \sqrt{\varepsilon})$, and $\alpha > 1$, there exists a (d, ε) entanglement distillation protocol for ρ_{AB} such that

$$\log_2 d \geq \widetilde{H}_\alpha(A|E)_\psi - \frac{1}{\alpha - 1}\log_2\left(\frac{1}{(\sqrt{\varepsilon}-\eta)^2}\right) - \log_2\left(\frac{1}{1-(\sqrt{\varepsilon}-\eta)^2}\right) + 4\log_2\eta, \quad (13.2.21)$$

where

$$\widetilde{H}_\alpha(A|E)_\psi = -\inf_{\sigma_E} \widetilde{D}_\alpha(\psi_{AE} \| \mathbb{1}_A \otimes \sigma_E) \quad (13.2.22)$$

is the sandwiched Rényi conditional entropy. Applying this inequality to the state $\rho_{AB}^{\otimes n}$ for all $n \geq 1$ leads to the following:

Proposition 13.20

For every state ρ_{AB} and $\varepsilon \in (0,1)$, there exists an (n, d, ε) entanglement distillation protocol for ρ_{AB} such that the rate $\frac{\log_2 d}{n}$ satisfies

$$\frac{\log_2 d}{n} \geq \widetilde{H}_\alpha(A|E)_\psi - \frac{2\alpha - 1}{n(\alpha-1)}\log_2\left(\frac{4}{\varepsilon}\right) - \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon}{4}}\right), \quad (13.2.23)$$

for all $n \geq 1$ and $\alpha > 1$, where ψ_{ABE} is a purification of ρ_{AB}. In general,

$$E_D^{n,\varepsilon}(A;B) \geq \sup_{\mathcal{L}} \frac{1}{n}\widetilde{H}_\alpha(A'|E')_\phi - \frac{2\alpha-1}{n(\alpha-1)}\log_2\left(\frac{4}{\varepsilon}\right) - \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon}{4}}\right), \quad (13.2.24)$$

for all $n \geq 1$ and $\alpha > 1$, where the optimization is over every LOCC channel $\mathcal{L}_{A^nB^n \to A'B'}$, such that $\phi_{A'B'E'}$ is a purification of $\mathcal{L}_{A^nB^n\to A'B'}(\rho_{AB}^{\otimes n})$.

Proof: Let ψ_{ABE} be a purification of ρ_{AB}, and use the tensor-product purification $\psi_{ABE}^{\otimes n}$ for $\rho_{AB}^{\otimes n}$. Also, let $\eta = \frac{\sqrt{\varepsilon}}{2}$. Substituting all of this into the inequality in

(13.2.21) and simplifying leads to

$$\frac{\log_2 d}{n} \geq \frac{1}{n}\widetilde{H}_\alpha(A^n|E^n)_{\psi^{\otimes n}} - \frac{1}{n(\alpha-1)}\log_2\left(\frac{4}{\varepsilon}\right) - \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon}{4}}\right) - \frac{2}{n}\log_2\left(\frac{4}{\varepsilon}\right), \quad (13.2.25)$$

Then, optimizing over every LOCC channel $\mathcal{L}_{A^nB^n\to A'B'}$, and using the definition of $E_D^{n,\varepsilon}(A;B)_\rho$ in (13.2.4), we obtain (13.2.24).

By employing additivity of the sandwiched Rényi conditional entropy for all $\alpha > 1$, we have that

$$\widetilde{H}_\alpha(A^n|E^n)_{\psi^{\otimes n}} = n\widetilde{H}_\alpha(A|E)_\psi. \quad (13.2.26)$$

Note that the proof of additivity follows similarly to the proof of Proposition 11.21. This leads to (13.2.23). ■

With the inequality in (13.2.23), we can prove that the coherent information is an achievable rate for entanglement distillation.

Theorem 13.21 Achievability of Coherent Information for Entanglement Distillation

The coherent information $I(A\rangle B)_\rho$ of a bipartite state ρ_{AB} is an achievable rate for entanglement distillation for ρ_{AB}. In other words, $E_D(A;B)_\rho \geq I(A\rangle B)_\rho$ for every bipartite state ρ_{AB}.

Proof: Let ψ_{ABE} be a purification of ρ_{AB}. Fix $\varepsilon \in (0,1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta = \delta_1 + \delta_2. \quad (13.2.27)$$

Set $\alpha \in (1,\infty)$ such that

$$\delta_1 \geq I(A\rangle B)_\rho - \widetilde{H}_\alpha(A|E)_\psi. \quad (13.2.28)$$

Note that this is possible because $\widetilde{H}_\alpha(A|E)_\psi$ increases monotonically with decreasing α (this follows from Proposition 7.23), so that

$$\lim_{\alpha\to 1^+}\widetilde{H}_\alpha(A|E)_\psi = \sup_{\alpha\in(1,\infty)}\widetilde{H}_\alpha(A|E)_\psi. \quad (13.2.29)$$

Also,

$$\lim_{\alpha\to 1^+} \widetilde{H}_\alpha(A|E)_\psi = \sup_{\alpha\in(1,\infty)} \widetilde{H}_\alpha(A|E)_\psi \tag{13.2.30}$$
$$= \sup_{\alpha\in(1,\infty)} \left(-\inf_{\sigma_E} \widetilde{D}_\alpha(\psi_{AE}\|\mathbb{1}_A\otimes\sigma_E)\right) \tag{13.2.31}$$
$$= -\inf_{\alpha\in(1,\infty)} \inf_{\sigma_E} \widetilde{D}_\alpha(\psi_{AE}\|\mathbb{1}_A\otimes\sigma_E) \tag{13.2.32}$$
$$= -\inf_{\sigma_E} \inf_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\psi_{AE}\|\mathbb{1}_A\otimes\sigma_E) \tag{13.2.33}$$
$$= -\inf_{\sigma_E} D(\psi_{AE}\|\mathbb{1}_A\otimes\sigma_E) \tag{13.2.34}$$
$$= H(A|E)_\psi, \tag{13.2.35}$$

where the fifth equality follows from Proposition 7.30. Then, by definition of conditional entropy, and the fact that ψ_{ABE} is a pure state, we find that

$$H(A|E)_\psi = H(AE)_\psi - H(E)_\psi = H(B)_\psi - H(AB)_\psi = I(A\rangle B)_\rho. \tag{13.2.36}$$

With $\alpha \in (1,\infty)$ chosen such that (13.2.28) holds, take n large enough so that

$$\delta_2 \geq \frac{2\alpha-1}{n(\alpha-1)}\log_2\!\left(\frac{4}{\varepsilon}\right) + \frac{1}{n}\log_2\!\left(\frac{1}{1-\frac{\varepsilon}{4}}\right). \tag{13.2.37}$$

Now, we use the fact that for the n and ε chosen above there exists an (n,d,ε) protocol such that

$$\frac{\log_2 d}{n} \geq \widetilde{H}_\alpha(A|E)_\psi - \frac{2\alpha-1}{n(\alpha-1)}\log_2\!\left(\frac{4}{\varepsilon}\right) - \frac{1}{n}\log_2\!\left(\frac{1}{1-\frac{\varepsilon}{4}}\right). \tag{13.2.38}$$

(This follows from Proposition 13.20 above.) Rearranging the right-hand side of this inequality, and using (13.2.27), (13.2.28), and (13.2.37), we find that

$$\frac{\log_2 d}{n} \geq I(A\rangle B)_\rho - \left(I(A\rangle B)_\rho - \widetilde{H}_\alpha(A|E)_\psi + \frac{2\alpha-1}{n(\alpha-1)}\log_2\!\left(\frac{4}{\varepsilon}\right) + \frac{1}{n}\log_2\!\left(\frac{1}{1-\frac{\varepsilon}{4}}\right)\right) \tag{13.2.39}$$
$$\geq I(A\rangle B)_\rho - (\delta_1+\delta_2) \tag{13.2.40}$$
$$= I(A\rangle B)_\rho - \delta. \tag{13.2.41}$$

We thus have that there exists an (n, d, ε) entanglement distillation protocol with rate $\frac{\log_2 d}{n} \geq I(A\rangle B)_\rho - \delta$. Therefore, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement distillation protocol with $R = I(A\rangle B)_\rho$ for all sufficiently large n such that (13.2.37) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(I(A\rangle B)_\rho - \delta)}, \varepsilon)$ entanglement distillation protocol. This means that, by definition, $I(A\rangle B)_\rho$ is an achievable rate. ■

Proof of the Achievability Part of Theorem 13.19

Let $\mathcal{L}_{A^k B^k \to A'B'}$ be an arbitrary LOCC channel with $k \geq 1$, let

$$\omega_{A'B'} = \mathcal{L}_{A^k B^k \to A'B'}(\rho_{AB}^{\otimes k}), \tag{13.2.42}$$

and let $\phi_{A'B'E'}$ be a purification of $\omega_{A'B'}$. Fix $\varepsilon \in (0, 1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta = \delta_1 + \delta_2. \tag{13.2.43}$$

Set $\alpha \in (1, \infty)$ such that

$$\delta_1 \geq \frac{1}{k} I(A'\rangle B')_\omega - \frac{1}{k} \widetilde{H}_\alpha(A'|E')_\phi, \tag{13.2.44}$$

which is possible based on the arguments given in the proof of Theorem 13.21 above. Then, with this choice of α, take n large enough so that

$$\delta_2 \geq \frac{2\alpha - 1}{kn(\alpha - 1)} \log_2\left(\frac{4}{\varepsilon}\right) + \frac{1}{kn} \log_2\left(\frac{1}{1 - \frac{\varepsilon}{4}}\right). \tag{13.2.45}$$

Now, we use the fact that, for the chosen n and ε, there exists an (n, d, ε) entanglement distillation protocol such that (13.2.23) holds, i.e.,

$$\frac{\log_2 d}{n} \geq \widetilde{H}_\alpha(A'|E')_\phi - \frac{2\alpha - 1}{n(\alpha - 1)} \log_2\left(\frac{4}{\varepsilon}\right) - \frac{1}{n} \log_2\left(\frac{1}{1 - \frac{\varepsilon}{4}}\right). \tag{13.2.46}$$

Dividing both sides by k gives

$$\frac{\log_2 d}{kn} \geq \frac{1}{k} \widetilde{H}_\alpha(A'|E')_\phi - \frac{2\alpha - 1}{kn(\alpha - 1)} \log_2\left(\frac{4}{\varepsilon}\right) - \frac{1}{kn} \log_2\left(\frac{1}{1 - \frac{\varepsilon}{4}}\right). \tag{13.2.47}$$

Rearranging the right-hand side of this inequality, and using (13.2.43)–(13.2.45), we find that

$$\frac{\log_2 d}{kn} \geq \frac{1}{k} I(A'\rangle B')_\omega - \left(\frac{1}{k} I(A'\rangle B')_\omega - \frac{1}{k} \widetilde{H}_\alpha(A'|E')_\phi\right.$$

$$+\frac{2\alpha-1}{kn(\alpha-1)}\log_2\left(\frac{4}{\varepsilon}\right)+\frac{1}{kn}\log_2\left(\frac{1}{1-\frac{\varepsilon}{4}}\right)\Bigg) \tag{13.2.48}$$

$$\geq \frac{1}{k}I(A'\rangle B')_\omega-(\delta_1+\delta_2) \tag{13.2.49}$$

$$=\frac{1}{k}I(A'\rangle B')_\omega-\delta. \tag{13.2.50}$$

Thus, there exists a (kn, d, ε) entanglement distillation protocol with rate $\frac{\log_2 d}{kn} \geq \frac{1}{k}I(A'\rangle B')_\omega - \delta$. Therefore, letting $n' \equiv kn$, there exists an $(n', 2^{n'(R-\delta)}, \varepsilon)$ entanglement distillation protocol with $R = \frac{1}{k}I(A'\rangle B')_\omega$ for all sufficiently large n such that (13.2.45) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(\frac{1}{k}I(A'\rangle B')_\omega-\delta)}, \varepsilon)$ entanglement distillation protocol. This means that $\frac{1}{k}I(A'\rangle B')_\omega$ is an achievable rate. (Recall that $\omega_{A'B'} = \mathcal{L}_{A^kB^k\to A'B'}(\rho_{AB}^{\otimes k})$.)

Now, since in the arguments above the LOCC channel $\mathcal{L}_{A^kB^k\to A'B'}$ is arbitrary, we conclude that

$$\frac{1}{k}\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes k})} \tag{13.2.51}$$

is an achievable rate. Finally, since the number k of copies of ρ_{AB} is arbitrary, we conclude that

$$\lim_{k\to\infty}\frac{1}{k}\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes k})} \tag{13.2.52}$$

is an achievable rate.

13.2.2 Proof of the Weak Converse

In order to prove the weak converse part of Theorem 13.19, we make use of Corollary 13.8, specifically (13.1.44): given a bipartite state ρ_{AB}, for every (d, ε) entanglement distillation protocol for ρ_{AB}, with $\varepsilon \in [0, 1/2)$, it holds that

$$\log_2 d \leq \frac{1}{1-2\varepsilon}\left(\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho)} + h_2(\varepsilon)\right), \tag{13.2.53}$$

where the optimization is with respect to every LOCC channel $\mathcal{L}_{AB\to A'B'}$. Applying this inequality to the state $\rho_{AB}^{\otimes n}$ immediately leads to the following.

Proposition 13.22

Let ρ_{AB} be a bipartite state, and let $n \geq 1$ and $\varepsilon \in [0, 1/2)$. For an (n, d, ε) entanglement distillation protocol for ρ_{AB} with corresponding LOCC channel $\mathcal{L}_{A^n B^n \to A' B'}$, the rate $\frac{\log_2 d}{n}$ satisfies

$$\frac{\log_2 d}{n} \leq \frac{1}{1 - 2\varepsilon} \left(\sup_{\mathcal{L}} \frac{1}{n} I(A' \rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \frac{1}{n} h_2(\varepsilon) \right). \tag{13.2.54}$$

Consequently,

$$E_D^{n,\varepsilon}(A; B)_\rho \leq \frac{1}{1 - 2\varepsilon} \left(\frac{1}{n} \sup_{\mathcal{L}} I(A' \rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \frac{1}{n} h_2(\varepsilon) \right), \tag{13.2.55}$$

where the optimization is over every LOCC channel $\mathcal{L}_{A^n B^n \to A' B'}$.

Proof: The inequality in (13.2.54) is immediate from (13.1.44) in Corollary 13.8 by applying that inequality to the state $\rho_{AB}^{\otimes n}$ and dividing both sides by n. The inequality in (13.2.55) follows immediately by definition of $E_D^{n,\varepsilon}$ in (13.2.4). ■

Proof of the Weak Converse Part of Theorem 13.19

Suppose that R is an achievable rate for entanglement distillation for the bipartite state ρ_{AB}. Then, by definition, for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB}. For all such protocols for which $\varepsilon \in (0, 1/2)$, the inequality in (13.2.54) holds, so that

$$R - \delta \leq \frac{1}{1 - 2\varepsilon} \left(\frac{1}{n} \sup_{\mathcal{L}} I(A' \rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \frac{1}{n} h_2(\varepsilon) \right). \tag{13.2.56}$$

Since the inequality holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$R \leq \lim_{n \to \infty} \frac{1}{1 - 2\varepsilon} \left(\frac{1}{n} \sup_{\mathcal{L}} I(A' \rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \frac{1}{n} h_2(\varepsilon) \right) + \delta \tag{13.2.57}$$

$$= \frac{1}{1 - 2\varepsilon} \lim_{n \to \infty} \frac{1}{n} \sup_{\mathcal{L}} I(A' \rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \delta. \tag{13.2.58}$$

Then, since this inequality holds for all $\varepsilon \in (0, 1/2)$, $\delta > 0$, we conclude that

$$R \leq \lim_{\varepsilon,\delta\to 0}\left(\frac{1}{1-2\varepsilon}\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})} + \delta\right) \tag{13.2.59}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})}. \tag{13.2.60}$$

We have thus shown that the quantity $\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})}$ is a weak converse rate for entanglement distillation for ρ_{AB}.

13.2.3 Rains Relative Entropy Strong Converse Upper Bound

As stated previously, the expression in (13.2.13) for distillable entanglement involves both a limit over an unbounded number of copies of the initial state ρ_{AB}, as well as an optimization over all two-way LOCC channels. Computing the distillable entanglement is therefore intractable in general. In this section, we use the one-shot upper bound established in Section 13.1.1 to show that the Rains relative entropy is a strong converse upper bound on the distillable entanglement of a bipartite state.

We start by recalling the upper bound in (13.1.45), which tells us that

$$\log_2 d \leq \widetilde{R}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{13.2.61}$$

for an arbitrary (d,ε) entanglement distillation protocol, where $\varepsilon \in (0,1)$. Recall that

$$\widetilde{R}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}). \tag{13.2.62}$$

The upper bound above is a consequence of the fact that PPT′ operators are useless for entanglement distillation, in the sense that for every $\sigma_{AB} \in \mathrm{PPT}'(A:B)$, the bound $\mathrm{Tr}[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{d}$ holds.

Applying the upper bound in (13.2.61) to the state $\rho_{AB}^{\otimes n}$ leads to the following result:

Corollary 13.23

Let ρ_{AB} be a bipartite state, let $n \geq 1$, $\varepsilon \in [0, 1)$, and $\alpha > 1$. For an (n, d, ε) entanglement distillation protocol, the following bound holds

$$\frac{\log_2 d}{n} \leq \widetilde{R}_\alpha(A; B)_\rho + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{1 - \varepsilon}\right). \tag{13.2.63}$$

Consequently,

$$E_D^{n,\varepsilon}(A; B)_\rho \leq \widetilde{R}_\alpha(A; B)_\rho + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{1 - \varepsilon}\right). \tag{13.2.64}$$

Proof: An (n, d, ε) entanglement distillation protocol for ρ_{AB} is a (d, ε) entanglement distillation protocol for $\rho_{AB}^{\otimes n}$. Therefore, applying the inequality in (13.2.61) to the state $\rho_{AB}^{\otimes n}$ and dividing both sides by n leads to

$$\frac{\log_2 d}{n} \leq \frac{1}{n} \widetilde{R}_\alpha(A^n; B^n)_{\rho^{\otimes n}} + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{1 - \varepsilon}\right). \tag{13.2.65}$$

Now, by subadditivity of the sandwiched Rényi Rains relative entropy (see (9.3.18)), we have that

$$\widetilde{R}_\alpha(A^n; B^n)_{\rho^{\otimes n}} \leq n \widetilde{R}_\alpha(A; B)_\rho. \tag{13.2.66}$$

Therefore,

$$\frac{\log_2 d}{n} \leq \widetilde{R}_\alpha(A; B)_\rho + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{1 - \varepsilon}\right), \tag{13.2.67}$$

as required. Since this inequality holds for all (n, d, ε) protocols, we obtain (13.2.64) by optimizing over all protocols $(d, \mathcal{L}_{AB \to \hat{A}\hat{B}})$, with $d_{\hat{A}} = d_{\hat{B}} = d \geq 1$. ■

Given an $\varepsilon \in (0, 1)$, the inequality in (13.2.63) gives us a bound on the rate of an arbitrary (n, d, ε) entanglement distillation protocol for a state ρ_{AB}. If instead we fix the rate to be r, so that $d = 2^{nr}$, then the inequality in (13.2.63) is as follows:

$$r \leq \widetilde{R}_\alpha(A; B)_\rho + \frac{\alpha}{n(\alpha - 1)} \log_2\left(\frac{1}{1 - \varepsilon}\right) \tag{13.2.68}$$

for all $\alpha > 1$. Rearranging this inequality gives us the following lower bound on ε:

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha - 1}{\alpha}\right)\left(r - \widetilde{R}_\alpha(A; B)_\rho\right)} \tag{13.2.69}$$

for all $\alpha > 1$.

Theorem 13.24 Strong Converse Upper Bound on Distillable Entanglement

Let ρ_{AB} be a bipartite state. The Rains relative entropy $R(A;B)_\rho$ is a strong converse rate for entanglement distillation for ρ_{AB}, i.e.,

$$\widetilde{E}_D(A;B)_\rho \leq R(A;B)_\rho, \tag{13.2.70}$$

where we recall that $R(A;B)_\rho$ is defined as

$$R(A;B)_\rho = \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} D(\rho_{AB} \| \sigma_{AB}). \tag{13.2.71}$$

Proof: Let $\varepsilon \in [0,1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \tag{13.2.72}$$

Set $\alpha \in (1, \infty)$ such that

$$\delta_1 \geq \widetilde{R}_\alpha(A;B)_\rho - R(A;B)_\rho, \tag{13.2.73}$$

which is possible because $\widetilde{R}_\alpha(A;B)_\rho$ is monotonically increasing in α (which follows from Proposition 7.31) and because $\lim_{\alpha \to 1^+} \widetilde{R}_\alpha(A;B)_\rho = R(A;B)_\rho$ (see Appendix 10.A for a proof). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{13.2.74}$$

Now, with the values of n and ε as above, an arbitrary (n, d, ε) entanglement distillation protocol for ρ_{AB} satisfies (13.2.63), i.e.,

$$\frac{\log_2 d}{n} \leq \widetilde{R}_\alpha(A;B)_\rho + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right) \tag{13.2.75}$$

for all $\alpha \in (1, \infty)$. Rearranging the right-hand side of this inequality, and using (13.2.72)–(13.2.74), we obtain

$$\frac{\log_2 d}{n} \leq R(A;B)_\rho + \widetilde{R}_\alpha(A;B)_\rho - R(A;B)_\rho + \frac{\alpha}{n(\alpha-1)} \log_2\left(\frac{1}{1-\varepsilon}\right) \tag{13.2.76}$$

$$\leq R(A;B)_\rho + \delta_1 + \delta_2 \quad (13.2.77)$$
$$= R(A;B)_\rho + \delta' \quad (13.2.78)$$
$$< R(A;B)_\rho + \delta. \quad (13.2.79)$$

So we have that $\frac{\log_2 d}{n} < R(A;B)_\rho + \delta$ for all (n,d,ε) entanglement distillation protocols for ρ_{AB} with sufficiently large n such that (13.2.74) holds. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(R(A;B)_\rho+\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB} for all sufficiently large n such that (13.2.74) holds. For if it were to exist, there would be a $d \geq 1$ such that $\log_2 d = n(R(A;B)_\rho + \delta)$, which we have just seen is not possible. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R(A;B)_\rho+\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB}. This means that $R(A;B)_\rho$ is a strong converse rate, so that $\widetilde{E}_D(A;B)_\rho \leq R(A;B)_\rho$. ■

Given that the Rains relative entropy is a strong converse rate for distillable entanglement, by following arguments analogous to those in the proof above, we can conclude that $\frac{1}{k}R(A^k;B^k)_{\rho^{\otimes k}}$ is a strong converse rate for all $k \geq 2$. Therefore, the regularized quantity

$$R^{\text{reg}}(A;B)_\rho \coloneqq \lim_{n\to\infty} \frac{1}{n} R(A^n;B^n)_{\rho^{\otimes n}}, \quad (13.2.80)$$

is a strong converse rate for entanglement distillation for ρ_{AB}, so that

$$\widetilde{E}_D(A;B)_\rho \leq R^{\text{reg}}(A;B)_\rho. \quad (13.2.81)$$

By subadditivity of Rains relative entropy (see (9.3.18)),

$$R^{\text{reg}}(A;B)_\rho \leq R(A;B)_\rho, \quad (13.2.82)$$

so that the regularized quantity in general gives a tighter upper bound on distillable entanglement.

13.2.3.1 The Strong Converse from a Different Point of View

We now show that the Rains relative entropy is a strong converse rate using the equivalent characterization of a strong converse rate in (13.2.9). In other words, given a bipartite state ρ_{AB}, we show that for an arbitrary sequence $\{(n, 2^{nr}, \varepsilon_n)\}_{n\in\mathbb{N}}$

of (n, d, ε) protocols with rates $r > R(A;B)_\rho$, it holds that $\lim_{n\to\infty} \varepsilon_n = 1$. Indeed, for every element of the sequence, the inequality in (13.2.68) applies; namely,

$$r \le \widetilde{R}_\alpha(A;B)_\rho + \frac{\alpha}{n(\alpha-1)} \log_2\!\left(\frac{1}{1-\varepsilon_n}\right) \tag{13.2.83}$$

for all $\alpha > 1$. Rearranging this inequality gives us the following lower bound on ε_n:

$$\varepsilon_n \ge 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(r-\widetilde{R}_\alpha(A;B)_\rho\right)} \tag{13.2.84}$$

for all $\alpha > 1$. Now, since $r > R(A;B)_\rho$, $\lim_{\alpha\to 1^+} \widetilde{R}_\alpha(A;B)_\rho = R(A;B)_\rho$ (see Appendix 10.A), and because $\widetilde{R}_\alpha(A;B)_\rho$ is monotonically increasing in α (this follows from Proposition 7.31), there exists an $\alpha^* > 1$ such that $r > \widetilde{R}_{\alpha^*}(A;B)_\rho$. Applying the inequality in (13.2.84) to this value of α, we find that

$$\varepsilon_n \ge 1 - 2^{-n\left(\frac{\alpha^*-1}{\alpha^*}\right)\left(r-\widetilde{R}_{\alpha^*}(A;B)_\rho\right)}. \tag{13.2.85}$$

Then, taking the limit $n \to \infty$ on both sides of this inequality, we conclude that $\lim_{n\to\infty} \varepsilon_n \ge 1$. But $\varepsilon_n \le 1$ for all n because ε_n is a probability by definition. So we obtain $\lim_{n\to\infty} \varepsilon_n = 1$. Since the rate $r > R(A;B)_\rho$ is arbitrary, we conclude that $R(A;B)_\rho$ is a strong converse rate for entanglement distillation for ρ_{AB}. We also see from (13.2.85) that the sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ approaches one at an exponential rate.

13.2.4 Squashed Entanglement Weak Converse Upper Bound

In this section, we establish the squashed entanglement of a bipartite state as a weak converse upper bound on its distillable entanglement. The main idea is to apply the one-shot bound from Theorem 13.9 and the additivity of the squashed entanglement (Proposition 9.32) in order to arrive at this conclusion.

Corollary 13.25

Let ρ_{AB} be a bipartite state, let $n \ge 1$, and let $\varepsilon \in [0,1)$. For an (n, d, ε) entanglement distillation protocol, the following bound holds

$$\frac{1}{n}\log_2 d \le \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\text{sq}}(A;B)_\rho + \frac{g_2(\sqrt{\varepsilon})}{n}\right). \tag{13.2.86}$$

Proof: An (n, d, ε) entanglement distillation protocol for ρ_{AB} is a (d, ε) entanglement distillation protocol for $\rho_{AB}^{\otimes n}$. Therefore, applying the inequality in (13.9) to the state $\rho_{AB}^{\otimes n}$ and dividing both sides by n leads to

$$\frac{1}{n}\log_2 d \leq \frac{1}{1-\sqrt{\varepsilon}}\left(\frac{1}{n}E_{\text{sq}}(A^n;B^n)_{\rho^{\otimes n}} + \frac{g_2(\sqrt{\varepsilon})}{n}\right). \tag{13.2.87}$$

Now, by additivity of the squashed entanglement (see (9.4.8)), we have that

$$E_{\text{sq}}(A^n;B^n)_{\rho^{\otimes n}} = nE_{\text{sq}}(A;B)_\rho. \tag{13.2.88}$$

This concludes the proof. ■

We now provide a proof of (13.2.15), the statement that the squashed entanglement is a weak converse rate for entanglement distillation. Suppose that R is an achievable rate for entanglement distillation for the bipartite state ρ_{AB}. Then, by definition, for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ entanglement distillation protocol for ρ_{AB}. For all such protocols, the inequality in (13.2.86) holds, so that

$$R - \delta \leq \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\text{sq}}(A;B)_\rho + \frac{1}{n}g_2(\sqrt{\varepsilon})\right). \tag{13.2.89}$$

Since the inequality holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$R \leq \lim_{n\to\infty} \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\text{sq}}(A;B)_\rho + \frac{1}{n}g_2(\sqrt{\varepsilon})\right) + \delta \tag{13.2.90}$$

$$= \frac{1}{1-\sqrt{\varepsilon}}E_{\text{sq}}(A;B)_\rho + \delta. \tag{13.2.91}$$

Then, since this inequality holds for all $\varepsilon \in (0,1]$ and $\delta > 0$, we conclude that

$$R \leq \lim_{\varepsilon,\delta\to 0}\left(\frac{1}{1-\sqrt{\varepsilon}}E_{\text{sq}}(A;B)_\rho + \delta\right) \tag{13.2.92}$$

$$= E_{\text{sq}}(A;B)_\rho. \tag{13.2.93}$$

We have thus shown that the squashed entanglement $E_{\text{sq}}(A;B)_\rho$ is a weak converse rate for entanglement distillation.

13.2.5 One-Way Entanglement Distillation

In Section 13.1.2, we considered a one-way entanglement distillation protocol to derive a lower bound on the one-shot distillable entanglement of a bipartite state. In the asymptotic setting, this leads to the coherent information lower bound on distillable entanglement, i.e.,

$$E_D(A;B)_\rho \geq I(A\rangle B)_\rho = H(B)_\rho - H(AB)_\rho, \tag{13.2.94}$$

which holds for every bipartite state ρ_{AB}. By simply reversing the roles of Alice and Bob in the protocol, it follows that

$$E_D(A;B)_\rho \geq I(B\rangle A)_\rho = H(A)_\rho - H(AB)_\rho. \tag{13.2.95}$$

The quantity on the right-hand side of the above inequality is sometimes called *reverse coherent information*. Thus, in general, we have the following lower bound on distillable entanglement:

$$E_D(A;B)_\rho \geq \max\{I(A\rangle B)_\rho, I(B\rangle A)_\rho\}, \tag{13.2.96}$$

which holds for every bipartite state ρ_{AB}.

The coherent information lower bound can be improved by first applying a two-way LOCC channel to n copies of the given state, and then performing a one-way entanglement distillation protocol at the coherent information rate. This leads to

$$E_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}} I(A'\rangle B')_{\mathcal{L}(\rho^{\otimes n})} = \lim_{n\to\infty} \frac{1}{n} D^{\leftrightarrow}(\rho_{AB}^{\otimes n}) \tag{13.2.97}$$

where the optimization is over every two-way LOCC channel $\mathcal{L}_{A^nB^n\to A'B'}$.

If we restrict the optimization in (13.2.97) above to one-way LOCC channels of the form $\mathcal{L}_{A^nB^n\to A'B'}$, then we obtain what is called the *one-way distillable entanglement*, denoted by $E_D^{\rightarrow}(A;B)_\rho$, and defined operationally in a similar way to the distillable entanglement $E_D(A;B)_\rho$, but with the free operations allowed restricted to one-way LOCC. A key result is the following equality:

$$E_D^{\rightarrow}(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{\rightarrow}} I(A'\rangle B')_{\mathcal{L}^{\rightarrow}(\rho^{\otimes n})} = \lim_{n\to\infty} \frac{1}{n} D^{\rightarrow}(\rho_{AB}^{\otimes n}), \tag{13.2.98}$$

where

$$D^{\rightarrow}(\rho_{AB}) := \sup_{\mathcal{L}^{\rightarrow}} I(A'\rangle B')_{\mathcal{L}^{\rightarrow}(\rho)}. \tag{13.2.99}$$

Like the distillable entanglement, the one-way distillable entanglement is an operational quantity of interest in entanglement theory. Furthermore, the equality in (13.2.98) can be proved along similar lines to how we proved (13.2.13).

In what follows, we show that this expression for one-way distillable entanglement can be simplified.

Theorem 13.26 One-Way Distillable Entanglement of a Bipartite State

The one-way distillable entanglement of a bipartite state ρ_{AB} is given by

$$E_D^{\rightarrow}(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_V I(A'\rangle XB^n)_{V\rho^{\otimes n}V^\dagger}, \tag{13.2.100}$$

where $V_{A^n\to A'XE}$ is an isometry of the form

$$V_{A^n\to A'XE} = \sum_{x\in\mathcal{X}} K^x_{A^n\to A'} \otimes |x\rangle_X \otimes |x\rangle_E, \tag{13.2.101}$$

with $\sum_{x\in\mathcal{X}}(K^x_{A^n})^\dagger K^x_{A^n} = \mathbb{1}_{A^n}$, $d_{A'} = d_A^n$, $d_X \leq d_A^2$, and $d_E = |\mathcal{X}| = d_X$. Additionally,

$$D^{\rightarrow}(\rho_{AB}) = \sup_V I(A'\rangle XB)_{V\rho V^\dagger}, \tag{13.2.102}$$

with the optimization over isometries V as in (13.2.101), with $n = 1$.

This theorem tells us that, to determine the one-way distillable entanglement of a bipartite state, it suffices to optimize over one-way LOCC channels that consist of only a quantum instrument for Alice, with each of the corresponding maps containing just one Kraus operator. Furthermore, it suffices to take $A' = A^n$ and $B' = B^n$.

Proof: We start by recalling from Section 4.6.2 (see also the beginning of Section 13.1.2) that every one-way LOCC channel $\mathcal{L}^{\rightarrow}_{A^nB^n\to A'B'}$ can be expressed as

$$\begin{aligned} \omega_{A'B'} &:= \mathcal{L}^{\rightarrow}_{A^nB^n\to A'B'}(\rho_{A^nB^n}) && (13.2.103)\\ &= \sum_{x\in\mathcal{X}} (\mathcal{E}^x_{A^n\to A'} \otimes \mathcal{D}^x_{B^n\to B'})(\rho_{A^nB^n}) && (13.2.104)\\ &= \left(\mathcal{D}_{X_BB^n\to B'} \circ \mathcal{C}_{X_A\to X_B} \circ \mathcal{E}_{A^n\to A'X_A}\right)(\rho_{A^nB^n}), && (13.2.105)\end{aligned}$$

where $\mathcal{X}$ is some finite alphabet, $d_{X_A} = d_{X_B} = |\mathcal{X}|$, $\{\mathcal{E}^x_{A^n\to A'}\}_{x\in\mathcal{X}}$ is a set of completely positive maps such that $\sum_{x\in\mathcal{X}}\mathcal{E}^x_{A^n\to A'}$ is trace preserving, and $\{\mathcal{D}^x_{B^n\to B'}\}_{x\in\mathcal{X}}$ is a set of channels. In particular,

$$\mathcal{E}_{A^n\to A'X_A}(\rho_{A^nB^n}) = \sum_{x\in\mathcal{X}}\mathcal{E}^x_{A^n\to A'}(\rho_{A^nB^n})\otimes|x\rangle\langle x|_{X_A}, \tag{13.2.106}$$

$$\mathcal{D}_{X_BB^n\to B'}(|x\rangle\langle x|_{X_B}\otimes\rho_{A^nB^n}) = \mathcal{D}^x_{B^n\to B'}(\rho_{A^nB^n}). \tag{13.2.107}$$

For every $n\geq 1$, if we restrict the optimization in (13.2.98) such that $|\mathcal{X}| = d^2_{A^n} = d^{2n}_A$, $d = d_{A'} = d^n_A$, $\mathcal{D}^x_{B^n\to B'} = \mathrm{id}_{B^n}$ for all $x\in\mathcal{X}$, and $\mathcal{E}^x_{A^n\to A'}(\cdot) = K^x_{A^n}(\cdot)K^x_{A^n}$ for all $x\in\mathcal{X}$ such that $\sum_{x\in\mathcal{X}}(K^x_{A^n})^\dagger K^x_{A^n} = \mathbb{1}_{A^n}$, then the LOCC channel $\mathcal{L}^{\rightarrow}_{A^nB^n\to A'B'}$ reduces to

$$\mathcal{L}^{\rightarrow}_{A^nB^n\to A'B'}(\rho_{A^nB^n}) = \sum_{x\in\mathcal{X}}K^x_{A^n}\rho_{A^nB^n}(K^x_{A^n})^\dagger\otimes|x\rangle\langle x|_X \tag{13.2.108}$$

for every state $\rho_{A^nB^n}$, and it has an isometric extension of the form

$$V_{A^n\to A^nXE} = \sum_{x\in\mathcal{X}}K^x_{A^n}\otimes|x\rangle_X\otimes|x\rangle_E. \tag{13.2.109}$$

We thus obtain

$$E^{\rightarrow}_D(A;B)_\rho \geq \lim_{n\to\infty}\frac{1}{n}\sup_V I(A^n\rangle XB^n)_\omega \tag{13.2.110}$$

The rest of the proof is devoted to proving the reverse inequality. Let $\mathcal{L}^{\rightarrow}_{A^nB^n\to A'B'}$ be an arbitrary one-way LOCC channel of the form in (13.2.103)–(13.2.105). For every state $\rho_{A^nB^n}$, let

$$p_x := \mathrm{Tr}[\mathcal{E}^x_{A^n\to A'}(\rho_{A^nB^n})], \tag{13.2.111}$$

$$\rho^x_{A'B^n} := \frac{1}{p_x}\mathcal{E}^x_{A^n\to A'}(\rho_{A^nB^n}), \tag{13.2.112}$$

$$\rho^x_{B^n} := \frac{1}{p_x}\mathrm{Tr}_{A'}[\mathcal{E}^x_{A^n\to A'}(\rho_{A^nB^n})], \tag{13.2.113}$$

for all $x\in\mathcal{X}$. Then, using the data-processing inequality for coherent information in (7.3.17) and the direct-sum property for quantum relative entropy, we obtain

$$I(A'\rangle B')_\omega \leq I(A'\rangle B^nX_B)_{\mathcal{E}(\rho)} \tag{13.2.114}$$

$$= \sum_{x\in\mathcal{X}}p_xD(\rho^x_{A'B^n}\|\mathbb{1}_{A'}\otimes\rho^x_{B^n}) \tag{13.2.115}$$

$$= \sum_{x \in \mathcal{X}} p_x I(A'\rangle B^n)_{\rho^x}. \quad (13.2.116)$$

Now,

$$\mathcal{E}_{A^n \to A' X_A}(\rho_{A^n B^n}) = \sum_{x \in \mathcal{X}} \mathcal{E}^x_{A^n \to A'}(\rho_{AB}) \otimes |x\rangle\langle x|_{X_A} \quad (13.2.117)$$

$$= \sum_{x \in \mathcal{X}} p_x \rho^x_{A'B} \otimes |x\rangle\langle x|_{X_A}. \quad (13.2.118)$$

Suppose that $\mathcal{E}^x_{A^n \to A'}$ has the following Kraus representation:

$$\mathcal{E}^x_{A^n \to A'}(\cdot) = \sum_{y \in \mathcal{Y}} K^{x,y}_{A^n \to A'}(\cdot)(K^{x,y}_{A^n \to A'})^\dagger, \quad (13.2.119)$$

where $\mathcal{Y}$ is some finite alphabet. Let

$$q_{x,y} \coloneqq \operatorname{Tr}[K^{x,y}_{A^n \to A'} \rho_{A^n B^n} (K^{x,y}_{A^n \to A'})^\dagger], \quad (13.2.120)$$

and observe that $p_x = \sum_{y \in \mathcal{Y}} q_{x,y}$. Therefore, for each $x \in \mathcal{X}$, the values $r_{y|x} \coloneqq \frac{q_{x,y}}{p_x}$ constitute a probability distribution on $\mathcal{Y}$, in the sense that $r_{y|x} \geq 0$ for all $y \in \mathcal{Y}$, and $\sum_{y \in \mathcal{Y}} r_{y|x} = 1$. Using this, and letting

$$\rho^{x,y}_{A'B^n} \coloneqq \frac{1}{q_{x,y}} K^{x,y}_{A^n \to A'} \rho_{A^n B^n} (K^{x,y}_{A^n \to A'})^\dagger, \quad (13.2.121)$$

so that

$$p_x \rho^x_{A'B^n} = \sum_{y \in \mathcal{Y}} q_{x,y} \rho^{x,y}_{A'B^n}, \quad (13.2.122)$$

we find that

$$p_x I(A'\rangle B^n)_{\rho^x} \leq p_x \sum_{y \in \mathcal{Y}} r_{y|x} I(A'\rangle B^n)_{\rho^{x,y}}, \quad (13.2.123)$$

for all $x \in \mathcal{X}$, where the inequality follows from convexity of coherent information (see (7.2.121)). Without loss of generality, we can take $A' \equiv A^n$: if A' has a dimension smaller than that of A^n, we can always first isometrically embed A' into A^n. The coherent information remains unchanged under this isometric embedding.

Combining the last inequality above with the one in (13.2.116), we conclude that

$$I(A'\rangle B')_\omega \leq \sum_{x \in \mathcal{X}} p_x \sum_{y \in \mathcal{Y}} r_{y|x} I(A'\rangle B^n)_{\rho^{x,y}} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} q_{x,y} I(A'\rangle B^n)_{\rho^{x,y}}. \quad (13.2.124)$$

Then assigning the superindex $z = (x, y)$ with $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, we finally have

$$I(A'\rangle B')_\omega \leq \sup_{\mathcal{E}} I(A'\rangle B^n Z_B)_{\mathcal{E}(\rho^{\otimes n})}, \tag{13.2.125}$$

where the optimization is over channels $\mathcal{E}_{A^n \to A' Z_B}$ of the form

$$\mathcal{E}_{A^n \to A' Z_B}(\cdot) := \sum_{z \in \mathcal{X}} K^z_{A^n \to A'}(\cdot)(K^z_{A^n \to A'})^\dagger \otimes |z\rangle\langle z|_{Z_B}, \tag{13.2.126}$$

such that $\sum_{z \in \mathcal{Z}} (K^z_{A^n \to A'})^\dagger K^z_{A^n \to A'} = \mathbb{1}_{A^n}$. (This is effectively an optimization over operators $\{K^z_{A^n \to A'}\}_{z \in \mathcal{Z}}$ such that $\sum_{z \in \mathcal{Z}} (K^z_{A^n \to A'})^\dagger K^z_{A^n \to A'} = \mathbb{1}_{A^n}$.) The channel $\mathcal{E}_{A^n \to A' Z_B}$ has an isometric extension of the form

$$V_{A^n \to A' Z_B E} = \sum_{z \in \mathcal{Z}} K^z_{A^n \to A'} \otimes |z\rangle_{Z_B} \otimes |z\rangle_E, \tag{13.2.127}$$

where $d_E = d_{Z_B} = |\mathcal{Z}|$. Since the number of Kraus operators need not exceed $d^2_{A^n} = d^{2n}_A$ (see Theorem 4.3), we can take $d_Z = d^{2n}_A$ without loss of generality. We can thus optimize over all isometries of the form in (13.2.127). Altogether, we have that

$$I(A'\rangle B')_\omega \leq \sup_V I(A'\rangle Z B^n)_{V\rho^{\otimes n} V^\dagger} \tag{13.2.128}$$

for every one-way LOCC channel $\mathcal{L}^{\to}_{A^n B^n \to A' B'}$ and all $n \geq 1$. Optimizing over all one-way LOCC channels on the left-hand side of the inequality above, and taking the limit $n \to \infty$ leads us to conclude that

$$E^{\to}_D(A; B)_\rho \leq \lim_{n \to \infty} \frac{1}{n} \sup_V I(A'\rangle Z B^n)_{V\rho^{\otimes n} V^\dagger}. \tag{13.2.129}$$

Combining this with (13.2.110) and reassigning Z as X finishes the proof. ■

Lemma 13.27

For every bipartite state ρ_{AB}, the optimized coherent information lower bound on distillable entanglement is non-negative, i.e., $D^{\to}(\rho_{AB}) \geq 0$.

Proof: Let $\psi_{ABR} = |\psi\rangle\langle\psi|_{ABR}$ be a purification of ρ_{AB}, and consider the following Schmidt decomposition of $|\psi\rangle_{ABR}$:

$$|\psi\rangle_{ABR} = \sum_{k=0}^{r-1} \sqrt{\lambda_k} |\phi_k\rangle_A \otimes |\varphi_k\rangle_{BR}. \tag{13.2.130}$$

Then, let

$$V_{A\to A'XE} := \sum_{k=0}^{r-1} |k\rangle_{A'}\langle\phi_k|_A \otimes |k\rangle_X \otimes |k\rangle_E. \tag{13.2.131}$$

It is then straightforward to show that $I(A'\rangle XB)_{V\rho V^\dagger} = 0$. Since V is an example of an isometry in the optimization for $D^{\to}(\rho_{AB})$, we conclude that $D^{\to}(\rho_{AB}) \geq I(A'\rangle XB)_{V\rho V^\dagger} = 0$. ■

13.3 Examples

We now consider classes of bipartite states and evaluate the upper and lower bounds on their distillable entanglement that we have established in this chapter. In some cases, the distillable entanglement can be determined exactly because the upper and lower bounds coincide.

13.3.1 Pure States

The simplest example for which distillable entanglement can be determined exactly is the class of pure bipartite states. In this case, the coherent information lower bound and the Rains relative entropy upper bound coincide and are equal to the entropy of the reduced state. Indeed, for the coherent information, the joint entropy $H(AB)_\psi = 0$ for every pure state ψ_{AB}, so that

$$I(A\rangle B)_\psi = H(B)_\psi - H(AB)_\psi = H(B)_\psi = H(A)_\psi, \tag{13.3.1}$$

where the last equality follows from the Schmidt decomposition theorem (Theorem 2.2) to see that the reduced states ψ_A and ψ_B have the same non-zero eigenvalues, and thus the same value for the entropy. On the other hand, Proposition 9.20 states that the relative entropy of entanglement $E_R(A;B)_\psi = H(A)_\psi$, and we also know that $E_R(A;B)_\psi \geq R(A;B)_\psi$ (see (9.1.149)). We thus have the following:

Theorem 13.28 Distillable Entanglement for Pure States

The distillable entanglement of a pure bipartite state ψ_{AB} is equal to the entropy of the reduced state on A, i.e.,

$$E_D(A;B)_\psi = H(A)_\psi. \tag{13.3.2}$$

13.3.2 Degradable and Anti-Degradable States

In this section, we define two classes of states for which the one-way distillable entanglement takes on a simple form.

Definition 13.29 Degradable and Anti-Degradable Bipartite States

Given a bipartite state ρ_{AB} with purification ψ_{ABE}, we call it *degradable* if there exists a quantum channel $\mathcal{D}_{B\to E'}$ such that the state $\tau_{AE'E} := \mathcal{D}_{B\to E'}(\psi_{ABE})$ satisfies

$$\tau_{AE'} = \tau_{AE} = \psi_{AE}. \tag{13.3.3}$$

We call ρ_{AB} *anti-degradable* if there exists a quantum channel $\mathcal{A}_{E\to B'}$ such that the state $\omega_{ABB'} := \mathcal{A}_{E\to B'}(\psi_{ABE})$ satisfies

$$\omega_{AB'} = \omega_{AB} = \rho_{AB}. \tag{13.3.4}$$

Remark: Degradable and anti-degradable states are the state counterparts of degradable and anti-degradable channels; see Definition 4.6. In fact, observe that the Choi state of a degradable channel is a degradable state, and the Choi state of an anti-degradable channel is an anti-degradable state.

Anti-degradable states are also sometimes called symmetrically extendible states or two-extendible states (please consult the Bibliographic Notes in Section 13.5).

Intuitively, a degradable state is one for which the system B can be used to simulate (via a quantum channel $\mathcal{D}_{B\to E'}$) the correlations between A and E. Analogously, an anti-degrdable state is one for which the system E can be used to simulate (via a quantum channel $\mathcal{A}_{E\to B'}$) the correlations between A and B.

An anti-degradable state ρ_{AB} is one for which the environment E (corresponding to the purifying system of ρ_{AB}) *cannot* be decoupled from A and B through LOCC from A to B alone. Indeed, recall the task of decoupling from Section 13.1.2 (in particular, see Figure 13.2). Since a channel can always be applied to E in order to simulate the correlations between A and B, from the point of view of A, the systems B and E become indistinguishable, so that A and B cannot be (perfectly) decoupled from E. Given that decoupling is not possible for anti-degradable states, we might expect that anti-degradable states have zero one-way distillable entanglement. This is indeed true, as we now show.

Theorem 13.30 One-Way Distillable Entanglement for Anti-Degradable States

For an anti-degradable state ρ_{AB}, the one-way distillable entanglement is equal to zero, i.e., $E_D^{\rightarrow}(A;B)_\rho = 0$.

Proof: Let $V_{A\to A'XE}$ be an arbitrary isometry in the optimization for $D^{\rightarrow}(\rho_{AB})$. Also, let ψ_{ABR} be a purification of ρ_{AB}. Then, because ρ_{AB} is anti-degradable, there exists a channel $\mathcal{A}_{R\to B}$ such that

$$\rho_{AB} = \mathcal{A}_{R\to B}(\psi_{AR}). \tag{13.3.5}$$

Now, let

$$\omega_{A'XEBR} = V_{A\to A'XE}\psi_{ABR}V^\dagger_{A\to A'XE}, \tag{13.3.6}$$

which is a pure state. Then, using the fact that

$$\begin{aligned} \omega_{A'XB} &= \mathrm{Tr}_E[V_{A\to A'XE}\rho_{AB}V^\dagger_{A\to A'XE}] && (13.3.7)\\ &= (\mathcal{A}_{R\to B}\circ \mathrm{Tr}_E)(V_{A\to A'XE}\psi_{AR}V^\dagger_{A\to A'XE}) && (13.3.8)\\ &= \mathcal{A}_{R\to B}(\omega_{A'XR}), && (13.3.9)\end{aligned}$$

and that $\omega_{XB} = \mathcal{A}_{R\to B}(\omega_{XR})$, we find that

$$\begin{aligned} I(A'\rangle XB)_\omega &\le I(A'\rangle RX)_\omega && (13.3.10)\\ &= H(RX)_\omega - H(A'RX)_\omega && (13.3.11)\\ &= H(A'EB)_\omega - H(EB)_\omega && (13.3.12)\\ &= H(A'XB)_\omega - H(XB)_\omega && (13.3.13)\\ &= -I(A'\rangle XB)_\omega, && (13.3.14)\end{aligned}$$

where we used the data-processing inequality in (7.3.17), and for the subsequent equalities we used the fact that $\omega_{A'XEBR}$ is a pure state that is symmetric in X and

E. We thus have $I(A'\rangle XB)_{V\rho V^\dagger} \leq 0$ for every isometry V used in the optimization for $D^{\to}(\rho_{AB})$, implying that $D^{\to}(\rho_{AB}) \leq 0$. However, since $D^{\to}(\rho_{AB}) \geq 0$ by Lemma 13.27, we obtain $D^{\to}(\rho_{AB}) = 0$. The statement that $E_D^{\to}(A;B)_\rho = 0$ follows by repeating the same argument for n copies of ρ_{AB} and using the fact that $\rho_{AB}^{\otimes n}$ is an anti-degradable state if ρ_{AB} is. ■

Theorem 13.31 One-Way Distillable Entanglement for Degradable States

For a degradable state ρ_{AB}, we have

$$D^{\to}(\rho_{AB}) = I(A\rangle B)_\rho. \tag{13.3.15}$$

Consequently, $D^{\to}(\rho_{AB}^{\otimes n}) = nD^{\to}(\rho_{AB})$, and thus the one-way distillable entanglement of a degradable state ρ_{AB} is equal to its coherent information:

$$E_D^{\to}(A;B)_\rho = I(A\rangle B)_\rho. \tag{13.3.16}$$

Proof: First, observe that if we pick the isometry V in (13.2.102) to be $V_{A\to A'XE} = \mathbb{1}_A \otimes |0,0\rangle_{XE}$ (so that X and E are one-dimensional systems), then we obtain $D^{\to}(\rho_{AB}) \geq I(A\rangle B)_\rho$. Then, we conclude that $D^{\to}(\rho_{AB}^{\otimes n}) \geq nI(A\rangle B)_\rho$ because coherent information is additive for product states; thus, $E_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} D^{\to}(\rho_{AB}^{\otimes n}) \geq I(A\rangle B)_\rho$.

We now prove the reverse inequality. Let $V_{A\to A'XE}$ be an arbitrary isometry in the optimization for $D^{\to}(\rho_{AB})$. Also, let ψ_{ABR} be a purification of ρ_{AB}. Then, since ρ_{AB} is degradable, there exists a channel $\mathcal{D}_{B\to R'}$ such that the state $\tau_{AR'R} := \mathcal{D}_{B\to R'}(\psi_{ABR})$ satisfies

$$\tau_{AR'} = \tau_{AR} = \psi_{AR}. \tag{13.3.17}$$

Let $W_{B\to R'F}$ be an isometric extension of $\mathcal{D}_{B\to R'}$, and let

$$|\varphi\rangle_{AR'FR} = W_{B\to R'F}|\psi\rangle_{ABR}, \tag{13.3.18}$$
$$|\omega\rangle_{A'XEBR} = V_{A\to A'XE}|\psi\rangle_{ABR}, \tag{13.3.19}$$
$$|\phi\rangle_{A'XER'FR} = W_{B\to R'F}|\omega\rangle_{A'XEBR} = V_{A\to A'XE}|\varphi\rangle_{AR'FR}. \tag{13.3.20}$$

Then, by invariance of entropy under the isometry $W_{B\to R'F}$,

$$I(A'\rangle XB)_\omega = H(XB)_\omega - H(A'XB)_\omega \tag{13.3.21}$$

$$= H(XR'F)_\phi - H(A'XR'F)_\phi \tag{13.3.22}$$
$$= H(XR'F)_\phi - H(ER)_\phi \tag{13.3.23}$$
$$= H(XR'F)_\phi - H(XR)_\phi, \tag{13.3.24}$$

where the second-to-last line follows because $\phi_{A'XER'FR}$ is a pure state, so that $H(A'XR'F)_\phi = H(ER)_\phi$, and then for the last line we used the fact that $\phi_{A'XER'FR}$ is symmetric in X and E by definition of $V_{A\to A'XE}$. Next, due to the fact that $\tau_{AR'} = \tau_{AR}$, it holds that $\phi_{XR'} = \phi_{XR}$. We thus obtain

$$I(A'\rangle XB)_\omega = H(XR'F)_\phi - H(XR')_\phi \tag{13.3.25}$$
$$= -I(F\rangle XR')_\phi \tag{13.3.26}$$
$$\leq -I(F\rangle R')_\phi \tag{13.3.27}$$
$$= H(FR')_\phi - H(R')_\phi \tag{13.3.28}$$
$$= H(FR')_\phi - H(R)_\phi, \tag{13.3.29}$$

where the inequality follows from the data-processing inequality with the partial trace channel Tr_X, and the last equality follows because $\phi_R = \phi_{R'}$, due to the degradability of ρ_{AB}. Finally, observe that

$$\phi_{R'F} = W_{B\to R'F}\rho_B W^\dagger_{B\to R'F}, \tag{13.3.30}$$

so that $H(R'F)_\phi = H(B)_\rho$ by isometric invariance of entropy. Also, $\phi_R = \mathrm{Tr}_{AB}[\psi_{ABR}]$, which implies that $H(R)_\phi = H(AB)_\rho$. We thus obtain

$$I(A'\rangle XB)_\omega \leq I(A\rangle B)_\rho \tag{13.3.31}$$

for every isometry $V_{A\to A'XE}$. This implies that $D^\to(\rho_{AB}) \leq I(A\rangle B)_\rho$, i.e.,

$$D^\to(\rho_{AB}) = I(A\rangle B)_\rho. \tag{13.3.32}$$

In other words, the trivial isometry $V_{A\to A'XE} = \mathbb{1}_A \otimes |0,0\rangle_{XE}$ is optimal for $D^\to(\rho_{AB})$ when ρ_{AB} is a degradable state. Thus, by additivity of coherent information for product states, we obtain $E_D(A;B)_\rho = I(A\rangle B)_\rho$, as required. ■

13.4 Summary

In this chapter, we studied the task of entanglement distillation, in which the goal is for Alice and Bob to convert many copies of a shared entangled state ρ_{AB} to some (smaller) number of ebits, i.e., copies of a two-qubit maximally entangled state using local operations and classical communication. The largest rate at which this can be done, given arbitrarily many copies of ρ_{AB} and such that the error vanishes, is called distillable entanglement, and we denote it by $E_D(A;B)_\rho$. We started with the one-shot setting, in which we allow for some error in the distillation protocol, and we determined both upper and lower bounds on the number of (approximate) ebits that can be obtained. Then, in the asymptotic setting, we showed that the coherent information $I(A\rangle B)_\rho$ is a lower bound on distillable entanglement for every bipartite state ρ_{AB}. We also found that two different entanglement measures are upper bounds on distillable entanglement, namely, the Rains relative entropy and squashed entanglement. These are the best known upper bounds on distillable entanglement.

By first performing entanglement distillation to transform their mixed entangled states to approximately pure maximally entangled states and then performing the quantum teleportation protocol, Alice can transmit any quantum state to Bob. In this sense, entanglement distillation can be used to realize a near-ideal quantum channel between Alice and Bob. This fact underlies achievable strategies for quantum communication, which is the task of perfectly transmitting an arbitrary quantum state from Alice to Bob when the resource is a quantum channel $\mathcal{N}_{A\to B}$ connecting Alice and Bob rather than a shared bipartite state ρ_{AB}. Quantum communication is the subject of the next chapter.

13.5 Bibliographic Notes

Although our focus in this book is on communication, maximally entangled states are useful resources for many other quantum information processing tasks, (see, e.g., Horodecki et al. (2009b) for applications of entanglement in quantum computing), which makes entanglement distillation a relevant topic in its own right.

The concept of entanglement distillation was initially developed by Bennett et al. (1996b,c), who also provided one-way and two-way protocols for distillation from two-qubit Bell-diagonal states. The precise mathematical definition of distillable entanglement was given by Rains (1998, 1999a,b); Horodecki et al. (2000); Plenio and Virmani (2007). Relaxing the set of allowed operations from LOCC channels to separable and completely PPT-preserving channels as a means to obtain upper bounds on distillable entanglement was considered by Rains (1998, 1999a,b).

Berta (2008); Buscemi and Datta (2010b); Brandao and Datta (2011); Wilde et al. (2017) have considered lower bounds on distillable entanglement in the one-shot setting. The lower bound that we present in Proposition 13.10 is the one given by Wilde et al. (2017, Proposition 21), which makes use of the one-shot decoupling results obtained by Dupuis et al. (2014). In particular, the proof of Theorem 13.11 provided in Appendix 13.A comes directly from the proof of (Dupuis et al., 2014, Theorem 3.3). The notion of decoupling has played an important role in the development of quantum information theory. It was originally proposed by Schumacher and Westmoreland (2002) in the context of understanding approximate quantum error correction and quantum communication. It was then developed in much more detail by Horodecki et al. (2005b, 2007) in the context of state merging and by Hayden et al. (2008a) for understanding the coherent information lower bound on quantum capacity. Dupuis (2010) developed the method in more detail in his PhD thesis for a variety of information-processing tasks, and this culminated in the general decoupling theorem presented as (Dupuis et al., 2014, Theorem 3.3).

For the SDP formulations of conditional min- and max-entropy, as well as their smoothed variants, see (Tomamichel, 2015, Chapter 6). For more information about unitary designs and about Haar measure integration over unitaries, we refer to (Collins and Śniady, 2006; Roy and Scott, 2009). The one-shot upper bound that we present in Proposition 13.6 and Theorem 13.7 based on the fact that PPT' operators are useless for entanglement distillation was determined by Tomamichel et al. (2016).

In the asymptotic setting, Devetak and Winter (2005) used random coding arguments to establish the coherent information lower bound (also called the "hashing inequality") on the distillable entanglement of a bipartite quantum state. The corresponding hashing protocol was presented by Bennett et al. (1996c) for two-qubit Bell-diagonal states. Devetak and Winter (2005) also determined that the general expression in Theorem 13.19 is an achievable rate for entanglement distillation from a bipartite state, and they also proved the converse. Horodecki et al. (2000) conjectured this formula earlier, conditioned on the hashing inequality being true.

Theorem 13.26 is due to Devetak and Winter (2005). The other results in Sections 13.2.5 and 13.3.2 on one-way entanglement distillation were obtained by Leditzky et al. (2018), who in the same work used the concepts of approximate degradablity and approximate anti-degradability of bipartite states to derive upper bounds on distillable entanglement. We note that anti-degradable quantum states,

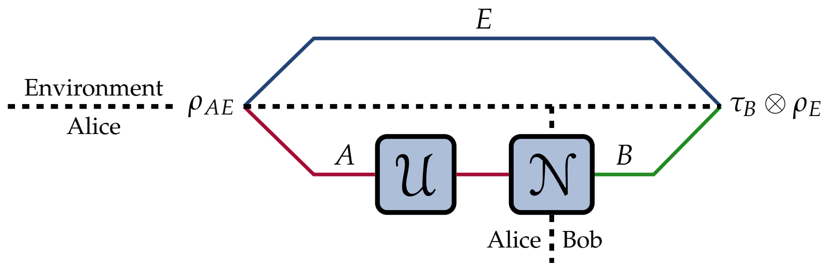


FIGURE 13.4: Given a bipartite state ρ_{AE} and a quantum channel $\mathcal{N}_{A\to B}$, the goal of decoupling is to obtain a state $\tau_B \otimes \rho_E$ that is in tensor product with the environment E, where $\rho_E = \mathrm{Tr}_A[\rho_{AE}]$. To assist with the task, Alice is allowed to apply an arbitrary unitary U to her system A.

as defined by Leditzky et al. (2018), are also known as *symmetrically extendible* states, *two-extendible* states, or *two-shareable* states (Werner, 1989a; Doherty et al., 2004; Yang, 2006).

PPT entangled states (i.e., bound entangled states) were discovered by Horodecki (1997), and Horodecki et al. (1998) showed that PPT states are useless for entanglement distillation. A major open and challenging question, which is alluded to in Section 13.2.0.1 (see also Figure 13.3) is whether there exist NPT (negative partial transpose) bound entangled states. For discussions concerning NPT bound entanglement, we refer to (Horodecki and Horodecki, 1999; DiVincenzo et al., 2000; Dür et al., 2000).

Appendix 13.A One-Shot Decoupling and Proof of Theorem 13.11

The key insight needed to obtain the lower bound on one-shot distillable entanglement in Theorem 13.10 is that entanglement distillation can be thought of in terms of decoupling. The general scenario of decoupling is depicted in Figure 13.4. Given a quantum channel $\mathcal{N}_{A\to B}$ and a bipartite state ρ_{AE}, the goal of decoupling is to obtain a state $\mathcal{N}_{A\to B}(\rho_{AE})$ that is decoupled from the system E, i.e., a state approximately of the form $\tau_B \otimes \rho_E$, where $\rho_E = \mathrm{Tr}_A[\rho_{AE}]$ and τ_B is some state. Note that the reduced state of E at the output is the same as the reduced state of E at the input because we apply a channel only to the system A and such a channel is trace preserving. We also allow Alice an arbitrary unitary that she can apply to her system before sending it through the channel $\mathcal{N}_{A\to B}$, so that we require

$$\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) \approx_\varepsilon \tau_B \otimes \rho_E, \tag{13.A.1}$$

for some state τ_B, up to some error ε. Of course, exact equality in (13.A.1) cannot be obtained in general, and we make the goal instead to obtain an output state $\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger)$ that is as close as possible to the state $\Phi_B^{\mathcal{N}} \otimes \rho_E$, where $\Phi_B^{\mathcal{N}} = \mathrm{Tr}_A[\Phi_{AB}^{\mathcal{N}}]$ and $\Phi_{AB}^{\mathcal{N}}$ is the Choi state of $\mathcal{N}_{A\to B}$. The choice for τ_B given by the reduced Choi state might seem arbitrary, but it is taken for analytical considerations and leads to a good bound for our purposes. By using trace distance as our measure of closeness, the goal is to determine an upper bound on the following quantity:

$$\min_{U_A} \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \tag{13.A.2}$$

Note that the minimum over all unitaries never exceeds the average, meaning that

$$\begin{aligned} \min_{U_A} & \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \\ & \leq \int_{U_A} \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \, \mathrm{d}U_A, \end{aligned} \tag{13.A.3}$$

where the integral over all unitaries U_A is with respect to the Haar measure, and it can be thought of as a uniform average over the continuous set of all unitaries U_A acting on the system A. Theorem 13.11 provides an upper bound on the right-hand side of the inequality above, and we restate the result here for convenience:

$$\int_{U_A} \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \, \mathrm{d}U_A \leq 2^{-\frac{1}{2}\widetilde{H}_2(A|E)_\rho - \frac{1}{2}\widetilde{H}_2(A|B)_{\Phi^{\mathcal{N}}}}, \tag{13.A.4}$$

where we recall that the sandwiched Rényi conditional entropy of order two of a bipartite state ω_{CD} is defined as

$$\widetilde{H}_2(C|D)_\omega = -\inf_{\sigma_D} \widetilde{D}_2(\omega_{CD} \| \mathbb{1}_C \otimes \sigma_D) \tag{13.A.5}$$

$$= -\inf_{\sigma_D} \log_2 \mathrm{Tr}\left[\left(\sigma_D^{-\frac{1}{4}} \omega_{CD} \sigma_D^{-\frac{1}{4}} \right)^2 \right], \tag{13.A.6}$$

and the optimization is over every state σ_D.

Proof of Theorem 13.11

Let

$$M_{BE} := \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E, \tag{13.A.7}$$

$$\sigma_{BE} := \tau_B \otimes \zeta_E, \tag{13.A.8}$$

with τ_B and ζ_E arbitrary positive definite states. By the variational characterization of the trace norm in (2.2.115), we have that

$$\|M_{BE}\|_1 = \max_{U_{BE}} |\text{Tr}[U_{BE} M_{BE}]|\,, \tag{13.A.9}$$

where the optimization is over every unitary U_{BE}. Using the Cauchy–Schwarz inequality (see (2.2.32)), and suppressing system labels for brevity, we obtain

$$\|M\|_1 = \max_U |\text{Tr}[UM]| \tag{13.A.10}$$

$$= \max_U \left|\text{Tr}\left[\left(\sigma^{\frac{1}{4}} U \sigma^{\frac{1}{4}}\right)\left(\sigma^{-\frac{1}{4}} M \sigma^{-\frac{1}{4}}\right)\right]\right| \tag{13.A.11}$$

$$\leq \max_U \sqrt{\text{Tr}\left[\left(\sigma^{\frac{1}{4}} U \sigma^{\frac{1}{4}}\right)\left(\sigma^{\frac{1}{4}} U^\dagger \sigma^{\frac{1}{4}}\right)\right] \text{Tr}\left[\sigma^{-\frac{1}{4}} M \sigma^{-\frac{1}{2}} M^\dagger \sigma^{-\frac{1}{4}}\right]} \tag{13.A.12}$$

$$= \sqrt{\max_U \text{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right] \text{Tr}\left[\sigma^{-\frac{1}{4}} M \sigma^{-\frac{1}{2}} M^\dagger \sigma^{-\frac{1}{4}}\right]}. \tag{13.A.13}$$

Since $\sigma^{\frac{1}{2}}$ and $U\sigma^{\frac{1}{2}}U^\dagger$ are positive definite for every unitary U, by the Cauchy–Schwarz inequality, we conclude that

$$\text{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right] = \left|\text{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right]\right| \tag{13.A.14}$$

$$\leq \sqrt{\text{Tr}\left[\sigma^{\frac{1}{2}} \sigma^{\frac{1}{2}}\right] \text{Tr}\left[U \sigma^{\frac{1}{2}} U^\dagger U \sigma^{\frac{1}{2}} U^\dagger\right]} \tag{13.A.15}$$

$$= \text{Tr}[\sigma] \tag{13.A.16}$$

$$= 1 \tag{13.A.17}$$

for every unitary U, which implies that

$$\max_U \text{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right] \leq \text{Tr}[\sigma] = 1. \tag{13.A.18}$$

On the other hand, by taking $U = \mathbb{1}$ in the optimization over U, we obtain

$$\max_U \text{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right] \geq \text{Tr}[\sigma] = 1, \tag{13.A.19}$$

which means that

$$\max_U \mathrm{Tr}\left[\sigma^{\frac{1}{2}} U \sigma^{\frac{1}{2}} U^\dagger\right] = \mathrm{Tr}[\sigma] = 1. \tag{13.A.20}$$

Therefore,

$$\|M\|_1 \leq \sqrt{\mathrm{Tr}\left[\sigma^{-\frac{1}{4}} M \sigma^{-\frac{1}{2}} M^\dagger \sigma^{-\frac{1}{4}}\right]} \tag{13.A.21}$$

$$= \sqrt{\mathrm{Tr}\left[\left(\sigma^{-\frac{1}{4}} M \sigma^{-\frac{1}{4}}\right)^2\right]}, \tag{13.A.22}$$

where the last line follows because M is Hermitian. So we have that

$$\left\|\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E\right\|_1 \\ \leq \sqrt{\mathrm{Tr}\left[\left((\tau_B \otimes \zeta_E)^{-\frac{1}{4}} (\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E)(\tau_B \otimes \zeta_E)^{-\frac{1}{4}}\right)^2\right]}. \tag{13.A.23}$$

Now, define

$$\widetilde{\mathcal{N}}_{A\to B}(\cdot) := \tau_B^{-\frac{1}{4}} \mathcal{N}_{A\to B}(\cdot) \tau_B^{-\frac{1}{4}}, \tag{13.A.24}$$

$$\widetilde{\rho}_{AE} := \zeta_E^{-\frac{1}{4}} \rho_{AE} \zeta_E^{-\frac{1}{4}}. \tag{13.A.25}$$

Using these definitions, we can write the inequality above as

$$\left\|\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E\right\|_1 \\ \leq \sqrt{\mathrm{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A \widetilde{\rho}_{AE} U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right]}. \tag{13.A.26}$$

Taking the integral over unitaries U_A on both sides of this inequality, and using Jensen's inequality (see (2.3.21)), which applies because the square root function is concave, we obtain

$$\int_{U_A} \left\|\mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E\right\|_1 \, \mathrm{d}U_A \\ \leq \int_{U_A} \sqrt{\mathrm{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A \widetilde{\rho}_{AE} U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right]} \, \mathrm{d}U_A \tag{13.A.27}$$

$$\leq \sqrt{\int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right] \, \mathrm{d}U_A}. \tag{13.A.28}$$

Expanding the integral on the right-hand side of the inequality above leads to

$$\int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right] \, \mathrm{d}U_A$$

$$= \int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^2\right] \, \mathrm{d}U_A \tag{13.A.29}$$

$$- 2\int_{U_A} \operatorname{Tr}\left[\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)(\Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E)\right] \, \mathrm{d}U_A + \operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E)^2] \tag{13.A.30}$$

$$= \int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^2\right] \, \mathrm{d}U_A$$

$$- 2\operatorname{Tr}\left[\widetilde{\mathcal{N}}_{A\to B}\left(\int_{U_A} U_A\widetilde{\rho}_{AE}U_A^\dagger \, \mathrm{d}U_A\right)(\Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E)\right] + \operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E)^2]. \tag{13.A.31}$$

Now, using (13.1.83), we have that

$$\widetilde{\mathcal{N}}_{A\to B}\left(\int_{U_A} U_A\widetilde{\rho}_{AE}U_A^\dagger \, \mathrm{d}U_A\right) = \widetilde{\mathcal{N}}_{A\to B}(\pi_A \otimes \operatorname{Tr}_A[\widetilde{\rho}_{AE}]) = \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E. \tag{13.A.32}$$

Therefore,

$$\int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right] \, \mathrm{d}U_A$$

$$= \int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^2\right] \, \mathrm{d}U_A - \operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2]\operatorname{Tr}[\widetilde{\rho}_E^2]. \tag{13.A.33}$$

We now use the fact that

$$\operatorname{Tr}[X^2] = \operatorname{Tr}[X^{\otimes 2}F] \tag{13.A.34}$$

for every operator X, where F is the swap operator defined in (2.5.12). In (13.A.33) above, we have $X \equiv X_{BE} = \widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)$, which is a bipartite operator. The corresponding swap operator is $F_{BE} = F_B \otimes F_E$, with F_B the swap operator acting on two copies of $\mathcal{H}_B$ and F_E the swap operator acting on two copies of $\mathcal{H}_E$. Therefore,

$$\operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^2\right]$$

$$= \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^{\otimes 2}(F_B\otimes F_E)\right] \tag{13.A.35}$$

$$= \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B}(U_A^{\otimes 2}\widetilde{\rho}^{\otimes 2}_{AE}(U_A^\dagger)^{\otimes 2})\right)(F_B\otimes F_E)\right] \tag{13.A.36}$$

$$= \operatorname{Tr}\left[\widetilde{\rho}^{\otimes 2}_{AE}\left(\left((U_A^\dagger)^{\otimes 2}(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B})^\dagger(F_B)U_A^{\otimes 2}\right)\otimes F_E\right)\right], \tag{13.A.37}$$

where the last line follows from the definition of the adjoint of a channel. We thus have

$$\int_{U_A}\operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A\widetilde{\rho}_{AE}U_A^\dagger)\right)^2\right]\mathrm{d}U_A$$
$$= \operatorname{Tr}\left[\widetilde{\rho}^{\otimes 2}_{AE}\left(\left(\int_{U_A}(U_A^\dagger)^{\otimes 2}(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B})^\dagger(F_B)U_A^{\otimes 2}\,\mathrm{d}U_A\right)\otimes F_E\right)\right]. \tag{13.A.38}$$

Now, we use the following known fact (a standard result in Schur–Weyl duality): for every operator X acting on $\mathbb{C}^d\otimes\mathbb{C}^d$, with $d\geq 1$,

$$\int_U (U^\dagger)^{\otimes 2}XU^{\otimes 2}\,\mathrm{d}U = \alpha\mathbb{1}+\beta F, \tag{13.A.39}$$

where F is again the swap operator, and

$$\alpha = \frac{\operatorname{Tr}[X]}{d^2-1}-\frac{\operatorname{Tr}[XF]}{d(d^2-1)}, \tag{13.A.40}$$

$$\beta = \frac{\operatorname{Tr}[XF]}{d^2-1}-\frac{\operatorname{Tr}[X]}{d(d^2-1)}. \tag{13.A.41}$$

Taking $X\equiv(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B})^\dagger(F_B)$, which is an operator acting on two copies of $\mathcal{H}_A$, we obtain

$$\operatorname{Tr}[(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B})^\dagger(F_B)] = \operatorname{Tr}[\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B}(\mathbb{1}_A^{\otimes 2})F_B] \tag{13.A.42}$$

$$= d_A^2\operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^{\otimes 2}F_B] \tag{13.A.43}$$

$$= d_A^2\operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2], \tag{13.A.44}$$

where the last line follows from (13.A.34). By similar reasoning, we obtain

$$\operatorname{Tr}[F_A(\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B})^\dagger(F_B)] = \operatorname{Tr}[\widetilde{\mathcal{N}}^{\otimes 2}_{A\to B}(F_A)F_B] \tag{13.A.45}$$

$$= d_A^2\operatorname{Tr}[(F_A\otimes F_B)(\Phi_{AB}^{\widetilde{\mathcal{N}}})^{\otimes 2}] \tag{13.A.46}$$

$$= d_A^2 \operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2], \tag{13.A.47}$$

where the second equality follows by expressing the action of $\widetilde{\mathcal{N}}_{A\to B}^{\otimes 2}$ on F_A with the Choi state $\Phi_{AB}^{\widetilde{\mathcal{N}}}$, using (4.2.5). To obtain the last equality, we again used (13.A.34). We thus have

$$\alpha = \frac{\operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2]}{d_A^2 - 1}\left(d_A^2 - d_A \frac{\operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]}{\operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2]}\right), \tag{13.A.48}$$

$$\beta = \frac{\operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]}{d_A^2 - 1}\left(d_A^2 - d_A \frac{\operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2]}{\operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]}\right). \tag{13.A.49}$$

We now make use of the following general fact, whose proof we provide below in Lemma 13.32: for every non-zero positive semi-definite operator P_{AB} with $P_B \coloneqq \operatorname{Tr}_A[P_{AB}]$, the following inequalities hold

$$\frac{1}{d_A} \leq \frac{\operatorname{Tr}[P_{AB}^2]}{\operatorname{Tr}[P_B^2]} \leq d_A. \tag{13.A.50}$$

Applying these inequalities to the expressions for α and β above, we obtain

$$\alpha \leq \operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2], \quad \beta \leq \operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]. \tag{13.A.51}$$

We thus have that

$$\int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A \widetilde{\rho}_{AE} U_A^\dagger)\right)^2\right] \mathrm{d}U_A \tag{13.A.52}$$

$$= \operatorname{Tr}\left[\widetilde{\rho}_{AE}^{\otimes 2}\left((\alpha \mathbb{1}_A^{\otimes 2} + \beta F_A) \otimes F_E\right)\right] \tag{13.A.53}$$

$$= \alpha \operatorname{Tr}[\widetilde{\rho}_E^2] + \beta \operatorname{Tr}[\widetilde{\rho}_{AE}^2] \tag{13.A.54}$$

$$\leq \operatorname{Tr}[(\Phi_B^{\widetilde{\mathcal{N}}})^2]\operatorname{Tr}[\widetilde{\rho}_E^2] + \operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]\operatorname{Tr}[\widetilde{\rho}_{AE}^2]. \tag{13.A.55}$$

Combining this with (13.A.33), we find that

$$\int_{U_A} \operatorname{Tr}\left[\left(\widetilde{\mathcal{N}}_{A\to B}(U_A \widetilde{\rho}_{AE} U_A^\dagger) - \Phi_B^{\widetilde{\mathcal{N}}} \otimes \widetilde{\rho}_E\right)^2\right] \mathrm{d}U_A$$
$$\leq \operatorname{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]\operatorname{Tr}[\widetilde{\rho}_{AE}^2]. \tag{13.A.56}$$

Then, by (13.A.28), and recalling the definitions in (13.A.24) and (13.A.25), we obtain

$$\int_{U_A} \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \, \mathrm{d}U_A$$

$$\leq \sqrt{\mathrm{Tr}[(\Phi_{AB}^{\widetilde{\mathcal{N}}})^2]\mathrm{Tr}[\widetilde{\rho}_{AE}^2]} \tag{13.A.57}$$

$$= \left(\mathrm{Tr}\left[\left(\tau_B^{-\frac{1}{4}} \Phi_{AB}^{\mathcal{N}} \tau_B^{-\frac{1}{4}}\right)^2\right]\right)^{\frac{1}{2}} \left(\mathrm{Tr}\left[\left(\zeta^{-\frac{1}{4}} \rho_{AE} \zeta_E^{-\frac{1}{4}}\right)^2\right]\right)^{\frac{1}{4}}. \tag{13.A.58}$$

This inequality holds for all states τ_B and ζ_E, which means that

$$\int_{U_A} \left\| \mathcal{N}_{A\to B}(U_A \rho_{AE} U_A^\dagger) - \Phi_B^{\mathcal{N}} \otimes \rho_E \right\|_1 \, \mathrm{d}U_A$$

$$\leq \left(\inf_{\tau_B} \mathrm{Tr}\left[\left(\tau_B^{-\frac{1}{4}} \Phi_{AB}^{\mathcal{N}} \tau_B^{-\frac{1}{4}}\right)^2\right]\right)^{\frac{1}{2}} \left(\inf_{\zeta_E} \mathrm{Tr}\left[\left(\zeta_E^{-\frac{1}{4}} \rho_{AE} \zeta_E^{-\frac{1}{4}}\right)^2\right]\right)^{\frac{1}{2}} \tag{13.A.59}$$

$$= 2^{-\frac{1}{2}H_2(A|B)_{\Phi^{\mathcal{N}}} - \frac{1}{2}\widetilde{H}_2(A|E)_\rho} \tag{13.A.60}$$

$$= 2^{-\frac{1}{2}\widetilde{H}_2(A|E)_\rho - \frac{1}{2}H_2(A|B)_{\Phi^{\mathcal{N}}}}, \tag{13.A.61}$$

which completes the proof. ■

Lemma 13.32

For every non-zero positive semi-definite operator P_{AB}, with $P_B = \mathrm{Tr}_A[P_{AB}]$, it holds that

$$\frac{1}{d_A} \leq \frac{\mathrm{Tr}[P_{AB}^2]}{\mathrm{Tr}[P_B^2]} \leq d_A. \tag{13.A.62}$$

PROOF: Letting A' denote a copy of A, and applying the Cauchy–Schwarz inequality (see (2.2.32)), we find that

$$\mathrm{Tr}[P_B^2] = \mathrm{Tr}[(P_{AB} \otimes \mathbb{1}_{A'})(P_{A'B} \otimes \mathbb{1}_A)] \tag{13.A.63}$$

$$\leq \sqrt{\mathrm{Tr}[(P_{AB} \otimes \mathbb{1}_{A'})^2]\mathrm{Tr}[(P_{A'B} \otimes \mathbb{1}_A)^2]} \tag{13.A.64}$$

$$= \mathrm{Tr}[P_{AB}^2 \otimes \mathbb{1}_{A'}] \tag{13.A.65}$$

$$= d_A \mathrm{Tr}[P_{AB}^2]. \tag{13.A.66}$$

The other inequality follows from the operator inequality $P_{AB} \leq d_A \mathbb{1}_A \otimes P_B$, after sandwiching it by $P_{AB}^{\frac{1}{2}}$ and taking a full trace. This operator inequality follows in turn because $\frac{1}{d_A^2} \sum_{i=0}^{d^2-1} U_A^i P_{AB} (U_A^i)^\dagger = \pi_A \otimes P_B$ (Lemma 3.15), for $\{U_A^i\}_i$ the set of Heisenberg–Weyl operators, and by noticing that all terms in the sum are positive semi-definite and one term in the sum is $\frac{1}{d_A^2} P_{AB}$. ■

Chapter 14

Quantum Communication

In the previous chapter, we considered entanglement distillation, which is the task of taking many copies of a mixed entangled state ρ_{AB} shared by Alice and Bob and transforming them to a maximally entangled state $\Phi_{\hat{A}\hat{B}}$ of Schmidt rank $d \geq 2$. Using the quantum teleportation protocol, the maximally entangled state resulting from entanglement distillation can be used for quantum communication, in the sense that Alice can transfer an arbitrary state of $\log_2 d$ qubits to Bob.

Now, if Alice and Bob are distantly separated, then how do they obtain many copies of the shared entangled state ρ_{AB} in the first place? Typically, one of the parties, say Alice, prepares two quantum systems in an entangled state and sends one of them through a quantum channel $\mathcal{N}_{A \to B}$ to Bob, thereby establishing the shared entangled state. Rather than use the shared entangled state as the resource for communication, it is more natural to use the quantum channel itself as the resource, as it could in principle lead to better strategies and higher rates. This is the scenario that we consider in this chapter.

Recall that in the case of classical communication from Chapter 12, we considered messages from a set $\mathcal{M}$, and the goal was to find upper and lower bounds on the maximum number $\log_2 |\mathcal{M}|$ of transmitted bits over a quantum channel for a given error ε. Now, in the case of quantum communication, the goal is to transmit a given number of *qubits*, rather than bits, for a given error ε. Formally, suppose that the sender, Alice, holds a quantum system A' with dimension $d \geq 1$ that she would like to transmit over the channel $\mathcal{N}$ to Bob, the receiver. In general, the state of this system could be entangled with the state of some other system R (of arbitrary dimension) to which Alice does not have access, and so we suppose that

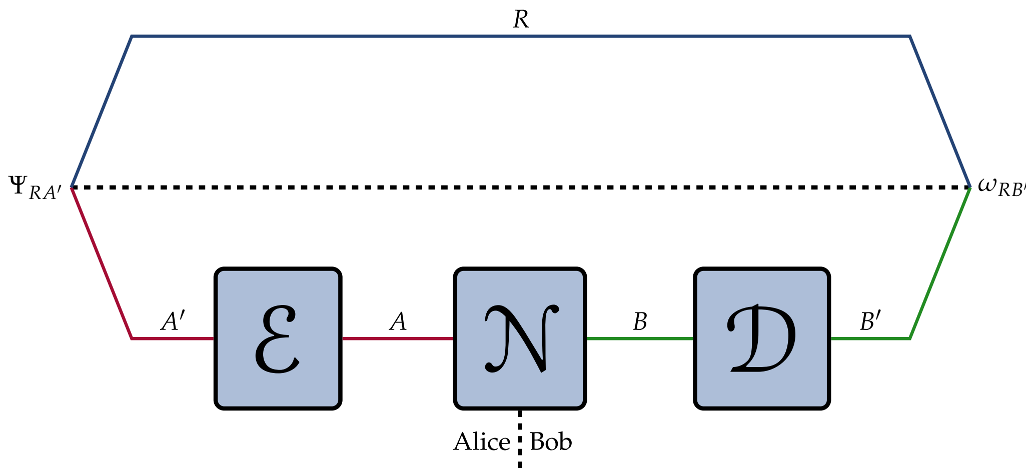


FIGURE 14.1: Depiction of a quantum communication protocol for one use of the quantum channel $\mathcal{N}$. Alice shares the maximally entangled state $\Psi_{RA'}$ with an inaccessible reference system R. She uses the channel $\mathcal{E}_{A'\to A}$ to encode her system A' into a system A, which is sent through the channel $\mathcal{N}_{A\to B}$. Bob then applies the decoding channel $\mathcal{D}_{B\to B'}$, such that the final state is $\omega_{RB'}$ shared between Bob and the reference system.

the joint state is a pure state $\Psi_{RA'}$ with Schmidt rank d. Note that, by the Schmidt decomposition theorem (Theorem 2.2), the dimension of R need not exceed the dimension of A', which is d. The goal is to determine the largest value of $\log_2 d$ (which can be thought of as the number of qubits in the system A') for which the A' part of an arbitrary entangled state $\Psi_{RA'}$ can be transmitted with error at most ε. This general quantum communication scenario is known as *strong subspace transmission*. As usual, Alice and Bob are allowed local encoding and decoding channels, respectively, to help with this task; see Figure 14.1 for a depiction of a one-shot protocol for quantum communication. In the asymptotic setting, they are also allowed as many uses of the channel $\mathcal{N}$ as desired. The quantum capacity of $\mathcal{N}$, denoted by $Q(\mathcal{N})$, is then the largest value of $\frac{1}{n}\log_2 d$ such that the A' part of an arbitrary pure state $\Psi_{RA'}$ can be transmitted to Bob with error that vanishes as the number n of channel uses increases.

Note that the notion of quantum communication presented above (strong subspace transmission) is completely general and includes as special cases the following information-processing tasks:

1. *Entanglement transmission*: Here, Alice's system A' is in the maximally entangled state $\Phi_{RA'}$ with the reference system R, and the goal is to transmit

the system A' to Bob. This is a special case of strong subspace transmission in which $\Psi_{RA'} = \Phi_{RA'}$.

2. *Entanglement generation*: Alice prepares a pure entangled state $\Psi_{A'A}$, with $d_{A'} = d \geq 1$ and A the input system to the channel $\mathcal{N}$. The goal is to transmit the system A to Bob such that the resulting state shared by Alice and Bob is the maximally entangled state of Schmidt rank d. We show in Appendix 14.A how entanglement generation is related to the notion of quantum communication that we consider here.

   Note that entanglement generation is similar to entanglement distillation, and we elaborate more upon this similarity in Section 14.1.3.

3. *Subspace transmission*: In this scenario, Alice wishes to send a system A', in an arbitrary pure state $\varphi_{A'}$, to Bob. This is a special case of the protocol above in which the system R is not entangled with Alice's system, so that $\Psi_{RA'} = \phi_R \otimes \varphi_{A'}$ for some pure state ϕ_R on R.

   Note that subspace transmission can be accomplished by first performing an entanglement transmission or entanglement generation protocol and then performing the quantum teleportation protocol (however, this approach requires the use of a forward classical channel).

We discuss these alternative notions of quantum communication in more detail, as well as prove some relationships between them, in Appendix 14.A.

We start our development of quantum communication with the one-shot setting. Recall that mutual-information channel measures appear in the one-shot upper bounds for entanglement-assisted classical communication and that Holevo-information channel measures appear in the one-shot upper bounds for classical communication. In the case of quantum communication, we find that coherent-information channel measures appear in the one-shot upper bounds. The one-shot lower bound that we obtain is based on the one-shot, one-way entanglement distillation protocol in Section 13.1.2 of the previous chapter, along with an argument for the removal of the classical communication used in that protocol. We then move on to the asymptotic setting, and we find that the quantum capacity of a quantum channel $\mathcal{N}$ is equal to its regularized coherent information, i.e.,

$$Q(\mathcal{N}) = I^c_{\text{reg}}(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} I^c(\mathcal{N}^{\otimes n}), \tag{14.0.1}$$

where we recall the definition of the coherent information $I^c(\mathcal{N})$ of a channel from (7.11.107). Thus, as with the classical capacity, the quantum capacity is difficult to compute in general. We then find tractable upper bounds on the quantum capacity, and for this purpose, the channel entanglement measures defined in Chapter 10 play an important role.

14.1 One-Shot Setting

A (strong subspace) quantum communication protocol for a quantum channel $\mathcal{N}$ in the one-shot setting is illustrated in Figure 14.1. It is defined by the three elements $(d, \mathcal{E}_{A'\to A}, \mathcal{D}_{B\to B'})$, in which d is the dimension of the system A', $\mathcal{E}_{A'\to A}$ is an encoding channel with $d_{A'} = d$, and $\mathcal{D}_{B\to B'}$ is a decoding channel with $d_{B'} = d_{A'} = d$. We call the pair $(\mathcal{E}, \mathcal{D})$ of encoding and decoding channels a *quantum communication code*[1] for $\mathcal{N}$.

Remark: In a strong subspace transmission protocol, the goal is to transmit one share of a pure state $\Psi_{RA'}$, with corresponding state vector $|\Psi\rangle_{RA'}$. Note that the state vector $|\Psi\rangle_{RA'}$ has a Schmidt decomposition of the form

$$|\Psi\rangle_{RA'} = \sum_{x=1}^{d} \sqrt{p(x)}|\xi_x\rangle_R \otimes |\zeta_x\rangle_{A'}, \tag{14.1.1}$$

where $\{p(x)\}_{x=1}^{d}$ are the Schmidt coefficients and $\{|\xi_x\rangle_R\}_{x=1}^{d}$, $\{|\zeta_x\rangle_{A'}\}_{x=1}^{d}$ are orthonormal sets of vectors for R and A', respectively. When written in this form, the state vector $|\Psi\rangle_{RA'}$ can be understood as a coherent version of the initial state $\Phi^{p}_{MM'}$ for classical and entanglement-assisted classical communication (see, e.g., (11.1.2)). The key difference in the classical-communication case is that there is a fixed orthonormal basis $\{|m\rangle\}_{m\in\mathcal{M}}$ corresponding to the messages m in the message set $\mathcal{M}$. In quantum communication, the goal is to transmit a state of a quantum system, which means that there is no particular basis used for communication. The encoding and decoding channels should thus be defined so that they can reliably transmit states of the system in an *arbitrary* basis.

The protocol proceeds as follows: we start with the entangled state $\Psi_{RA'}$, where the system A' belongs to Alice and the system R is an arbitrary reference system inaccessible to Alice. Alice then sends the system A' through the encoding channel $\mathcal{E}_{A'\to A}$ and sends A through the channel $\mathcal{N}_{A\to B}$. Once Bob receives the system B, he applies the decoding channel $\mathcal{D}_{B\to B'}$ to it. The final state of the protocol is

[1] The quantum communication codes that we consider in this chapter are essentially equivalent to codes for performing *approximate quantum error correction*. Please consult the Bibliographic Notes in Section 14.5 for more information.

therefore

$$\omega_{RB'} \coloneqq (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Psi_{RA'}). \tag{14.1.2}$$

Let us now quantify the reliability of the protocol described above, i.e., how close the final state $\omega_{RB'}$ is to the initial state $\Psi_{RA'}$. In Chapter 6, we discussed two measures of closeness for states:

- Normalized trace distance, using which the distance between the initial and final states is $\frac{1}{2}\left\|\Psi_{RA'} - \omega_{RB'}\right\|_1$. The lower the normalized trace distance, the more reliable the protocol is.
- Fidelity, in which case we have

$$F(\Psi_{RA'}, \omega_{RB'}) = \left\|\sqrt{\Psi_{RA'}}\sqrt{\omega_{RB'}}\right\|_1^2 = \langle\Psi|_{RA'}\omega_{RB'}|\Psi\rangle_{RA'} \tag{14.1.3}$$

  as the closeness measure between the initial and final states of the protocol. The higher the fidelity, the more reliable the protocol is.

These two measures of closeness are arguably equivalent to each other, in the sense that one can be used to bound the other via the inequality (6.2.88) shown in Theorem 6.14, which we restate here: for all states ρ and σ,

$$1 - \sqrt{F(\rho,\sigma)} \leq \frac{1}{2}\left\|\rho - \sigma\right\|_1 \leq \sqrt{1 - F(\rho,\sigma)}. \tag{14.1.4}$$

Now, our figure of merit for the quantum communication protocol should not be based on just one particular initial state, in this case $\Psi_{RA'}$. Recall that the task of quantum communication is to reliably transmit one share of an *arbitrary* pure state through the channel $\mathcal{N}_{A\to B}$. Intuitively, therefore, the closer the overall channel $\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}$ is to the identity channel $\mathrm{id}_{A'\to B'}$, the better the code $(\mathcal{E}, \mathcal{D})$ is at the quantum communication task, and so our figure of merit should quantify this distance. One method to determine this distance is to calculate how well a given code can transmit one share of a state in the worst case, i.e., by either the highest value of the trace distance or by the lowest value of the fidelity. If a code can be designed such that, in the worst case, the fidelity (trace distance) is high (low), then by definition any other state will do just as well or better. We are thus led to define the following two figures of merit:

1. *Worst-case trace distance*: We define this as

$$\sup_{\Psi_{RA'}} \frac{1}{2}\left\|\Psi_{RA'} - (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Psi_{RA'})\right\|_1 . \tag{14.1.5}$$

Recalling Definition 6.18, we see that worse-case trace distance is equal to the *diamond distance* between the identity channel $\mathrm{id}_{A'\to B'}$ and the channel $\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}$:

$$\frac{1}{2}\left\|\mathrm{id}_{A'\to B'} - \mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}\right\|_{\diamond}. \tag{14.1.6}$$

2. *Worst-case fidelity*: We define this as

$$\inf_{\Psi_{RA'}} \langle\Psi|_{RA'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Psi_{RA'})|\Psi\rangle_{RA'}, \tag{14.1.7}$$

which is the same as the channel fidelity $F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})$ from Definition 6.22.

These two figures of merit are arguably equivalent, as mentioned before, due to the inequality in (14.1.4) relating the trace distance and the fidelity. For the rest of this chapter, we exclusively use the worst-case fidelity of the code as the figure of merit, and we define the *error probability of the quantum communication code* $(\mathcal{E}, \mathcal{D})$ for $\mathcal{N}$ as

$$\begin{aligned} p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) &\coloneqq 1 - F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) && (14.1.8)\\ &= \sup_{\Psi_{RA'}} \left\{1 - \langle\Psi|_{RA'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Psi_{RA'})|\Psi\rangle_{RA'}\right\}. && (14.1.9) \end{aligned}$$

One can view this quantity as the quantum analogue of the maximum error probability for the classical communication tasks of Chapters 11 and 12, which is why we use the same notation for it as in those chapters.

Definition 14.1 (d, ε) Quantum Communication Protocol

Let $(d, \mathcal{E}, \mathcal{D})$ be the elements of a quantum communication protocol for the quantum channel $\mathcal{N}$. The protocol is called a (d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \le \varepsilon$.

As alluded to at the beginning of this chapter, a special case of interest in quantum communication is entanglement transmission, which is when the state $\Psi_{RA'}$ is fixed to be the maximally entangled state $\Phi_{RA'} = |\Phi\rangle\!\langle\Phi|_{RA'}$, where we recall that

$$|\Phi\rangle_{RA'} = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1} |i, i\rangle_{RA'}. \tag{14.1.10}$$

Now, since this state is a particular state in the optimization in (14.1.9) for the error probability $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N})$, we conclude that

$$p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \geq 1 - \langle\Phi|_{RA'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'})|\Phi\rangle_{RA'} \tag{14.1.11}$$
$$= 1 - F_e(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}) \tag{14.1.12}$$
$$=: \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}), \tag{14.1.13}$$

where in the second line we have identified the entanglement fidelity of the channel $\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}$, as stated in Definition 6.21. In the last line, we have defined the quantity $\overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N})$. As the notation suggests, this quantity is a quantum analogue of the average error probability for classical and entanglement-assisted classical communication. In classical and entanglement-assisted classical communication, the average error probability corresponds to taking a uniform distribution over the messages being sent. Similarly, in quantum communication, the average error probability can be thought of as taking a uniform distribution for the Schmidt coefficients in (14.1.1), which by definition gives a maximally entangled state.

Another way of writing the average error probability for a quantum communication code is via what is known as the *entanglement test*, which we introduced in the previous chapter. It is analogous to the comparator test that we defined in Chapters 11 and 12 in the context of classical communication. The entanglement test is defined by the POVM $\{\Phi_{RB'}, \mathbb{1}_{RB'} - \Phi_{RB'}\}$. The outcomes of the entanglement test tell us whether the state being measured is the maximally entangled state $\Phi_{RB'}$. Since the state $\Phi_{RB'}$ is pure, using (6.2.2), the probability that the state $\omega_{RB'}$ at the end of the protocol is in the maximally entangled state, i.e., the probability that the state "passes the entanglement test," is

$$\text{Tr}[\Phi_{RB'}\omega_{RB'}] = \langle\Phi|_{RB'}\omega_{RB'}|\Phi\rangle_{RB'} \tag{14.1.14}$$
$$= F(\Phi_{RB'}, \omega_{RB'}) \tag{14.1.15}$$
$$= 1 - \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}). \tag{14.1.16}$$

As stated at the beginning of this chapter, the goal of quantum communication is to determine the maximum number $\log_2 d$ of qubits that can be transmitted over a quantum channel $\mathcal{N}$, in the sense that the A' part of an arbitrary pure state $\Psi_{RA'}$, with $d_{A'} = d$, can be transmitted over the channel with error at most $\varepsilon \in (0, 1]$. We call this maximum number of transmitted qubits the *one-shot quantum capacity of* $\mathcal{N}$.

Definition 14.2 One-Shot Quantum Capacity of a Quantum Channel

Given a quantum channel $\mathcal{N}$ and $\varepsilon \in (0,1]$, the *one-shot ε-error quantum capacity of* $\mathcal{N}$, denoted by $Q^{\varepsilon}(\mathcal{N})$, is defined to be the maximum number $\log_2 d$ of transmitted qubits among all (d,ε) quantum communication protocols over $\mathcal{N}$. In other words,

$$Q^{\varepsilon}(\mathcal{N}) := \sup_{(d,\mathcal{E},\mathcal{D})} \{\log_2 d : p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) \leq \varepsilon\}, \tag{14.1.17}$$

where the optimization is with respect to $d \in \mathbb{N}$, $d \geq 1$, encoding channels $\mathcal{E}$ with input system dimension d, and decoding channels $\mathcal{D}$ with output system dimension d.

In addition to finding, for a given $\varepsilon \in (0,1]$, the maximum number of transmitted qubits among all (d,ε) quantum communication protocols over $\mathcal{N}_{A\to B}$, we can consider the following complementary problem: for a given dimension $d \geq 1$, find the smallest possible error among all (d,ε) quantum communication protocols for $\mathcal{N}_{A\to B}$, which we denote by $\varepsilon^*_Q(d;\mathcal{N})$. In other words, the complementary problem is to determine

$$\varepsilon^*_Q(d;\mathcal{N}) := \inf_{\mathcal{E},\mathcal{D}} \{p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) : d_{A'} = d_{B'} = d\}, \tag{14.1.18}$$

where the optimization is with respect to all encoding channels $\mathcal{E}_{A'\to A}$ and decoding channels $\mathcal{D}_{B\to B'}$ such that $d_{A'} = d_{B'} = d$. In this book, we focus primarily on the problem of optimizing the number of transmitted qubits rather than the error, and so our primary quantity of interest is the one-shot quantum capacity $Q^{\varepsilon}(\mathcal{N})$.

14.1.1 Protocol for a Useless Channel

Consider an arbitrary (d,ε) quantum communication protocol for a channel $\mathcal{N}_{A\to B}$, with $\varepsilon \in (0,1]$ and with encoding and decoding channels $\mathcal{E}$ and $\mathcal{D}$, respectively. This means that $p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) \leq \varepsilon$. By the arguments in (14.1.11)–(14.1.13), this protocol realizes a (d,ε) entanglement transmission protocol, in the sense that

$$\text{Tr}[\Phi_{RB'}\rho_{RB'}] \geq 1-\varepsilon, \tag{14.1.19}$$

where $\Phi_{RA'}$ is the maximally entangled state defined in (14.1.10) and

$$\rho_{RB'} := (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'}). \tag{14.1.20}$$

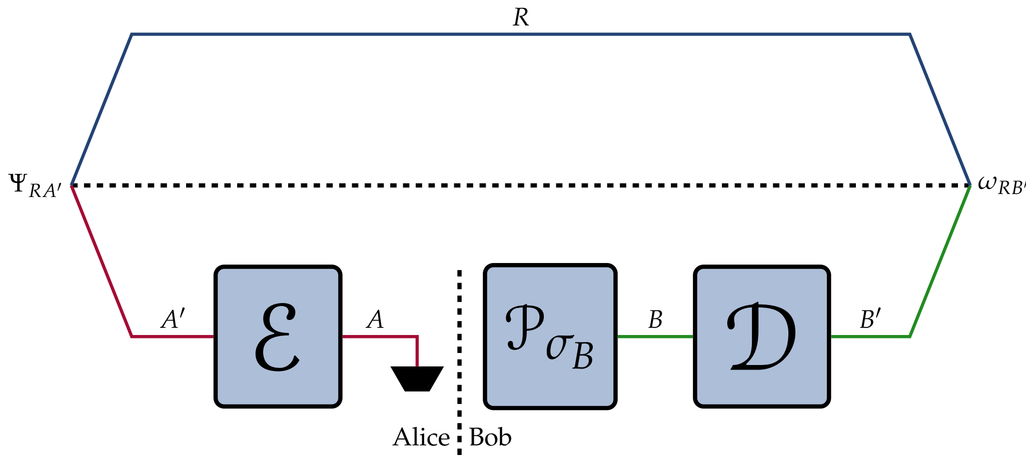


FIGURE 14.2: Depiction of a protocol that is useless for entanglement transmission. The encoded half of Alice's share of the pure state $\Psi_{RA'}$ is discarded and replaced by an arbitrary (but fixed) state σ_B.

Consider now the same protocol but over the useless channel depicted in Figure 14.2. This useless channel is exactly the same as the one considered in Chapters 11 and 12; namely, it is the replacement channel for some state σ_B. For the initial state $\Phi_{RA'}$, the state at the end of the protocol for the replacement channel is

$$\tau_{RB'} = (\mathcal{D}_{B\to B'} \circ \mathcal{R}^{\sigma_B}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'}) = \pi_R \otimes \mathcal{D}_{B\to B'}(\sigma_B). \tag{14.1.21}$$

As in classical communication and entanglement-assisted classical communication, we now use the hypothesis testing relative entropy to compare the state $\rho_{RB'}$ obtained at the end of the quantum communication protocol over the channel $\mathcal{N}$ with the state $\tau_{RB'}$ obtained at the end of the quantum communication protocol for the replacement channel $\mathcal{R}^{\sigma_B}$. In particular, we make use of Lemma 13.3, because the state $\tau_{RB'}$ satisfies $\mathrm{Tr}_{B'}[\tau_{RB}] = \pi_R$ and due to (14.1.19). Therefore, using (13.1.9) in Lemma 13.3, we conclude that

$$\log_2 d \le \frac{1}{2} I_H^{\varepsilon}(R;B')_\rho. \tag{14.1.22}$$

Another bound from Lemma 13.3, namely, the one in (13.1.8), is a more general upper bound that requires only the assumption in (14.1.19) and does not have the interpretation of being a comparison between a quantum communication protocol for $\mathcal{N}$ and a quantum communication protocol for $\mathcal{R}^{\sigma_B}_{A\to B}$. Applying this bound gives

$$\log_2 d \le I_H^{\varepsilon}(R\rangle B')_\rho. \tag{14.1.23}$$

This latter bound is the one that we employ in this chapter because it leads to a formula for the quantum capacity of some channels of interest in applications. This inequality tells us that, given an arbitrary (d, ε) quantum communication protocol with corresponding code $(\mathcal{E}, \mathcal{D})$, the ε-hypothesis testing coherent information $I_H^\varepsilon(R\rangle B')_\rho$, with $\rho_{RB'}$ given by (14.1.20), is an upper bound on the maximum number of qubits that can be transmitted over the channel with error at most ε. Note that a different choice for the encoding and decoding generally produces a different value for the upper bound. We would like an upper bound that applies regardless of the specific protocol. In other words, we would like an upper bound that is a function of the channel $\mathcal{N}_{A\to B}$ only.

14.1.2 Upper Bound on the Number of Transmitted Qubits

We now establish a general upper bound on the number of transmitted qubits in an arbitrary quantum communication protocol. This bound holds independently of the encoding and decoding channels used in the protocol and depends only on the given communication channel $\mathcal{N}_{A\to B}$ and the error ε.

Theorem 14.3 Upper Bound on One-Shot Quantum Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For a (d, ε) quantum communication protocol for $\mathcal{N}_{A\to B}$, with $\varepsilon \in (0, 1]$, the number of qubits transmitted over $\mathcal{N}$ is bounded from above by the ε-hypothesis testing coherent information of $\mathcal{N}$ defined in (7.11.96), i.e.,

$$\log_2 d \le I_H^{c,\varepsilon}(\mathcal{N}). \tag{14.1.24}$$

Consequently, for the one-shot quantum capacity of $\mathcal{N}$,

$$Q^\varepsilon(\mathcal{N}) \le I_H^{c,\varepsilon}(\mathcal{N}). \tag{14.1.25}$$

PROOF: Let $\mathcal{E}$ and $\mathcal{D}$ be the encoding and decoding channels, respectively, for a (d, ε) quantum communication protocol for $\mathcal{N}$. Then, by (14.1.23), we have that

$$\log_2 d \le I_H^\varepsilon(R\rangle B')_\rho = \inf_{\sigma_{B'}} D_H^\varepsilon(\rho_{RB'} \| \mathbb{1}_R \otimes \sigma_{B'}), \tag{14.1.26}$$

where $\rho_{RB'} = (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'})$ is the state defined in (14.1.20). By restricting the optimization in the definition of $I_H^\varepsilon(R\rangle B')_\rho$ over every state $\sigma_{B'}$

to the set $\{\mathcal{D}_{B\to B'}(\tau_B) : \tau_B \in \mathrm{D}(\mathcal{H}_B)\}$, we obtain

$$\begin{aligned} I_H^\varepsilon(R\rangle B')_\rho \\ &= \inf_{\sigma_{B'}} D_H^\varepsilon(\rho_{RB'}\|\mathbb{1}_R\otimes\sigma_{B'}) &(14.1.27)\\ &\le \inf_{\tau_B} D_H^\varepsilon((\mathcal{D}_{B\to B'}\circ\mathcal{N}_{A\to B}\circ\mathcal{E}_{A'\to A})(\Phi_{RA'})\|\mathbb{1}_R\otimes\mathcal{D}_{B\to B'}(\tau_B)) &(14.1.28)\\ &\le \inf_{\tau_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathbb{1}_R\otimes\tau_B) &(14.1.29)\end{aligned}$$

where the second inequality follows from the data-processing inequality for hypothesis testing relative entropy and we let $\rho_{RA} := \mathcal{E}_{A'\to A}(\Phi_{RA'})$. We now take the supremum over every state ρ_{RA}, which effectively corresponds to taking the supremum over all encoding channels, and since it suffices to consider only pure states when optimizing the coherent information (see the arguments after Definition 7.85), we conclude that

$$\begin{aligned} I_H^\varepsilon(R\rangle B')_\rho &\le \inf_{\tau_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathbb{1}_R\otimes\tau_B) &(14.1.30)\\ &\le \sup_{\psi_{RA}}\inf_{\tau_B} D_H^\varepsilon(\mathcal{N}_{A\to B}(\psi_{RA})\|\mathbb{1}_R\otimes\tau_B) &(14.1.31)\\ &= I_H^{c,\varepsilon}(\mathcal{N}), &(14.1.32)\end{aligned}$$

as required. ■

As an immediate consequence of Theorem 14.3 and Propositions 7.70 and 7.71, we obtain the following:

Corollary 14.4

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon\in[0,1)$. For all (d,ε) quantum communication protocols for $\mathcal{N}$, the following bounds hold:

$$(1-2\varepsilon)\log_2 d \le I^c(\mathcal{N}) + h_2(\varepsilon), \tag{14.1.33}$$

$$\log_2 d \le \widetilde{I}_\alpha^c(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\,\alpha>1, \tag{14.1.34}$$

where $I^c(\mathcal{N})$ is the coherent information of $\mathcal{N}$, as defined in (7.11.107), and $\widetilde{I}_\alpha^c(\mathcal{N})$ is the sandwiched Rényi coherent information of $\mathcal{N}$, as defined in (7.11.100).

The proof of (14.1.33) is analogous to the proof of (13.1.44). The proof of (14.1.34) follows by combining Theorem 14.3 with Proposition 7.71.

Since the bounds in (14.1.33) and (14.1.34) hold for an arbitrary (d, ε) quantum communication protocol for $\mathcal{N}$, we have that

$$(1 - 2\varepsilon)Q^{\varepsilon}(\mathcal{N}) \leq I^{c}(\mathcal{N}) + h_2(\varepsilon), \tag{14.1.35}$$

$$Q^{\varepsilon}(\mathcal{N}) \leq \widetilde{I}^{c}_{\alpha}(\mathcal{N}) + \frac{\alpha}{\alpha - 1}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{14.1.36}$$

for all $\varepsilon \in [0, 1)$.

Let us summarize the steps that we took to arrive at the bounds in (14.1.33) and (14.1.34):

1. We first compared a quantum communication protocol for $\mathcal{N}$ with the same protocol for a useless channel by using the hypothesis testing relative entropy. This led us to Lemma 13.3, and the resulting upper bound in (14.1.23).

2. We then used the data-processing inequality for the hypothesis testing relative entropy to remove the decoding channel from the bound in (14.1.23). This is done in (14.1.29) in the proof of Theorem 14.3.

3. Finally, we optimized over all encoding channels in (14.1.30)–(14.1.32) to obtain Theorem 14.3, in which the bound is a function solely of the channel and the error probability. Using Propositions 7.70 and 7.71, which relate hypothesis testing relative entropy to quantum relative entropy and sandwiched Rényi relative entropy, we arrived at Corollary 14.4.

14.1.3 Lower Bound on the Number of Transmitted Qubits via Entanglement Distillation

Having derived upper bounds on the number of transmitted qubits for an arbitrary quantum communication protocol, let us now determine a lower bound on the number of transmitted qubits. As with the other communication scenarios that we have considered so far, in order to obtain a lower bound on the number qubits that can be transmitted, we need to devise an explicit (d, ε) quantum communication protocol for all $\varepsilon \in (0, 1)$. The protocol we consider is based on the one-shot, one-way entanglement distillation protocol from Proposition 13.10 in Chapter 13, which establishes that, for an arbitrary bipartite state ρ_{AB} and for all $\varepsilon \in (0, 1]$ and $\eta \in [0, \sqrt{\varepsilon})$, there exists a (d, ε) one-way entanglement distillation protocol for ρ_{AB} such that

$$\log_2 d = -H_{\max}^{\sqrt{\varepsilon}-\eta}(A|B)_\rho + 4 \log_2 \eta. \tag{14.1.37}$$

The goal in this section is to show that entanglement distillation can be used to develop a quantum communication strategy. Specifically, we show that the existence of a (d, ε) one-way entanglement distillation protocol for the bipartite state $\omega_{AB} = \mathcal{N}_{A' \to B}(\psi_{AA'})$ implies the existence of a (d', ε') quantum communication protocol, with d' and ε' being functions of d and ε. The claim is as follows:

Theorem 14.5 Lower Bound on One-Shot Quantum Capacity

Let $\mathcal{N}_{A \to B}$ be a quantum channel. For all $\varepsilon \in (0, 1)$, $\eta \in [0, \varepsilon\sqrt{\delta}/4)$, and $\delta \in (0, 1)$, there exists a (d, ε) quantum communication protocol for $\mathcal{N}_{A \to B}$ such that

$$\log_2 d = \sup_{\psi_{AA'}} \left(-H_{\max}^{\frac{\varepsilon\sqrt{\delta}}{4}-\eta}(A|B)_\omega \right) + \log_2(\eta^4(1 - \delta)), \tag{14.1.38}$$

where $\omega_{AB} = \mathcal{N}_{A' \to B}(\psi_{AA'})$. Consequently,

$$Q^\varepsilon(\mathcal{N}) \geq \sup_{\psi_{AA'}} \left(-H_{\max}^{\frac{\varepsilon\sqrt{\delta}}{4}-\eta}(A|B)_\omega \right) + \log_2(\eta^4(1 - \delta)) \tag{14.1.39}$$

for all $\eta \in [0, \varepsilon\sqrt{\delta}/4)$ and $\delta \in (0, 1)$, where $\omega_{AB} = \mathcal{N}_{A' \to B}(\psi_{AA'})$.

The first step in the proof of Theorem 14.5 is to observe that one-way entanglement distillation is an example of entanglement generation, albeit with forward (i.e., sender to receiver) classical communication, which we introduced at the beginning of this chapter and formally define below. We then show that forward classical communication does not help for entanglement generation, even in the non-asymptotic setting. One-way entanglement distillation thus implies entanglement generation. We then show that entanglement generation implies entanglement transmission, which we defined at the beginning of this chapter. Finally, we show that entanglement transmission implies quantum communication.

Before proceeding with the proof of Theorem 14.5, let us formally define entanglement generation (with and without one-way LOCC assistance) and entanglement transmission.

- *Entanglement generation*: An entanglement generation protocol for $\mathcal{N}_{A\to B}$ is defined by the three elements $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$, where $\Psi_{A'A}$ is a pure state with $d_{A'} = d$, and $\mathcal{D}_{B\to B'}$ is a decoding channel with $d_{B'} = d$. The goal of the protocol is to transmit the system A such that the final state

$$\sigma_{A'B'} \coloneqq (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B})(\Psi_{A'A}) \tag{14.1.40}$$

  is close in fidelity to a maximally entangled state of Schmidt rank d. The *entanglement generation error* of the protocol is given by

$$p_{\text{err}}^{(\text{EG})}(\Psi_{A'A}, \mathcal{D}; \mathcal{N}) \coloneqq 1 - \langle\Phi|_{A'B'}\sigma_{A'B'}|\Phi\rangle_{A'B'} \tag{14.1.41}$$

$$= 1 - F(\Phi_{A'B'}, \sigma_{A'B'}). \tag{14.1.42}$$

  We call the protocol $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$ a (d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{EG})}(\Psi_{A'A}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

  Note that an entanglement generation protocol $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$ over $\mathcal{N}_{A\to B}$ is an example of an entanglement distillation protocol $(d, \mathcal{L}_{AB\to\hat{A}\hat{B}})$ for the state $\rho_{A'B} = \mathcal{N}_{A\to B}(\Psi_{A'A})$, with $\hat{A} \equiv A'$, $\hat{B} \equiv B'$, and $\mathcal{L}_{AB\to\hat{A}\hat{B}} \equiv \mathcal{D}_{B\to B'}$.

- *Entanglement generation assisted by one-way LOCC*: An entanglement generation protocol for $\mathcal{N}_{A\to B}$ assisted by one-way LOCC from A to B is defined by $(d, \Psi_{A'A}, \{\mathcal{E}^x_{A'A\to A'A}\}_x, \{\mathcal{D}^x_{B\to B'}\}_x)$, where $d \geq 1$, $\Psi_{A'A}$ is a pure state with $d_{A'} = d$, $\{\mathcal{E}^x_{A'A\to A'A}\}_{x\in\mathcal{X}}$ is a set of completely positive maps indexed by a finite alphabet $\mathcal{X}$ such that $\sum_{x\in\mathcal{X}} \mathcal{E}^x_{A'A\to A'A}$ is trace preserving, and $\{\mathcal{D}^x_{B\to B'}\}_{x\in\mathcal{X}}$ is a set of quantum channels indexed by $\mathcal{X}$, with $d_{B'} = d$. The goal of the protocol is to transmit the system A such that the final state

$$\sigma^{\to}_{A'B'} \coloneqq \sum_{x\in\mathcal{X}} (\mathcal{D}^x_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}^x_{A'A\to A'A})(\Psi_{A'A}) \tag{14.1.43}$$

  is close in fidelity to a maximally entangled state of Schmidt rank d. The error of the protocol is given by

$$p_{\text{err}}^{(\text{EG}),\to}(\Psi_{A'A}, \{\mathcal{E}^x\}_x, \{\mathcal{D}^x\}_x; \mathcal{N}) = 1 - F(\Phi_{A'B'}, \sigma^{\to}_{A'B'}). \tag{14.1.44}$$

  We call the protocol $(d, \Psi_{A'A}, \{\mathcal{E}^x_{A'A\to A'A}\}_x, \{\mathcal{D}^x_{B\to B'}\}_x)$ a (d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{EG}),\to}(\Psi_{A'A}, \{\mathcal{E}^x\}_x, \{\mathcal{D}^x\}_x; \mathcal{N}) \leq \varepsilon$.

- *Entanglement transmission*: An entanglement transmission protocol for $\mathcal{N}_{A\to B}$ consists of the three elements $(d, \mathcal{E}, \mathcal{D})$, where $d \geq 1$, $\mathcal{E}_{A'\to A}$ is an encoding channel with $d_{A'} = d$, and $\mathcal{D}_{B\to B'}$ is a decoding channel with $d_{B'} = d$. The goal of the protocol is to transmit the A' system of a maximally entangled state $\Phi_{RA'}$ of Schmidt rank d such that the final state

$$\omega_{RB'} := (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'}) \tag{14.1.45}$$

is close to the initial maximally entangled state. The *entanglement transmission error* of the protocol is

$$p_{\text{err}}^{(\text{ET})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) := 1 - \langle\Phi|_{RB'}\omega_{RB'}|\Phi\rangle_{RB'} \tag{14.1.46}$$

$$= 1 - F_e(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}), \tag{14.1.47}$$

where we recall the entanglement fidelity of a channel from Definition 6.21. We call the protocol $(d, \mathcal{E}, \mathcal{D})$ a (d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{ET})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

Observe that the error criterion for entanglement transmission is the same as the average error criterion for quantum communication (see (14.1.11)–(14.1.13)). This means that the existence of an arbitrary (d, ε) quantum communication protocol implies the existence of a (d, ε) entanglement transmission protocol. Also observe that the existence of an arbitrary (d, ε) entanglement transmission protocol implies the existence of a (d, ε) entanglement generation protocol with respect to the state $\Psi_{RA} \equiv \mathcal{E}_{A'\to A}(\Phi_{RA'})$ (with the systems R, A, and A' belonging to Alice).

14.1.3.1 Proof of Theorem 14.5

We start by showing that an arbitrary entanglement distillation protocol for the state $\mathcal{N}_{A\to B}(\psi_{A'A})$, with $\psi_{A'A}$ a pure state, has the same performance parameters as an entanglement generation protocol with one-way LOCC assistance.

Consider an arbitrary (d, ε) entanglement distillation protocol for $\mathcal{N}_{A\to B}(\psi_{A'A})$ given by a one-way LOCC channel $\mathcal{L}_{A'B\to\hat{A}\hat{B}}$, with $d_{\hat{A}} = d_{\hat{B}} = d$. In general, this LOCC channel has the form $\mathcal{L}_{A'B\to\hat{A}\hat{B}} = \sum_{x\in\mathcal{X}} \mathcal{E}^x_{A'\to\hat{A}} \otimes \mathcal{D}^x_{B\to\hat{B}}$, where $\mathcal{X}$ is some finite alphabet, $\{\mathcal{E}^x_{A'\to\hat{A}}\}_{x\in\mathcal{X}}$ is a set of completely positive maps such that $\sum_{x\in\mathcal{X}} \mathcal{E}^x_{A'\to\hat{A}}$ is trace preserving, and $\{\mathcal{D}^x_{B\to\hat{B}}\}_{x\in\mathcal{X}}$ is a set of channels. The output state of the entanglement distillation protocol is

$$\sum_{x\in\mathcal{X}}(\mathcal{E}^x_{A'\to\hat{A}}\otimes\mathcal{D}^x_{B\to\hat{B}})(\mathcal{N}_{A\to B}(\psi_{A'A}))$$
$$=\sum_{x\in\mathcal{X}}(\mathcal{D}^x_{B\to\hat{B}}\circ\mathcal{N}_{A\to B}\circ\mathcal{E}^x_{A'\to\hat{A}})(\psi_{A'A}),\quad(14.1.48)$$

which has the form of a state at the output of an entanglement generation protocol with one-way LOCC assistance. We thus have that a (d,ε) entanglement distillation protocol for $\mathcal{N}_{A\to B}(\psi_{A'A})$ is equivalent to a (d,ε) entanglement generation protocol for $\mathcal{N}$ with one-way LOCC assistance. We now show that one-way LOCC assistance does not help for entanglement generation.

Lemma 14.6

Given a (d,ε) entanglement generation protocol for a channel $\mathcal{N}$, assisted by one-way LOCC, with $d\geq 1$ and $\varepsilon\in[0,1]$, there exists a (d,ε) entanglement generation protocol for $\mathcal{N}$ (without one-way LOCC assistance).

Proof: Consider an arbitrary (d,ε) entanglement generation protocol assisted by one-way LOCC. The output state of such a protocol has the form in (14.1.43), i.e.,

$$\sigma^{\rightarrow}_{A'B'}=\sum_{x\in\mathcal{X}}(\mathcal{D}^x_{B\to B'}\circ\mathcal{N}_{A\to B}\circ\mathcal{E}^x_{A'A\to A'A})(\Psi_{A'A}),\quad(14.1.49)$$

and by definition we have

$$F(\Phi_{A'B'},\sigma^{\rightarrow}_{A'B'})=\operatorname{Tr}[\Phi_{A'B'}\sigma^{\rightarrow}_{A'B'}]\geq 1-\varepsilon.\quad(14.1.50)$$

Now, let

$$p(x):=\operatorname{Tr}[\mathcal{E}^x_{A'A\to A'A}(\Psi_{A'A})],\quad(14.1.51)$$
$$\rho^x_{A'A}:=\frac{1}{p(x)}\mathcal{E}^x_{A'A\to A'A}(\Psi_{A'A}).\quad(14.1.52)$$

Using this, we can write $\sigma^{\rightarrow}_{A'B'}$ as

$$\sigma^{\rightarrow}_{A'B'}=\sum_{x\in\mathcal{X}}p(x)(\mathcal{D}^x_{B\to B'}\circ\mathcal{N}_{A\to B})(\rho^x_{A'A}).\quad(14.1.53)$$

For every $x\in\mathcal{X}$, let $\rho^x_{A'A}$ have the following spectral decomposition:

$$\rho^x_{A'A}=\sum_{k=1}^{r_x}q(k|x)\phi^{x,k}_{A'A},\quad(14.1.54)$$

where $r_x = \operatorname{rank}(\rho^x_{A'A})$. We thus have that

$$\sigma^{\to}_{A'B'} = \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_x} p(x)q(k|x)(\mathcal{D}^x_{B\to B'} \circ \mathcal{N}_{A\to B})(\phi^{x,k}_{A'A}). \tag{14.1.55}$$

Then, letting $\sigma^{x,k}_{A'B'} \coloneqq (\mathcal{D}^x_{B\to B'} \circ \mathcal{N}_{A\to B})(\phi^{x,k}_{A'A})$, we conclude that

$$\operatorname{Tr}[\Phi_{A'B'}\sigma^{\to}_{A'B'}] = \sum_{x\in\mathcal{X}} \sum_{k=1}^{r_x} p(x)q(k|x)\operatorname{Tr}[\Phi_{A'B'}\sigma^{x,k}_{A'B'}] \tag{14.1.56}$$

$$\leq \max_{\substack{x\in\mathcal{X},\\ 1\leq k\leq r_x}} \operatorname{Tr}[\Phi_{A'B'}\sigma^{x,k}_{A'B'}] \tag{14.1.57}$$

$$= \max_{\substack{x\in\mathcal{X},\\ 1\leq k\leq r_x}} \operatorname{Tr}[\Phi_{A'B'}(\mathcal{D}^x_{B\to B'} \circ \mathcal{N}_{A\to B})(\phi^{x,k}_{A'A})]. \tag{14.1.58}$$

In other words, there exists a pair $(\phi^{x,k}_{A'A}, \mathcal{D}^x_{B\to B'})$ (namely, the one that achieves the maximum on the right-hand side of the inequality above) such that

$$\operatorname{Tr}[\Phi_{A'B'}\sigma^{x,k}_{A'B'}] \geq \operatorname{Tr}[\Phi_{A'B'}\sigma_{A'B'}] \geq 1-\varepsilon. \tag{14.1.59}$$

Therefore, the triple $(d, \phi^{k,x}_{A'A}, \mathcal{D}^x_{B\to B'})$ constitutes a (d, ε) entanglement generation protocol (without one-way LOCC assistance). ■

As we have seen in Chapter 13, forward classical communication certainly helps in general for entanglement distillation. However, it does not help for entanglement generation because the resource for entanglement generation is a quantum channel, whereas for entanglement distillation the resource is a bipartite quantum state. Having a quantum channel as the resource is more powerful than having a bipartite quantum state because, when using a quantum channel, there is an extra degree of freedom in the input state to the channel. The proof of Lemma 14.6 demonstrates that the forward classical communication in an arbitrary (d, ε) is not needed, and the proof essentially relies on convexity of the entanglement fidelity performance criterion.

We now show that entanglement generation implies entanglement transmission, up to a transformation of the performance parameters.

Lemma 14.7 Entanglement Generation to Entanglement Transmission

Given a (d, ε) entanglement generation protocol for a channel $\mathcal{N}_{A\to B}$, with $d \geq 1$ and $\varepsilon \in [0, 1]$, there exists a $(d, 4\varepsilon)$ entanglement transmission protocol for $\mathcal{N}$.

Proof: Let $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$ be the elements of a (d, ε) entanglement generation protocol for $\mathcal{N}_{A\to B}$, with $d_{A'} = d_{B'} = d$. This implies that the output state

$$\sigma_{A'B'} = (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B})(\Psi_{A'A}) \tag{14.1.60}$$

satisfies

$$F(\Phi_{A'B'}, \sigma_{A'B'}) \geq 1 - \varepsilon. \tag{14.1.61}$$

We now construct an entanglement transmission protocol. To this end, let $A' \equiv R$ be a reference system inaccessible to both Alice and Bob. By the data-processing inequality for fidelity (Theorem 6.9) with respect to the partial trace channel $\mathrm{Tr}_{B'}$, we have that

$$F(\Phi_R, \Psi_R) = F(\mathrm{Tr}_{B'}[\Phi_{RB'}], \mathrm{Tr}_{B'}[\sigma_{RB'}]) \geq F(\Phi_{RB'}, \sigma_{RB'}) \geq 1 - \varepsilon. \tag{14.1.62}$$

Next, by Uhlmann's theorem (Theorem 6.8), there exists an isometric channel $\mathcal{U}_{A'\to A}$ such that

$$F(\Phi_R, \Psi_R) = F(\mathcal{U}_{A'\to A}(\Phi_{RA'}), \Psi_{RA'}) \geq 1 - \varepsilon. \tag{14.1.63}$$

We let this isometric channel $\mathcal{U}_{A'\to A}$ be the encoding channel for the entanglement transmission protocol, and we let

$$\omega_{RB'} = (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{U}_{A'\to A})(\Phi_{RA'}). \tag{14.1.64}$$

Next, using the sine distance (Definition 6.16), by definition of the (d, ε) entanglement generation protocol, we have that $P(\Phi_{RB'}, \sigma_{RB'}) \leq \sqrt{\varepsilon}$. Similarly, from (14.1.63) we have that

$$P(\Psi_{RA}, \mathcal{U}_{A'\to A}(\Phi_{RA'})) \leq \sqrt{\varepsilon}. \tag{14.1.65}$$

Therefore, by the triangle inequality for the sine distance (Lemma 6.17), we conclude that

$$P(\Phi_{RB'}, \omega_{RB'}) \leq P(\Phi_{RB'}, \sigma_{RB'}) + P(\sigma_{RB'}, \omega_{RB'}) \tag{14.1.66}$$

$$\leq \sqrt{\varepsilon} + \sqrt{\varepsilon} \tag{14.1.67}$$
$$\leq 2\sqrt{\varepsilon}, \tag{14.1.68}$$

where the second inequality follows from the data-processing inequality for sine distance (see (6.2.114)) and (14.1.65) to see that

$$P(\sigma_{RB'}, \omega_{RB'}) \tag{14.1.69}$$
$$= P((\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B})(\Psi_{RA}), (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{U}_{A'\to A})(\Phi_{RA'})) \tag{14.1.70}$$
$$\leq P(\Psi_{RA}, \mathcal{U}_{A'\to A}(\Phi_{RA'})) \tag{14.1.71}$$
$$\leq \sqrt{\varepsilon}. \tag{14.1.72}$$

Therefore, by definition of the sine distance, we conclude that

$$1 - F(\Phi_{RB'}, \omega_{RB'}) \leq 4\varepsilon, \tag{14.1.73}$$

so that $(d, \mathcal{U}_{A'\to A}, \mathcal{D}_{B\to B'})$ constitutes a $(d, 4\varepsilon)$ entanglement transmission protocol, as required. ■

Finally, we show that entanglement transmission implies quantum communication, up to a transformation of the performance parameters. We could alternatively call this statement "quantum expurgation," because the arguments in the proof are analogous to the expurgation arguments applied in the proof of the lower bound for one-shot classical communication in Proposition 12.5.

Lemma 14.8 Entanglement Transmission to Quantum Communication

Given a (d, ε) entanglement transmission protocol for a channel $\mathcal{N}_{A\to B}$, with $d \geq 1$ and $\varepsilon \in [0,1]$, for all $\delta \in (0,1)$, there exists a $(\lfloor (1-\delta)d \rfloor, 2\sqrt{\varepsilon/\delta})$ quantum communication protocol for $\mathcal{N}$.

Proof: Suppose that a (d, ε) entanglement transmission code for $\mathcal{N}_{A\to B}$ exists, and let $\mathcal{E}_{A'\to A}$ and $\mathcal{D}_{B\to B'}$ be the corresponding encoding and decoding channels, respectively, with $d_{A'} = d_{B'} = d$. The condition $\overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ then holds, namely,

$$1 - \text{Tr}[\Phi_{RB'}\omega_{RB'}] \leq \varepsilon, \tag{14.1.74}$$

where

$$\omega_{RB'} = (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'}). \tag{14.1.75}$$

Let

$$\mathcal{C}_{A'\to B'} := \mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A}. \tag{14.1.76}$$

We proceed with the following algorithm:

1. Set $k = d$ and $\mathcal{H}_d = \mathcal{H}_{A'}$. Suppose for now that $(1-\delta)d$ is a positive integer.

2. Set $|\phi_k\rangle \in \mathcal{H}_k$ to be a state vector that achieves the minimum fidelity of $\mathcal{C}_{A'\to B'}$:

$$|\phi_k\rangle := \arg\min_{|\phi\rangle\in\mathcal{H}_k} \langle\phi|\mathcal{C}_{A'\to B'}(|\phi\rangle\langle\phi|)|\phi\rangle, \tag{14.1.77}$$

   and set the fidelity F_k of $|\phi_k\rangle$ as follows:

$$F_k := \min_{|\phi\rangle\in\mathcal{H}_k} \langle\phi|\mathcal{C}_{A'\to B'}(|\phi\rangle\langle\phi|)|\phi\rangle \tag{14.1.78}$$

$$= \langle\phi_k|\mathcal{C}_{A'\to B'}(|\phi_k\rangle\langle\phi_k|)|\phi_k\rangle. \tag{14.1.79}$$

3. Set

$$\mathcal{H}_{k-1} := \operatorname{span}\{|\psi\rangle \in \mathcal{H}_k : |\langle\psi|\phi_k\rangle| = 0\}. \tag{14.1.80}$$

   That is, $\mathcal{H}_{k-1}$ is set to the orthogonal complement of $|\phi_k\rangle$ in $\mathcal{H}_k$, so that $\mathcal{H}_k = \mathcal{H}_{k-1} \oplus \operatorname{span}\{|\phi_k\rangle\}$. Set $k \to k-1$.

4. Repeat steps 2-3 until $k = (1-\delta)d$ after step 3.

The idea behind this algorithm is to successively remove minimum fidelity states from $\mathcal{H}_{A'}$ until $k = (1-\delta)d$. By the structure of the algorithm and some analysis given below, we are then guaranteed that for this k and lower that

$$1 - \min_{|\phi\rangle\in\mathcal{H}_k} \langle\phi|\mathcal{C}(|\phi\rangle\langle\phi|)|\phi\rangle \leq \varepsilon/\delta. \tag{14.1.81}$$

That is, the subspace $\mathcal{H}_k$ is good for quantum communication of states at the channel input with fidelity at least $1-\varepsilon/\delta$ (to be precise, the subspace $\mathcal{H}_k$ is good for subspace transmission as defined in the introduction of this chapter). Furthermore, the algorithm implies that

$$F_d \leq F_{d-1} \leq \cdots \leq F_{(1-\delta)d}, \tag{14.1.82}$$

$$\mathcal{H}_d \supseteq \mathcal{H}_{d-1} \supseteq \cdots \supseteq \mathcal{H}_{(1-\delta)d}. \tag{14.1.83}$$

Also, $\{|\phi_k\rangle\}_{k=1}^{\ell}$ is an orthonormal basis for $\mathcal{H}_\ell$, where $\ell \in \{1,\ldots,d\}$. Note that the unit vectors $|\phi_k\rangle$, $k \in \{(1-\delta)d-1,\ldots,1\}$ can be generated by repeating the algorithm above exhaustively.

We now analyze the claims above by employing Markov's inequality and some other tools. From (14.1.74), we have that

$$F(\Phi_{RB'}, \mathcal{C}_{A'\to B'}(\Phi_{RA'})) \geq 1 - \varepsilon. \tag{14.1.84}$$

Since $\{|\phi_k\rangle\}_{k=1}^d$ is an orthonormal basis for $\mathcal{H}_d$, we can write

$$|\Phi\rangle_{RA'} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |\overline{\phi_k}\rangle_R \otimes |\phi_k\rangle_{A'}, \tag{14.1.85}$$

where complex conjugation is taken with respect to the basis $\{|i\rangle\}_{i=0}^{d-1}$ used in (14.1.10). A consequence of the data-processing inequality for fidelity under the dephasing channel $\omega_R \mapsto \sum_{k=1}^d |\overline{\phi_k}\rangle\langle\overline{\phi_k}|_R \omega_R |\overline{\phi_k}\rangle\langle\overline{\phi_k}|_R$ and convexity of the square function is that

$$F(\Phi_{RB'}, (\mathrm{id}_R \otimes \mathcal{C}_{A'\to B'})(\Phi_{RA'})) \leq \frac{1}{d} \sum_{k=1}^{d} \langle\phi_k|\mathcal{C}_{A'\to B'}(|\phi_k\rangle\langle\phi_k|)|\phi_k\rangle \tag{14.1.86}$$

$$= \frac{1}{d} \sum_{k=1}^{d} F_k. \tag{14.1.87}$$

This means that

$$\frac{1}{d} \sum_{k=1}^{d} F_k \geq 1 - \varepsilon \Longleftrightarrow \frac{1}{d} \sum_{k=1}^{d} (1 - F_k) \leq \varepsilon. \tag{14.1.88}$$

Now, taking K to be a uniform random variable with realizations $k \in \{1, \ldots, d\}$ and applying Markov's inequality (see (2.3.20)), we find that

$$\Pr[1 - F_K \geq \varepsilon/\delta] \leq \frac{\mathbb{E}[1 - F_K]}{\varepsilon/\delta} \leq \frac{\varepsilon}{\varepsilon/\delta} = \delta. \tag{14.1.89}$$

So this implies that $(1 - \delta)d$ of the F_k values are such that $F_k \geq 1 - \varepsilon/\delta$. Since they are ordered as given in (14.1.82), we conclude that $\mathcal{H}_{(1-\delta)d}$, which by definition has dimension $(1 - \delta)d$, is a good subspace for quantum communication in the following sense (subspace transmission):

$$\min_{|\phi\rangle \in \mathcal{H}_{(1-\delta)d}} \langle\phi|\mathcal{C}_{A'\to B'}(|\phi\rangle\langle\phi|)|\phi\rangle \geq 1 - \varepsilon/\delta. \tag{14.1.90}$$

Now, applying Proposition 6.25 to (14.1.90), we conclude that

$$\min_{|\psi\rangle\in\mathcal{H}'_{(1-\delta)d}\otimes\mathcal{H}_{(1-\delta)d}} \langle\psi|(\mathrm{id}_{\mathcal{H}'_{(1-\delta)d}}\otimes\mathcal{C}_{A'\to B'})(|\psi\rangle\langle\psi|)|\phi\rangle \geq 1-2\sqrt{\varepsilon/\delta}, \quad (14.1.91)$$

i.e., $p^*_{\mathrm{err}}(\mathcal{E},\mathcal{D};\mathcal{N}) \leq 2\sqrt{\varepsilon/\delta}$, which is the criterion for strong subspace transmission (the strongest notion of quantum communication).

To finish off the proof, suppose that $(1-\delta)d$ is not an integer. Then there exists a $\delta' > \delta$ such that $(1-\delta')d = \lfloor(1-\delta)d\rfloor$ is a positive integer. By the above reasoning, there exists a code satisfying (14.1.91), except with δ replaced by δ', and with the code dimension equal to $\lfloor(1-\delta)d\rfloor$. We also have that $1-2\sqrt{\varepsilon/\delta'} > 1-2\sqrt{\varepsilon/\delta}$. This concludes the proof. ■

We now return to the proof of Theorem 14.5. To finish it off, we combine the results of Lemmas 14.6, 14.7, and 14.8 to conclude that the existence of a (d,ε) entanglement distillation protocol for $\omega_{AB} = \mathcal{N}_{A'\to B}(\psi_{AA'})$ implies the existence of a (d',ε') quantum communication protocol, where

$$d' = (1-\delta)d, \quad \varepsilon' = 4\sqrt{\frac{\varepsilon}{\delta}}, \quad \delta \in (0,1). \quad (14.1.92)$$

Recalling that d is given by (14.1.37), we conclude that

$$\log_2 d' = -H_{\max}^{\frac{\varepsilon'\sqrt{\delta}}{4}-\eta}(A|B)_\omega + \log_2(\eta^4(1-\delta)). \quad (14.1.93)$$

Then, since the pure state $\psi_{AA'}$ used in (14.1.93) is arbitrary, we conclude that there exists a (d',ε') quantum communication protocol satisfying

$$\log_2 d' = \sup_{\psi_{AA'}}\left(-H_{\max}^{\frac{\varepsilon'\sqrt{\delta}}{4}-\eta}(A|B)_\omega\right) + \log_2(\eta^4(1-\delta)) \quad (14.1.94)$$

for all $\eta \in [0, \varepsilon'\sqrt{\delta}/4)$ and $\delta \in (0,1)$. This is precisely the statement in (14.1.38), and so the proof of Theorem 14.5 is complete.

Applying the relation between smooth conditional min- and max-entropy in (13.1.72) to the result of Theorem 14.5, and combining it with (7.8.83), we obtain the following.

Corollary 14.9

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{V}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending $\mathcal{N}_{A\to B}$. For all $\varepsilon \in (0,1)$, $\delta \in (0,1)$, $\eta \in [0, \varepsilon\sqrt{\delta}/4)$, and $\alpha > 1$, there exists a (d,ε) quantum communication protocol for $\mathcal{N}$ with

$$\log_2 d \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{\alpha-1}\log_2\left(\frac{1}{f(\varepsilon,\delta,\eta)}\right) - \log_2\left(\frac{1}{1-f(\varepsilon,\delta,\eta)}\right) + \log_2(\eta^4(1-\delta)), \quad (14.1.95)$$

where $f(\varepsilon,\delta,\eta) \coloneqq \left(\frac{\varepsilon\sqrt{\delta}}{4}-\eta\right)^2$, $\phi_{AE} = \mathrm{Tr}_B[\mathcal{V}^{\mathcal{N}}_{A'\to BE}(\psi_{AA'})] = \mathcal{N}^c_{A'\to E}(\psi_{AA'})$, and $\psi_{AA'}$ is a pure state with the dimension of A' equal to the dimension of A.

Since the inequality in (14.1.95) holds for all (d,ε) quantum communication protocols, we have that

$$Q^\varepsilon(\mathcal{N}) \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{\alpha-1}\log_2\left(\frac{1}{f(\varepsilon,\delta,\eta)}\right) - \log_2\left(\frac{1}{1-f(\varepsilon,\delta,\eta)}\right) + \log_2(\eta^4(1-\delta)), \quad (14.1.96)$$

where

$$\phi_{AE} = \mathrm{Tr}_B[\mathcal{V}^{\mathcal{N}}_{A'\to BE}(\psi_{AA'})] = \mathcal{N}^c_{A'\to E}(\psi_{AA'}), \quad (14.1.97)$$

$f(\varepsilon,\delta,\eta)$ is defined just above, $\eta \in [0, \varepsilon\sqrt{\delta}/4)$, $\delta \in (0,1)$, and $\alpha > 1$.

14.1.3.2 Remark on Forward Classical Communication

To summarize what we did in this section, we used the result from Proposition 13.10 on one-shot entanglement distillation to prove the existence of a quantum communication protocol in the one-shot setting. Note that the entanglement distillation protocol of Proposition 13.10 involves one-way classical communication, while quantum communication (as we defined it at the beginning of this chapter) does not. In other words, in this section we managed to remove the one-way classical communication from the entanglement distillation protocol and thereby argue for

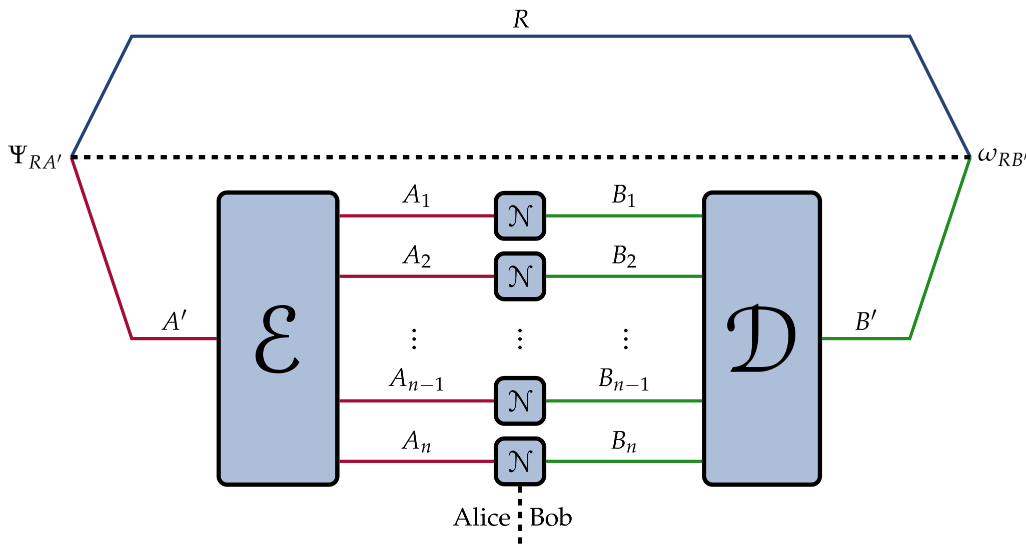


FIGURE 14.3: A general quantum communication protocol for $n \geq 1$ memoryless/unassisted uses of a quantum channel $\mathcal{N}$. Alice uses the channel $\mathcal{E}$ to encode her share A' of the pure state $\Psi_{RA'}$ into n quantum systems $A_1, A_2, \ldots, A_n$. She then sends each one of these through the channel $\mathcal{N}$. Bob finally applies a joint decoding channel $\mathcal{D}$ on the systems $B_1, B_2, \ldots, B_n$, resulting in the state $\omega_{RB'}$ given by (14.2.1).

the existence of a quantum communication protocol. More generally, it holds that forward classical communication (i.e., from the sender to the receiver) does not enhance the corresponding quantum capacity of the channel. In other words, if $Q^{\rightarrow}(\mathcal{N})$ denotes the quantum capacity of the channel $\mathcal{N}$ when classical communication from the sender to the receiver is allowed as part of the protocol, then $Q^{\rightarrow}(\mathcal{N}) = Q(\mathcal{N})$. This is a direct consequence of the chain of reasoning given in Lemmas 14.6, 14.7, and 14.8, and we return to this point in Chapter 19 when we consider LOCC-assisted quantum communication.

14.2 Quantum Capacity of a Quantum Channel

We now consider the asymptotic setting. In this scenario, depicted in Figure 14.3, the quantum system A' to be transmitted to Bob is encoded into n copies $A_1, \ldots, A_n$ of a quantum system A, for $n \geq 1$. Each of these systems is then sent independently through the channel $\mathcal{N}$. We call this the asymptotic setting because the number n can be arbitrarily large.

Analysis of the asymptotic setting is almost exactly the same as that of the one-shot setting. This is due to the fact that n independent uses of the channel $\mathcal{N}$ can be regarded as a single use of the channel $\mathcal{N}^{\otimes n}$. So the only change that needs to be made is to replace $\mathcal{N}$ with $\mathcal{N}^{\otimes n}$ and to define the encoding and decoding channels as acting on n systems instead of just one. In particular, the state at the end of the protocol becomes

$$\omega_{RB'} = (\mathcal{D}_{B^n\to B'} \circ \mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{E}_{A'\to A^n})(\Psi_{RA'}). \tag{14.2.1}$$

Then, just as in the one-shot setting, we define the error probability of the code $(\mathcal{E}, \mathcal{D})$ for n independent uses of $\mathcal{N}$ as

$$p_{\text{err}}^*(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) = 1 - F(\mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}). \tag{14.2.2}$$

Definition 14.10 **(n, d, ε) Quantum Communication Protocol**

Let $(d, \mathcal{E}_{A'\to A^n}, \mathcal{D}_{B^n\to B'})$ be the elements of a quantum communication protocol for n independent uses of the channel $\mathcal{N}_{A\to B}$, where $d_{A'} = d_{B'} = d$. The protocol is called an (n, d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^*(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) \leq \varepsilon$.

The *rate* of an (n, d, ε) quantum communication protocol is defined as the number of qubits transmitted per channel use, i.e.,

$$R(n, d) := \frac{\log_2 d}{n}. \tag{14.2.3}$$

Observe that the rate depends only on the dimension d of the system A' of the pure state $\Psi_{RA'}$ to be transmitted and on the number of channel uses. In particular, it does not directly depend on the communication channel nor on the encoding and decoding channels. For a given $\varepsilon \in [0, 1]$ and $n \geq 1$, the highest rate among all (n, d, ε) protocols is denoted by $Q^{n,\varepsilon}(\mathcal{N})$, and it is defined as

$$Q^{n,\varepsilon}(\mathcal{N}) := \frac{1}{n} Q^{\varepsilon}(\mathcal{N}^{\otimes n}) = \sup_{(d,\mathcal{E},\mathcal{D})} \left\{ \frac{\log_2 d}{n} : p_{\text{err}}^*(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) \leq \varepsilon \right\}, \tag{14.2.4}$$

where in the second equality we use the definition of the one-shot quantum capacity Q^{ε} given in (14.1.17), and the supremum is over all $d \geq 1$, encoding channels $\mathcal{E}$ with input system dimension d, and decoding channels $\mathcal{D}$ with output system dimension d.

Definition 14.11 Achievable Rate for Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for quantum communication over* $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ quantum communication protocol for $\mathcal{N}$.

As we prove in Appendix A,

$$R \text{ achievable rate} \iff \lim_{n\to\infty} \varepsilon_Q^*(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) = 0 \quad \forall\, \delta > 0. \tag{14.2.5}$$

In other words, a rate R is achievable if for all $\delta > 0$, the optimal error probability for a sequence of protocols with rate $R - \delta$ vanishes as the number n of uses of $\mathcal{N}$ increases.

Definition 14.12 Quantum Capacity of a Quantum Channel

The *quantum capacity of a quantum channel* $\mathcal{N}$, denoted by $Q(\mathcal{N})$, is defined to be the supremum of all achievable rates, i.e.,

$$Q(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\} \tag{14.2.6}$$

An equivalent definition of quantum capacity is

$$Q(\mathcal{N}) = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} Q^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{14.2.7}$$

We prove this in Appendix A.

Definition 14.13 Weak Converse Rate for Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *weak converse rate for quantum communication over* $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

We show in Appendix A that

$$R \text{ weak converse rate} \iff \lim_{n\to\infty} \varepsilon_Q^*(2^{n(R-\delta)}; \mathcal{N}^{\otimes n}) > 0 \quad \forall\, \delta > 0. \tag{14.2.8}$$

In other words, a weak converse rate is a rate for which the optimal error probability cannot be made to vanish, even in the limit of a large number of channel uses.

Definition 14.14 Strong Converse Rate for Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for quantum communication over* $\mathcal{N}$ if for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ quantum communication protocol for $\mathcal{N}$.

We show in Appendix A that

$$R \text{ strong converse rate} \iff \lim_{n\to\infty} \varepsilon_Q^*(2^{n(R+\delta)}; \mathcal{N}^{\otimes n}) = 1 \quad \forall\, \delta > 0. \tag{14.2.9}$$

Unlike the weak converse, in which the optimal error is required to simply be bounded away from zero as the number n of channel uses increases, in order to have a strong converse rate, the optimal error has to converge to one as n increases. By comparing (14.2.8) and (14.2.9), we conclude that every strong converse rate is a weak converse rate.

Definition 14.15 Strong Converse Quantum Capacity of a Quantum Channel

The *strong converse quantum capacity* of a quantum channel $\mathcal{N}$, denoted by $\widetilde{Q}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{Q}(\mathcal{N}) := \inf\{R : \ R \text{ is a strong converse rate for } \mathcal{N}\} \tag{14.2.10}$$

We can also write the strong converse quantum capacity as

$$\widetilde{Q}(\mathcal{N}) = \sup_{\varepsilon\in[0,1)} \limsup_{n\to\infty} \frac{1}{n} Q^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{14.2.11}$$

See Appendix A for a proof. We also show in Appendix A that

$$Q(\mathcal{N}) \leq \widetilde{Q}(\mathcal{N}) \tag{14.2.12}$$

for every quantum channel $\mathcal{N}$.

We now state the main theorem of this chapter, which gives an expression for the quantum capacity of a quantum channel.

Theorem 14.16 Quantum Capacity

The quantum capacity of a quantum channel $\mathcal{N}_{A\to B}$ is equal to the regularized coherent information $I^c_{\text{reg}}(\mathcal{N})$ of $\mathcal{N}$, i.e.,

$$Q(\mathcal{N}) = I^c_{\text{reg}}(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} I^c(\mathcal{N}^{\otimes n}). \tag{14.2.13}$$

Recall from (7.11.107) that the channel coherent information is defined as

$$I^c(\mathcal{N}) = \sup_{\psi_{RA}} I(R\rangle B)_\omega, \tag{14.2.14}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$ and ψ_{RA} is a pure state with the dimension of R equal to the dimension of A.

Observe that the expression in (14.2.13) for the quantum capacity of a quantum channel is somewhat similar to the expression in (12.2.14) for the classical capacity of a quantum channel, in the sense that both capacities involve a regularization of a corresponding channel measure. In the case of quantum communication, we obtain the regularization of the channel's coherent information, whereas in the case of classical communication, we obtain the regularization of Holevo information. Due to the regularization, which involves a limit of an arbitrarily large number of uses of the channel, the quantum capacity is in general difficult to compute.

We show below in Section 14.2.3 that the coherent information is always *superadditive*, meaning that $I^c(\mathcal{N}^{\otimes n}) \geq nI^c(\mathcal{N})$ for every channel $\mathcal{N}$. This means that the coherent information is always a lower bound on the quantum capacity of a channel $\mathcal{N}$:

$$Q(\mathcal{N}) \geq I^c(\mathcal{N}) \text{ for every channel } \mathcal{N}. \tag{14.2.15}$$

If the coherent information happens to be additive for a particular channel, then the regularization in (14.2.13) is not required. The coherent information is known to be additive for all degradable and anti-degradable channels. (See Definition 4.6.) In fact, for anti-degradable channels, the coherent information is equal to zero, a fact that we prove in Section 14.3.2 below. Examples of degradable and anti-degradable channels include the amplitude damping channel (Section 4.5.1) and the quantum erasure channel (Section 4.5.2). For all such channels, we thus have $Q(\mathcal{N}) = I^c(\mathcal{N})$.

Also, just as with classical communication, in this case Theorem 14.16 only makes a statement about the quantum capacity and not about the strong converse quantum capacity. One way of attempting to prove that the coherent information of a channel is equal to its strong converse quantum capacity involves proving that the sandwiched Rényi coherent information is additive for the channel. Unfortunately, this quantity has not been shown to be additive for *any* quantum channel thus far, which means that this approach is not known to be useful for obtaining strong converse quantum capacities. We consider another approach to strong converse quantum capacities in Section 14.2.4, which leads to a strong converse theorem for dephasing channels (see (4.5.35)). In terms of a general statement about the converse, the best we can say generally is that the regularized coherent information is a weak converse rate for all quantum channels.

There are two ingredients to the proof of Theorem 14.16:

1. *Achievability*: We show that $I^c_{\text{reg}}(\mathcal{N})$ is an achievable rate, which involves explicitly constructing a quantum communication protocol. The developments in Section 14.1.3 on a lower bound for one-shot quantum capacity can be used (via the substitution $\mathcal{N} \to \mathcal{N}^{\otimes n}$) to argue for the existence of a quantum communication protocol for $\mathcal{N}$ in the asymptotic setting at the rate $I^c_{\text{reg}}(\mathcal{N})$.

   The achievability part of the proof establishes that $Q(\mathcal{N}) \geq I^c_{\text{reg}}(\mathcal{N})$.

2. *Weak Converse*: We show that $I^c_{\text{reg}}(\mathcal{N})$ is a weak converse rate, from which it follows that $Q(\mathcal{N}) \leq I^c_{\text{reg}}(\mathcal{N})$. To show that $I^c_{\text{reg}}(\mathcal{N})$ is a weak converse rate, we use the one-shot upper bounds from Section 14.1.2 to conclude that every achievable rate R satisfies $R \leq I^c_{\text{reg}}(\mathcal{N})$.

We first establish in Section 14.2.1 that the quantity $I^c_{\text{reg}}(\mathcal{N})$ is an achievable rate for quantum communication over $\mathcal{N}$. Then, in Section 14.2.2, we prove that $I^c_{\text{reg}}(\mathcal{N})$ is a weak converse rate.

14.2.1 Proof of Achievability

In this section, we prove that $I^c_{\text{reg}}(\mathcal{N})$ is an achievable rate for quantum communication over a channel $\mathcal{N}$.

First, recall from Corollary 14.9 that for all $\varepsilon \in (0,1)$, $\delta \in (0,1)$, $\eta \in [0, \varepsilon\sqrt{\delta}/4)$, and $\alpha > 1$, there exists a (d, ε) quantum communication protocol for $\mathcal{N}_{A\to B}$ with

$$\log_2 d \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{\alpha - 1} \log_2\!\left(\frac{1}{f(\varepsilon,\delta,\eta)}\right)$$

$$- \log_2\left(\frac{1}{1 - f(\varepsilon, \delta, \eta)}\right) + \log_2(\eta^4(1-\delta)), \quad (14.2.16)$$

where $f(\varepsilon, \delta, \eta) := \left(\frac{\varepsilon\sqrt{\delta}}{4} - \eta\right)^2$, $\phi_{AE} = \mathcal{N}^c_{A' \to E}(\psi_{AA'})$, and $\psi_{AA'}$ is a pure state with the dimension of A' equal to the dimension of A. Note that

$$\widetilde{H}_\alpha(A|E)_\phi = -\inf_{\sigma_E} \widetilde{D}_\alpha(\phi_{AE} \| \mathbb{1}_A \otimes \sigma_E), \quad (14.2.17)$$

where the optimization is with respect to states σ_E. A simple corollary of (14.2.16) is the following.

Corollary 14.17 Lower Bound for Quantum Communication in the Asymptotic Setting

Let $\mathcal{N}_{A \to B}$ be a quantum channel. For all $n \in \mathbb{N}$, $\varepsilon \in (0, 1)$, and $\alpha > 1$, there exists an (n, d, ε) quantum communication protocol for $\mathcal{N}$ with

$$\frac{\log_2 d}{n} \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{n(\alpha - 1)} \log_2\left(\frac{128}{\varepsilon^2}\right) - \frac{1}{n} \log_2\left(\frac{1}{1 - \frac{\varepsilon^2}{128}}\right) - \frac{4}{n} \log_2\left(\frac{1}{\varepsilon}\right) - \frac{15}{n}, \quad (14.2.18)$$

where $\phi_{AE} = \mathcal{N}^c_{A' \to E}(\psi_{AA'})$ and $\psi_{AA'}$ is a pure state with the dimension of A' equal to the dimension of A.

Proof: Applying the inequality in (14.2.16) to the channel $\mathcal{N}^{\otimes n}$, letting $\delta = 1/2$, and letting $\eta = \frac{\varepsilon}{8\sqrt{2}}$ leads to

$$\frac{\log_2 d}{n} \geq \sup_{\Psi_{A^n A'^n}} H(A^n|E^n)_\Phi - \frac{1}{n(\alpha - 1)} \log_2\left(\frac{128}{\varepsilon^2}\right) - \frac{1}{n} \log_2\left(\frac{1}{1 - \frac{\varepsilon^2}{128}}\right) - \frac{4}{n} \log_2\left(\frac{1}{\varepsilon}\right) - \frac{15}{n}, \quad (14.2.19)$$

where $\Phi_{A^n E^n} = (\mathcal{N}^c_{A' \to E})^{\otimes n}(\Psi_{A^n A'^n})$, and the optimization is with respect to pure states $\Psi_{A^n A'^n}$ with $d_A = d_{A'}$. Now, let $\psi_{AA'}$ be an arbitrary pure state, and let

$\phi_{AE} = \mathcal{N}^c_{A'\to E}(\psi_{AA'})$. Then, restricting the optimization in the definition of $\widetilde{H}_\alpha(A^n|E^n)_{\phi^{\otimes n}}$ to product states (see (14.2.17)) leads to

$$\widetilde{H}_\alpha(A^n|E^n)_{\phi^{\otimes n}} \geq n\widetilde{H}_\alpha(A|E)_\phi. \tag{14.2.20}$$

In other words,

$$\sup_{\Psi_{A^nA'^n}} \widetilde{H}_\alpha(A^n|E^n)_\Phi \geq \widetilde{H}_\alpha(A^n|E^n)_{\phi^{\otimes n}} \geq n\widetilde{H}_\alpha(A|E)_\phi \tag{14.2.21}$$

for every pure state $\psi_{AA'}$, which implies that

$$\sup_{\Psi_{A^nA'^n}} \widetilde{H}_\alpha(A^n|E^n)_\Phi \geq n \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi. \tag{14.2.22}$$

Therefore, the inequality in (14.2.19) simplifies to (14.2.18), as required. ■

The inequality in (14.2.18) implies that

$$Q^{n,\varepsilon}(\mathcal{N}) \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{n(\alpha-1)}\log_2\left(\frac{128}{\varepsilon^2}\right) \\ - \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon^2}{128}}\right) - \frac{4}{n}\log_2\left(\frac{1}{\varepsilon}\right) - \frac{15}{n}, \tag{14.2.23}$$

for all $n \geq 1$, $\varepsilon \in (0,1)$, and $\alpha > 1$, where $d_{A'} = d_A$ and $\phi_{AE} = \mathcal{N}^c_{A'\to E}(\psi_{AA'})$.

We can now use (14.2.18) to prove that the regularized coherent information $I^c_{\text{reg}}(\mathcal{N})$ is an achievable rate for quantum communication over $\mathcal{N}$.

Proof of the Achievability Part of Theorem 14.16

Fix $\varepsilon \in (0,1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta = \delta_1 + \delta_2. \tag{14.2.24}$$

Set $\alpha \in (1,\infty)$ such that

$$\delta_1 \geq I^c(\mathcal{N}) - \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi, \tag{14.2.25}$$

where $\psi_{AA'}$ is a pure state with the dimension of A' equal to the dimension of A and $\phi_{AE} = \mathcal{N}^c_{A'\to E}(\psi_{AA'})$. Note that this is possible because $\widetilde{H}_\alpha(A|E)_\phi$ increases monotonically with decreasing α (this follows from Proposition 7.23), so that

$$\lim_{\alpha\to 1^+} \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi = \sup_{\alpha\in(1,\infty)} \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi \tag{14.2.26}$$

$$= \sup_{\alpha\in(1,\infty)} \sup_{\psi_{AA'}} \left(- \inf_{\sigma_E} \widetilde{D}_\alpha(\phi_{AE} \| \mathbb{1}_A \otimes \sigma_E) \right) \tag{14.2.27}$$

$$= - \inf_{\alpha\in(1,\infty)} \inf_{\psi_{AA'}} \inf_{\sigma_E} \widetilde{D}_\alpha(\phi_{AE} \| \mathbb{1}_A \otimes \sigma_E) \tag{14.2.28}$$

$$= - \inf_{\psi_{AA'}} \inf_{\sigma_E} \inf_{\alpha\in(1,\infty)} \widetilde{D}_\alpha(\phi_{AE} \| \mathbb{1} \otimes \sigma_E) \tag{14.2.29}$$

$$= - \inf_{\psi_{AA'}} \inf_{\sigma_E} D(\phi_{AE} \| \mathbb{1}_A \otimes \sigma_E) \tag{14.2.30}$$

$$= \sup_{\psi_{AA'}} \left(- \inf_{\sigma_E} D(\phi_{AE} \| \mathbb{1}_A \otimes \sigma_E) \right) \tag{14.2.31}$$

$$= \sup_{\psi_{AA'}} H(A|E)_\phi, \tag{14.2.32}$$

where the fifth equality follows from Proposition 7.22. Now, let $\mathcal{V}^{\mathcal{N}}_{A'\to BE}$ be an isometric channel extending $\mathcal{N}_{A'\to B}$ such that $\phi_{AE} = \mathcal{N}^c_{A'\to E}(\psi_{AA'}) = \mathrm{Tr}_B[\mathcal{V}^{\mathcal{N}}_{A'\to BE}(\psi_{AA'})]$. Since the state $\mathcal{V}^{\mathcal{N}}_{A'\to BE}(\psi_{AA'})$ is pure, we can view it as a purification of $\omega_{AB} = \mathcal{N}_{A'\to B}(\psi_{AA'})$, so that

$$H(A|E)_\phi = H(AE)_\phi - H(E)_\phi \tag{14.2.33}$$

$$= H(B)_\omega - H(AB)_\omega \tag{14.2.34}$$

$$= I(A\rangle B)_\omega \tag{14.2.35}$$

for every pure state $\psi_{AA'}$. Therefore,

$$\sup_{\psi_{AA'}} H(A|E)_\phi = \sup_{\psi_{AA'}} I(A\rangle B)_\omega = I^c(\mathcal{N}). \tag{14.2.36}$$

With $\alpha \in (1,\infty)$ chosen such that (14.2.25) holds, take n large enough so that

$$\delta_2 \geq \frac{1}{n(\alpha-1)} \log_2\!\left(\frac{128}{\varepsilon^2}\right) + \frac{1}{n}\log_2\!\left(\frac{1}{1-\frac{\varepsilon^2}{128}}\right) + \frac{4}{n}\log_2\!\left(\frac{1}{\varepsilon}\right) + \frac{15}{n}. \tag{14.2.37}$$

Now, we use the fact that for the n and ε chosen above there exists an (n, d, ε) protocol such that

$$\frac{\log_2 d}{n} \geq \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi - \frac{1}{n(\alpha-1)} \log_2\left(\frac{128}{\varepsilon^2}\right) - \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon^2}{128}}\right) - \frac{4}{n}\log_2\left(\frac{1}{\varepsilon}\right) - \frac{15}{n}, \quad (14.2.38)$$

which holds due to Corollary 14.17. Rearranging the right-hand side of this inequality, and using (14.2.24), (14.2.25), and (14.2.37), we find that

$$\frac{\log_2 d}{n} \geq I^c(\mathcal{N}) - \left(I^c(\mathcal{N}) - \sup_{\psi_{AA'}} \widetilde{H}_\alpha(A|E)_\phi + \frac{1}{n(\alpha-1)}\log_2\left(\frac{128}{\varepsilon^2}\right) + \frac{1}{n}\log_2\left(\frac{1}{1-\frac{\varepsilon^2}{128}}\right) + \frac{4}{n}\log_2\left(\frac{1}{\varepsilon}\right) + \frac{15}{n}\right) \quad (14.2.39)$$

$$\geq I^c(\mathcal{N}) - (\delta_1 + \delta_2) \quad (14.2.40)$$

$$= I^c(\mathcal{N}) - \delta. \quad (14.2.41)$$

Thus, there exists an (n, d, ε) quantum communication protocol with rate $\frac{\log_2 d}{n} \geq I^c(\mathcal{N}) - \delta$. Therefore, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ quantum communication protocol with $R = I^c(\mathcal{N})$ for all sufficiently large n such that (14.2.37) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(I^c(\mathcal{N})-\delta)}, \varepsilon)$ quantum communication protocol. This means, by definition, that $I^c(\mathcal{N})$ is an achievable rate for quantum communication over $\mathcal{N}$.

Now, we can repeat the arguments above for the tensor-power channel $\mathcal{N}^{\otimes k}$ for all $k \geq 1$, and so we conclude that $\frac{1}{k}I^c(\mathcal{N}^{\otimes k})$ is an achievable rate (the arguments are similar to the arguments in the proof of the achievability part of Proposition 13.19). Since k is arbitrary, we conclude that $\lim_{k\to\infty} \frac{1}{k} I^c(\mathcal{N}^{\otimes k}) = I^c_{\text{reg}}(\mathcal{N})$ is an achievable rate for quantum communication over $\mathcal{N}$. ■

14.2.2 Proof of the Weak Converse

We now show that the regularized coherent information $I^c_{\text{reg}}(\mathcal{N})$ is a weak converse rate for quantum communication over $\mathcal{N}$. This establishes that $Q(\mathcal{N}) \leq I^c_{\text{reg}}(\mathcal{N})$ and therefore that $Q(\mathcal{N}) = I^c_{\text{reg}}(\mathcal{N})$, completing the proof of Theorem 14.16.

Let us first recall from Theorem 14.4 that for every quantum channel $\mathcal{N}$, we have the following: for all $\varepsilon \in [0, 1)$ and (d, ε) quantum communication protocols

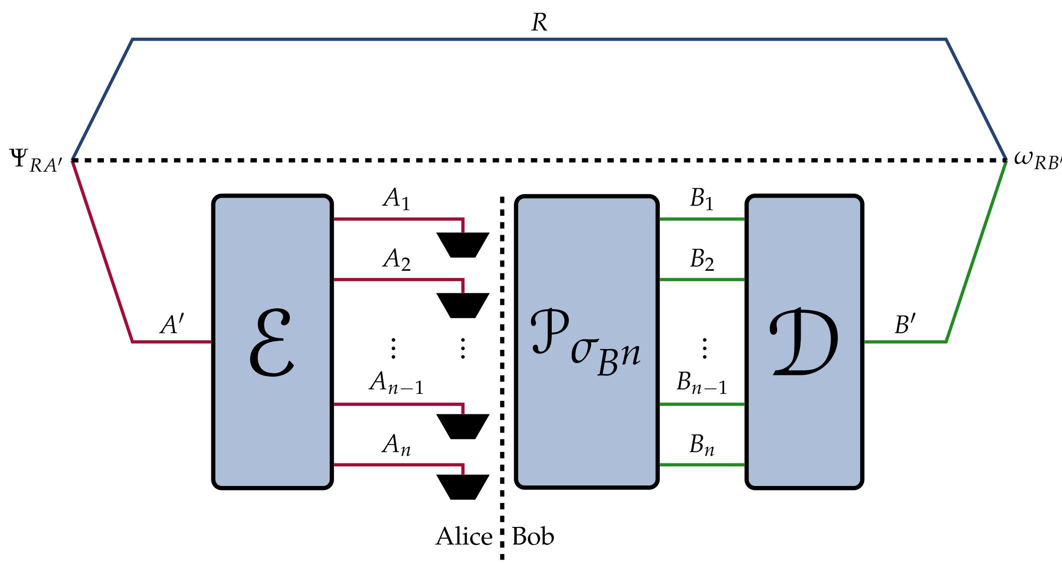


FIGURE 14.4: Depiction of a protocol that is useless for quantum communication in the asymptotic setting. The state encoding Alice's share of the pure state $\Psi_{RA'}$ is discarded and replaced by an arbitrary (but fixed) state σ_{B^n}.

for $\mathcal{N}$,

$$(1-2\varepsilon)\log_2 d \leq I^c(\mathcal{N}) + h_2(\varepsilon), \tag{14.2.42}$$

$$\log_2 d \leq \widetilde{I}^c_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{14.2.43}$$

To obtain these inequalities, we considered a quantum communication protocol for a useless channel. The useless channel in the asymptotic setting is analogous to the one in Figure 14.2 and is shown in Figure 14.4. Applying (14.2.42) and (14.2.43) to the channel $\mathcal{N}^{\otimes n}$ leads to the following.

Corollary 14.18 Upper Bounds for Quantum Communication in the Asymptotic Setting

Let $\mathcal{N}$ be a quantum channel. For all $\varepsilon \in [0,1)$, $n \in \mathbb{N}$, and (n,d,ε) quantum communication protocols over n uses of $\mathcal{N}$, the number of transmitted qubits is bounded from above as follows:

$$(1-2\varepsilon)\frac{\log_2 d}{n} \leq \frac{1}{n}I^c(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon), \tag{14.2.44}$$

$$\frac{\log_2 d}{n} \leq \frac{1}{n}\widetilde{I}^c_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1. \tag{14.2.45}$$

Proof: Since the inequalities in (14.2.42) and (14.2.43) of Theorem 14.4 hold for every channel $\mathcal{N}$, they hold for the channel $\mathcal{N}^{\otimes n}$. Therefore, applying (14.2.42) and (14.2.43) to $\mathcal{N}^{\otimes n}$ and dividing both sides by n, we obtain the desired result. ■

The inequalities in the corollary above give us, for all $\varepsilon \in [0, 1)$ and $n \in \mathbb{N}$, an upper bound on the rate of an arbitrary (n, d, ε) quantum communication protocol. If instead we fix a particular rate R by letting $d = 2^{nR}$, then we can obtain a lower bound on the error probability of an $(n, 2^{nR}, \varepsilon)$ quantum communication protocol. Specifically, using (14.2.45), we find that

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{n}\widetilde{I}^c_\alpha(\mathcal{N}^{\otimes n})\right)} \tag{14.2.46}$$

for all $\alpha > 1$.

The inequalities in (14.2.44) and (14.2.45) imply that

$$(1-2\varepsilon)Q^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{n}I^c(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon), \tag{14.2.47}$$

$$Q^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{n}\widetilde{I}^c_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{14.2.48}$$

with $n \geq 1$ and $\varepsilon \in [0, 1)$.

Using (14.2.44), we can now prove the weak converse part of Theorem 14.16.

Proof of the Weak Converse Part of Theorem 14.16

Suppose that R is an achievable rate. Then, by definition, for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ quantum communication protocol for $\mathcal{N}$. For all such protocols, the inequality (14.2.44) in Corollary 14.18 holds, so that

$$(1-2\varepsilon)(R-\delta) \leq \frac{1}{n}I^c(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon). \tag{14.2.49}$$

Since this bound holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$(1-2\varepsilon)R \leq \lim_{n\to\infty}\left(\frac{1}{n}I^c(\mathcal{N}^{\otimes n}) + \frac{1}{n}h_2(\varepsilon)\right) + (1-2\varepsilon)\delta, \tag{14.2.50}$$

$$= \lim_{n\to\infty}\frac{1}{n}I^c(\mathcal{N}^{\otimes n}) + (1-2\varepsilon)\delta. \tag{14.2.51}$$

Then, since this inequality holds for all $\varepsilon \in (0, 1/2)$ and $\delta > 0$, we obtain

$$R \leq \lim_{\varepsilon,\delta\to 0} \left\{ \frac{1}{1-2\varepsilon} \lim_{n\to\infty} \frac{1}{n} I^c(\mathcal{N}) + \delta \right\} = \lim_{n\to\infty} \frac{1}{n} I^c(\mathcal{N}^{\otimes n}) = I^c_{\text{reg}}(\mathcal{N}). \quad (14.2.52)$$

We have thus shown that if R is an achievable rate, then $R \leq I^c_{\text{reg}}(\mathcal{N})$. The contrapositive of this statement is that if $R > I^c_{\text{reg}}(\mathcal{N})$, then R is not an achievable rate. By definition, therefore, $I^c_{\text{reg}}(\mathcal{N})$ is a weak converse rate.

14.2.3 The Additivity Question

Although we have shown that the quantum capacity $Q(\mathcal{N})$ of a channel $\mathcal{N}$ is given by its regularized coherent information $I^c_{\text{reg}}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} I^c(\mathcal{N}^{\otimes n})$, without the additivity of $I^c(\mathcal{N})$, this result is not particularly helpful since it is not clear whether the regularized coherent information can be computed in general.

The coherent information is always *superadditive*, meaning that for two arbitrary quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$,

$$I^c(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq I^c(\mathcal{N}_1) + I^c(\mathcal{N}_2). \quad (14.2.53)$$

This follows from the fact that coherent information is additive for product states $\tau_{A_1B_1} \otimes \omega_{A_2B_2}$:

$$I(A_1A_2\rangle B_1B_2)_{\tau\otimes\omega} = I(A_1\rangle B_1)_\tau + I(A_2\rangle B_2)_\omega, \quad (14.2.54)$$

which is a consequence of (7.1.6) and the additivity of entropy for product states (see (7.2.104)).

Now, let $\psi_{R_1R_2A_1A_2}$, $\phi_{R_1A_1}$, $\varphi_{R_2A_2}$ be arbitrary pure states, where A_1 and A_2 are input systems to the channels $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, and $d_{R_1} = d_{A_1}$ and $d_{R_2} = d_{A_2}$. Then, letting

$$\rho_{R_1R_2B_1B_2} := ((\mathcal{N}_1)_{A_1\to B_1} \otimes (\mathcal{N}_2)_{A_2\to B_2})(\psi_{R_1R_2A_1A_2}), \quad (14.2.55)$$

$$\tau_{R_1B_1} := (\mathcal{N}_1)_{A_1\to B_1}(\phi_{R_1A_1}), \quad (14.2.56)$$

$$\omega_{R_2B_2} := (\mathcal{N}_2)_{A_2\to B_2}(\varphi_{R_2A_2}), \quad (14.2.57)$$

and restricting the optimization in the definition of coherent information of a channel to pure product states, we find that

$$I^c(\mathcal{N}_1 \otimes \mathcal{N}_2) = \sup_{\psi_{R_1R_2A_1A_2}} I(R_1R_2\rangle B_1B_2)_\rho \quad (14.2.58)$$

$$\geq \sup_{\phi_{R_1A_1}\otimes\varphi_{R_2A_2}} I(R_1R_2\rangle B_1B_2)_{\tau\otimes\omega} \tag{14.2.59}$$

$$= \sup_{\phi_{R_1A_1}\otimes\varphi_{R_2A_2}} \{I(R_1\rangle A_1)_\tau + I(R_2\rangle B_2)_\omega\} \tag{14.2.60}$$

$$= \sup_{\phi_{R_1A_1}} I(R_1\rangle B_1)_\tau + \sup_{\varphi_{R_2A_2}} I(R_2\rangle B_2)_\omega \tag{14.2.61}$$

$$= I^c(\mathcal{N}_1) + I^c(\mathcal{N}_2), \tag{14.2.62}$$

which is precisely (14.2.53). The reverse inequality does not hold in general, but it does for degradable channels (see Section 14.3.1 below).

For the sandwiched Rényi coherent information of a bipartite state ρ_{AB}, which is defined as

$$\widetilde{I}^c_\alpha(A\rangle B)_\rho = \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB}\|\mathbb{1}_A\otimes\sigma_B), \tag{14.2.63}$$

where the optimization is over states σ_B, the following additivity equality holds for all product states $\tau_{A_1B_1}\otimes\omega_{A_2B_2}$ and $\alpha\in(1,\infty)$:

$$\widetilde{I}_\alpha(A_1A_2\rangle B_1B_2)_{\tau\otimes\omega} = \widetilde{I}_\alpha(A_1\rangle B_1)_\tau + \widetilde{I}_\alpha(A_2\rangle B_2)_\omega. \tag{14.2.64}$$

This equality follows by reasoning similar to that given for the proof of Proposition 11.21. By the same reasoning given in (14.2.58)–(14.2.62), we conclude that

$$I^c_\alpha(\mathcal{N}_1\otimes\mathcal{N}_2) \geq I^c_\alpha(\mathcal{N}_1) + I^c_\alpha(\mathcal{N}_2) \tag{14.2.65}$$

for all $\alpha\in(1,\infty)$, where $I^c_\alpha(\mathcal{N})$ is the sandwiched Rényi coherent information of the channel $I^c_\alpha(\mathcal{N})$. Whether the reverse inequality holds, even for particular classes of channels, is an open question.

14.2.4 Rains Information Strong Converse Upper Bound

Except for channels for which the coherent information is known to be additive (such as the class of degradable channels; see Section 14.3.1 below), the quantum capacity of a channel is difficult to compute. This prompts us to find tractable upper bounds on quantum capacity. This search for tractable upper bounds is entirely analogous to what was done in Section 12.2.5 for classical communication in order to obtain tractable strong converse upper bounds on classical capacity.

Recall that in the previous chapter on entanglement distillation, our approach to obtaining strong converse upper bounds on distillable entanglement consisted of comparing the state at the output of an entanglement distillation protocol with one that is useless for entanglement distillation. We considered the set of PPT′ operators as the useless set, and we obtained state entanglement measures as upper bounds in the one-shot and asymptotic settings. Now, observe that entanglement transmission is similar to entanglement distillation in the sense that, like entanglement distillation, the error criterion for entanglement transmission involves comparing the output state of the protocol to the maximally entangled state. This suggests that the state entanglement measures defined in Section 9.3, and in particular the results of Proposition 13.6 and Corollary 13.7, are relevant. However, the main resource that we are considering in this chapter is a quantum channel and not a quantum state, and so we have an extra degree of freedom in the input state to the channel, which we can optimize. This suggests that the channel entanglement measures from Chapter 10 are relevant, and this is indeed what we find.

Proposition 14.19

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For an arbitrary (d, ε) quantum communication protocol for $\mathcal{N}_{A\to B}$, the number $\log_2 d$ of qubits transmitted over $\mathcal{N}$ satisfies

$$\log_2 d \leq R_H^{\varepsilon}(\mathcal{N}), \tag{14.2.66}$$

where

$$R_H^{\varepsilon}(\mathcal{N}) = \sup_{\psi_{SA}} R_H^{\varepsilon}(S;B)_{\omega} \tag{14.2.67}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in \mathrm{PPT}'(S:B)} D(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}) \tag{14.2.68}$$

is the ε-hypothesis testing Rains information of the channel $\mathcal{N}$, defined in (10.4.8). Consequently, the one-shot quantum capacity does not exceed the ε-hypothesis testing Rains information of the channel $\mathcal{N}$:

$$Q^{\varepsilon}(\mathcal{N}) \leq R_H^{\varepsilon}(\mathcal{N}). \tag{14.2.69}$$

REMARK: Note that in the expression for $R_H^{\varepsilon}(\mathcal{N})$ above it suffices to optimize over pure states ψ_{RA}, with the dimension of S equal to the dimension of A. We showed this in (10.1.3)–(10.1.6) immediately after Definition 10.1.

PROOF: By the arguments in (14.1.11)–(14.1.13), a (d, ε) quantum communication protocol is a (d, ε) entanglement transmission protocol. As such, we conclude that the state $\rho_{RB'}$ defined in (14.1.19) satisfies $\text{Tr}[\Phi_{RB'}\rho_{RB'}] \geq 1 - \varepsilon$. We can therefore apply Proposition 13.6 to conclude that

$$\log_2 d \leq R_H^\varepsilon(S; B')_\rho. \tag{14.2.70}$$

Note that

$$R_H^\varepsilon(S; B')_\rho$$
$$= \inf_{\sigma_{SB'} \in \text{PPT}'(S:B')} D_H^\varepsilon(\rho_{SB'} \| \sigma_{SB'}) \tag{14.2.71}$$
$$= \inf_{\sigma_{SB'} \in \text{PPT}'(S:B')} D_H^\varepsilon((\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{SA'}) \| \sigma_{SB'}). \tag{14.2.72}$$

Now, since every local channel is completely PPT preserving (this follows immediately from Proposition 4.29 and Lemma 4.30), we conclude that the channel $\mathcal{D}_{B\to B'} \equiv \text{id}_S \otimes \mathcal{D}_{B\to B'}$ is completely PPT preserving, so that the set

$$\{\mathcal{D}_{B\to B'}(\tau_{SB}) : \tau_{SB} \in \text{PPT}'(S:B)\} \tag{14.2.73}$$

is a subset of $\text{PPT}'(S:B')$. Thus, by restricting the optimization over all operators $\sigma_{SB'} \in \text{PPT}'(S:B')$ to the outputs $\mathcal{D}_{B\to B'}(\tau_{SB})$ of the decoding channel $\mathcal{D}_{B\to B'}$ acting on operators $\tau_{SB} \in \text{PPT}'(S:B)$, we obtain

$$R_H^\varepsilon(S; B')_\omega$$
$$\leq \inf_{\tau_{SB} \in \text{PPT}'(S:B)} D_H^\varepsilon((\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{SA'}) \| \mathcal{D}_{B\to B'}(\tau_{SB})) \tag{14.2.74}$$
$$\leq \inf_{\tau_{SB} \in \text{PPT}'(S:B)} D_H^\varepsilon((\mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{SA'}) \| \tau_{SB}) \tag{14.2.75}$$
$$= \inf_{\tau_{SB} \in \text{PPT}'(S:B)} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{SA}) \| \tau_{SB}), \tag{14.2.76}$$

where the second inequality follows from the data-processing inequality for hypothesis testing relative entropy, and the equality follows by letting $\rho_{SA} = \mathcal{E}_{A'\to A}(\Phi_{SA'})$. Finally, after optimizing over all states ρ_{SA}, we obtain

$$R_H^\varepsilon(S; B')_\rho \leq \sup_{\rho_{SA}} \inf_{\tau_{SB} \in \text{PPT}'(S:B)} D_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{SA}) \| \tau_{SB}) \tag{14.2.77}$$
$$= R_H^\varepsilon(\mathcal{N}), \tag{14.2.78}$$

so that, by (14.2.70), we conclude that

$$\log_2 d \leq R_H^\varepsilon(\mathcal{N}), \tag{14.2.79}$$

as required. ■

The result of Proposition 14.19 is analogous to the result of Theorem 14.3. Combining it with Proposition 7.71 leads to the following:

Corollary 14.20 One-Shot Rains Upper Bound for Quantum Communication

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all (d,ε) quantum communication protocols over $\mathcal{N}$, we have that

$$\log_2 d \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{14.2.80}$$

where

$$\widetilde{R}_\alpha(\mathcal{N}) = \sup_{\psi_{SA}} \widetilde{R}_\alpha(S;B)_\omega \tag{14.2.81}$$

$$= \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\text{PPT}'(S:B)} \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}) \tag{14.2.82}$$

is the *sandwiched Rényi Rains information of* $\mathcal{N}$, defined in (10.4.10).

Remark: Note that in the expression for $\widetilde{R}_\alpha(\mathcal{N})$ above it suffices to optimize over pure states ψ_{SA}, with the dimension of S equal to the dimension of A. We showed this in (10.1.3)–(10.1.6) immediately after Definition 10.1.

Since the inequality in (14.2.80) holds for all (d,ε) quantum communication protocols, we conclude the following upper bound on the one-shot quantum capacity:

$$Q^\varepsilon(\mathcal{N}) \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{14.2.83}$$

for all $\alpha > 1$.

For n channel uses, the bound in (14.2.80) becomes

$$\frac{\log_2 d}{n} \leq \frac{1}{n}\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \quad \forall\, \alpha > 1, \tag{14.2.84}$$

which holds for an arbitrary (n,d,ε) quantum communication protocol that employs n uses of the channel $\mathcal{N}$, where $n \geq 1$ and $\varepsilon \in [0,1)$. We can simplify this inequality by making use of the following fact.

Proposition 14.21 Weak Subadditivity of Rényi Rains Information of a Channel

Let $\mathcal{N}_{A\to B}$ be a quantum channel, with d_A the dimension of the input system A. For all $\alpha > 1$ and $n \in \mathbb{N}$, we have

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}) \leq n\widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha(d_A^2 - 1)}{\alpha - 1}\log_2(n+1). \tag{14.2.85}$$

Proof: Throughout this proof, for convenience we make use of the alternate notation

$$\widetilde{R}_\alpha(\rho_{AB}) \equiv \widetilde{R}_\alpha(A;B)_\rho \tag{14.2.86}$$

where ρ_{AB} is a bipartite state.

Let ψ_{SA^n} be an arbitrary pure state, with the dimension of S equal to the dimension of A^n, and let $\rho_{A^n} := \mathrm{Tr}_S[\psi_{SA^n}]$. We start by observing that the channel $\mathcal{N}^{\otimes n}$ is covariant with respect to the symmetric group $\mathcal{S}_n$. In particular, if we let $\{W^\pi_{A^n}\}_{\pi\in\mathcal{S}_n}$ and $\{W^\pi_{B^n}\}_{\pi\in\mathcal{S}_n}$ be the unitary representations of $\mathcal{S}_n$, defined in (2.5.1), acting on $\mathcal{H}_A^{\otimes n}$ and $\mathcal{H}_B^{\otimes n}$, respectively, then for every state ρ_{A^n}, we have that

$$\mathcal{N}^{\otimes n}(W^\pi_{A^n}\rho_{A^n}W^{\pi\dagger}_{A^n}) = W^\pi_{B^n}\mathcal{N}^{\otimes n}(\rho_{A^n})W^{\pi\dagger}_{B^n} \tag{14.2.87}$$

for all $\pi \in \mathcal{S}_n$. Consequently, by Proposition 10.12, in particular by (10.4.19), we find that

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\psi_{SA^n})) \leq \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\psi^{\overline{\rho}}_{SA^n})), \tag{14.2.88}$$

where the state $\overline{\rho}_{A^n}$ is defined as

$$\overline{\rho}_{A^n} = \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} W^\pi_{A^n}\rho_{A^n}W^{\pi\dagger}_{A^n}, \tag{14.2.89}$$

and $\psi^{\overline{\rho}}_{SA^n}$ is a purification of $\overline{\rho}_{A^n}$.

Since the state $\overline{\rho}_{A^n}$ is permutation invariant by definition, by Lemma 3.13, it has a purification $|\phi^{\overline{\rho}}\rangle_{\hat{A}^nA^n} \in \mathrm{Sym}_n(\mathcal{H}_{\hat{A}A})$, where the dimension of $\hat{A}$ is equal to the dimension of A. Consequently, there exists an isometry $V_{S\to\hat{A}^n}$ such that

$$V_{S\to\hat{A}^n}|\psi^{\overline{\rho}}\rangle_{SA^n} = |\phi^{\overline{\rho}}\rangle_{\hat{A}^nA^n}. \tag{14.2.90}$$

Therefore, by isometric invariance of the sandwiched Rényi Rains relative entropy, we obtain

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\psi^{\overline{\rho}}_{SA^n})) = \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\phi^{\overline{\rho}}_{\hat{A}^n A^n})). \tag{14.2.91}$$

Now, since $\phi^{\overline{\rho}}_{\hat{A}^n A^n}$ is a state, the operator inequality $\phi^{\overline{\rho}}_{\hat{A}^n A^n} \leq \mathbb{1}_{\hat{A}^n A^n}$ holds. Multiplying both sides of this inequality from the left and right by the projection $\Pi_{\text{Sym}_n(\mathcal{H}_{\hat{A}A})} \equiv \Pi_{\hat{A}^n A^n}$ onto the symmetric subspace of $\mathcal{H}^{\otimes n}_{\hat{A}A}$, we obtain

$$\Pi_{\hat{A}^n A^n} |\phi^{\overline{\rho}}\rangle\langle\phi^{\overline{\rho}}|_{\hat{A}^n A^n} \Pi_{\hat{A}^n A^n} \leq \Pi^2_{\hat{A}^n A^n} = \Pi_{\hat{A}^n A^n}. \tag{14.2.92}$$

But $|\phi^{\overline{\rho}}\rangle_{\hat{A}^n A^n} \in \text{Sym}_n(\mathcal{H}_{\hat{A}A})$, which means that

$$\Pi_{\hat{A}^n A^n} |\phi^{\overline{\rho}}\rangle\langle\phi^{\overline{\rho}}|_{\hat{A}^n A^n} \Pi_{\hat{A}^n A^n} = |\phi^{\overline{\rho}}\rangle\langle\phi^{\overline{\rho}}|_{\hat{A}^n A^n}. \tag{14.2.93}$$

Therefore,

$$\phi^{\overline{\rho}}_{\hat{A}^n A^n} \leq \Pi_{\hat{A}^n A^n} = \binom{d_A^2 + n - 1}{n} \int \phi^{\otimes n}_{\hat{A}A}\, \mathrm{d}\phi, \tag{14.2.94}$$

where the equality follows from (2.5.18). Now, note that

$$\binom{n + d_A^2 - 1}{n} = \frac{(n + d_A^2 - 1)(n + d_A^2 - 2)\cdots(n+2)(n+1)}{(d_A^2 - 1)(d_A^2 - 2)\cdots 2 \cdot 1} \tag{14.2.95}$$

$$= \frac{n + d_A^2 - 1}{d_A^2 - 1} \cdot \frac{n + d_A^2 - 2}{d_A^2 - 2} \cdots \frac{n+2}{2} \cdot \frac{n+1}{1}. \tag{14.2.96}$$

Then, using the fact that $\frac{n+k}{k} \leq n + 1$ for all $k \geq 1$, and applying this inequality to each factor on the right-hand side of the above equation, we obtain

$$\binom{n + d_A^2 - 1}{n} \leq (n+1)^{d_A^2 - 1} \tag{14.2.97}$$

Therefore,

$$\phi^{\overline{\rho}}_{\hat{A}^n A^n} \leq (n+1)^{d_A^2 - 1} \int \phi^{\otimes n}_{\hat{A}A}\, \mathrm{d}\phi \equiv (n+1)^{d_A^2 - 1} \xi_{\hat{A}^n A^n}. \tag{14.2.98}$$

Next, we use (7.5.44) to obtain

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\phi^{\overline{\rho}}_{\hat{A}^n A^n}) \leq \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\xi_{\hat{A}^n A^n})) + \frac{\alpha}{\alpha - 1} \log_2 (n+1)^{d_A^2 - 1}. \tag{14.2.99}$$

Then, by quasi-convexity of the Rényi Rains relative entropy (Proposition 9.26), we obtain

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\xi_{\hat{A}^n A^n})) \leq \sup_{\phi_{\hat{A}A}} \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\phi^{\otimes n}_{\hat{A}A})). \tag{14.2.100}$$

Then, using subadditivity of the sandwiched Rényi Rains relative entropy for tensor-product states, as given by (9.3.18), we find that

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\phi^{\otimes n}_{\hat{A}A})) \leq n\widetilde{R}_\alpha(\mathcal{N}_{A\to B}(\phi_{\hat{A}A})). \tag{14.2.101}$$

Putting everything together, we finally obtain

$$\begin{aligned} &\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\psi_{SA^n})) \\ &\quad\leq n\sup_{\phi_{\hat{A}A}} \widetilde{R}_\alpha(\mathcal{N}_{A\to B}(\phi_{\hat{A}A})) + \frac{\alpha(d_A^2-1)}{\alpha-1}\log_2(n+1) &(14.2.102)\\ &\quad= n\widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha(d_A^2-1)}{\alpha-1}\log_2(n+1). &(14.2.103)\end{aligned}$$

Since the pure state ψ_{SA^n} is arbitrary, we conclude that

$$\widetilde{R}_\alpha(\mathcal{N}^{\otimes n}) = \sup_{\psi_{SA^n}} \widetilde{R}_\alpha(\mathcal{N}^{\otimes n}_{A\to B}(\psi_{SA^n})) \leq n\widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha(d_A^2-1)}{\alpha-1}\log_2(n+1), \tag{14.2.104}$$

as required. ■

Combining (14.2.85) with (14.2.84) leads to the following upper bound on the rate of an arbitrary (n, d, ε) quantum communication protocol for a quantum channel $\mathcal{N}_{A\to B}$:

$$\frac{\log_2 d}{n} \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{(n+1)^{d_A^2-1}}{1-\varepsilon}\right) \tag{14.2.105}$$

for all $\alpha > 1$. Consequently, the following bound holds for the n-shot quantum capacity:

$$Q^{n,\varepsilon}(\mathcal{N}) \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{(n+1)^{d_A^2-1}}{1-\varepsilon}\right) \tag{14.2.106}$$

for all $\alpha > 1$.

With this bound, we are now ready to state the main result of this section, which is that the Rains information of a channel is an upper bound on the strong converse capacity of an arbitrary quantum channel $\mathcal{N}$.

Theorem 14.22 Strong Converse Upper Bound on Quantum Capacity

The Rains information $R(\mathcal{N})$ of a quantum channel $\mathcal{N}_{A\to B}$ is a strong converse rate for quantum communication over $\mathcal{N}$. In other words, $\widetilde{Q}(\mathcal{N}) \leq R(\mathcal{N})$ for every quantum channel $\mathcal{N}$.

Recall from (10.4.6) that

$$R(\mathcal{N}) = \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\text{PPT}'(S:B)} D(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}), \tag{14.2.107}$$

where the supremum is with respect to pure states ψ_{SA} with $d_S = d_A$.

Proof: Let $\varepsilon \in [0,1)$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that

$$\delta > \delta_1 + \delta_2 =: \delta'. \tag{14.2.108}$$

Set $\alpha \in (1,\infty)$ such that

$$\delta_1 \geq \widetilde{R}_\alpha(\mathcal{N}) - R(\mathcal{N}), \tag{14.2.109}$$

which is possible because $\widetilde{R}_\alpha(\mathcal{N})$ is monotonically increasing in α (which follows from Proposition 7.31) and because $\lim_{\alpha\to 1^+} \widetilde{R}_\alpha(\mathcal{N}) = R(\mathcal{N})$ (see Appendix 10.A for a proof). With this value of α, take n large enough so that

$$\delta_2 \geq \frac{\alpha}{n(\alpha-1)} \log_2\!\left(\frac{(n+1)^{d_A^2-1}}{1-\varepsilon}\right), \tag{14.2.110}$$

where d_A is the dimension of the input space of the channel $\mathcal{N}$.

Now, with the values of n and ε as above, an (n,d,ε) quantum communication protocol satisfies (14.2.105), i.e.,

$$\frac{\log_2 d}{n} \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\!\left(\frac{(n+1)^{d_A^2-1}}{1-\varepsilon}\right), \tag{14.2.111}$$

for all $\alpha > 1$. Rearranging the right-hand side of this inequality, and using (14.2.108)–(14.2.110), we obtain

$$\frac{\log_2 d}{n} \leq R(\mathcal{N}) + \widetilde{R}_\alpha(\mathcal{N}) - R(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)} \log_2\!\left(\frac{(n+1)^{d_A^2-1}}{1-\varepsilon}\right) \tag{14.2.112}$$

$$\leq R(\mathcal{N}) + \delta_1 + \delta_2 \tag{14.2.113}$$
$$= R(\mathcal{N}) + \delta' \tag{14.2.114}$$
$$< R(\mathcal{N}) + \delta. \tag{14.2.115}$$

So we have that $R(\mathcal{N})+\delta > \frac{\log_2 d}{n}$ for all (n, d, ε) quantum communication protocols with sufficiently large n. Due to this strict inequality, it follows that there cannot exist an $(n, 2^{n(R(\mathcal{N})+\delta)}, \varepsilon)$ quantum communication protocol for all sufficiently large n such that (14.2.110) holds, for if it did there would exist a d such that $\log_2 d = n(R(\mathcal{N}) + \delta)$, which we have just seen is not possible. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in [0, 1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R(\mathcal{N})+\delta)}, \varepsilon)$ quantum communication protocol. This means that $R(\mathcal{N})$ is a strong converse rate, and thus that $\widetilde{Q}(\mathcal{N}) \leq R(\mathcal{N})$. ■

14.2.4.1 The Strong Converse from a Different Point of View

Let us now show that the Rains relative entropy of a quantum channel $\mathcal{N}$ is a strong converse rate according to the definition of a strong converse rate in Appendix A. To this end, consider a sequence $\{(n, 2^{nr}, \varepsilon_n)\}_{n\in\mathbb{N}}$ of (n, d, ε) quantum communication protocols, with each element of the sequence having an arbitrary (but fixed) rate $r > R(\mathcal{N})$. For each element of the sequence, the inequality in (14.2.105) holds, which means that

$$r \leq \widetilde{R}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha - 1)} \log_2\!\left(\frac{(n+1)^{d_A^2-1}}{1 - \varepsilon_n}\right) \tag{14.2.116}$$

for all $\alpha > 1$. Rearranging this inequality leads to the following lower bound on the error probabilities ε_n:

$$\varepsilon_n \geq 1 - (n+1)^{d_A^2-1} \cdot 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(r-\widetilde{R}_\alpha(\mathcal{N})\right)}. \tag{14.2.117}$$

Now, since $r > R(\mathcal{N})$, $\lim_{\alpha\to 1^+} \widetilde{R}_\alpha(\mathcal{N}) = R(\mathcal{N})$ (see Appendix 10.A for a proof), and since the sandwiched Rényi Rains relative entropy is monotonically increasing in α (see Proposition 7.31), there exists an $\alpha^* > 1$ such that $R > \widetilde{R}_{\alpha^*}(\mathcal{N})$. Applying the inequality in (14.2.117) to this value of α, we find that

$$\varepsilon_n \geq 1 - (n+1)^{d_A^2-1} \cdot 2^{-n\left(\frac{\alpha^*-1}{\alpha^*}\right)\left(r-\widetilde{R}_{\alpha^*}(\mathcal{N})\right)}. \tag{14.2.118}$$

Then, taking the limit $n \to \infty$ on both sides of this inequality, we conclude that $\lim_{n\to\infty} \varepsilon_n \geq 1$. However, $\varepsilon_n \leq 1$ for all n since ε_n is a probability by definition. So we conclude that $\lim_{n\to\infty} \varepsilon_n = 1$. Since the rate $r > R(\mathcal{N})$ is arbitrary, we conclude that $R(\mathcal{N})$ is a strong converse rate. We also see from (14.2.118) that the sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ approaches one at an exponential rate.

14.2.5 Squashed Entanglement Weak Converse Bound

One of the results of Chapter 19 is that the squashed entanglement of a quantum channel (see Definition 10.14) is a weak converse rate for quantum communication assisted by LOCC. Since the LOCC-assisted quantum capacity is an upper bound on the unassisted quantum capacity considered in this chapter, we conclude that the squashed entanglement of a channel is a weak converse rate for unassisted quantum communication.

We present some statements along these lines briefly here, indicating that the squashed entanglement gives an upper bound on the one-shot quantum capacity, the n-shot quantum capacity, as well as the asymptotic quantum capacity. Complete proofs of these statements are available in Chapter 19, and they follow from the fact that the assistance of LOCC can only increase rates of quantum communication. Propositions 14.23 and 14.24 below are a direct consequence of Theorem 19.4, and Theorem 14.25 below is a consequence of Theorem 19.15.

Proposition 14.23 One-Shot Squashed Entanglement Upper Bound

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0, 1)$. For all (d, ε) quantum communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\log_2 d \leq \frac{1}{1-\sqrt{\varepsilon}} \left(E_{\text{sq}}(\mathcal{N}) + g_2(\sqrt{\varepsilon})\right), \tag{14.2.119}$$

where $E_{\text{sq}}(\mathcal{N})$ is the squashed entanglement of the channel $\mathcal{N}$ (see Definition 10.14) and $g_2(\delta) \coloneqq (\delta + 1)\log_2(\delta + 1) - \delta \log_2 \delta$. Consequently, for the one-shot quantum capacity of $\mathcal{N}$,

$$Q^{\varepsilon}(\mathcal{N}) \leq \frac{1}{1-\sqrt{\varepsilon}} \left(E_{\text{sq}}(\mathcal{N}) + g_2(\sqrt{\varepsilon})\right). \tag{14.2.120}$$

Proposition 14.24 ***n*-Shot Squashed Entanglement Upper Bound**

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all (n,d,ε) quantum communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\frac{1}{n}\log_2 d \leq \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\text{sq}}(\mathcal{N}) + \frac{g_2(\sqrt{\varepsilon})}{n}\right). \tag{14.2.121}$$

Consequently, for the n-shot quantum capacity of $\mathcal{N}$,

$$Q^{n,\varepsilon}(\mathcal{N}) \leq \frac{1}{1-\sqrt{\varepsilon}}\left(E_{\text{sq}}(\mathcal{N}) + \frac{g_2(\sqrt{\varepsilon})}{n}\right). \tag{14.2.122}$$

Theorem 14.25 **Weak Converse Upper Bound on Quantum Capacity**

The squashed entanglement $E_{\text{sq}}(\mathcal{N})$ of a quantum channel $\mathcal{N}_{A\to B}$ is a weak converse rate for quantum communication over $\mathcal{N}$. In other words, $Q(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N})$ for every quantum channel $\mathcal{N}$.

14.3 Examples

We now consider the quantum capacity for particular classes of quantum channels. As remarked earlier, computing the quantum capacity of an arbitrary channel is a difficult task. This task is made more difficult by the fact that, in some cases, the coherent information is known to be *strictly superadditive*, meaning that

$$I^c(\mathcal{N}^{\otimes n}) > nI^c(\mathcal{N}). \tag{14.3.1}$$

This fact confirms that regularization of the coherent information really is needed in general in order to compute the quantum capacity, and that additivity of coherent information is simply not true for all channels. Another interesting phenomenon related to quantum capacity is *superactivation*, which is when two channels $\mathcal{N}_1$ and $\mathcal{N}_2$, each with zero quantum capacity, i.e., $Q(\mathcal{N}_1) = Q(\mathcal{N}_2) = 0$, can combine to have non-zero quantum capacity, i.e., $Q(\mathcal{N}_1 \otimes \mathcal{N}_2) > 0$. Please consult the Bibliographic Notes in Section 14.5 for more information about strict superadditivity and superactivation.

In this section, we show that coherent information is additive for all degradable channels, which means that regularization is not needed in order to compute their capacities. The same turns out to be true for generalized dephasing channels, and we prove this by showing that the Rains relative entropy of those channels coincides with their coherent information. We also show that anti-degradable channels have zero quantum capacity. Finally, we evaluate the upper and lower bounds established in this chapter for the generalized amplitude damping channel.

Before starting, let us first recall the definition of coherent information of a channel:

$$I^c(\mathcal{N}) = \sup_{\psi_{RA}} I(R\rangle B)_\omega = \sup_{\rho}\{H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho))\}, \tag{14.3.2}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$ and the second equality is explained in (7.11.111)–(7.11.113). We let

$$I^c(\rho, \mathcal{N}) := H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho)). \tag{14.3.3}$$

14.3.1 Degradable Channels

Recall from Definition 4.6 that a channel $\mathcal{N}_{A\to B}$ is degradable if there exists a channel $\mathcal{D}_{B\to E}$ such that

$$\mathcal{N}^c = \mathcal{D} \circ \mathcal{N}, \tag{14.3.4}$$

where $\mathcal{N}^c$ is a channel complementary to $\mathcal{N}$ (see Definition 4.5) and $d_E \geq \text{rank}(\Gamma^{\mathcal{N}}_{AB})$. In particular, if $V_{A\to BE}$ is an isometric extension of $\mathcal{N}$, so that

$$\mathcal{N}(\rho) = \text{Tr}_E[V\rho V^\dagger] \tag{14.3.5}$$

for every state ρ, then

$$\mathcal{N}^c(\rho) = \text{Tr}_B[V\rho V^\dagger]. \tag{14.3.6}$$

We now show that the coherent information is additive for degradable quantum channels, meaning that

$$I^c(\mathcal{N} \otimes \mathcal{M}) = I^c(\mathcal{N}) + I^c(\mathcal{M}) \tag{14.3.7}$$

for all degradable quantum channels $\mathcal{N}$ and $\mathcal{M}$. Consequently, regularization is unnecessary, and we conclude that the quantum capacity of a degradable channel is equal to its coherent information:

$$Q(\mathcal{N}) = I^c(\mathcal{N}) \text{ for every degradable channel } \mathcal{N}. \tag{14.3.8}$$

Proposition 14.26 Additivity of Coherent Information for Degradable Channels

Let $\mathcal{N}$ and $\mathcal{M}$ be degradable channels. Then, the coherent information is additive, i.e.,

$$I^c(\mathcal{N} \otimes \mathcal{M}) = I^c(\mathcal{N}) + I^c(\mathcal{M}). \tag{14.3.9}$$

PROOF: As shown in Section 14.2.3, we always have superadditivity of coherent information, so that $I^c(\mathcal{N} \otimes \mathcal{M}) \geq I^c(\mathcal{N}) + I^c(\mathcal{M})$. So we prove that the reverse inequality also holds for the case of degradable channels.

Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be the degrading channels for $\mathcal{N}$ and $\mathcal{M}$, respectively, meaning that

$$\mathcal{N}^c = \mathcal{D}_1 \circ \mathcal{N}, \quad \mathcal{M}^c = \mathcal{D}_2 \circ \mathcal{M}. \tag{14.3.10}$$

Now, let $\rho_{A_1A_2}$ be an arbitrary state on the input systems A_1 and A_2 of the channels $\mathcal{N}$ and $\mathcal{M}$, respectively. Using (7.11.103) and (7.11.104), along with the fact that $(\mathcal{N} \otimes \mathcal{M})^c = \mathcal{N}^c \otimes \mathcal{M}^c$, we find that

$$H(\mathcal{N}^c(\rho_{A_1})) + H(\mathcal{M}^c(\rho_{A_2}) - H((\mathcal{N} \otimes \mathcal{M})^c(\rho_{A_1A_2}) \tag{14.3.11}$$

$$= D((\mathcal{N} \otimes \mathcal{M})^c(\rho_{A_1A_2}) \| \mathcal{N}^c(\rho_{A_1}) \otimes \mathcal{M}^c(\rho_{A_2})) \tag{14.3.12}$$

$$= D((\mathcal{D}_1 \circ \mathcal{N} \otimes \mathcal{D}_2 \circ \mathcal{M})(\rho_{A_1A_2}) \| (\mathcal{D}_1 \circ \mathcal{N})(\rho_{A_1}) \otimes (\mathcal{D}_2 \circ \mathcal{M})(\rho_{A_2})) \tag{14.3.13}$$

$$\leq D((\mathcal{N} \otimes \mathcal{M})(\rho_{A_1A_2}) \| \mathcal{N}(\rho_{A_1}) \otimes \mathcal{M}(\rho_{A_2})) \tag{14.3.14}$$

$$= H(\mathcal{N}(\rho_{A_1})) + H(\mathcal{M}(\rho_{A_2})) - H((\mathcal{N} \otimes \mathcal{M})(\rho_{A_1A_2})), \tag{14.3.15}$$

where the third equality follows from (14.3.10), the inequality follows from the data-processing inequality for quantum relative entropy, and the last equality from (7.11.103) and (7.11.104). Rearranging this inequality and applying subadditivity of the entropy $H((\mathcal{N} \otimes \mathcal{M})(\rho_{A_1A_2}))$ gives

$$H((\mathcal{N} \otimes \mathcal{M})(\rho_{A_1A_2})) - H((\mathcal{N} \otimes \mathcal{M})^c(\rho_{A_1A_2})) \tag{14.3.16}$$

$$\leq H(\mathcal{N}(\rho_{A_1})) - H(\mathcal{N}^c(\rho_{A_2})) + H(\mathcal{M}(\rho_{A_2})) - H(\mathcal{M}^c(\rho_{A_2})) \tag{14.3.17}$$

$$\leq \sup_{\rho_{A_1}} \{H(\mathcal{N}(\rho_{A_1})) - H(\mathcal{N}^c(\rho_{A_1}))\} \tag{14.3.18}$$

$$+ \sup_{\rho_{A_2}} \{H(\mathcal{M}(\rho_{A_2})) - H(\mathcal{M}^c(\rho_{A_2}))\} \tag{14.3.19}$$

$$= I^c(\mathcal{N}) + I^c(\mathcal{M}) \tag{14.3.20}$$

Since the state $\rho_{A_1A_2}$ is arbitrary, we conclude that

$$I^c(\mathcal{N}\otimes\mathcal{M}) = \sup_{\rho_{A_1A_2}} \{H((\mathcal{N}\otimes\mathcal{M})(\rho_{A_1A_2})) - H((\mathcal{N}\otimes\mathcal{N})^c(\rho_{A_1A_2}))\} \quad (14.3.21)$$

$$\leq I^c(\mathcal{N}) + I^c(\mathcal{M}), \quad (14.3.22)$$

as required. ■

Another useful fact about a degradable channel $\mathcal{N}$ is that the coherent information $I^c(\rho,\mathcal{N})$ defined in (14.3.3) is concave in the input state ρ.

Lemma 14.27

For a degradable channel $\mathcal{N}$, the function $\rho\mapsto I^c(\rho,\mathcal{N})$ is concave in the input state ρ. In other words, for every finite alphabet $\mathcal{X}$, probability distribution $p:\mathcal{X}\to[0,1]$, and set $\{\rho_A^x\}_{x\in\mathcal{X}}$ of states,

$$I^c\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x,\mathcal{N}\right) \geq \sum_{x\in\mathcal{X}} p(x) I^c(\rho_A^x,\mathcal{N}). \quad (14.3.23)$$

Proof: Let $\mathcal{D}$ be a degrading channel corresponding to $\mathcal{N}$, so that $\mathcal{N}^c = \mathcal{D}\circ\mathcal{N}$. Next, define the following states:

$$\omega_{XB} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{N}_{A\to B}(\rho_A^x), \quad (14.3.24)$$

$$\tau_{XE} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{N}^c(\rho_A^x) = \mathcal{D}_{B\to E}(\omega_{XB}) \quad (14.3.25)$$

Then, by noting that $\tau_{XE} = \mathcal{D}_{B\to E}(\omega_{XB})$ and applying the data-processing inequality for quantum mutual information (see (7.2.202)), we obtain

$$I(X;E)_\tau \leq I(X;B)_\omega . \quad (14.3.26)$$

Then, using (7.1.10) and rearranging leads to

$$H(B)_\omega - H(E)_\tau \geq H(B|X)_\omega - H(E|X)_\tau, \quad (14.3.27)$$

which is the desired inequality in (14.3.23). Indeed, the left-hand side of the inequality above is simply $I^c\left(\sum_{x\in\mathcal{X}} p(x)\rho_A^x,\mathcal{N}\right)$. For the right-hand side, we find

that

$$H(B|X)_\omega = \sum_{x\in\mathcal{X}} p(x) H(\mathcal{N}(\rho_A^x)), \tag{14.3.28}$$

$$H(E|X)_\tau = \sum_{x\in\mathcal{X}} p(x) H(\mathcal{N}^c(\rho_A^x)), \tag{14.3.29}$$

because ω_{XB} and τ_{XE} are classical-quantum states. Therefore, the right-hand side of (14.3.27) is equal to $\sum_{x\in\mathcal{X}} p(x) I^c(\rho_A^x, \mathcal{N})$. ■

14.3.1.1 Generalized Dephasing Channels

While additivity of coherent information for degradable channels allows for a tractable formula for their quantum capacity, the question about the *strong converse* quantum capacity $\widetilde{Q}(\mathcal{N})$ remains. In other words, is it the case that $\widetilde{Q}(\mathcal{N}) = I^c(\mathcal{N})$ for all degradable quantum channels $\mathcal{N}$? We answer this question here for a particular class of degradable channels.

We consider the class of degradable channels $\mathcal{N}$ called *generalized dephasing channels*. Such channels are defined by the following isometric extension:

$$V^{\mathcal{N}}_{A\to BE} = \sum_{i=0}^{d-1} |i\rangle_B\langle i|_A \otimes |\psi_i\rangle_E, \tag{14.3.30}$$

where $d \geq 1$ and where the state vectors $\{|\psi_i\rangle\}_{i=0}^{d-1}$ are arbitrary (not necessarily orthonormal). Recalling the discussion in Section 4.4.7 on Hadamard channels, in particular (4.4.102), we see that generalized dephasing channels are Hadamard channels, as in (4.4.95), with V therein set to $\mathbb{1}$.

For a state ρ_A, we have that

$$\mathcal{N}_{A\to B}(\rho_A) = \mathrm{Tr}_E[V^{\mathcal{N}}_{A\to BE}\rho_A V^{\mathcal{N}\dagger}_{A\to BE}] = \sum_{i,j=0}^{d-1} \langle i|\rho_A|j\rangle\langle\psi_i|\psi_j\rangle|i\rangle\langle j|_B, \tag{14.3.31}$$

and

$$\mathcal{N}^c_{A\to E}(\rho_A) = \mathrm{Tr}_B[V^{\mathcal{N}}_{A\to BE}\rho_A V^{\mathcal{N}\dagger}_{A\to BE}] = \sum_{i=0}^{d-1} \langle i|\rho_A|i\rangle|\psi_i\rangle\langle\psi_i|_E. \tag{14.3.32}$$

Then, it is straightforward to see that

$$\mathcal{N}^c \circ \mathcal{N}(\rho) = \mathcal{N}^c(\rho) \tag{14.3.33}$$

for every state ρ. This implies that generalized dephasing channels $\mathcal{N}$ are degradable, with $\mathcal{N}^c$ being the degrading channel.

We now show that $\widetilde{Q}(\mathcal{N}) = I^c(\mathcal{N})$ for every generalized dephasing channel $\mathcal{N}$. We do this by showing that the Rains information $R(\mathcal{N})$ of a generalized dephasing channel is equal to its coherent information.

Theorem 14.28 Quantum Capacity of Generalized Dephasing Channels

For every generalized dephasing channel $\mathcal{N}$ (defined by the isometric extension in (14.3.30)), the following equalities hold

$$Q(\mathcal{N}) = \widetilde{Q}(\mathcal{N}) = R(\mathcal{N}) = I^c(\mathcal{N}), \tag{14.3.34}$$

which establish the coherent information as the quantum capacity and strong converse quantum capacity.

Proof: It suffices to show that $R(\mathcal{N}) = I^c(\mathcal{N})$. Note that the inequality $I^c(\mathcal{N}) \leq R(\mathcal{N})$ holds for every quantum channel $\mathcal{N}$ by combining the result of Theorem 14.16 with the result of Theorem 14.22. We now show that the reverse inequality holds for all generalized dephasing channels.

We start by observing that every generalized dephasing channel $\mathcal{N}$ is covariant with respect to the operators $\{Z(j)\}_{j=0}^{d-1}$ defined in (3.2.48):

$$\mathcal{N}(Z(j)\rho Z(j)^\dagger) = Z(j)\mathcal{N}(\rho)Z(j)^\dagger \tag{14.3.35}$$

for every state ρ and for all $0 \leq j \leq d-1$, where

$$Z(j) = \sum_{k=0}^{d-1} e^{\frac{2\pi i k j}{d}} |k\rangle\langle k|. \tag{14.3.36}$$

Then, for every state ρ, the average state

$$\overline{\rho} := \frac{1}{d}\sum_{j=0}^{d-1} Z(j)\rho Z(j)^\dagger = \sum_{k=0}^{d-1} \langle k|\rho|k\rangle |k\rangle\langle k| \tag{14.3.37}$$

is diagonal in the basis $\{|i\rangle\}_{i=0}^{d-1}$. Since the quantities $\langle k|\rho|k\rangle$ are probabilities, we conclude that for every state ρ, its corresponding average state $\overline{\rho}$ has a purification of the following form:

$$|\phi^{\overline{\rho}}\rangle_{RA} = \sum_{i=0}^{d-1} \sqrt{p(i)}|i\rangle_R \otimes |i\rangle_A =: |\psi^p\rangle_{RA}, \tag{14.3.38}$$

where $p : \{0, 1, \dots, d-1\} \to [0, 1]$ is a probability distribution. Therefore, by Proposition 10.12, when calculating the Rains information $R(\mathcal{N})$, it suffices to optimize over the pure states ψ^p_{RA}:

$$R(\mathcal{N}) = \sup_{\psi^p_{RA}} \inf_{\sigma_{RB} \in \mathrm{PPT}'(R:B)} D(\mathcal{N}_{A\to B}(\psi^p_{RA}) \| \sigma_{RB}). \tag{14.3.39}$$

Now, restricting the optimization in the definition of the coherent information $I^c(\mathcal{N})$ to the pure states in (14.3.38), we obtain

$$I^c(\mathcal{N}) = \sup_{\psi_{RA}} \inf_{\sigma_B} D(\mathcal{N}_{A\to B}(\psi_{RA}) \| \mathbb{1}_R \otimes \sigma_B) \tag{14.3.40}$$

$$\geq \sup_{\psi^p_{RA}} \inf_{\sigma_B} D(\mathcal{N}_{A\to B}(\psi^p_{RA}) \| \mathbb{1}_R \otimes \sigma_B) \tag{14.3.41}$$

$$\geq \sup_{\psi^p_{RA}} \inf_{\sigma_B} D(\Delta(\mathcal{N}_{A\to B}(\psi^p_{RA})) \| \Delta(\mathbb{1}_R \otimes \sigma_B)), \tag{14.3.42}$$

where the last line follows from the data-processing inequality for quantum relative entropy, and we introduced the following channel:

$$\Delta(\rho) := \Pi\rho\Pi + (\mathbb{1} - \Pi)\rho(\mathbb{1} - \Pi), \quad \Pi = \sum_{i=0}^{d-1} |i\rangle\langle i|_R \otimes |i\rangle\langle i|_B. \tag{14.3.43}$$

Now, it is straightforward to check that

$$\Pi\mathcal{N}_{A\to B}(\psi^p_{RA})\Pi = \mathcal{N}_{A\to B}(\psi^p_{RA}), \tag{14.3.44}$$

which implies that

$$\Delta(\mathcal{N}_{A\to B}(\psi^p_{RA})) = \Pi\mathcal{N}_{A\to B}(\psi^p_{RA})\Pi. \tag{14.3.45}$$

Therefore, because Π and $\mathbb{1} - \Pi$ project onto orthogonal subspaces, we obtain

$$D(\Delta(\mathcal{N}_{A\to B}(\psi^p_{RA})) \| \Delta(\mathbb{1}_R \otimes \sigma_B))$$

$$= D(\Pi\mathcal{N}_{A\to B}(\psi^p_{RA})\Pi\|\Pi(\mathbb{1}_R\otimes\sigma_B)\Pi) \tag{14.3.46}$$

$$= D\left(\Pi\mathcal{N}_{A\to B}(\psi^p_{RA})\Pi\middle\|\sum_{i=0}^{d-1}q(i)|i\rangle\langle i|_R\otimes|i\rangle\langle i|_B\right), \tag{14.3.47}$$

where the last line follows because

$$\Pi(\mathbb{1}_R\otimes\sigma_B)\Pi = \sum_{i=0}^{d-1}q(i)|i\rangle\langle i|_R\otimes|i\rangle\langle i|_B, \tag{14.3.48}$$

with the probability distribution $q(i) \coloneqq \langle i|\sigma_B|i\rangle$. Note that the right-hand side of the equation above is a state in $\mathrm{PPT}'(R\!:\!B)$. Therefore, we have

$$I^c(\mathcal{N}) \geq \sup_{\psi^p_{RA}}\inf_{\sigma_B} D(\Delta(\mathcal{N}_{A\to B}(\psi^p_{RA}))\|\Delta(\mathbb{1}_R\otimes\sigma_B)) \tag{14.3.49}$$

$$= \sup_{\psi^p_{RA}}\inf_{\sigma_B} D(\Pi\mathcal{N}_{A\to B}(\psi^p_{RA})\Pi\|\Pi(\mathbb{1}_R\otimes\sigma_B)\Pi) \tag{14.3.50}$$

$$= \sup_{\psi^p_{RA}}\inf_{\sigma_B} D\left(\mathcal{N}_{A\to B}(\psi^p_{RA})\middle\|\sum_{i=0}^{d-1}\langle i|\sigma_B|i\rangle|i\rangle\langle i|_R\otimes|i\rangle\langle i|_B\right) \tag{14.3.51}$$

$$\geq \sup_{\psi^p_{RA}}\inf_{\sigma_{RB}\in\mathrm{PPT}'(R:B)} D(\mathcal{N}_{A\to B}(\psi^p_{RA})\|\sigma_{RB}) \tag{14.3.52}$$

$$= R(\mathcal{N}), \tag{14.3.53}$$

completing the proof. ■

14.3.2 Anti-Degradable Channels

Let us now consider anti-degradable channels. Recall from Definition 4.6 that a channel $\mathcal{N}_{A\to B}$ is anti-degradable if there exists an anti-degrading channel $\mathcal{A}_{E\to B}$ such that

$$\mathcal{N} = \mathcal{A}\circ\mathcal{N}^c, \tag{14.3.54}$$

where $\mathcal{N}^c$ is a channel complementary to $\mathcal{N}$ and $d_E \geq \mathrm{rank}(\Gamma^{\mathcal{N}}_{AB})$.

Proposition 14.29 Coherent Information for Anti-Degradable Channels

The coherent information vanishes for all anti-degradable channels, i.e., $I^c(\mathcal{N}) = 0$ for every anti-degradable channel $\mathcal{N}$. Therefore, $Q(\mathcal{N}) = 0$ for all anti-degradable channels.

Proof: Let $\mathcal{N}_{A\to B}$ have the following Stinespring representation: $\mathcal{N}(\rho_A) = \operatorname{Tr}_E[V_{A\to BE}\rho_A V^\dagger_{A\to BE}]$. Then, for every pure state ψ_{RA}, the state vector

$$|\phi\rangle_{RBE} \coloneqq V_{A\to BE}|\psi\rangle_{RA} \tag{14.3.55}$$

is such that

$$I(R\rangle B)_\phi = \frac{1}{2}\left(I(R;B)_\phi - I(R;E)_\phi\right). \tag{14.3.56}$$

To see this, let us first note that $H(E)_\phi = H(RB)_\phi$ and $H(RE)_\phi = H(B)_\phi$. These identities hold because ϕ_{RBE} is a pure state, implying that the reduced states ϕ_E and ϕ_{RB} have the same spectrum and the reduced states ϕ_{RE} and ϕ_B have the same spectrum. This, along with (7.11.103), leads to

$$\frac{1}{2}\left(I(R;B)_\phi - I(R;E)_\phi\right) \tag{14.3.57}$$

$$= \frac{1}{2}\left(H(R)_\phi + H(B)_\phi - H(RB)_\phi - H(R)_\phi - H(E)_\phi + H(RE)_\phi\right) \tag{14.3.58}$$

$$= H(B)_\phi - H(RB)_\phi \tag{14.3.59}$$

$$= I(R\rangle B)_\phi. \tag{14.3.60}$$

Next, using (7.11.104), noting that $\phi_{RB} = \mathcal{N}_{A\to B}(\psi_{RA}) =: \omega_{RB}$, and using the fact that $\mathcal{N}$ is anti-degradable, so that $\mathcal{N} = \mathcal{A}\circ\mathcal{N}^c$, we have

$$I(R;B)_\phi \le I(R;E)_\phi, \tag{14.3.61}$$

where the inequality follows from the data-processing inequality for mutual information under local channels (see (7.2.202)) and the facts that $\mathcal{N}_{A\to B}(\psi_{RA}) = \mathcal{A}_{E\to B}\circ\mathcal{N}^c_{A\to E}(\psi_{RA})$ and the reduced state $\phi_{RE} = \mathcal{N}^c_{A\to E}(\psi_{RA})$. Therefore, from (14.3.56) we conclude that

$$I(R\rangle B)_\omega \le 0 \tag{14.3.62}$$

for every pure state ψ_{RA}. This implies that

$$I^c(\mathcal{N}) = \sup_{\psi_{RA}} I(R\rangle B)_\omega = 0, \tag{14.3.63}$$

as required. ■

14.3.3 Generalized Amplitude Damping Channel

Let us recall the definition of the generalized amplitude damping channel (GADC) from (4.5.10):

$$\mathcal{A}_{\gamma,N}(\rho) = A_1 \rho A_1^\dagger + A_2 \rho A_2^\dagger + A_3 \rho A_3^\dagger + A_4 \rho A_4^\dagger, \tag{14.3.64}$$

where

$$A_1 = \sqrt{1-N}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad A_2 = \sqrt{1-N}\begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{14.3.65}$$

$$A_3 = \sqrt{N}\begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix}, \quad A_4 = \sqrt{N}\begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}, \tag{14.3.66}$$

and $\gamma, N \in [0,1]$. It is straightforward to show that

$$\mathcal{A}_{\gamma,N}(\rho) = X \mathcal{A}_{\gamma,1-N}(X\rho X)X \tag{14.3.67}$$

for every state ρ and all $\gamma, N \in [0,1]$. In other words, the GADC $\mathcal{A}_{\gamma,N}$ is related to the GADC $\mathcal{A}_{\gamma,1-N}$ via a simple pre- and post-processing by the Pauli unitary $X = |0\rangle\langle 1| + |1\rangle\langle 0|$. The information-theoretic aspects of the GADC are thus invariant under the interchange $N \leftrightarrow 1-N$, which means that we can, without loss of generality, restrict the parameter N to the interval $[0, 1/2]$.

For $N = 0$, the GADC reduces to the amplitude damping channel $\mathcal{A}_\gamma$ defined in (4.5.1), which is degradable. Indeed, we first note that

$$\mathcal{A}^c_{\gamma,0} = \mathcal{A}_{1-\gamma,0}, \tag{14.3.68}$$

where the complementary channel $\mathcal{A}^c_{\gamma,0}$ (recall Definition 4.5) is defined via the following isometric extension:

$$V^{\gamma,N} := A_1 \otimes |0\rangle + A_2 \otimes |1\rangle + A_3 \otimes |2\rangle + A_4 \otimes |3\rangle. \tag{14.3.69}$$

We now use the fact that, for all $\gamma_1, \gamma_2, N_1, N_2 \in [0,1]$,

$$\mathcal{A}_{\gamma,N} = \mathcal{A}_{\gamma_2,N_2} \circ \mathcal{A}_{\gamma_1,N_1}, \tag{14.3.70}$$

where $\gamma = \gamma_1 + \gamma_2 - \gamma_1\gamma_2$ and $N = \frac{\gamma_1(1-\gamma_2)N_1 + \gamma_2 N_2}{\gamma_1+\gamma_2-\gamma_1\gamma_2}$. From this fact, it follows that the defining condition for degradability, namely, $\mathcal{D}_{\gamma,0} \circ \mathcal{A}_{\gamma,0} = \mathcal{A}^c_{\gamma,0} = \mathcal{A}_{1-\gamma,0}$, is satisfied by the quantum channel $\mathcal{D}_{\gamma,0} := \mathcal{A}_{\frac{1-2\gamma}{1-\gamma},0}$. It can be shown that for $N > 0$, the GADC $\mathcal{A}_{\gamma,N}$ is not degradable for all $\gamma \in (0,1]$ (please consult the Bibliographic Notes in Section 14.5).

Since $\mathcal{A}_{\gamma,0}$ is degradable, its coherent information is additive, which means that its quantum capacity is equal to its coherent information, i.e.,

$$Q(\mathcal{A}_{\gamma,0}) = I^c(\mathcal{A}_{\gamma,0}) = \sup_{\rho} \left\{ H(\mathcal{A}_{\gamma,0}(\rho)) - H(\mathcal{A}^c_{\gamma,0}(\rho)) \right\} \tag{14.3.71}$$

$$= \sup_{\rho} I^c(\rho, \mathcal{A}_{\gamma,0}), \tag{14.3.72}$$

where we have used the expression in (14.3.2). Now, as explained in Section 11.3.2, the GADC is covariant with respect to the Pauli operator Z. Furthermore, by Lemma 14.27, the function $\rho \mapsto I^c(\rho, \mathcal{A}_{\gamma,0})$ is concave. Therefore, for every state ρ,

$$I^c\left(\frac{1}{2}\rho + \frac{1}{2}Z\rho Z, \mathcal{A}_{\gamma,0}\right) \geq \frac{1}{2}I^c(\rho, \mathcal{A}_{\gamma,0}) + \frac{1}{2}I^c(Z\rho Z, \mathcal{A}_{\gamma,0}). \tag{14.3.73}$$

Now, using the fact that $\mathcal{A}_{\gamma,0}$ is covariant with respect to Z, and the fact that $\mathcal{A}^c_{\gamma,0} = \mathcal{A}_{1-\gamma,0}$, we obtain

$$I^c(Z\rho Z, \mathcal{A}_{\gamma,0}) = H(\mathcal{A}_{\gamma,0}(Z\rho Z)) - H(\mathcal{A}^c_{\gamma,0}(Z\rho Z)) \tag{14.3.74}$$

$$= H(Z\mathcal{A}_{\gamma,0}(\rho)Z) - H(Z\mathcal{A}_{1-\gamma,0}(\rho)Z) \tag{14.3.75}$$

$$= H(\mathcal{A}_{\gamma,0}(\rho)) - H(\mathcal{A}_{1-\gamma,0}(\rho)) \tag{14.3.76}$$

$$= H(\mathcal{A}_{\gamma,0}(\rho)) - H(\mathcal{A}^c_{\gamma,0}(\rho)) \tag{14.3.77}$$

$$= I^c(\rho, \mathcal{A}_{\gamma,0}). \tag{14.3.78}$$

Therefore,

$$I^c\left(\frac{1}{2}\rho + \frac{1}{2}Z\rho Z, \mathcal{A}_{\gamma,0}\right) \geq I^c(\rho, \mathcal{A}_{\gamma,0}) \tag{14.3.79}$$

for every state ρ. Recalling from (4.5.28) that the state $\frac{1}{2}\rho + \frac{1}{2}Z\rho Z$ results from the action of the completely dephasing channel on ρ, which means that it is diagonal in the standard basis, we find that

$$\max_{p\in[0,1]} I^c((1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|, \mathcal{A}_{\gamma,0}) \geq I^c(\rho, \mathcal{A}_{\gamma,0}) \tag{14.3.80}$$

for every state ρ, which means that

$$Q(\mathcal{A}_{\gamma,0}) = I^c(\mathcal{A}_{\gamma,0}) = \max_{p\in[0,1]} I^c((1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|, \mathcal{A}_{\gamma,0}) \tag{14.3.81}$$

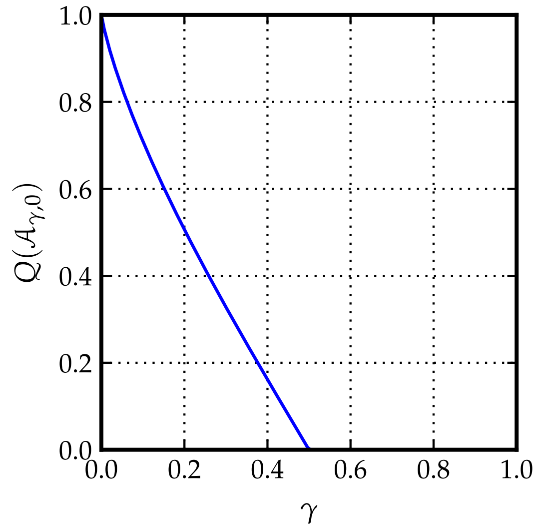


FIGURE 14.5: Quantum capacity of the amplitude damping channel, as given by (14.3.82). The capacity is equal to zero for $\gamma \geq \frac{1}{2}$ because in this parameter range the channel is anti-degradable.

$$= \max_{p\in[0,1]} \{h_2((1-\gamma)p) - h_2(\gamma p)\}, \tag{14.3.82}$$

where in the last line we have evaluated $I^c((1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|, \mathcal{A}_{\gamma,0})$. See Figure 14.5 for a plot of the quantum capacity of the amplitude damping channel $\mathcal{A}_{\gamma,0}$. Note that the capacity vanishes at $\gamma = \frac{1}{2}$, which is due to the fact that for $\gamma \geq \frac{1}{2}$ the amplitude damping channel $\mathcal{A}_{\gamma,0}$ (and more generally the GADC $\mathcal{A}_{\gamma,N}$ for $N \in [0,1]$) is anti-degradable. From Proposition 14.29, we thus have that $Q(\mathcal{A}_{\gamma,N}) = 0$ for all $N \in [0,1]$ and $\gamma \geq \frac{1}{2}$.

Let us now consider the coherent information of the GADC $\mathcal{A}_{\gamma,N}$ for $N > 0$. In this case, the coherent information $I^c(\mathcal{A}_{\gamma,N})$ is a lower bound on the quantum capacity of the GADC. As with the amplitude damping channel, it can be shown that for the GADC $\mathcal{A}_{\gamma,N}$ with $N > 0$ it suffices to optimize over states diagonal in the standard basis in order to compute the coherent information:

$$I^c(\mathcal{A}_{\gamma,N}) = \max_{p\in[0,1]} I^c((1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|, \mathcal{A}_{\gamma,N}), \tag{14.3.83}$$

for all $\gamma \in (0,1)$ and all $N > 0$. The proof of this is more involved, since for $N > 0$ the GADC is not degradable, meaning that we cannot use Lemma 14.27. Please consult the Bibliographic Notes in Section 14.5 for a source of the proof.

In Figure 14.6, we plot the coherent information lower bound given by (14.3.83).

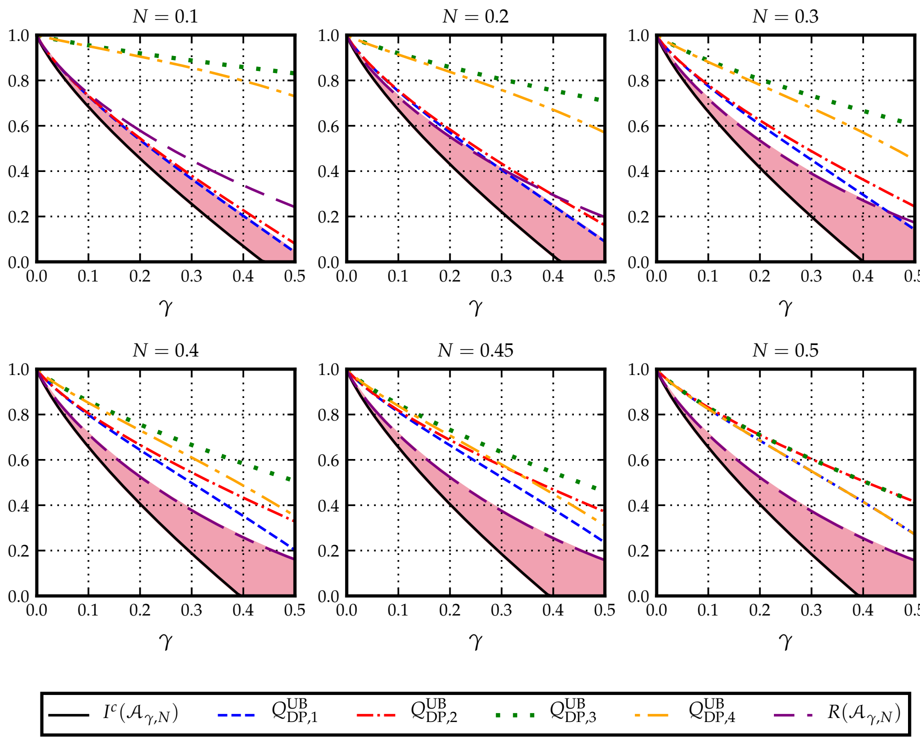


FIGURE 14.6: The coherent information lower bound $I^c(\mathcal{A}_{\gamma,N})$ and four upper bounds on the quantum capacity of the generalized amplitude damping channel $\mathcal{A}_{\gamma,N}$. The quantum capacity lies within the shaded region.

We also plot the Rains information upper bound $R(\mathcal{A}_{\gamma,N})$ as well as four other upper bounds that are based on the following identities, which follow from (14.3.70):

$$\mathcal{A}_{\gamma,N} = \mathcal{A}_{\gamma N,1} \circ \mathcal{A}_{\frac{\gamma(1-N)}{1-\gamma N},0}, \tag{14.3.84}$$

$$\mathcal{A}_{\gamma,N} = \mathcal{A}_{\gamma(1-N),0} \circ \mathcal{A}_{\frac{\gamma N}{1-\gamma(1-N)},1}. \tag{14.3.85}$$

It then follows that

$$Q(\mathcal{A}_{\gamma,N}) \leq Q(\mathcal{A}_{\frac{\gamma(1-N)}{1-\gamma N},0}) =: Q^{\text{UB}}_{\text{DP},1}(\gamma, N), \tag{14.3.86}$$

$$Q(\mathcal{A}_{\gamma,N}) \leq Q(\mathcal{A}_{\gamma(1-N),0}) =: Q^{\text{UB}}_{\text{DP},2}(\gamma, N), \tag{14.3.87}$$

$$Q(\mathcal{A}_{\gamma,N}) \leq Q(\mathcal{A}_{\gamma N,0}) =: Q^{\text{UB}}_{\text{DP},3}(\gamma, N), \tag{14.3.88}$$

$$Q(\mathcal{A}_{\gamma,N}) \leq Q(\mathcal{A}_{\frac{\gamma N}{1-\gamma(1-N)},0}) =: Q^{\text{UB}}_{\text{DP},4}(\gamma, N). \tag{14.3.89}$$

Note that the right-hand side of each inequality can be calculated using (14.3.82). We have also made use of (14.3.67), which implies that $Q(\mathcal{A}_{\gamma,1}) = Q(\mathcal{A}_{\gamma,0})$. These inequalities hold due to the fact that, for the composition of two quantum channels $\mathcal{N}$ and $\mathcal{M}$,

$$Q(\mathcal{N} \circ \mathcal{M}) \leq Q(\mathcal{M}) \quad \text{and} \quad Q(\mathcal{N} \circ \mathcal{M}) \leq Q(\mathcal{N}). \tag{14.3.90}$$

The first inequality holds by the data-processing inequality. The second inequality can be viewed as a lower bound on the quantum capacity of the channel $\mathcal{N}$ that arises from a coding strategy consisting of some encoding followed by many uses of the channel $\mathcal{M}$.

14.4 Summary

In this chapter, we studied quantum communication. Given a quantum channel $\mathcal{N}_{A \to B}$ connecting Alice and Bob, the goal in quantum communication is to determine the highest rate, called the quantum capacity and denoted by $Q(\mathcal{N})$, at which the A' part of an arbitrary pure state $\Psi_{RA'}$ can be transmitted to Bob without error. At the disposal of Alice and Bob are local encoding and decoding channels, as well as an arbitrary number of (unassisted) uses of the channel $\mathcal{N}_{A \to B}$. By unassisted, we mean that Alice and Bob are not allowed to communicate with each other between channel uses. We found that the coherent information $I^c(\mathcal{N})$ of $\mathcal{N}$ is always a lower bound on its quantum capacity, and that, in general, computing the exact value of the capacity involves a regularization, so that $Q(\mathcal{N}) = I^c_{\text{reg}}(\mathcal{N})$.

Starting with the one-shot setting, in which only one use of the channel is allowed and there is some tolerable non-zero error, we determined both upper and lower bounds on the number of qubits that can be transmitted. The one-shot upper bound involves the hypothesis testing relative entropy in a way similar to how it is involved in classical communication and entanglement distillation. Specifically, we establish the hypothesis testing coherent information as an upper bound. This leads to the coherent information (hence regularized coherent information) weak converse upper bound in the asymptotic setting. To obtain a lower bound, we used the results of Chapter 13 on entanglement distillation. We found that we could take the entanglement distillation protocol developed in that chapter and convert it to a suitable quantum communication protocol. We proved that this lower bound is optimal when applied to the asymptotic setting, in the sense that it leads to the coherent information (hence regularized coherent information) as an achievable rate, which matches the upper bound. For degradable channels, we showed that the coherent information is additive, meaning that $Q(\mathcal{N}) = I^c(\mathcal{N})$ for all degradable channels. We also showed that anti-degradable channels have zero quantum capacity.

With the goal of obtaining tractable estimates of quantum capacity for general channels, we found that the Rains information $R(\mathcal{N})$ of $\mathcal{N}$ is a strong converse upper bound on the quantum capacity of $\mathcal{N}$. This allowed us to conclude that the quantum capacity of the generalized dephasing channel is equal to its coherent information, because its Rains information and coherent information coincide. We also looked ahead to Chapter 19 and concluded from the results there that the squashed entanglement of a quantum channel is an upper bound on quantum capacity.

14.5 Bibliographic Notes

The problem of determining the capacity of a quantum channel for transmitting quantum information, in a manner analogous to Shannon's channel capacity theorem, was proposed by Shor (1995). The notion of quantum communication that we consider in this chapter, as well as the notion of entanglement transmission, was defined by Schumacher (1996). The notion of subspace transmission was defined by Barnum et al. (2000) (see also (Bennett et al., 1997)), and the notion of entanglement generation was defined by Devetak (2005). These different notions of quantum communication, and the connections between them, have been examined by Kretschmann and Werner (2004), where they also proved that the capacities for these variations are all equal to each other.

Upper and lower bounds on one-shot quantum capacity have been established by Buscemi and Datta (2010a); Datta and Hsieh (2013); Beigi et al. (2016); Tomamichel et al. (2016); Anshu et al. (2019); Wang et al. (2019b). The approach of using hypothesis testing relative entropy for obtaining an upper bound on one-shot quantum capacity (specifically, Theorem 14.3) comes from work by Matthews and Wehner (2014). The lower bound on the one-shot quantum capacity in Theorem 14.5 comes from work on one-shot decoupling (Dupuis et al., 2014), which was then used by Wilde et al. (2017, Proposition 21) to obtain a lower bound on the one-shot distillable entanglement. The various code conversions in Lemmas 14.6, 14.7, and 14.8 are available in a number of works, including Barnum et al. (2000); Kretschmann and Werner (2004); Klesse (2007); Watrous (2018) (see also Wilde and Qi (2018)). The one-shot Rains upper bound in Corollary 14.20 was obtained by Tomamichel et al. (2017).

In the asymptotic setting, Schumacher (1996); Schumacher and Nielsen (1996); Barnum et al. (1998, 2000) established coherent information as an upper bound on quantum capacity, and Lloyd (1997); Shor (2002b); Devetak (2005) established the lower bound. (See also the proofs of Klesse (2008); Hayden et al. (2008b).) Decoupling as a method for understanding quantum capacity was initially studied by Schumacher and Westmoreland (2002) and developed in further detail by Hayden et al. (2008a). The Rains information strong converse upper bound (Theorem 14.22) was established by Tomamichel et al. (2017). Weak subadditivity of Rényi Rains information of a channel (Proposition 14.21) is also due to Tomamichel et al. (2017). We also mention that upper bounds on quantum capacity based on approximate degradability and approximate anti-degradability of channels have been established by Sutter et al. (2017) (see also (Leditzky et al., 2018)).

Additivity of coherent information for degradable channels was shown by Devetak and Shor (2005). Smith and Yard (2008); Smith et al. (2011) demonstrated the phenomenon of superactivation of quantum capacity, and DiVincenzo et al. (1998); Smith and Smolin (2007); Cubitt et al. (2015); Elkouss and Strelchuk (2015) demonstrated superadditivity of coherent information. Lemma 14.27 was presented in (Yard et al., 2008, Lemma 5). The quantum capacity of generalized dephasing channels was established by Devetak and Shor (2005) and the strong converse by Tomamichel et al. (2017). See (Morgan and Winter, 2014) for the pretty-strong converse for the quantum capacity of degradable channels. The generalized amplitude damping channel (GADC) has been studied in detail by Khatri et al. (2020). The fact that this channel is not degradable for $N \in (0,1)$ and $\gamma \in (0,1]$ follows from (Cubitt et al., 2008, Theorem 4). The expression in (14.3.82) for the quantum capacity of the amplitude damping channel was given by Giovannetti and Fazio (2005). For a proof of (14.3.83), see (García-Patrón et al., 2009, Appendix). A proof of anti-degradability of the GADC $\mathcal{A}_{\gamma,N}$ for all $\gamma \geq \frac{1}{2}$ can be found in (Khatri et al., 2020, Proposition 2).

Appendix 14.A Alternative Notions of Quantum Communication

At the beginning of this chapter, we considered three alternative notions of quantum communication, and we described how they are implied by the notion of quantum communication as defined at the beginning of Section 14.1. We now precisely define these other notions of quantum communication, and we show how the notion of quantum communication considered in the chapter (strong subspace transmission) implies all three alternatives.

Definition 14.30 Entanglement Transmission

An *entanglement transmission protocol* for $\mathcal{N}_{A\to B}$ consists of the three elements $(d, \mathcal{E}, \mathcal{D})$, where $d \geq 1$, $\mathcal{E}_{A'\to A}$ is an encoding channel with $d_{A'} = d$, and $\mathcal{D}_{B\to B'}$ is a decoding channel with $d_{B'} = d$. The goal of the protocol is to transmit the A' system of a maximally entangled state $\Phi_{RA'}$ of Schmidt rank d such that the final state

$$\omega_{RB'} := (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'}) \tag{14.A.1}$$

is close to the initial maximally entangled state. The *entanglement transmission error* of the protocol is

$$\begin{aligned} p_{\text{err}}^{(\text{ET})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) &:= 1 - \langle\Phi|_{RB'}\omega_{RB'}|\Phi\rangle_{RB'} && (14.A.2) \\ &= 1 - F_e(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}), && (14.A.3) \end{aligned}$$

where we recall the entanglement fidelity of a channel from Definition 6.21. We call the protocol $(d, \mathcal{E}, \mathcal{D})$ a (d, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{ET})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

It is straightforward to see that if there exists a (d, ε) quantum communication protocol for a quantum channel $\mathcal{N}$ (as per Definition 14.1), then there exists a (d, ε) entanglement transmission protocol. Indeed, for a (d, ε) quantum communication protocol with encoding and decoding channel $\mathcal{E}$ and $\mathcal{D}$, we have that

$$\begin{aligned} &p_{\text{err}}^{*}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \\ &\quad = \max_{|\Psi\rangle_{RA'}} \{1 - \langle\Psi|_{RA'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Psi_{RA'})|\Psi\rangle_{RA'}\} && (14.A.4) \\ &\quad = 1 - F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) && (14.A.5) \\ &\quad \leq \varepsilon. && (14.A.6) \end{aligned}$$

However, since the maximally entangled state $\Phi_{AB'}$ is a particular pure state in the optimization for $F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})$, we conclude that

$$p_{\text{err}}^{(\text{ET})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) = 1 - F_e(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) \leq 1 - F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) \leq \varepsilon. \tag{14.A.7}$$

So the elements $(d, \mathcal{E}, \mathcal{D})$ form a (d, ε) entanglement transmission protocol.

Definition 14.31 Entanglement Generation

An *entanglement generation protocol* for $\mathcal{N}_{A\to B}$ is defined by the three elements $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$, where $\Psi_{A'A}$ is a pure state with $d_{A'} = d$, and $\mathcal{D}_{B\to B'}$ is a decoding channel with $d_{B'} = d$. The goal of the protocol is to transmit the system A such that the final state

$$\sigma_{A'B'} := (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B})(\Psi_{A'A}) \tag{14.A.8}$$

is close in fidelity to a maximally entangled state of Schmidt rank d. The *entanglement generation error* of the protocol is given by

$$p_{\text{err}}^{(\text{EG})}(\Psi_{A'A}, \mathcal{D}; \mathcal{N}) := 1 - \langle\Phi|_{A'B'}\sigma_{A'B'}|\Phi\rangle_{A'B'} \tag{14.A.9}$$

$$= 1 - F(\Phi_{A'B'}, \sigma_{A'B'}). \tag{14.A.10}$$

We call the protocol $(d, \Psi_{A'A}, \mathcal{D}_{B\to B'})$ a *(d, ε) protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{EG})}(\Psi_{A'A}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

Consider a (d, ε) quantum communication protocol for $\mathcal{N}_{A\to B}$ given by the elements $(d, \mathcal{E}_{A'\to A}, \mathcal{D}_{B\to B'})$, where $d_{A'} = d_{B'} = d$. Then, by the arguments above, the same elements constitute a (d, ε) entanglement transmission protocol, so that

$$\langle\Phi|_{RB'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\Phi_{RA'})|\Phi\rangle_{RB'} \geq 1 - \varepsilon. \tag{14.A.11}$$

Now, let the system $R \equiv \tilde{A}$ belong to Alice, and let $\Psi_{\tilde{A}A} := \mathcal{E}_{A'\to A}(\Phi_{\tilde{A}A})$. Then,

$$\langle\Phi|_{\tilde{A}B'}(\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B})(\Psi_{\tilde{A}A})|\Phi\rangle_{RB'} \geq 1 - \varepsilon. \tag{14.A.12}$$

Therefore, by definition, the elements $(d, \Psi_{\tilde{A}A}, \mathcal{D}_{B\to B'})$ constitute a (d, ε) entanglement generation protocol.

Definition 14.32 Subspace Transmission

A *subspace transmission protocol* over the quantum channel $\mathcal{N}_{A\to B}$ consists of the three elements $(d, \mathcal{E}, \mathcal{D})$, where $d \geq 1$ and $\mathcal{E}$ and $\mathcal{D}$ are encoding and decoding channels. The goal of the protocol is to transmit an arbitrary pure state $\psi_{A'}$ such that the final state

$$\omega_{B'} := (\mathcal{D}_{B\to B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{A'\to A})(\psi_{A'}) \tag{14.A.13}$$

is close in fidelity to the initial state. The *state transmission error* of the protocol is

$$p_{\text{err}}^{(\text{ST})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) := 1 - \min_{\psi} \langle\psi|\mathcal{N}(\psi)|\psi\rangle = 1 - F_{\min}(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}), \tag{14.A.14}$$

where we recall the minimum fidelity of a channel defined in (6.4.5). We call the protocol $(d, \mathcal{E}, \mathcal{D})$ a (d, ε) protocol, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}^{(\text{ST})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

REMARK: An alternative way to define the error criterion for a subspace transmission code would be to use the average fidelity, defined in (6.4.3); please consult the Bibliographic Notes in Section 14.5.

Given a (d, ε) quantum communication protocol for the channel $\mathcal{N}$ with the elements $(d, \mathcal{E}, \mathcal{D})$, the equality $p_{\text{err}}^{*}(\mathcal{E}, \mathcal{D}; \mathcal{N}) = 1 - F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})$ holds, where $F(\cdot)$ is the channel fidelity defined in (6.4.6). Then, restricting the optimization in $F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})$ to pure states $\Psi_{RA'} = |\Psi\rangle\langle\Psi|_{RA'}$ such that $|\Psi\rangle_{RA'} = |\phi\rangle_R \otimes |\psi\rangle_{A'}$, we obtain

$$\begin{aligned}
&p_{\text{err}}^{(\text{ST})}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \\
&\quad = 1 - F_{\min}(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) && (14.A.15)\\
&\quad = 1 - \min_{|\psi\rangle} \langle\psi|\mathcal{N}(\psi)|\psi\rangle && (14.A.16)\\
&\quad = 1 - \min_{|\phi\rangle, |\psi\rangle} \left((\langle\phi|_R \otimes \langle\psi|_{A'})(\phi_R \otimes \mathcal{N}(\psi_{A'}))(|\phi\rangle_R \otimes |\psi\rangle_{A'}) \right) && (14.A.17)\\
&\quad \leq 1 - F(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) && (14.A.18)\\
&\quad \leq \varepsilon. && (14.A.19)
\end{aligned}$$

So the elements $(d, \mathcal{E}, \mathcal{D})$ form a (d, ε) subspace transmission protocol.

Chapter 15

Secret Key Distillation

This chapter considers the task of secret key distillation. The setting of this task is that Alice and Bob share a bipartite quantum state ρ_{AB}, and the goal is for them to perform local operations and public communication in order to transform ρ_{AB} to a state that approximates an ideal secret key. Some questions are in order: What is an ideal secret key and for whom is it secret? How much secret key can they extract from this state? These are the main questions addressed in this chapter.

The information-theoretic model we assume is that the physical laboratories of Alice and Bob are secure, so that system A of the state ρ_{AB} is physically secured in Alice's laboratory and system B is physically secured in Bob's. We suppose that an eavesdropper Eve possesses a system E that purifies ρ_{AB}. That is, if ψ_{ABE} is a purification of ρ_{AB}, then we suppose that system E of ψ_{ABE} is in Eve's possession. This model gives the eavesdropper a lot of power. Indeed, if $\omega_{ABE'}$ is an arbitrary extension of the state ρ_{AB}, then as a consequence of Proposition 4.4, Eve can transform ψ_{ABE} to $\omega_{ABE'}$ by means of a channel acting on her system E. We also assume that any classical data transmitted between Alice and Bob is public, so that Eve has access to all of it.

An ideal secret key of $\log_2 K$ secret bits is a tripartite state of the following form:

$$\overline{\Phi}_{AB} \otimes \sigma_E, \tag{15.0.1}$$

where

$$\overline{\Phi}_{AB} := \frac{1}{K} \sum_{i=0}^{K-1} |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B. \tag{15.0.2}$$

There are three salient aspects of such a tripartite key state:

1. The key value is uniformly random and thus hard to guess.
2. The key values in the registers of Alice and Bob are perfectly correlated. That is, if Alice measures the key value to be $i \in \{0, \ldots, K-1\}$, then Bob is guaranteed to measure the same value.
3. The overall state is a product state between systems AB and E. This means that Eve's system E is of no use in guessing the key value.

The goal of a secret-key distillation protocol is for Alice and Bob to transform the initial state ψ_{ABE}, by means of local operations and public classical communication, to a state that approximates an ideal key state of the form in (15.0.1).

A secret key is useful in a communication task called the one-time pad protocol (also known as the Vernam cipher). In this protocol, we suppose that Alice has a message $m \in \{0, \ldots, K-1\}$ that she would like to send to Bob. By making use of the key, Alice can calculate $\tilde{m} \coloneqq m \oplus i$, where i is the key value and the addition is modulo K, and then send the encrypted message $\tilde{m}$ over a public classical channel. Since the key is ideal, no one else besides Alice and Bob knows the precise key value i, and the encrypted message $\tilde{m}$ is uniformly random, which means that it is hard to guess (i.e., there is a $1/K$ chance that an eavesdropper could guess it, which becomes small as K becomes large). When Bob receives the encrypted message $\tilde{m}$, he can calculate $m = \tilde{m} \ominus i$ and decrypt the message m because he knows the key value i. This is one of the main uses of a secret key and in turn why we are interested in secret key distillation.

It turns out that there are strong connections between entanglement distillation from Chapter 13 and secret key distillation. They are not precisely the same tasks but there are strong links, and the structure of this chapter follows the structure of Chapter 13 quite closely. The main reason for the strong connection is that the maximally entangled state $\Phi_{AB} = \frac{1}{K}\sum_{i,j=0}^{K-1} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B$ can be used to generate an ideal key state. To see this, consider that the state Φ_{AB} is unextendible, so that the only possible extension of it is a tensor-product extension of the form $\Phi_{AB} \otimes \sigma_E$. Then, if Alice and Bob perform local measurement channels on their systems A and B, with respect to the computational basis, they can realize the ideal tripartite key state of the form in (15.0.1). Thus, if one can generate maximally entangled states, then one can generate key states. However, the converse is not true in general, and this is what distinguishes secret key distillation from entanglement distillation.

Similar to what we have done in previous chapters, here we establish lower and upper bounds on the number of secret key bits that can be distilled from a bipartite state ρ_{AB}. The lower bounds are given in terms of the private information of the state, and the upper bounds are given in terms of not only the private information but also the squashed entanglement and the relative entropy of entanglement. The fact that we can use entanglement measures as bounds further highlights the connection between secret key distillation and entanglement.

15.1 One-Shot Setting

The one-shot setting for secret key distillation begins with Alice and Bob sharing a state ρ_{AB}, and we assume that the eavesdropper Eve has access to a system E of a purification of ρ_{AB}. For concreteness, let ψ_{ABE} denote the purification of ρ_{AB}, with system A of ψ_{ABE} held by Alice, B by Bob, and E by Eve. Keep in mind that all purifications of ρ_{AB} are related by an isometric channel acting on the E system, so that Eve can reach all purifications easily by performing an isometric channel on her system E. The model we assume is that the laboratory of Alice is physically secure and the quantum system A is fully contained in it. Similarly, we assume that the laboratory of Bob is physically secure and contains the system B. However, if the state ρ_{AB} is mixed, then the purifying degrees of freedom in E are available to Eve (if, on the other hand, ρ_{AB} is pure, then an arbitrary purification ψ_{ABE} is always a tensor-product state of the systems AB and E and, in this sense, it is understood that Eve does not really have access to purifying degrees of freedom). This approach gives the most power to the eavesdropper for the setting of secret key distillation.

In a secret-key distillation protocol, Alice and Bob are allowed to use local operations and public classical communication (abbreviated as LOPC). An LOPC channel is similar to an LOCC channel (as discussed in Section 4.6.2), but the critical difference is that Eve gets a copy of all of the classical data exchanged. Recall that a generic LOCC channel $\mathcal{L}_{AB\to A'B'}$ can be written as follows, as discussed in Definition 4.22:

$$\mathcal{L}_{AB\to A'B'} = \sum_{z\in\mathcal{Z}} \mathcal{E}^z_{A\to A'} \otimes \mathcal{F}^z_{B\to B'}, \tag{15.1.1}$$

where $\mathcal{Z}$ is a finite alphabet and $\{\mathcal{E}^z_{A\to A'}\}_{z\in\mathcal{Z}}$ and $\{\mathcal{F}^z_{B\to B'}\}_{z\in\mathcal{Z}}$ are sets of completely positive maps such that the sum map $\mathcal{L}_{AB\to A'B'}$ is trace preserving. Then an LOPC channel is the following enlargement of $\mathcal{L}_{AB\to A'B'}$:

$$\mathcal{L}_{AB\to A'B'Z} = \sum_{z\in\mathcal{Z}} \mathcal{E}^z_{A\to A'} \otimes \mathcal{F}^z_{B\to B'} \otimes |z\rangle\langle z|_Z, \tag{15.1.2}$$

such that Eve has access to the system Z, which contains all of the classical data exchanged to realize $\mathcal{L}_{AB\to A'B'}$.

The goal of a secret-key distillation protocol is for Alice and Bob to produce an approximation of an ideal secret-key state, which is defined as follows:

Definition 15.1 Tripartite Key State

A state γ_{ABE} is a tripartite key state of size K, or containing $\log_2 K$ bits of secrecy, if local measurements of the A and B systems lead to the same uniformly random outcome and the system E is product with the measurement outcomes. That is, after Alice and Bob send their systems through local dephasing (measurement) channels

$$\mathcal{M}_{A,B}(\cdot) := \sum_{i=0}^{K-1} |i\rangle\langle i|_{A,B}(\cdot)|i\rangle\langle i|_{A,B}, \tag{15.1.3}$$

the resulting state on AB and E is as follows:

$$(\mathcal{M}_A \otimes \mathcal{M}_B)(\gamma_{ABE}) = \overline{\Phi}_{AB} \otimes \sigma_E, \tag{15.1.4}$$

for some state σ_E and where $\overline{\Phi}_{AB}$ is the maximally classically correlated state

$$\overline{\Phi}_{AB} := \frac{1}{K}\sum_{i=0}^{K-1} |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B. \tag{15.1.5}$$

As stated in Definition 15.1, the defining aspect of an ideal tripartite key state is that the systems A and B of Alice and Bob are perfectly correlated and uniformly random. This property makes the actual key value, which ends up being observed by both Alice and Bob, hard to guess if there are many key values. Furthermore, the fact that the overall state is such that it is tensor product between AB and E implies that Eve's system cannot provide any help at all in guessing the key value.

With the notions above in place, we can now formally define a secret-key distillation protocol. Such a protocol for the state ρ_{AB} is defined by the pair

$(K, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ})$, where $K \in \mathbb{N}$ and $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}$ is an LOPC channel as defined in (15.1.2), with $d_{K_A} = d_{K_B} = K$. The *key distillation error* $p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB})$ of the protocol is given by the infidelity, defined as

$$p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}) \coloneqq \inf_{\gamma_{K_AK_BEZ}} \left(1 - F(\gamma_{K_AK_BEZ}, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}(\psi_{ABE}))\right), \tag{15.1.6}$$

where the optimization is with respect to every tripartite key state $\gamma_{K_AK_BEZ}$ of size K, which is of the form in Definition 15.1 under the identifications $K_A \leftrightarrow A$, $K_B \leftrightarrow B$, and $EZ \leftrightarrow E$. Furthermore, ψ_{ABE} is a purification of ρ_{AB}. Note that the key distillation error $p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB})$ is invariant under the choice of a purification ψ_{ABE} because it involves an optimization over every tripartite key state $\gamma_{K_AK_BEZ}$, purifications are related by isometric channels, the fidelity is invariant under isometric channels, and $\mathcal{V}_E(\gamma_{K_AK_BEZ})$ is an ideal tripartite key state if $\gamma_{K_AK_BEZ}$ is, where $\mathcal{V}_E$ is an isometric channel. The optimization in (15.1.6) guarantees the existence of at least one state σ_{EZ} of the eavesdropper such that the actual state $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}(\psi_{ABE})$ of the protocol approximates an ideal tripartite key state, in the following sense

$$(\mathcal{M}_{K_A} \otimes \mathcal{M}_{K_B})(\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}(\psi_{ABE})) \approx (\mathcal{M}_{K_A} \otimes \mathcal{M}_{K_B})(\gamma_{K_AK_BEZ}) \tag{15.1.7}$$

$$= \overline{\Phi}_{K_AK_B} \otimes \sigma_{EZ}, \tag{15.1.8}$$

if the key distillation error $p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB})$ is small.

At this point, it might not be clear why we employ the infidelity error criterion in (15.1.6) rather than the normalized trace distance. We did so in Chapter 13 in the context of entanglement distillation because it corresponded to the operational notion of an entanglement test (see (13.1.4)). We later show how the infidelity error criterion corresponds to the operational notion of a "privacy test," which justifies its use in the context of secret key distillation.

Definition 15.2 (K, ε) secret-key distillation protocol

A secret key distillation protocol $(K, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ})$ for the state ρ_{AB} is called a (K, ε) *protocol*, with $\varepsilon \in [0, 1]$, if $p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}) \leq \varepsilon$.

Given $\varepsilon \in [0, 1]$, the largest number $\log_2 K$ of ε-approximate secret-key bits that can be extracted from a state ρ_{AB} among all (K, ε) secret-key distillation protocols is called the *one-shot ε-distillable key of ρ_{AB}*.

Definition 15.3 One-Shot Distillable Key

Given a bipartite state ρ_{AB} and $\varepsilon \in [0,1]$, the *one-shot distillable key of* ρ_{AB}, denoted by $K_D^\varepsilon(\rho_{AB}) \equiv K_D^\varepsilon(A;B)_\rho$, is defined as

$$K_D^\varepsilon(A;B)_\rho := \sup_{(K,\mathcal{L}^{\leftrightarrow})} \left\{\log_2 K : p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB}) \leq \varepsilon\right\}, \tag{15.1.9}$$

where the optimization is over all $K \in \mathbb{N}$ and every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}$ with $d_{K_A} = d_{K_B} = K$.

Calculating the one-shot distillable key is difficult computationally because it involves optimizing over the key size K and over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}$, with $d_{K_A} = d_{K_B} = K$. We thus try to estimate the one-shot distillable key by determining upper and lower bounds on it. Section 15.1.3 introduces upper bounds on the one-shot distillable key. Before doing so, we first clarify how secret-key distillation protocols can be thought of from a different perspective as bipartite private-state distillation protocols.

15.1.1 Tripartite Key States and Bipartite Private States

An important insight for secret key distillation is that there is a way to describe the whole theory exclusively in terms of a bipartite scenario. This is related to the assumption that the eavesdropper Eve possesses a full purification ψ_{ABE} of the original state ρ_{AB}, along with the structure of quantum mechanics.

To motivate this concept, consider that an approximate tripartite state γ_{ABE} (as described in Definition 15.1) is generated at the end of a key distillation protocol, and it is such that all that the eavesdropper possesses is only available in the system E (in this context, let us make the same identifications $K_A \leftrightarrow A$, $K_B \leftrightarrow B$, and $EZ \leftrightarrow E$ discussed around (15.1.6)). As such, we can consider a purification of the state γ_{ABE} of the form $\gamma_{AA'BB'E}$, in which the joint system $A'B'$ constitutes the purifying system. Since a secret-key distillation protocol involves only three parties, and we already argued that the system E is all that Eve possesses, it follows that Alice and Bob jointly possess the purifying system, which can be split among them as $A'B'$. The reduced state $\gamma_{AA'BB'} = \text{Tr}_E[\gamma_{AA'BB'E}]$ is then a bipartite state because all systems involved are in possession of Alice and Bob. If the original state γ_{ABE} is a tripartite key state according to Definition 15.1, then by constructing

$\gamma_{AA'BB'}$ according to this procedure, the resulting state is called a bipartite private state, and it has a particular structure. Conversely, if $\gamma_{AA'BB'}$ is a state with the structure of a bipartite private state, then it follows that by purifying this state to $\gamma_{AA'BB'E}$ with an E system and tracing over systems A' and B', we arrive at a tripartite key state. So there is an equivalence between these two viewpoints (tripartite picture of key distillation and bipartite picture of private state distillation). We develop this correspondence in detail in what follows.

Before starting, we briefly mention that the equivalence between the tripartite and bipartite pictures of key distillation implies that we can bring the tools of entanglement theory (Chapter 9) to bear on the problem of establishing upper bounds on the number of approximate secret-key bits that can be generated in a key distillation protocol. This is one of the main applications of this correspondence, and we note here that it has led to other insights in quantum information theory.

Definition 15.4 Bipartite Private State

A state $\gamma_{ABA'B'}$ is a bipartite private state of size K, containing $\log_2 K$ bits of secrecy, if after purifying $\gamma_{ABA'B'}$ to a pure state $\gamma_{ABA'B'E}$ with purifying system E and tracing over the systems $A'B'$, the resulting state γ_{ABE} is a tripartite key state of size K. The systems A and B are called key systems, and the systems A' and B' are called shield systems.

Theorem 15.5

A state $\gamma_{ABA'B'}$ is a bipartite private state if and only if it has the following form:

$$\gamma_{ABA'B'} = U_{ABA'B'} \left(\Phi_{AB} \otimes \theta_{A'B'}\right) U^{\dagger}_{ABA'B'}, \tag{15.1.10}$$

where Φ_{AB} is a maximally entangled state of Schmidt rank K:

$$\Phi_{AB} := \frac{1}{K} \sum_{i,j=0}^{K-1} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B, \tag{15.1.11}$$

$\theta_{A'B'}$ is some state, and $U_{ABA'B'}$ is a global twisting unitary of the following form:

$$U_{ABA'B'} := \sum_{i,j=0}^{K-1} |i\rangle\langle i|_A \otimes |j\rangle\langle j|_B \otimes U^{ij}_{A'B'}. \tag{15.1.12}$$

In the above, $U^{ij}_{A'B'}$ is a unitary operator for all $i, j \in \{0, \ldots, K-1\}$.

PROOF: Suppose that $\gamma_{ABA'B'}$ has the form in (15.1.10). A particular purification of $\gamma_{ABA'B'}$ is

$$\begin{aligned} &|\phi^\gamma\rangle_{ABA'B'E} \\ &= U_{ABA'B'}|\Phi\rangle_{AB} \otimes |\psi^\theta\rangle_{A'B'E} && (15.1.13) \\ &= \left(\sum_{i,j=0}^{K-1} |i\rangle\langle i|_A \otimes |j\rangle\langle j|_B \otimes U^{ij}_{A'B'}\right)\left(\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1} |k\rangle_A|k\rangle_B \otimes |\psi^\theta\rangle_{A'B'E}\right) && (15.1.14) \\ &= \frac{1}{\sqrt{K}}\sum_{k=0}^{K-1} |k\rangle_A|k\rangle_B \otimes U^{kk}_{A'B'}|\psi^\theta\rangle_{A'B'E}, && (15.1.15) \end{aligned}$$

where $|\psi^\theta\rangle_{A'B'E}$ purifies $\theta_{A'B'}$. The local dephasing channels in (15.1.3) lead to the following state

$$\begin{aligned} &(\mathcal{M}_A \otimes \mathcal{M}_B)(|\phi^\gamma\rangle\langle\phi^\gamma|_{ABA'B'E}) \\ &\qquad = \frac{1}{K}\sum_{k=0}^{K-1} |k\rangle\langle k|_A \otimes |k\rangle\langle k|_B \otimes U^{kk}_{A'B'}|\psi^\theta\rangle\langle\psi^\theta|_{A'B'E}\left(U^{kk}_{A'B'}\right)^\dagger. && (15.1.16) \end{aligned}$$

Taking a partial trace over the $A'B'$ systems leads to

$$\begin{aligned} &\mathrm{Tr}_{A'B'}\left[\frac{1}{K}\sum_{k=0}^{K-1} |k\rangle\langle k|_A \otimes |k\rangle\langle k|_B \otimes U^{kk}_{A'B'}|\psi^\theta\rangle\langle\psi^\theta|_{A'B'E}\left(U^{kk}_{A'B'}\right)^\dagger\right] \\ &= \frac{1}{K}\sum_{k=0}^{K-1} |k\rangle\langle k|_A \otimes |k\rangle\langle k|_B \otimes \mathrm{Tr}_{A'B'}\left[U^{kk}_{A'B'}|\psi^\theta\rangle\langle\psi^\theta|_{A'B'E}\left(U^{kk}_{A'B'}\right)^\dagger\right] && (15.1.17) \\ &= \frac{1}{K}\sum_{k=0}^{K-1} |k\rangle\langle k|_A \otimes |k\rangle\langle k|_B \otimes \mathrm{Tr}_{A'B'}\left[\left(U^{kk}_{A'B'}\right)^\dagger U^{kk}_{A'B'}|\psi^\theta\rangle\langle\psi^\theta|_{A'B'E}\right] && (15.1.18) \\ &= \overline{\Phi}_{AB} \otimes \rho_E. && (15.1.19) \end{aligned}$$

Thus, the particular purification $|\phi^\gamma\rangle_{ABA'B'E}$ leads to a tripartite key state on systems ABE. Now, in the development above, we chose a particular purification of $\gamma_{ABA'B'}$. However, given that all purifications are related by isometries acting on the purifying system, every purification can be written as $V_{E\to E'}|\phi^{\gamma}\rangle_{ABA'B'E}$ for some isometry $V_{E\to E'}$. Then repeating the calculation above gives that the reduced state on ABE' after local dephasing channels on A and B is

$$\overline{\Phi}_{AB} \otimes V_{E\to E'}\rho_E(V_{E\to E'})^{\dagger}, \tag{15.1.20}$$

so that there is no correlation between the measurement outcomes of Alice and Bob and the system E'. Furthermore, the measurement outcomes are perfectly correlated and uniformly random. So we conclude that a state $\gamma_{ABA'B'}$ of the form in (15.1.10) is a bipartite private state.

Conversely, suppose now that $\gamma_{ABA'B'}$ is a bipartite private state held by Alice and Bob, and let $|\phi^{\gamma}\rangle_{ABA'B'E}$ be a purification of it, with E the purifying system. Expanding the state in the basis of the local measurements of Alice and Bob gives

$$|\phi^{\gamma}\rangle_{ABA'B'E} = \sum_{i,j=0}^{K-1} \alpha_{i,j}|i\rangle_A|j\rangle_B|\phi_{i,j}\rangle_{A'B'E}, \tag{15.1.21}$$

for some states $|\phi_{i,j}\rangle_{A'B'E}$ and probability amplitudes $\{\alpha_{i,j}\}_{i,j}$. However, in order for the measurement outcomes of Alice and Bob to be perfectly correlated and uniformly random, it is necessary that

$$|\alpha_{i,j}|^2 = \begin{cases} \frac{1}{K} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \tag{15.1.22}$$

(Any other values for the amplitudes $\alpha_{i,j}$ would lead to a different distribution upon measurement of the A and B systems.) So the global state should have the following form:

$$|\phi^{\gamma}\rangle_{ABA'B'E} = \sum_{i=0}^{K-1} \frac{1}{\sqrt{K}}|i\rangle_A|i\rangle_B e^{i\varphi_i}|\phi_{i,i}\rangle_{A'B'E}. \tag{15.1.23}$$

In order for the reduced density operator on E to be independent of the measurement outcomes of Alice and Bob, it is necessary for it to be a fixed state with no dependence on i:

$$\mathrm{Tr}_{A'B'}[|\phi_{i,i}\rangle\langle\phi_{i,i}|_{A'B'E}] = \sigma_E. \tag{15.1.24}$$

In such a case, then all of the states $|\phi_{i,i}\rangle_{A'B'E}$ are purifications of the same state σ_E, so that there exists a unitary $U^i_{A'B'}$ relating each $|\phi_{i,i}\rangle_{A'B'E}$ to a fixed purification $|\phi^{\sigma}\rangle_{A'B'E}$ of σ:

$$e^{i\varphi_i}|\phi_{i,i}\rangle_{A'B'E} = U^i_{A'B'}|\phi^{\sigma}\rangle_{A'B'E}. \tag{15.1.25}$$

Thus, we can write the global state as

$$\frac{1}{\sqrt{K}}\sum_{i=0}^{K-1}|i\rangle_A|i\rangle_B U^{i,i}_{A'B'}|\phi^\sigma\rangle_{A'B'E}, \tag{15.1.26}$$

which is equivalent to

$$\left(\sum_{i,j=0}^{K-1}|i\rangle\langle i|_A\otimes|j\rangle\langle j|_B\otimes U^{i,j}_{A'B'}\right)|\Phi\rangle_{AB}\otimes|\phi^\sigma\rangle_{A'B'E}, \tag{15.1.27}$$

after setting $U^{i,j}_{A'B'} = U^i_{A'B'}$ for all $j \in \{0,\ldots,K-1\}$ (there is in fact full freedom in how the unitary $U^{i,j}_{A'B'}$ is chosen for $i \neq j$). One can now deduce that the reduced state on systems $ABA'B'$ has the form in (15.1.10). ■

Definition 15.6 ε-Approximate Tripartite Key State

Fix $\varepsilon \in [0,1]$. A state ρ_{ABE} is an ε-approximate tripartite key state if there exists a tripartite key state γ_{ABE}, as in Definition 15.1, such that

$$F(\rho_{ABE},\gamma_{ABE}) \geq 1-\varepsilon. \tag{15.1.28}$$

Similarly, a state $\rho_{ABA'B'}$ is an ε-approximate bipartite private state if there exists a bipartite private state $\gamma_{ABA'B'}$, as in Definition 15.4, such that

$$F(\rho_{ABA'B'},\gamma_{ABA'B'}) \geq 1-\varepsilon. \tag{15.1.29}$$

Approximate tripartite key states are in one-to-one correspondence with approximate bipartite private states, as summarized below:

Proposition 15.7

If $\rho_{ABA'B'}$ is an ε-approximate bipartite key state with K key values, then the state ρ_{ABE} is an ε-approximate tripartite key state with K key values, where $\rho_{ABE} = \mathrm{Tr}_{A'B'}[\psi^\rho_{ABA'B'E}]$ and $\psi^\rho_{ABA'B'E}$ is an arbitrary purification of $\rho_{ABA'B'}$. The converse statement is true as well.

Proof: Suppose that the inequality in (15.1.28) is satisfied. Let $\psi^\rho_{ABA'B'E}$ be a purification of ρ_{ABE}. Then by applying Uhlmann's theorem (Theorem 6.8), there exists a purification $\gamma_{ABA'B'E}$ of γ_{ABE} such that

$$F(\rho_{ABE}, \gamma_{ABE}) = F(\psi^{\rho}_{ABA'B'E}, \gamma_{ABA'B'E}). \tag{15.1.30}$$

Tracing over the E system and applying the data-processing inequality for fidelity (Theorem 6.9), we conclude that

$$F(\psi^{\rho}_{ABA'B'}, \gamma_{ABA'B'}) \geq 1 - \varepsilon. \tag{15.1.31}$$

Since γ_{ABE} is an ideal tripartite key state and the state $\gamma_{ABA'B'}$ arises from it via purification and tracing over system E, it follows from Definition 15.4 that $\gamma_{ABA'B'}$ is an ideal bipartite private state. In turn, according to Definition 15.6, it follows that $\psi^{\rho}_{ABA'B'}$ is an ε-approximate bipartite private state.

For the other implication, suppose that the inequality in (15.1.29) is satisfied. Let $\psi^{\rho}_{ABA'B'E}$ be a purification of $\rho_{ABA'B'}$. By applying Uhlmann's theorem (Theorem 6.8), there exists a purification $\gamma_{ABA'B'E}$ of the ideal bipartite private state $\gamma_{ABA'B'}$ such that

$$F(\rho_{ABA'B'}, \gamma_{ABA'B'}) = F(\psi^{\rho}_{ABA'B'E}, \gamma_{ABA'B'E}) \tag{15.1.32}$$

Tracing over the $A'B'$ systems and applying the data-processing inequality for fidelity, we conclude that

$$F(\psi^{\rho}_{ABE}, \gamma_{ABE}) \geq 1 - \varepsilon. \tag{15.1.33}$$

Since $\gamma_{ABA'B'}$ is an ideal bipartite private state and the state γ_{ABE} arises from it via purification and tracing over systems $A'B'$, it follows from Definition 15.4 that γ_{ABE} is an ideal tripartite key state. In turn, according to Definition 15.6, it follows that ψ^{ρ}_{ABE} is an ε-approximate tripartite key state. ■

15.1.2 Equivalence of Tripartite Key Distillation and Bipartite Private State Distillation

The equivalence between ideal and approximate tripartite key states and bipartite private states extends further, and it is a correspondence that allows us to consider secret key distillation in the bipartite picture. To this end, we define a bipartite private-state distillation protocol, and then we prove the equivalence.

A bipartite private-state distillation protocol for the state ρ_{AB} is defined by the pair $(K, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'})$, where $K \in \mathbb{N}$ and $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'}$ is an LOCC channel with $d_{K_A} = d_{K_B} = K$. The key distillation error $p^b_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB})$ of the protocol is given in terms of the infidelity, defined as

$$p^b_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}) \coloneqq \inf_{\gamma_{K_A K_B A' B'}} \left(1 - F(\gamma_{K_A K_B A' B'}, \mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B A' B'}(\rho_{AB}))\right), \tag{15.1.34}$$

where the optimization is with respect to every bipartite private state $\gamma_{K_A K_B A' B'}$ such that $d_{K_A} = d_{K_B} = K$.

Definition 15.8 (K, ε) Private-State Distillation Protocol

A bipartite private-state distillation protocol $(K, \mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B A' B'})$ for the state ρ_{AB} is called a (K, ε) protocol, with $\varepsilon \in [0, 1]$, if $p^b_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}) \leq \varepsilon$.

We now establish the main result of this section, which is the equivalence of tripartite key distillation and bipartite private-state distillation:

Theorem 15.9

Let $K \in \mathbb{N}$ and $\varepsilon \in [0, 1]$. Let ρ_{AB} be a bipartite state. There exists a (K, ε) tripartite key distillation protocol for ρ_{AB} if and only if there exists a (K, ε) bipartite private-state distillation protocol for ρ_{AB}.

Proof: We start by proving that there exists a (K, ε) bipartite private-state distillation protocol if there exists a (K, ε) tripartite key distillation protocol. Let ψ_{ABE} be a purification of ρ_{AB}, let $\mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B Z}$ be the LOPC channel realizing the key distillation, and let $\gamma_{K_A K_B E Z}$ be a tripartite key state such that

$$1 - F(\gamma_{K_A K_B E Z}, \mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B Z}(\psi_{ABE})) \leq \varepsilon. \tag{15.1.35}$$

The LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B Z}$ has the form in (15.1.2), so that

$$\mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B Z} = \sum_{z \in \mathcal{Z}} \mathcal{E}^z_{A \to K_A} \otimes \mathcal{F}^z_{B \to K_B} \otimes |z\rangle\langle z|_Z. \tag{15.1.36}$$

An isometric extension $U^{\mathcal{L}^{\leftrightarrow}}_{AB \to K_A K_B A' B' Z}$ of this LOPC channel is as follows:

$$U^{\mathcal{L}^{\leftrightarrow}}_{AB \to K_A K_B A' B' Z} \coloneqq \sum_{z \in \mathcal{Z}} V^{\mathcal{E}^z}_{A \to K_A A'} \otimes V^{\mathcal{F}^z}_{B \to K_B B'} \otimes |z\rangle_Z, \tag{15.1.37}$$

where $\{V^{\mathcal{E}^z}_{A\to K_A A'}\}_{z\in\mathcal{Z}}$ and $\{V^{\mathcal{F}^z}_{B\to K_B B'}\}_{z\in\mathcal{Z}}$ are sets of linear operators such that $U^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}$ is an isometry and

$$\mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z} = \mathcal{L}^{\leftrightarrow}_{AB\to K_A K_B Z}, \tag{15.1.38}$$

with

$$\mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}(\cdot) := U^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}(\cdot)(U^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z})^{\dagger}. \tag{15.1.39}$$

To meet these requirements, note that it is necessary for each $V^{\mathcal{E}^z}_{A\to K_A A'}$ and $V^{\mathcal{F}^z}_{B\to K_B B'}$ to be a contraction, i.e., satisfying

$$\left\|V^{\mathcal{E}^z}_{A\to K_A A'}\right\|_{\infty}, \left\|V^{\mathcal{F}^z}_{B\to K_B B'}\right\|_{\infty} \leq 1. \tag{15.1.40}$$

It then follows that the state $\mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}(\psi_{ABE})$ purifies $\mathcal{L}^{\leftrightarrow}_{AB\to K_A K_B Z}(\psi_{ABE})$, and by applying Uhlmann's theorem (Theorem 6.8), there exists a pure state $\gamma_{K_A K_B A'B'EZ}$ satisfying

$$\begin{aligned} F(\gamma_{K_A K_B EZ}, \mathcal{L}^{\leftrightarrow}_{AB\to K_A K_B Z}(\psi_{ABE})) \\ = F(\gamma_{K_A K_B A'B'EZ}, \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}(\psi_{ABE})). \end{aligned} \tag{15.1.41}$$

Now applying the same reasoning given in Proposition 15.7, we conclude that the following inequality holds

$$1 - F(\gamma_{K_A K_B A'B'}, (\mathrm{Tr}_Z \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z})(\rho_{AB})) \leq \varepsilon, \tag{15.1.42}$$

where $\gamma_{K_A K_B A'B'}$ is an ideal bipartite private state of size K. Note that the channel $\mathrm{Tr}_Z \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z}$ is an LOCC channel, because it has the following form:

$$\mathrm{Tr}_Z \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z} = \sum_{z\in\mathcal{Z}} \mathcal{V}^{\mathcal{E}^z}_{A\to K_A A'} \otimes \mathcal{V}^{\mathcal{F}^z}_{B\to K_B B'}. \tag{15.1.43}$$

Thus, there exists a (K, ε) bipartite private-state distillation protocol if there exists a (K, ε) tripartite key distillation protocol.

We now prove the opposite implication. Suppose that there exists a (K, ε) bipartite private-state distillation protocol. Let $\mathcal{L}^{\leftrightarrow}_{AB\to K_A K_B A'B'}$ be the LOCC channel realizing the private-state distillation, and let $\gamma_{K_A K_B A'B'}$ be an ideal bipartite private state satisfying

$$1 - F(\gamma_{K_A K_B A'B'}, \mathcal{L}^{\leftrightarrow}_{AB\to K_A K_B A'B'}(\rho_{AB})) \leq \varepsilon. \tag{15.1.44}$$

Let ψ_{ABE} be a purification of ρ_{AB}. Suppose that the LOCC channel $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'}$ has the following form:

$$\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'} = \sum_{z\in\mathcal{Z}} \mathcal{E}^z_{A\to K_AA'} \otimes \mathcal{F}^z_{B\to K_BB'}, \tag{15.1.45}$$

where $\mathcal{Z}$ is a finite alphabet and $\{\mathcal{E}^z_{A\to K_AA'}\}_{z\in\mathcal{Z}}$ and $\{\mathcal{F}^z_{B\to K_BB'}\}_{z\in\mathcal{Z}}$ are sets of completely positive maps such that $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'}$ is trace preserving. Without loss of generality, we can suppose that each completely positive map $\mathcal{E}^z_{A\to K_AA'}$ consists of a single Kraus operator, and we can suppose the same for $\mathcal{F}^z_{B\to K_BB'}$ (the reasoning here is similar to that given in the remark after Definition 3.5). Let us denote these as $E^z_{A\to K_AA'}$ and $F^z_{B\to K_BB'}$, respectively. Then an isometric extension of this LOCC channel is as follows:

$$U^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z} := \sum_{z\in\mathcal{Z}} E^z_{A\to K_AA'} \otimes F^z_{B\to K_BB'} \otimes |z\rangle_Z. \tag{15.1.46}$$

Thus, the state $\mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z}(\psi_{ABE})$ is a purification of $\mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BA'B'}(\rho_{AB})$. Applying Uhlmann's theorem (Theorem 6.8), it follows that there exists a purification $\gamma_{K_AK_BA'B'EZ}$ of $\gamma_{K_AK_BA'B'}$ satisfying

$$F(\gamma_{K_AK_BA'B'EZ}, \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z}(\psi_{ABE})) \geq 1-\varepsilon. \tag{15.1.47}$$

Tracing over systems $A'B'$ and applying the data-processing inequality for fidelity, we conclude that

$$F(\gamma_{K_AK_BEZ}, (\mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z})(\psi_{ABE})) \geq 1-\varepsilon. \tag{15.1.48}$$

Now by applying the same reasoning in Proposition 15.7, we conclude that the state $\gamma_{K_AK_BEZ}$ is an ideal tripartite key state. However, the channel

$$\mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z} \tag{15.1.49}$$

is not necessarily an LOPC channel due to the coherence of the Z system with the other systems. We can apply a completely dephasing channel $\overline{\Delta}_Z(\cdot) := \sum_{z\in\mathcal{Z}} |z\rangle\langle z|_Z(\cdot)|z\rangle\langle z|_Z$ to the Z system, and the fidelity does not decrease under the action of this channel, implying that

$$F(\overline{\Delta}_Z(\gamma_{K_AK_BEZ}), (\overline{\Delta}_Z \circ \mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z})(\psi_{ABE})) \geq 1-\varepsilon. \tag{15.1.50}$$

The state $\overline{\Delta}_Z(\gamma_{K_AK_BEZ})$ is a tripartite key state, and the channel

$$\overline{\Delta}_Z \circ \mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_AK_BA'B'Z} \tag{15.1.51}$$

is an LOPC channel, being explicitly written as follows:

$$\overline{\Delta}_Z \circ \mathrm{Tr}_{A'B'} \circ \mathcal{U}^{\mathcal{L}^{\leftrightarrow}}_{AB\to K_A K_B A'B'Z} = \sum_{z\in\mathcal{Z}} \mathcal{E}^z_{A\to K_A} \otimes \mathcal{F}^z_{B\to K_B} \otimes |z\rangle\langle z|_Z, \tag{15.1.52}$$

where

$$\mathcal{E}^z_{A\to K_A}(\cdot) := \mathrm{Tr}_{A'}[E^z_{A\to K_A A'}(\cdot)(E^z_{A\to K_A A'})^\dagger], \tag{15.1.53}$$

$$\mathcal{F}^z_{B\to K_B}(\cdot) := \mathrm{Tr}_{B'}[F^z_{B\to K_B B'}(\cdot)(F^z_{B\to K_B B'})^\dagger]. \tag{15.1.54}$$

Thus, we have proven that the existence of a (K, ε) bipartite private-state distillation protocol for ρ_{AB} implies the existence of a (K, ε) tripartite key distillation protocol. ■

15.1.2.1 Entanglement Distillation and Secret Key Distillation

The equivalence between tripartite key distillation and bipartite private-state distillation allows us to relate secret key distillation to entanglement distillation. Indeed, a maximally entangled state Φ_{AB} is a particular kind of bipartite private state in which the shield systems $A'B'$ are trivial and the twisting unitary $U_{ABA'B'}$ is the identity. This and Theorem 15.9 imply that a (K, ε) entanglement distillation protocol is a (K, ε) secret-key distillation protocol. However, the converse is not necessarily true because it is not generally possible to convert a bipartite private state of size K to a maximally entangled state of Schmidt rank K.

As a consequence of the discussion above, it follows that the one-shot distillable entanglement of a bipartite state ρ_{AB} is a lower bound on the one-shot distillable key of ρ_{AB}:

$$E^\varepsilon_D(A;B)_\rho \le K^\varepsilon_D(A;B)_\rho \tag{15.1.55}$$

for all $\varepsilon \in [0, 1]$. Since this relationship holds on the fundamental one-shot level, it also holds for the asymptotic quantities as well:

$$E_D(A;B)_\rho \le K_D(A;B)_\rho, \tag{15.1.56}$$

$$\widetilde{E}_D(A;B)_\rho \le \widetilde{K}_D(A;B)_\rho, \tag{15.1.57}$$

where $E_D(A;B)_\rho$ is the distillable entanglement of ρ_{AB} (Definition 13.15), $K_D(A;B)_\rho$ is the distillable key of ρ_{AB} (given later in Definition 15.28), and $\widetilde{E}_D(A;B)_\rho$ and $\widetilde{K}_D(A;B)_\rho$ are the strong converse quantities.

15.1.3 Upper Bounds on the Number of Secret-Key Bits

In this section, we provide three different upper bounds on one-shot distillable key, based on private information, relative entropy of entanglement, and squashed entanglement.

15.1.3.1 Private Information Upper Bound

Our study of upper bounds on one-shot distillable key begins with the private information and the following lemma.

Lemma 15.10

Let A and B be quantum systems with the same dimension $K \in \mathbb{N}$, let E be another quantum system of arbitrary dimension, and let $\varepsilon \in (0,1)$. Let ω_{ABE} be an ε-approximate tripartite key state of size K, as specified in Definition 15.6, and let $\omega^{\mathcal{M}}_{ABE} := (\mathcal{M}_A \otimes \mathcal{M}_B)(\omega_{ABE})$, where $\mathcal{M}_A$ and $\mathcal{M}_B$ are the measurement channels in Definition 15.6. Then the following inequality holds

$$\log_2 K \le I_H^{\sqrt{\varepsilon}+\delta}(A;B)_{\omega^{\mathcal{M}}} - I_{\max}^{\sqrt{\varepsilon}}(A;E)_{\omega^{\mathcal{M}}} + \log_2\left(\frac{1}{\delta}\right), \tag{15.1.58}$$

where $\delta \in (0, 1-\sqrt{\varepsilon})$ and

$$I_{\max}^{\sqrt{\varepsilon}}(A;E)_{\omega^{\mathcal{M}}} := \inf_{\widetilde{\omega}_{AE}: P(\widetilde{\omega}_{AE},\omega^{\mathcal{M}}_{AE}) \le \sqrt{\varepsilon}} \inf_{\tau_E} D_{\max}(\widetilde{\omega}_{AE} \| \widetilde{\omega}_A \otimes \tau_E). \tag{15.1.59}$$

Proof: Consider that the following condition holds from Definition 15.6:

$$F(\gamma_{ABE}, \omega_{ABE}) \ge 1 - \varepsilon, \tag{15.1.60}$$

where γ_{ABE} is an ideal tripartite key state. Applying the measurement channels $\mathcal{M}_A$ and $\mathcal{M}_B$ from Definition 15.1 and the data-processing inequality for fidelity, we conclude that

$$F(\overline{\Phi}_{AB} \otimes \sigma_E, (\mathcal{M}_A \otimes \mathcal{M}_B)(\omega_{ABE})) \ge 1 - \varepsilon. \tag{15.1.61}$$

Now tracing over system E and again applying the data-processing inequality for fidelity, we conclude that

$$F(\overline{\Phi}_{AB}, \omega^{\mathcal{M}}_{AB}) = F(\overline{\Phi}_{AB}, (\mathcal{M}_A \otimes \mathcal{M}_B)(\omega_{AB})) \ge 1 - \varepsilon. \tag{15.1.62}$$

Observe that the state $\omega^{\mathcal{M}}_{AB}$ is a classical state and can be written as

$$\omega^{\mathcal{M}}_{AB} = \sum_{i,j=0}^{K-1} p(i)q(j|i)|i\rangle\langle i|_A \otimes |j\rangle\langle j|_B, \tag{15.1.63}$$

for a probability distribution $p(i)$ and a conditional probability distribution $q(j|i)$. If we perform the comparator test $\{\Pi_{AB}, I_{AB} - \Pi_{AB}\}$ on systems AB of $\omega^{\mathcal{M}}_{AB}$, where

$$\Pi_{AB} \coloneqq \sum_{i=0}^{K-1} |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B, \tag{15.1.64}$$

then the probability of passing it is given by

$$\begin{aligned} &\mathrm{Tr}[\Pi_{AB}\omega^{\mathcal{M}}_{AB}] \\ &= \mathrm{Tr}\left[\left(\sum_{i'=0}^{K-1} |i'\rangle\langle i'|_A \otimes |i'\rangle\langle i'|_B\right)\left(\sum_{i,j=0}^{K-1} p(i)q(j|i)|i\rangle\langle i|_A \otimes |j\rangle\langle j|_B\right)\right] &(15.1.65)\\ &= \sum_{i=0}^{K-1} p(i)q(i|i) &(15.1.66)\end{aligned}$$

Now consider the following channel:

$$\mathcal{T}_{AB}(\tau_{AB}) \coloneqq \mathrm{Tr}[\Pi_{AB}\tau_{AB}]|1\rangle\langle 1| + \mathrm{Tr}[(I_{AB} - \Pi_{AB})\,\tau_{AB}]|0\rangle\langle 0|, \tag{15.1.67}$$

which outputs a classical flag register indicating if the comparator test is successful or not. Consider that

$$\begin{aligned} \mathcal{T}_{AB}(\overline{\Phi}_{AB}) &= |1\rangle\langle 1|, &(15.1.68)\\ \mathcal{T}_{AB}(\omega^{\mathcal{M}}_{AB}) &= \left(\sum_{i=0}^{K-1} p(i)q(i|i)\right)|1\rangle\langle 1| + \left(1 - \sum_{i=0}^{K-1} p(i)q(i|i)\right)|0\rangle\langle 0|. &(15.1.69)\end{aligned}$$

Employing the data-processing inequality for the fidelity and the findings above, we conclude that

$$\begin{aligned} 1 - \varepsilon &\leq F(\overline{\Phi}_{AB}, \omega^{\mathcal{M}}_{AB}) &(15.1.70)\\ &\leq F(\mathcal{T}_{AB}(\overline{\Phi}_{AB}), \mathcal{T}_{AB}(\omega^{\mathcal{M}}_{AB})) &(15.1.71)\end{aligned}$$

$$= \sum_{i=0}^{K-1} p(i)q(i|i). \tag{15.1.72}$$

Thus, we conclude that the probability of passing the comparator test satisfies

$$\operatorname{Tr}[\Pi_{AB}\omega_{AB}^{\mathcal{M}}] \geq 1 - \varepsilon. \tag{15.1.73}$$

Now let Π_A^δ be the projection onto the positive eigenspace of $\frac{1}{\delta}\overline{\Phi}_A - \omega_A^{\mathcal{M}}$, where $\delta \in (0,1)$. Consider that

$$\Pi_A^\delta \left(\frac{1}{\delta}\overline{\Phi}_A - \omega_A^{\mathcal{M}}\right) \Pi_A^\delta \geq 0 \qquad \Longrightarrow \qquad \Pi_A^\delta \omega_A^{\mathcal{M}} \Pi_A^\delta \leq \frac{1}{\delta}\Pi_A^\delta \overline{\Phi}_A \Pi_A^\delta, \tag{15.1.74}$$

and

$$\left(I_A - \Pi_A^\delta\right)\left(\frac{1}{\delta}\overline{\Phi}_A - \omega_A^{\mathcal{M}}\right)\left(I_A - \Pi_A^\delta\right) \leq 0 \tag{15.1.75}$$

$$\Longrightarrow \qquad \operatorname{Tr}[(I_A - \Pi_A^\delta)\overline{\Phi}_A] \leq \delta \operatorname{Tr}[(I_A - \Pi_A^\delta)\omega_A^{\mathcal{M}}] \leq \delta. \tag{15.1.76}$$

The latter inequality can be rewritten as

$$\operatorname{Tr}[\Pi_A^\delta \overline{\Phi}_A] \geq 1 - \delta. \tag{15.1.77}$$

Also, let σ_B be an arbitrary state, and consider that

$$\operatorname{Tr}[\Pi_A^\delta \Pi_{AB} \Pi_A^\delta (\omega_A^{\mathcal{M}} \otimes \sigma_B)] = \operatorname{Tr}[\Pi_{AB}(\Pi_A^\delta \omega_A^{\mathcal{M}} \Pi_A^\delta \otimes \sigma_B)] \tag{15.1.78}$$

$$\leq \frac{1}{\delta}\operatorname{Tr}[\Pi_{AB}(\Pi_A^\delta \overline{\Phi}_A \Pi_A^\delta \otimes \sigma_B)] \tag{15.1.79}$$

$$\leq \frac{1}{\delta}\operatorname{Tr}[\Pi_{AB}(\overline{\Phi}_A \otimes \sigma_B)] \tag{15.1.80}$$

$$= \frac{1}{\delta K}\operatorname{Tr}[\Pi_{AB}(I_A \otimes \sigma_B)] \tag{15.1.81}$$

$$= \frac{1}{\delta K}, \tag{15.1.82}$$

where the second inequality follows because Π_A^δ and $\overline{\Phi}_A$ commute. Then consider that

$$\operatorname{Tr}[(I_{AB} - \Pi_A^\delta \Pi_{AB} \Pi_A^\delta)\omega_{AB}^{\mathcal{M}}]$$

$$\leq \operatorname{Tr}[(I_{AB} - \Pi_A^\delta \Pi_{AB} \Pi_A^\delta)\overline{\Phi}_{AB}] + \frac{1}{2}\left\|\overline{\Phi}_{AB} - \omega_{AB}^{\mathcal{M}}\right\|_1 \tag{15.1.83}$$

$$\leq \text{Tr}[(I_{AB} - \Pi_{AB})\overline{\Phi}_{AB}] + \text{Tr}[(I_{AB} - \Pi_A^\delta \otimes I_B)\overline{\Phi}_{AB}] + \frac{1}{2}\left\|\overline{\Phi}_{AB} - \omega_{AB}^{\mathcal{M}}\right\|_1 \tag{15.1.84}$$

$$= \text{Tr}[(I_A - \Pi_A^\delta)\overline{\Phi}_A] + \frac{1}{2}\left\|\overline{\Phi}_{AB} - \omega_{AB}^{\mathcal{M}}\right\|_1 \tag{15.1.85}$$

$$\leq \delta + \sqrt{1 - F(\overline{\Phi}_{AB}, \omega_{AB}^{\mathcal{M}})} \tag{15.1.86}$$

$$\leq \delta + \sqrt{\varepsilon}. \tag{15.1.87}$$

The first inequality is a consequence of the variational characterization of the normalized trace distance from Theorem 6.1. The second inequality is a consequence of the following union bound for commuting projectors P and Q:

$$I - PQP \leq I - P + I - Q, \tag{15.1.88}$$

which in turn follows from $(I - P)\,(I - Q) \geq 0$. The third inequality follows from Theorem 6.14, and the last from (15.1.62). As such, the measurement operator $\Pi_A^\delta \Pi_{AB} \Pi_A^\delta$ is a particular measurement operator satisfying the contraints given in the optimization for the hypothesis testing relative entropy $D_H^{\sqrt{\varepsilon}+\delta}(\omega_{AB}^{\mathcal{M}} \| \omega_A^{\mathcal{M}} \otimes \sigma_B)$, and we thus conclude that

$$\log_2 \delta + \log_2 K = \log_2 \delta K \tag{15.1.89}$$

$$\leq -\log_2 \text{Tr}[\Pi_A^\delta \Pi_{AB} \Pi_A^\delta (\omega_A^{\mathcal{M}} \otimes \sigma_B)] \tag{15.1.90}$$

$$\leq D_H^{\sqrt{\varepsilon}+\delta}(\omega_{AB}^{\mathcal{M}} \| \omega_A^{\mathcal{M}} \otimes \sigma_B). \tag{15.1.91}$$

Since the bound holds for every state σ_B, we conclude that

$$\log_2 K \leq I_H^{\sqrt{\varepsilon}+\delta}(A;B)_{\omega^{\mathcal{M}}} + \log_2\left(\frac{1}{\delta}\right). \tag{15.1.92}$$

Now we aim to show that

$$I_{\max}^{\sqrt{\varepsilon}}(A;E)_\omega \leq 0. \tag{15.1.93}$$

Consider that (15.1.61) implies that

$$F(\overline{\Phi}_A \otimes \sigma_E, \omega_{AE}^{\mathcal{M}}) \geq 1 - \varepsilon. \tag{15.1.94}$$

Thus, the state $\overline{\Phi}_A \otimes \sigma_E$ is such that

$$P(\overline{\Phi}_A \otimes \sigma_E, \omega_{AE}^{\mathcal{M}}) \leq \sqrt{\varepsilon}. \tag{15.1.95}$$

Then

$$I_{\max}^{\sqrt{\varepsilon}}(A;E)_{\omega^{\mathcal{M}}} = \inf_{\widetilde{\omega}_{AE}:P(\widetilde{\omega}_{AE},\omega_{AE}^{\mathcal{M}})\leq\sqrt{\varepsilon}} \inf_{\tau_E} D_{\max}(\widetilde{\omega}_{AE}\|\widetilde{\omega}_A \otimes \tau_E) \tag{15.1.96}$$

$$\leq D_{\max}(\overline{\Phi}_A \otimes \sigma_E\|\overline{\Phi}_A \otimes \sigma_E) \tag{15.1.97}$$

$$= 0, \tag{15.1.98}$$

where the inequality follows from the choices $\widetilde{\omega}_{AE} = \overline{\Phi}_A \otimes \sigma_E$, $\tau_E = \sigma_E$, (15.1.95), and the fact that $D_{\max}(\rho\|\rho) = 0$ for every state ρ. ■

Note that the result of Lemma 15.10 is general and applies to every tripartite state that is close in fidelity to an ideal tripartite key state. Applying it to the state $\omega_{K_AK_BEZ} = \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BEZ}(\psi_{ABE})$ that is the final output of a (K,ε) tripartite key distillation protocol for a state ρ_{AB} with purification ψ_{ABE}, we obtain the following result:

Theorem 15.11 Upper Bound on One-Shot Distillable Key

Let ρ_{AB} be a bipartite state with purification ψ_{ABE}. For every (K,ε) tripartite key distillation protocol $(K, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BEZ})$ for ψ_{ABE}, with $\varepsilon \in (0,1)$ and $d_{K_A} = d_{K_B} = K$, the number of ε-approximate secret-key bits extracted at the end of the protocol is bounded from above by the LOPC-optimized private information of ρ_{AB}, i.e.,

$$\log_2 K \leq \sup_{\mathcal{L}} \left(I_H^{\sqrt{\varepsilon}+\delta}(X;B')_{\mathcal{L}(\psi)} - I_{\max}^{\sqrt{\varepsilon}}(X;EZ)_{\mathcal{L}(\psi)} \right) + \log_2\!\left(\frac{1}{\delta}\right), \tag{15.1.99}$$

where $\delta \in (0, 1-\sqrt{\varepsilon})$ and the optimization is over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$, where X and Z are classical systems. Consequently, for the one-shot ε-distillable key, the following bound holds

$$K_D^{\varepsilon}(A;B)_{\rho} \leq \sup_{\mathcal{L}} \left(I_H^{\sqrt{\varepsilon}+\delta}(X;B')_{\mathcal{L}(\psi)} - I_{\max}^{\sqrt{\varepsilon}}(X;EZ)_{\mathcal{L}(\psi)} \right) + \log_2\!\left(\frac{1}{\delta}\right). \tag{15.1.100}$$

Proof: For a (K,ε) tripartite key distillation protocol $(K, \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ})$ for ψ_{ABE}, by definition the state $\omega_{K_AK_BEZ} = \mathcal{L}^{\leftrightarrow}_{AB\to K_AK_BZ}(\psi_{ABE})$ satisfies

$$F(\gamma_{K_AK_BEZ}, \omega_{K_AK_BEZ}) \geq 1-\varepsilon, \tag{15.1.101}$$

where $\gamma_{K_A K_B EZ}$ is an ideal tripartite key state. Upon performing the local measurements $\mathcal{M}_{K_A}$ and $\mathcal{M}_{K_B}$ mentioned in Definition 15.1, we conclude that

$$F(\overline{\Phi}_{K_A K_B} \otimes \sigma_{EZ}, (\mathcal{M}_{K_A} \otimes \mathcal{M}_{K_B})(\omega_{K_A K_B EZ})) \geq 1 - \varepsilon. \tag{15.1.102}$$

Set

$$\omega^{\mathcal{M}}_{K_A K_B EZ} \coloneqq (\mathcal{M}_{K_A} \otimes \mathcal{M}_{K_B})(\omega_{K_A K_B EZ}). \tag{15.1.103}$$

Therefore, using (15.1.58), we conclude that

$$\log_2 K \leq I_H^{\sqrt{\varepsilon}+\delta}(K_A; K_B)_{\omega^{\mathcal{M}}} - I_{\max}^{\sqrt{\varepsilon}}(K_A; EZ)_{\omega^{\mathcal{M}}} + \log_2\left(\frac{1}{\delta}\right), \tag{15.1.104}$$

where $\delta \in (0, 1 - \sqrt{\varepsilon})$. Since $(\mathcal{M}_{K_A} \otimes \mathcal{M}_{K_B}) \circ \mathcal{L}^{\leftrightarrow}_{AB \to K_A K_B Z}$ is a particular LOPC channel of the form $\mathcal{L}^{\leftrightarrow}_{AB \to XB'Z}$, with X and Z classical systems, we conclude that

$$\begin{aligned} I_H^{\sqrt{\varepsilon}+\delta}&(K_A; K_B)_{\omega^{\mathcal{M}}} - I_{\max}^{\sqrt{\varepsilon}}(K_A; EZ)_{\omega^{\mathcal{M}}} \\ &\leq \sup_{\mathcal{L}} \left(I_H^{\sqrt{\varepsilon}+\delta}(X; B')_{\mathcal{L}(\psi)} - I_{\max}^{\sqrt{\varepsilon}}(X; EZ)_{\mathcal{L}(\psi)} \right). \end{aligned} \tag{15.1.105}$$

We thus conclude (15.1.99). Now employing the definition of the one-shot ε-distillable key in (15.1.9), we conclude (15.1.100). ■

15.1.3.2 Relative Entropy of Entanglement Upper Bound

We now consider an upper bound based on the relative entropy of entanglement (Section 9.2). In order to place an upper bound on the one-shot distillable key $K_D^{\varepsilon}(A; B)_\rho$ for a given state ρ_{AB} and $\varepsilon \in [0, 1]$, we consider state that are useless for key distillation. This approach is analogous to what we did previously for entanglement distillation in Section 13.1.1.

Which states are useless for key distillation? Suppose that a state σ_{AB} is separable, so that it can be written as

$$\sigma_{AB} = \sum_{x \in \mathcal{X}} p(x) \psi^x_A \otimes \varphi^x_B, \tag{15.1.106}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0, 1]$ is a probability distribution, and $\{\psi^x_A\}_{x \in \mathcal{X}}$ and $\{\varphi^x_B\}_{x \in \mathcal{X}}$ are sets of pure states. Consistent with the model of key distillation that we have discussed so far, the eavesdropper Eve is allowed to have access to the purifying system E of a purification of σ_{AB}, which in this case can be chosen as follows:

$$\psi_{ABE} = |\psi\rangle\langle\psi|_{ABE}, \tag{15.1.107}$$

$$|\psi\rangle_{ABE} := \sum_{x\in\mathcal{X}} \sqrt{p(x)}|\psi^x\rangle_A \otimes |\varphi^x\rangle_B \otimes |x\rangle_E. \tag{15.1.108}$$

Then Eve can measure the system E and obtain an outcome $x \in \mathcal{X}$ with probability $p(x)$, and the resulting state of Alice and Bob is the product state $\psi^x_A \otimes \varphi^x_B$. Being a product state, the resulting state $\psi^x_A \otimes \varphi^x_B$ of Alice and Bob has no correlation whatsoever and so cannot be used to generate a secret key. If Alice and Bob attempt to process this state using LOCC, the same problem arises. In the model of key distillation that we assume, Eve gets a copy of all classical data exchanged between Alice and Bob, and so the resulting state is still a product state and is useless for generating a secret key. As such, *all separable states are useless for key distillation*.

The intuition above is useful for reasoning about key distillation, but there is a way to make it precise by means of a construct called the "privacy test." In doing so, we exploit the equivalence between tripartite key distillation and bipartite private-state distillation discussed in Section 15.1.2 and identified in Theorem 15.9. The privacy test is analogous to the entanglement test used in Chapter 13, which we used to establish upper bounds on the number of approximate ebits that can be generated in an entanglement distillation protocol. Here we define a "privacy test" as a method for testing whether a given bipartite state is private. It forms an essential component in Proposition 15.15, which states that the ε-relative entropy of entanglement is an upper bound on the number of private bits in an ε-approximate bipartite private state.

Definition 15.12 Privacy Test

Let $\gamma_{ABA'B'}$ be a bipartite private state as given in Definition 15.4. A privacy test corresponding to $\gamma_{ABA'B'}$ (a γ-privacy test) is defined as the following measurement:

$$\{\Pi_{ABA'B'}, I_{ABA'B'} - \Pi_{ABA'B'}\}, \tag{15.1.109}$$

where

$$\Pi_{ABA'B'} := U_{ABA'B'} (\Phi_{AB} \otimes I_{A'B'}) U^\dagger_{ABA'B'} \tag{15.1.110}$$

and $U_{ABA'B'}$ is the twisting unitary specified in (15.1.12).

If one has access to the systems $ABA'B'$ of a bipartite state $\rho_{ABA'B'}$ and has a description of $\gamma_{ABA'B'}$ satisfying (15.1.29), then the γ-privacy test decides whether $\rho_{ABA'B'}$ is a private state with respect to $\gamma_{ABA'B'}$. The first outcome corresponds to the decision "yes, it is a γ-private state," and the second outcome corresponds to "no." Physically, this test is just untwisting the purported private state and projecting onto a maximally entangled state. The following lemma states that the probability for an ε-approximate bipartite private state to pass the γ-privacy test is not smaller than $1-\varepsilon$:

Lemma 15.13

Let $\varepsilon \in [0,1]$ and let $\rho_{ABA'B'}$ be an ε-approximate private state as given in Definition 15.6, with $\gamma_{ABA'B'}$ satisfying (15.1.29). The probability for $\rho_{ABA'B'}$ to pass the γ-privacy test is never smaller than $1-\varepsilon$:

$$\operatorname{Tr}[\Pi_{ABA'B'}\rho_{ABA'B'}] \geq 1-\varepsilon, \tag{15.1.111}$$

where $\Pi_{ABA'B'}$ is defined in (15.1.110).

Proof: One can see this bound explicitly by inspecting the following steps:

$$\begin{aligned}
&\operatorname{Tr}[\Pi_{ABA'B'}\rho_{ABA'B'}] \\
&= \operatorname{Tr}[U_{ABA'B'}\,(\Phi_{AB}\otimes I_{A'B'})\,U^{\dagger}_{ABA'B'}\rho_{ABA'B'}] &(15.1.112)\\
&= \operatorname{Tr}[(\Phi_{AB}\otimes I_{A'B'})\,U^{\dagger}_{ABA'B'}\rho_{ABA'B'}U_{ABA'B'}] &(15.1.113)\\
&= \langle\Phi|_{AB}\operatorname{Tr}_{A'B'}[U^{\dagger}_{ABA'B'}\rho_{ABA'B'}U_{ABA'B'}]|\Phi\rangle_{AB} &(15.1.114)\\
&= F(\Phi_{AB},\operatorname{Tr}_{A'B'}[U^{\dagger}_{ABA'B'}\rho_{ABA'B'}U_{ABA'B'}]) &(15.1.115)\\
&\geq F(\Phi_{AB}\otimes\theta_{A'B'},U^{\dagger}_{ABA'B'}\rho_{ABA'B'}U_{ABA'B'}) &(15.1.116)\\
&= F(U_{ABA'B'}(\Phi_{AB}\otimes\theta_{A'B'})U^{\dagger}_{ABA'B'},\rho_{ABA'B'}) &(15.1.117)\\
&= F(\gamma_{ABA'B'},\rho_{ABA'B'}) &(15.1.118)\\
&\geq 1-\varepsilon. &(15.1.119)
\end{aligned}$$

The third equality follows because Φ_{AB} is pure and by taking applying the definition of partial trace (over $A'B'$). The fourth equality follows from the expression in (6.2.2), for the fidelity between a pure state and a mixed state. The first inequality follows from the data-processing inequality for fidelity. The second-to-last equality follows from the unitary invariance of the fidelity, and the last equality follows because $\gamma_{ABA'B'}$ is an ideal private state, written as $\gamma_{ABA'B'} =$

$U_{ABA'B'}(\Phi_{AB} \otimes \theta_{A'B'})U^{\dagger}_{ABA'B'}$. ■

On the other hand, a separable state $\sigma_{ABA'B'} \in \mathrm{SEP}(AA' : BB')$ of the key and shield systems has a small chance of passing an arbitrary γ-privacy test:

Lemma 15.14

For a separable state $\sigma_{ABA'B'} \in \mathrm{SEP}(AA' : BB')$, the probability of passing an arbitrary γ-privacy test is not larger than $\frac{1}{K}$:

$$\mathrm{Tr}[\Pi_{ABA'B'}\sigma_{ABA'B'}] \leq \frac{1}{K}, \tag{15.1.120}$$

where K is the number of values that the secret key can take (i.e., $K = d_A = d_B$).

PROOF: The idea is to begin by establishing the bound for an arbitrary pure product state $|\phi\rangle_{AA'} \otimes |\varphi\rangle_{BB'}$, i.e., to show that

$$\mathrm{Tr}[\Pi_{ABA'B'}|\phi\rangle\langle\phi|_{AA'} \otimes |\varphi\rangle\langle\varphi|_{BB'}] \leq \frac{1}{K}. \tag{15.1.121}$$

We can expand these states with respect to the standard bases of A and B as follows:

$$|\phi\rangle_{AA'} \otimes |\varphi\rangle_{BB'} = \left[\sum_{i=1}^{K} \alpha_i |i\rangle_A \otimes |\phi_i\rangle_{A'}\right] \otimes \left[\sum_{j=1}^{K} \beta_j |j\rangle_B \otimes |\varphi_j\rangle_{B'}\right], \tag{15.1.122}$$

where $\sum_{i=1}^{K} |\alpha_i|^2 = \sum_{j=1}^{K} \left|\beta_j\right|^2 = 1$. We then find that

$$\begin{aligned}
&\mathrm{Tr}[\Pi_{ABA'B'}|\phi\rangle\langle\phi|_{AA'} \otimes |\varphi\rangle\langle\varphi|_{BB'}] \\
&\quad = \mathrm{Tr}[U_{ABA'B'}\,(\Phi_{AB} \otimes I_{A'B'})\,U^{\dagger}_{ABA'B'}|\phi\rangle\langle\phi|_{AA'} \otimes |\varphi\rangle\langle\varphi|_{BB'}] && (15.1.123)\\
&\quad = \left\|(\langle\Phi|_{AB} \otimes I_{A'B'})\,U^{\dagger}_{ABA'B'}|\phi\rangle_{AA'} \otimes |\varphi\rangle_{BB'}\right\|_2^2 && (15.1.124)\\
&\quad = \left\| \begin{array}{c} \frac{1}{\sqrt{K}}\left(\sum_{i=1}^{K}\langle i|_A \otimes \langle i|_B \otimes U^{ii\dagger}_{A'B'}\right) \times \\ \left(\sum_{i',j'=1}^{K} \alpha_{i'}\beta_{j'}|i'\rangle_A \otimes |j'\rangle_B \otimes |\phi_{i'}\rangle_{A'}|\varphi_{j'}\rangle_{B'}\right) \end{array} \right\|_2^2 && (15.1.125)\\
&\quad = \frac{1}{K}\left\|\sum_{i,i',j'=1}^{K} \alpha_{i'}\beta_{j'}\langle i|i'\rangle_A \otimes \langle i|j'\rangle_B \otimes U^{ii\dagger}_{A'B'}|\phi_{i'}\rangle_{A'}|\varphi_{j'}\rangle_{B'}\right\|_2^2 && (15.1.126)
\end{aligned}$$

$$= \frac{1}{K} \left\| \sum_{i=1}^{K} \alpha_i \beta_i U^{ii\dagger}_{A'B'} |\phi_i\rangle_{A'} |\varphi_i\rangle_{B'} \right\|_2^2 \tag{15.1.127}$$

$$= \frac{1}{K} \left\| \sum_{i=1}^{K} \alpha_i \beta_i |\xi_i\rangle_{A'B'} \right\|_2^2 \tag{15.1.128}$$

$$= \frac{1}{K} \sum_{i,j=1}^{K} \alpha_i \beta_i \alpha_j^* \beta_j^* \langle \xi_j | \xi_i \rangle_{A'B'}. \tag{15.1.129}$$

where $|\xi_i\rangle_{A'B'} := (U^{ii}_{A'B'})^\dagger |\phi_i\rangle_{A'} |\varphi_i\rangle_{B'}$ is a quantum state. The desired bound in (15.1.121) is then equivalent to

$$\sum_{i,j=1}^{K} \alpha_i \beta_i \alpha_j^* \beta_j^* \langle \xi_j | \xi_i \rangle_{A'B'} \leq 1. \tag{15.1.130}$$

Setting $\alpha_i = \sqrt{p_i} e^{i\theta_i}$ and $\beta_i = \sqrt{q_i} e^{i\eta_i}$, we find that

$$\sum_{i,j=1}^{K} \alpha_i \beta_i \alpha_j^* \beta_j^* \langle \xi_j | \xi_i \rangle_{A'B'} = \left| \sum_{i,j=1}^{K} \sqrt{p_i q_i p_j q_j} e^{i(\theta_i + \eta_i - \theta_j - \eta_j)} \langle \xi_j | \xi_i \rangle_{A'B'} \right| \tag{15.1.131}$$

$$\leq \sum_{i,j=1}^{K} \sqrt{p_i q_i p_j q_j} \left| \langle \xi_j | \xi_i \rangle_{A'B'} \right| \tag{15.1.132}$$

$$\leq \sum_{i,j=1}^{K} \sqrt{p_i q_i p_j q_j} \tag{15.1.133}$$

$$= \left[\sum_{i=1}^{K} \sqrt{p_i q_i} \right]^2 \leq 1, \tag{15.1.134}$$

where the last inequality holds for all probability distributions (this is just the statement that the classical fidelity cannot exceed one). The above reasoning thus establishes (15.1.120) for pure product states, and the bound for general separable states follows because every such state can be written as a convex combination of pure product states. ■

Recall from (9.2.3) that the ε-relative entropy of entanglement of a bipartite state ρ_{AB} is defined as

$$E_R^\varepsilon(A;B)_\rho := \inf_{\sigma_{AB} \in \mathrm{SEP}(A:B)} D_H^\varepsilon(\rho_{AB} \| \sigma_{AB}). \tag{15.1.135}$$

This quantity is an LOCC monotone, meaning that

$$E_R^{\varepsilon}(A;B)_\rho \geq E_R^{\varepsilon}(A';B')_\omega, \tag{15.1.136}$$

for $\omega_{A'B'} \coloneqq \mathcal{L}_{AB\to A'B'}(\rho_{AB})$, with $\mathcal{L}_{AB\to A'B'}$ an LOCC channel.

Proposition 15.15

Fix $\varepsilon \in [0,1]$. Let $\rho_{ABA'B'}$ be an ε-approximate bipartite private state, as given in Definition 15.6. Then the number $\log_2 K$ of private bits in such a state is bounded from above by the ε-relative entropy of entanglement of $\rho_{ABA'B'}$:

$$\log_2 K \leq E_R^{\varepsilon}(AA';BB')_\rho. \tag{15.1.137}$$

PROOF: Let $\sigma_{ABA'B'}$ be an arbitrary separable state in $\text{SEP}(AA' : BB')$. From Definition 15.6 and Lemma 15.13, we conclude that the γ-privacy test $\Pi_{ABA'B'}$ from (15.1.110) is a particular measurement operator satisfying the constraint $\text{Tr}[\Pi_{ABA'B'}\rho_{ABA'B'}] \geq 1-\varepsilon$ for $\beta_\varepsilon(\rho_{ABA'B'}\|\sigma_{ABA'B'})$. Applying Lemma 15.14 and the definition of β_ε, we conclude that

$$\beta_\varepsilon(\rho_{ABA'B'}\|\sigma_{ABA'B'}) \leq \text{Tr}[\Pi_{ABA'B'}\sigma_{ABA'B'}] \leq \frac{1}{K}. \tag{15.1.138}$$

Since the inequality holds for all separable states $\sigma_{ABA'B'} \in \text{SEP}(AA' : BB')$, we conclude that

$$\sup_{\sigma_{ABA'B'}\in\text{SEP}(AA':BB')} \beta_\varepsilon(\rho_{ABA'B'}\|\sigma_{ABA'B'}) \leq \frac{1}{K}. \tag{15.1.139}$$

Applying a negative logarithm and the definition in (15.1.135), we arrive at the inequality in (15.1.137). ■

A consequence of Proposition 15.15 is the following upper bound on the one-shot distillable key of ρ_{AB}:

Theorem 15.16 Relative Entropy of Entanglement Upper Bound on One-Shot Distillable Key

Let ρ_{AB} be a bipartite state. For every (K, ε) secret-key distillation protocol for ρ_{AB}, with $\varepsilon \in [0, 1]$, we have that

$$\log_2 K \leq E_R^{\varepsilon}(A; B)_{\rho}. \tag{15.1.140}$$

Consequently, for the one-shot distillable key, we have

$$K_D^{\varepsilon}(A; B)_{\rho} \leq E_R^{\varepsilon}(A; B)_{\rho}, \tag{15.1.141}$$

for every state ρ_{AB} and $\varepsilon \in [0, 1]$.

Proof: By Theorem 15.9, we can work in the picture of bipartite private-state distillation. Consider a (K, ε) private-state distillation protocol for ρ_{AB} with the corresponding LOCC channel $\mathcal{L}_{AB \to K_A K_B A' B'}$. Then, by definition, we have that

$$1 - F(\gamma_{K_A K_B A' B'}, \mathcal{L}_{AB \to K_A K_B A' B'}(\rho_{AB})) \leq \varepsilon \tag{15.1.142}$$

for some ideal bipartite private state $\gamma_{K_A K_B A' B'}$ of size K. Letting $\omega_{K_A K_B A' B'} := \mathcal{L}_{AB \to K_A K_B A' B'}(\rho_{AB})$, we have that $1 - F(\gamma_{K_A K_B A' B'}, \omega_{K_A K_B A' B'}) \leq \varepsilon$. The output state $\omega_{K_A K_B A' B'}$ of the private-state distillation protocol therefore satisfies the conditions of Proposition 15.15, which means that

$$\log_2 K \leq E_R^{\varepsilon}(K_A A'; K_B B')_{\omega}. \tag{15.1.143}$$

Now, as mentioned above and discussed in Section 9.2, E_R^{ε} is an entanglement measure. Thus, it satisfies the data-processing inequality under LOCC channels, which means that $E_R^{\varepsilon}(K_A A'; K_B B')_{\omega} \leq E_R^{\varepsilon}(A; B)_{\rho}$. We thus have $\log_2 K \leq E_R^{\varepsilon}(A; B)_{\rho}$. Since this inequality holds for all $K \in \mathbb{N}$ and for every LOCC channel $\mathcal{L}_{AB \to K_A K_B A' B'}$, by definition of the one-shot ε-distillable key, we obtain $K_D^{\varepsilon}(A; B)_{\rho} \leq E_R^{\varepsilon}(A; B)_{\rho}$, as required. ■

We then find the following upper bounds on the distillable key available in (K, ε) key distillation protocols:

Corollary 15.17

Let ρ_{AB} be a bipartite state, and let $\varepsilon \in [0,1)$. For every (K,ε) secret-key distillation protocol for ρ_{AB}, we have that

$$\left(1-2\sqrt{\varepsilon}-\delta\right)\log_2 K \leq \sup_{\mathcal{L}^{\leftrightarrow}} \left(I(X;B')_{\mathcal{L}^{\leftrightarrow}(\psi)} - I(X;EZ)_{\mathcal{L}^{\leftrightarrow}(\psi)}\right) + h_2(\sqrt{\varepsilon}+\delta) + \left(1-\sqrt{\varepsilon}-\delta\right)\log_2\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon}), \quad (15.1.144)$$

where $\delta \in \left(0, 1-\sqrt{\varepsilon}\right)$, ψ_{ABE} is a purification of ρ_{AB}, the information quantities are evaluated on the state $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}(\psi_{ABE})$, and the optimization is over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ with classical systems X and Z. The following bound holds for all $\alpha > 1$:

$$\log_2 K \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right), \quad (15.1.145)$$

where

$$\widetilde{E}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in \mathrm{SEP}(A;B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}) \quad (15.1.146)$$

is the sandwiched Rényi relative entropy of entanglement (see (9.2.4)).

Proof: Employing the same reasoning that led to (15.1.92) and (15.1.98), consider that the following bounds hold for a given (K,ε) secret-key distillation protocol:

$$\log_2 K \leq I_H^{\sqrt{\varepsilon}+\delta}(K_A;K_B)_{\omega^{\mathcal{M}}} + \log_2\left(\frac{1}{\delta}\right), \quad (15.1.147)$$

$$I_{\max}^{\sqrt{\varepsilon}}(K_A;EZ)_{\omega^{\mathcal{M}}} \leq 0. \quad (15.1.148)$$

where $\delta \in (0, 1-\sqrt{\varepsilon})$. Consider from Proposition 7.70 that

$$I_H^{\sqrt{\varepsilon}+\delta}(K_A;K_B)_{\omega^{\mathcal{M}}} \leq \frac{1}{1-\sqrt{\varepsilon}-\delta}\left(I(K_A;K_B)_{\omega^{\mathcal{M}}} + h_2(\sqrt{\varepsilon}+\delta)\right). \quad (15.1.149)$$

Combining (15.1.147) and (15.1.149), we obtain

$$\left(1-\sqrt{\varepsilon}-\delta\right)\log_2 K \leq I(K_A;K_B)_{\omega^{\mathcal{M}}} + h_2(\sqrt{\varepsilon}+\delta) + \left(1-\sqrt{\varepsilon}-\delta\right)\log_2\left(\frac{1}{\delta}\right). \quad (15.1.150)$$

Also, we have that

$$
\begin{aligned}
&I_{\max}^{\sqrt{\varepsilon}}(K_A;EZ)_{\omega^{\mathcal{M}}} \\
&= \inf_{\widetilde{\omega}_{K_AEZ}:P(\widetilde{\omega}_{K_AEZ},\omega^{\mathcal{M}})\leq\sqrt{\varepsilon}} \inf_{\tau_{EZ}} D_{\max}(\widetilde{\omega}_{K_AEZ}\|\widetilde{\omega}_{K_A}\otimes\tau_{EZ}) && (15.1.151)\\
&\geq \inf_{\widetilde{\omega}_{K_AEZ}:P(\widetilde{\omega}_{K_AEZ},\omega^{\mathcal{M}})\leq\sqrt{\varepsilon}} \inf_{\tau_{EZ}} D(\widetilde{\omega}_{K_AEZ}\|\widetilde{\omega}_{K_A}\otimes\tau_{EZ}) && (15.1.152)\\
&= \inf_{\widetilde{\omega}_{K_AEZ}:P(\widetilde{\omega}_{K_AEZ},\omega^{\mathcal{M}})\leq\sqrt{\varepsilon}} I(K_A;EZ)_{\widetilde{\omega}} && (15.1.153)\\
&\geq I(K_A;EZ)_{\omega^{\mathcal{M}}} - \sqrt{\varepsilon}\log_2 K - 2g_2(\sqrt{\varepsilon}). && (15.1.154)
\end{aligned}
$$

The first inequality follows because $D_{\max}(\rho\|\sigma) \geq D(\rho\|\sigma)$, and the second inequality is a consequence of Theorem 6.14 and (7.2.169). We then find that

$$
I_{\max}^{\sqrt{\varepsilon}}(K_A;EZ)_{\omega^{\mathcal{M}}} \geq I(K_A;EZ)_{\omega^{\mathcal{M}}} - \sqrt{\varepsilon}\log_2 K - 2g_2(\sqrt{\varepsilon}). \tag{15.1.155}
$$

Combining (15.1.148) and (15.1.155), we conclude that

$$
-\sqrt{\varepsilon}\log_2 K \leq -I(K_A;EZ)_{\omega^{\mathcal{M}}} + 2g_2(\sqrt{\varepsilon}). \tag{15.1.156}
$$

Adding (15.1.150) and (15.1.156) gives

$$
\begin{aligned}
\left(1-2\sqrt{\varepsilon}-\delta\right)\log_2 K &\leq I(K_A;K_B)_{\omega^{\mathcal{M}}} - I(K_A;EZ)_{\omega^{\mathcal{M}}} \\
&\quad + h_2(\sqrt{\varepsilon}+\delta) + \left(1-\sqrt{\varepsilon}-\delta\right)\log_2\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon}). && (15.1.157)
\end{aligned}
$$

Now by optimizing over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ with X and Z classical systems and observing that the state $\omega^{\mathcal{M}}_{K_AK_BEZ}$ results from the action of a particular LOPC channel on ψ_{ABE}, we conclude that

$$
I(K_A;K_B)_{\omega^{\mathcal{M}}} - I(K_A;EZ)_{\omega^{\mathcal{M}}} \leq \sup_{\mathcal{L}} \left(I(X;B')_{\mathcal{L}(\psi)} - I(X;EZ)_{\mathcal{L}(\psi)}\right), \tag{15.1.158}
$$

thus giving (15.1.144).

The inequality in (15.1.145) follows from Theorem 15.16 and (7.9.59) in Proposition 7.71. ■

Since the upper bounds in (15.1.144) and (15.1.145) hold for all (K,ε) secret-key distillation protocols, we conclude the following upper bounds on one-shot ε-distillable key:

$$\left(1 - 2\sqrt{\varepsilon} - \delta\right) K_D^{\varepsilon}(A;B)_\rho \leq \sup_{\mathcal{L}} \left(I(X;B')_{\mathcal{L}(\psi)} - I(X;EZ)_{\mathcal{L}(\psi)}\right) \\ + h_2(\sqrt{\varepsilon} + \delta) + \left(1 - \sqrt{\varepsilon} - \delta\right) \log_2\!\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon}), \quad (15.1.159)$$

$$\log_2 K \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha - 1} \log_2\!\left(\frac{1}{1-\varepsilon}\right), \qquad \forall \alpha > 1, \quad (15.1.160)$$

where $\delta \in (0, 1 - \sqrt{\varepsilon})$ and the optimization in (15.1.159) is over every LOPC channel $\mathcal{L}_{AB \to XB'Z}$.

15.1.3.3 Squashed Entanglement Upper Bound

We now turn to squashed entanglement and establish it as an upper bound on one-shot distillable key. Before doing so, we establish some preparatory lemmas.

We begin by establishing Lemma 15.19, which is an upper bound on the logarithm of the dimension K of a key system of an ε-approximate private state, as given in Definition 15.6, in terms of its squashed entanglement, plus another term depending only on ε and $\log_2 K$. In what follows, we suppose that $\gamma_{AA'BB'}$ is a private state with key systems AB and shield systems $A'B'$. Recall from Theorem 15.5 that a private state of $\log_2 K$ private bits can be written in the following form:

$$\gamma_{ABA'B'} = U_{ABA'B'} \left(\Phi_{AB} \otimes \sigma_{A'B'}\right) U^\dagger_{ABA'B'}, \quad (15.1.161)$$

where Φ_{AB} is a maximally entangled state of Schmidt rank K

$$\Phi_{AB} := \frac{1}{K} \sum_{i,j} |i\rangle\!\langle j|_A \otimes |i\rangle\!\langle j|_B, \quad (15.1.162)$$

and

$$U_{ABA'B'} = \sum_{i,j} |i\rangle\!\langle i|_A \otimes |j\rangle\!\langle j|_B \otimes U^{ij}_{A'B'} \quad (15.1.163)$$

is a controlled unitary known as a "twisting unitary," with each $U^{ij}_{A'B'}$ a unitary operator. Due to the fact that the maximally entangled state Φ_{AB} is unextendible, an arbitrary extension $\gamma_{AA'BB'E}$ of a private state $\gamma_{AA'BB'}$ necessarily has the following form:

$$\gamma_{AA'BB'E} = U_{AA'BB'} \left(\Phi_{AB} \otimes \sigma_{A'B'E}\right) U^\dagger_{AA'BB'}, \quad (15.1.164)$$

where $\sigma_{A'B'E}$ is an extension of $\sigma_{A'B'}$. We start with the following lemma, which applies to an arbitrary extension of a bipartite private state:

Lemma 15.18

Let $\gamma_{AA'BB'}$ be a bipartite private state, and let $\gamma_{AA'BB'E}$ be an extension of it, as given above. Then the following identity holds for every such extension:

$$2\log_2 K = I(A;BB'|E)_\gamma + I(A';B|AB'E)_\gamma. \tag{15.1.165}$$

Proof: First consider that the following identity holds as a consequence of two applications of the chain rule for conditional quantum mutual information (see (7.2.136)):

$$\begin{aligned} I(AA';BB'|E)_\gamma &= I(A;BB'|E)_\gamma + I(A';BB'|AE)_\gamma \\ &= I(A;BB'|E)_\gamma + I(A';B'|AE)_\gamma + I(A';B|B'AE)_\gamma. \end{aligned} \tag{15.1.166}$$

Combined with the following identity, which holds for an arbitrary extension $\gamma_{AA'BB'E}$ of a private state $\gamma_{AA'BB'}$,

$$I(AA';BB'|E)_\gamma = 2\log_2 K + I(A';B'|AE)_\gamma, \tag{15.1.167}$$

we recover the statement in (15.1.165). So it remains to prove (15.1.167).

By definition, we have that

$$I(AA';BB'|E)_\gamma = H(AA'E)_\gamma + H(BB'E)_\gamma - H(E)_\gamma - H(AA'BB'E)_\gamma. \tag{15.1.168}$$

By applying (15.1.162)–(15.1.164), we can write $\gamma_{AA'BB'E}$ as follows:

$$\gamma_{AA'BB'E} = \frac{1}{K}\sum_{i,j} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B \otimes U^{ii}_{A'B'}\sigma_{A'B'E}(U^{jj}_{A'B'})^\dagger. \tag{15.1.169}$$

Tracing over system B leads to the following state:

$$\gamma_{AA'B'E} = \frac{1}{K}\sum_{i} |i\rangle\langle i|_A \otimes \gamma^i_{A'B'E}, \tag{15.1.170}$$

where

$$\gamma^i_{A'B'E} := U^{ii}_{A'B'}\sigma_{A'B'E}(U^{ii}_{A'B'})^\dagger. \tag{15.1.171}$$

Similarly, tracing over system A of $\gamma_{AA'BB'E}$ leads to

$$\gamma_{BA'B'E} = \frac{1}{K}\sum_i |i\rangle\langle i|_B \otimes \gamma^i_{A'B'E}. \tag{15.1.172}$$

So these and the chain rule for conditional entropy (see (7.2.110)) imply that

$$H(AA'E)_\gamma = H(A)_\gamma + H(A'E|A)_\gamma = \log_2 K + H(A'E|A)_\gamma. \tag{15.1.173}$$

Similarly, we have that

$$H(BB'E)_\gamma = \log_2 K + H(B'E|B)_\gamma = \log_2 K + H(B'E|A)_\gamma, \tag{15.1.174}$$

where we have used the symmetries in (15.1.170)–(15.1.172). Since $\gamma_E = \gamma^i_E$ for all i (this is a consequence of $\gamma_{ABA'B'}$ being an ideal private state), we find that

$$H(E)_\gamma = \frac{1}{K}\sum_i H(E)_{\gamma^i} = H(E|A)_\gamma. \tag{15.1.175}$$

Finally, we have that

$$\begin{aligned} H(AA'BB'E)_\gamma &= H(ABA'B'E)_{\Phi\otimes\sigma} &(15.1.176)\\ &= H(AB)_\Phi + H(A'B'E)_\sigma &(15.1.177)\\ &= \frac{1}{K}\sum_i H(A'B'E)_{\gamma^i} &(15.1.178)\\ &= H(A'B'E|A)_\gamma. &(15.1.179)\end{aligned}$$

The first equality follows from unitary invariance of quantum entropy. The second equality follows because the entropy is additive for tensor-product states. The third equality follows because $H(AB)_\Phi = 0$ since Φ_{AB} is a pure state, and $\sigma_{A'B'E}$ is related to $\gamma^i_{A'B'E}$ by the unitary $U^{ii}_{A'B'}$. The final equality follows by applying (15.1.170), and the fact that conditional entropy is a convex combination of entropies for a classical-quantum state where the conditioning system is classical. Combining (15.1.168), (15.1.173), (15.1.174), (15.1.175), (15.1.179), and the fact that

$$I(A';B'|AE)_\gamma = H(A'E|A)_\gamma + H(B'E|A)_\gamma - H(E|A)_\gamma - H(A'B'E|A)_\gamma, \tag{15.1.180}$$

we recover (15.1.167). ■

We now establish the squashed entanglement upper bound for an approximate bipartite private state:

Proposition 15.19

Let $\gamma_{AA'BB'}$ be a private state, with key systems AB and shield systems $A'B'$, and let $\omega_{AA'BB'}$ be an ε-approximate private state, in the sense that

$$F(\gamma_{AA'BB'}, \omega_{AA'BB'}) \geq 1 - \varepsilon \tag{15.1.181}$$

for $\varepsilon \in [0, 1]$. Suppose that $d_A = d_B = K$. Then

$$(1 - 2\sqrt{\varepsilon}) \log_2 K \leq E_{\text{sq}}(AA'; BB')_\omega + 2g_2(\sqrt{\varepsilon}), \tag{15.1.182}$$

where

$$g_2(\delta) \coloneqq (\delta + 1) \log_2(\delta + 1) - \delta \log_2 \delta. \tag{15.1.183}$$

Proof: By applying Uhlmann's theorem for fidelity (Theorem 6.8) and the inequalities relating trace distance and fidelity from Theorem 6.14, for a given extension $\omega_{AA'BB'E}$ of $\omega_{AA'BB'}$, there exists an extension $\gamma_{AA'BB'E}$ of $\gamma_{AA'BB'}$ such that

$$\frac{1}{2} \|\gamma_{AA'BB'E} - \omega_{AA'BB'E}\|_1 \leq \sqrt{\varepsilon}. \tag{15.1.184}$$

Defining $f_1(\delta, K) \coloneqq 2\delta \log_2 K + 2g_2(\delta)$, we then find that

$$2 \log_2 K = I(A; BB'|E)_\gamma + I(A'; B|AB'E)_\gamma \tag{15.1.185}$$

$$\leq I(A; BB'|E)_\omega + I(A'; B|AB'E)_\omega + 2f_1(\sqrt{\varepsilon}, K) \tag{15.1.186}$$

$$\leq I(A; BB'|E)_\omega + I(A'; B|AB'E)_\omega + I(A'; B'|AE)_\omega + 2f_1(\sqrt{\varepsilon}, K) \tag{15.1.187}$$

$$= I(AA'; BB'|E)_\omega + 2f_1(\sqrt{\varepsilon}, K). \tag{15.1.188}$$

The first equality follows from Lemma 15.18. The first inequality follows from two applications of Proposition 7.10 (uniform continuity of conditional mutual information). The second inequality follows because $I(A'; B'|AE)_\omega \geq 0$ (this is strong subadditivity from Theorem 7.6). The last equality is a consequence of the chain rule for conditional mutual information, as used in (15.1.166). Since the inequality

$$2 \log_2 K \leq I(AA'; BB'|E)_\omega + 2f_1(\sqrt{\varepsilon}, K) \tag{15.1.189}$$

holds for an arbitrary extension of $\omega_{AA'BB'}$, the statement of the proposition follows. ■

We now put these statements together and arrive at the following squashed-entanglement upper bound on one-shot distillable key:

Theorem 15.20 Squashed Entanglement Upper Bound on One-Shot Distillable Key

Let ρ_{AB} be a bipartite state. For every (K, ε) secret-key distillation protocol for ρ_{AB}, with $\varepsilon \in [0, 1)$, we have that

$$\left(1 - 2\sqrt{\varepsilon}\right) \log_2 K \leq E_{\text{sq}}(A; B)_\rho + 2 g_2(\sqrt{\varepsilon}), \tag{15.1.190}$$

where $E_{\text{sq}}(A; B)_\rho$ is the squashed entanglement of ρ_{AB} (see Section 9.4) and $g_2(\delta) \coloneqq (\delta + 1) \log_2(\delta+1) - \delta \log_2 \delta$. Consequently, for the one-shot distillable key, we have

$$\left(1 - 2\sqrt{\varepsilon}\right) K_D^\varepsilon(A; B)_\rho \leq E_{\text{sq}}(A; B)_\rho + 2 g_2(\sqrt{\varepsilon}), \tag{15.1.191}$$

for every state ρ_{AB} and $\varepsilon \in [0, 1)$.

Proof: We exploit Theorem 15.9 and work in the bipartite picture of private-state distillation, instead of the tripartite picture of key distillation. With this in mind, consider a (K, ε) bipartite private-state distillation protocol for ρ_{AB} with the corresponding LOCC channel $\mathcal{L}_{AB \to K_A K_B A' B'}$. From the LOCC monotonicity of squashed entanglement (Theorem 9.33), we have that

$$E_{\text{sq}}(K_A A'; K_B B')_\omega \leq E_{\text{sq}}(A; B)_\rho, \tag{15.1.192}$$

where $\omega_{K_A A' K_B B'} \coloneqq \mathcal{L}_{AB \to K_A K_B A' B'}(\rho_{AB})$. Continuing, by the definition in (15.1.6) and applying Theorem 15.9, the following inequality holds

$$p_{\text{err}}(\mathcal{L}; \rho_{AB}) = 1 - F(\gamma_{K_A A' K_B B'}, \omega_{K_A A' K_B B'}) \leq \varepsilon \tag{15.1.193}$$

for some ideal bipartite private state $\gamma_{K_A A' K_B B'}$. As a consequence of Proposition 15.19, we find that

$$(1 - 2\sqrt{\varepsilon}) \log_2 K \leq E_{\text{sq}}(K_A A'; K_B B')_\omega + 2 g_2(\sqrt{\varepsilon}). \tag{15.1.194}$$

Combining (15.1.192) and (15.1.194), we conclude (15.1.190). Since this bound holds for all (K, ε) key distillation protocols, the bound in (15.1.191) follows after applying the definition in (15.1.9). ■

15.1.4 Lower Bound on the Number of Secret-Key Bits via Position-Based Coding and Convex Splitting

Having found upper bounds on one-shot distillable key, we now turn to establishing a lower bound. In order to establish a lower bound on distillable key, we have to find an explicit secret-key distillation protocol that works for an arbitrary bipartite state ρ_{AB} and an arbitrary error $\varepsilon \in (0,1)$. Recall that the goal of secret key distillation is for two parties, Alice and Bob, to make use of LOPC to transform a purification ψ_{ABE} of their shared state ρ_{AB} to an ideal key state of the form in Definition 15.1, with the key size K as large as possible, subject to the constraint that the error not exceed ε. Furthermore, we allow them to make use of public classical communication for free.

Before we get into the details, let us first slightly modify the model of secret key distillation, and we discuss later how the model we have already discussed can fit together with this alternative model. The alternative model consists of supposing that the state shared by Alice, Bob, and Eve, is a classical–quantum–quantum state ρ_{XBE} of the following form:

$$\rho_{XBE} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho^x_{BE}, \tag{15.1.195}$$

where $\mathcal{X}$ is a finite alphabet, $p : \mathcal{X} \to [0,1]$ is a probability distribution, and $\{\rho^x_{BE}\}_{x\in\mathcal{X}}$ is a set of states. The goal of a secret-key distillation protocol in this setting is to perform an LOPC channel $\mathcal{L}^{\leftrightarrow}_{XB\to K_AK_BZ}$ such that the final state $\omega_{K_AK_BEZ} \coloneqq \mathcal{L}^{\leftrightarrow}_{XB\to K_AK_BEZ}(\rho_{XBE})$ satisfies

$$p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{XBE}) \coloneqq \inf_{\sigma_{EZ}} \Big(1 - F(\overline{\Phi}_{K_AK_B}\otimes\sigma_{EZ}, \omega_{K_AK_BEZ})\Big) \le \varepsilon, \tag{15.1.196}$$

where the error $\varepsilon \in [0,1]$, the infimum is with respect to every state σ_{EZ}, and $\overline{\Phi}_{K_AK_B}$ is a maximally classically correlated state of size K

$$\overline{\Phi}_{K_AK_B} \coloneqq \frac{1}{K}\sum_{i=0}^{K-1} |i\rangle\langle i|_{K_A} \otimes |i\rangle\langle i|_{K_B}. \tag{15.1.197}$$

We then define the one-shot distillable key of ρ_{XBE} as follows:

$$K^{\varepsilon}_D(\rho_{XBE}) \coloneqq \sup_{(K,\mathcal{L}^{\leftrightarrow})} \left\{\log_2 K : p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{XBE}) \le \varepsilon\right\}, \tag{15.1.198}$$

where the optimization is over all $K \in \mathbb{N}$ and every LOPC channel $\mathcal{L}^{\leftrightarrow}_{XB\to K_A K_B Z}$ with $d_{K_A} = d_{K_B} = K$.

The main idea behind the lower bound is to exhibit a particular protocol that accomplishes the task of secret key distillation. The protocol we devise is simple to describe but more involved to analyze. Additionally, it really does take advantage of the fact that free public classical communication is allowed, in the sense that a large amount of public classical communication is employed. The protocol begins with Alice, Bob, and Eve sharing the state ρ_{XBE}, with Alice possessing system X, Bob B, and Eve E. Alice picks a value k uniformly at random from the set $[K] := \{1, \ldots, K\}$. This will end up being the value of the key. She also picks a value r uniformly at random from the set $[R] := \{1, \ldots, R\}$, where $R \in \mathbb{N}$. This variable plays the role of randomness that is used to confuse Eve about which key value k was chosen. Once k and r have been selected, Alice labels her system X of the state ρ_{XBE} by the pair (k, r), as $X_{k,r}$. Alice then prepares $KR - 1$ independent instances of the classical state

$$\sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|, \tag{15.1.199}$$

and labels the resulting systems as $X_{1,1}, \ldots, X_{k,r-1}, X_{k,r+1}, \ldots, X_{KR}$. Alice then sends the classical registers $X_{1,1}, \ldots, X_{K,R}$ in lexicographic order over a public classical communication channel, so that both Bob and Eve receive copies of them. At this point, for fixed values of k and r, the global shared state of Alice, Bob, and Eve is as follows:

$$\begin{aligned}\rho^{k,r}_{X^{KR}X'^{KR}X''^{KR}BE} := \rho_{X_{1,1}X'_{1,1}X''_{1,1}} \otimes \cdots \otimes \rho_{X_{k,r-1}X'_{k,r-1}X''_{k,r-1}} \otimes \\ \rho_{X_{k,r}X'_{k,r}X''_{k,r}BE} \otimes \rho_{X_{k,r+1}X'_{k,r+1}X''_{k,r+1}} \otimes \cdots \otimes \rho_{X_{K,R}X'_{K,R}X''_{K,R}},\end{aligned} \tag{15.1.200}$$

where Bob possesses all systems labeled as X' (in addition to his B system) and Eve possesses all systems labeled as X'' (in addition to her E system). Furthermore,

$$\rho_{X_{1,1}X'_{1,1}X''_{1,1}} = \cdots = \rho_{X_{k,r-1}X'_{k,r-1}X''_{k,r-1}} \tag{15.1.201}$$

$$= \rho_{X_{k,r+1}X'_{k,r+1}X''_{k,r+1}} = \cdots = \rho_{X_{K,R}X'_{K,R}X''_{K,R}} \tag{15.1.202}$$

$$= \sum_{x\in\mathcal{X}} p(x)|xxx\rangle\langle xxx|, \tag{15.1.203}$$

and

$$\rho_{X_{k,r}X'_{k,r}X''_{k,r}BE} = \sum_{x\in\mathcal{X}} p(x)|xxx\rangle\langle xxx| \otimes \rho^x_{BE}. \tag{15.1.204}$$

Thus, it is only the $X_{k,r}$ classical system that has correlation with Bob and Eve's systems BE and all others have no correlation whatsoever. The objective of the key distillation protocol is for Bob to identify the $X_{k,r}$ system that has correlation with his (and in this way, identify the key value), while the randomness variable r should have sufficient size R to severely reduce the chance that Eve can guess which X'' system is correlated with hers. The reduced state of Bob, for fixed k and r, is as follows:

$$\rho^{k,r}_{X'^{KR}B} = \rho_{X'_{1,1}} \otimes \cdots \otimes \rho_{X'_{k,r-1}} \otimes \rho_{X'_{k,r}B} \otimes \rho_{X'_{k,r+1}} \otimes \cdots \otimes \rho_{X'_{K,R}}, \tag{15.1.205}$$

while the reduced state of Eve, for a fixed value of k, is as follows:

$$\rho^{k}_{X''^{KR}E} \coloneqq \frac{1}{R}\sum_{r=1}^{R} \rho_{X''_{1,1}} \otimes \cdots \otimes \rho_{X''_{k,r-1}} \otimes \rho_{X''_{k,r}E} \otimes \rho_{X''_{k,r+1}} \otimes \cdots \otimes \rho_{X''_{K,R}}. \tag{15.1.206}$$

The idea behind confusing Eve is that if R is large enough, then it becomes difficult for Eve to determine which X'' system is correlated with her system E. What we show later is that if R is large enough, then her reduced state, for all key values k, is essentially indistinguishable from the following product state:

$$\rho_{X''_{1,1}} \otimes \cdots \otimes \rho_{X''_{K,R}} \otimes \rho_E, \tag{15.1.207}$$

where $\rho_E = \mathrm{Tr}_{XB}[\rho_{XBE}]$. If that is the case, then she can figure out essentially nothing about the key value k, leaving her no strategy other than to try and randomly guess it.

To analyze this protocol in detail, we employ two methods: position-based coding, as used previously in Section 11.1.3 in the context of classical communication, and another idea known as convex splitting. Looking at Bob's state in (15.1.205) and comparing it with that in (11.1.99), it is natural to employ position-based coding to figure out the value of k and r. Indeed, invoking Proposition 11.8 (in particular, (11.1.130)–(11.1.131)), if the following condition holds

$$\log_2 KR = \overline{I}^{\varepsilon-\eta}_H(X;B)_\rho - \log_2\!\left(\frac{4\varepsilon}{\eta^2}\right), \tag{15.1.208}$$

for $\eta \in (0,\varepsilon)$, and where $\overline{I}^{\varepsilon-\eta}_H(X;B)_\rho$ is the hypothesis testing mutual information defined in (7.11.88), then Bob can decode k and r with error probability no larger than ε. We would also like to guarantee that Eve's state in (15.1.206) is close to the product state in (15.1.207). This is where the convex-split lemma is useful, which states the following: If

$$\log_2 R = \overline{I}_{\max}^{\sqrt{\varepsilon}-\eta}(E;X)_\rho + \log_2\left(\frac{2}{\eta^2}\right), \tag{15.1.209}$$

then there exists a state $\widetilde{\rho}_E$ satisfying

$$1 - F(\rho^k_{X'^K R_E}, \rho_{X''_{1,1}} \otimes \cdots \otimes \rho_{X''_{K,R}} \otimes \widetilde{\rho}_E) \leq \varepsilon, \tag{15.1.210}$$

and $P(\widetilde{\rho}_E, \rho_E) \leq \sqrt{\varepsilon} - \eta$. Observe that the inequality above holds for all key values k. In the above, $\overline{I}_{\max}^{\delta}(E;X)_\rho$ is a smooth max-mutual information quantity defined for $\delta \in (0,1)$ as

$$\overline{I}_{\max}^{\delta}(E;X)_\rho \coloneqq \inf_{\widetilde{\rho}_{XE}:P(\widetilde{\rho}_{XE},\rho_{XE})\leq\delta} D_{\max}(\widetilde{\rho}_{XE} \| \rho_X \otimes \widetilde{\rho}_E). \tag{15.1.211}$$

Observe that $\overline{I}_{\max}^{\delta}(E;X)_\rho$ is different from the smooth max-mutual information quantity defined previously in (15.1.59). By suitably combining position-based coding with convex splitting and subtracting (15.1.209) from (15.1.208), we thus arrive at the conclusion that Alice and Bob can distill a key K of size

$$\log_2 K = \overline{I}_H^{\varepsilon-\eta}(X;B)_\rho - \overline{I}_{\max}^{\sqrt{\varepsilon}-\eta}(E;X)_\rho - \log_2\left(\frac{4\varepsilon}{\eta^2}\right) - \log_2\left(\frac{2}{\eta^2}\right), \tag{15.1.212}$$

and be guaranteed that

1. Bob can decode the key value k with error probability no larger than ε and
2. the key value is secure from Eve with security parameter ε (as given in (15.1.210)).

Having discussed the protocol for key distillation and some intuition justifying why the scheme works, we now formally state a lower bound on the one-shot distillable key of a state ρ_{XBE}:

Theorem 15.21

Let ρ_{XBE} be a classical–quantum–quantum state, with system X held by Alice, B by Bob, and E by Eve. For all $\varepsilon \in (0,1]$, $\varepsilon' = 1 - \sqrt{1-\varepsilon}$, $\delta \in (0, \varepsilon')$,

$\eta \in (0, \varepsilon' - \delta)$, and $\zeta \in (0, \delta)$, there exists a (K, ε) one-way key distillation protocol for ρ_{XBE} with

$$\log_2 K = \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho - \overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho - \log_2\left(\frac{4(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\left(\frac{2}{\zeta^2}\right), \quad (15.1.213)$$

where the hypothesis testing mutual information $\overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho$ is defined in (7.11.88) and the smooth max-mutual information $\overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho$ is defined in (15.1.211).

As discussed above, one of the main tools that we employ to prove this theorem is the smooth convex-split lemma, which we state here and prove in Appendix 15.A.

Lemma 15.22 Smooth convex split

Let ρ_{AE} be a state, and let $R \in \mathbb{N}$. Let $\tau_{A_1\cdots A_R E}$ denote the following state:

$$\tau_{A_1\cdots A_R E} \coloneqq \frac{1}{R}\sum_{r=1}^{R} \rho_{A_1} \otimes \cdots \otimes \rho_{A_{r-1}} \otimes \rho_{A_r E} \otimes \rho_{A_{r+1}} \otimes \cdots \otimes \rho_{A_R}. \quad (15.1.214)$$

Let $\varepsilon \in (0,1)$ and $\eta \in (0,\varepsilon)$. If

$$\log_2 R \geq \overline{I}_{\max}^{\varepsilon-\eta}(E;A)_\rho + \log_2\left(\frac{2}{\eta^2}\right), \quad (15.1.215)$$

then there exists a state $\widetilde{\rho}_E$ satisfying

$$P(\tau_{A_1\cdots A_R E}, \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E) \leq \varepsilon, \quad (15.1.216)$$

and $P(\widetilde{\rho}_E, \rho_E) \leq \varepsilon - \eta$.

We now prove Theorem 15.21.

Proof (Proof of Theorem 15.21): Fix $\varepsilon \in (0,1]$, $\delta \in (0,\varepsilon)$, $\eta \in (0, \varepsilon-\delta)$, and $\zeta \in (0,\delta)$. Alice performs the key distillation protocol discussed in the paragraph surrounding (15.1.199)–(15.1.207). The global state, for fixed k and r is as given in (15.1.200); the reduced state of Bob, for fixed k and r is as given in (15.1.205); and the reduced state of Eve, for fixed k, is given by (15.1.206). The overall global state, including Alice's classical registers that hold the key value k and the randomness value r, is as follows:

$$\rho_{K_A R_A X^{KR} X'^{KR} X''^{KR} BE} := \frac{1}{KR} \sum_{k=1}^{K} \sum_{r=1}^{R} |k\rangle\langle k|_{K_A} \otimes |r\rangle\langle r|_{R_A} \otimes \rho^{k,r}_{X^{KR} X'^{KR} X''^{KR} BE}, \tag{15.1.217}$$

where $\rho^{k,r}_{X^{KR} X'^{KR} X''^{KR} BE}$ is defined in (15.1.200). Tracing over the R_A and X^{KR} systems, the state becomes

$$\rho_{K_A X'^{KR} X''^{KR} BE} := \frac{1}{K} \sum_{k=1}^{K} |k\rangle\langle k|_{K_A} \otimes \frac{1}{R} \sum_{r=1}^{R} \rho^{k,r}_{X'^{KR} X''^{KR} BE} \tag{15.1.218}$$

By Proposition 11.8, we conclude that if

$$\log_2 KR = \overline{I}_H^{\varepsilon-\delta-\eta}(X;B)_\rho - \log_2\left(\frac{4(\varepsilon-\delta)}{\eta^2}\right), \tag{15.1.219}$$

then there exists a POVM $\{\Lambda^{k,r}_{X'^{KR} B}\}_{k\in[K], r\in[R]}$ such that

$$\mathrm{Tr}[\Lambda^{k,r}_{X'^{KR} B} \rho^{k,r}_{X'^{KR} B}] \geq 1 - (\varepsilon - \delta) \qquad \forall k \in [K], r \in [R]. \tag{15.1.220}$$

Let us define the measurement channel $\mathcal{M}'_{X'^{KR} B \to K_B R_B}$ as follows:

$$\mathcal{M}'_{X'^{KR} B \to K_B R_B}(\tau_{X'^{KR} B}) := \sum_{k=1}^{K} \sum_{r=1}^{R} \mathrm{Tr}[\Lambda^{k,r}_{X'^{KR} B} \tau_{X'^{KR} B}] |k\rangle\langle k|_{K_B} \otimes |r\rangle\langle r|_{R_B}, \tag{15.1.221}$$

and the reduced measurement channel $\mathcal{M}_{X'^{KR} B \to K_B}$ as

$$\mathcal{M}_{X'^{KR} B \to K_B}(\tau_{X'^{KR} B}) := (\mathrm{Tr}_{R_B} \circ \mathcal{M}'_{X'^{KR} B \to K_B R_B})(\tau_{X'^{KR} B}) \tag{15.1.222}$$

$$= \sum_{k=1}^{K} \mathrm{Tr}[\Lambda^{k,r}_{X'^{KR} B} \tau_{X'^{KR} B}] |k\rangle\langle k|_{K_B}. \tag{15.1.223}$$

Observe that

$$\frac{1}{2}\left\|\mathcal{M}'_{X'^{KR}B\to K_BR_B}(\rho^{k,r}_{X'^{KR}B}) - |k\rangle\langle k|_{K_B}\otimes|r\rangle\langle r|_{R_B}\right\|_1 \\ = 1 - \operatorname{Tr}[\Lambda^{k,r}_{X'^{KR}B}\rho^{k,r}_{X'^{KR}B}] \leq \varepsilon - \delta, \quad (15.1.224)$$

which follows from the same calculation given in (11.1.18)–(11.1.24). By the data-processing inequality for the trace distance, we conclude that

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}(\rho^{k,r}_{X'^{KR}B}) - |k\rangle\langle k|_{K_B}\right\|_1 \leq \varepsilon - \delta. \quad (15.1.225)$$

From convexity of trace distance, we conclude that

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}B}\right) - |k\rangle\langle k|_{K_B}\right\|_1 \leq \varepsilon - \delta. \quad (15.1.226)$$

The actual state at the end of the protocol is as follows:

$$\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE}), \quad (15.1.227)$$

and the ideal state to generate is

$$\overline{\Phi}_{K_AK_B}\otimes\widetilde{\rho}_{X''^{KR}E} = \frac{1}{K}\sum_{k=1}^{K}|k\rangle\langle k|_{K_A}\otimes|k\rangle\langle k|_{K_B}\otimes\widetilde{\rho}_{X''^{KR}E}, \quad (15.1.228)$$

where $\widetilde{\rho}_{X''^{KR}E}$ is some state of the eavesdropper Eve's systems $X''^{KR}E$. Thus, our goal is to find an upper bound on the following quantity

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE}) - \overline{\Phi}_{K_AK_B}\otimes\widetilde{\rho}_{X''^{KR}E}\right\|_1, \quad (15.1.229)$$

which we will convert at the end to an upper bound on

$$1 - F(\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE}), \overline{\Phi}_{K_AK_B}\otimes\widetilde{\rho}_{X''^{KR}E}). \quad (15.1.230)$$

To this end, let us first consider bounding the following intermediate quantity:

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE}) - \frac{1}{K}\sum_{k=1}^{K}|k\rangle\langle k|_{K_A}\otimes|k\rangle\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E}\right\|_1. \quad (15.1.231)$$

We find that

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_A X'^{KR} X''^{KR} BE}) - \frac{1}{K}\sum_{k=1}^{K}|k\rangle\langle k|_{K_A}\otimes|k\rangle\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E}\right\|_1$$

$$= \frac{1}{2}\left\| \begin{array}{c} \frac{1}{K}\sum_{k=1}^{K}|k\rangle\langle k|_{K_A}\otimes\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}X''^{KR}BE}\right) \\ -\frac{1}{K}\sum_{k=1}^{K}|k\rangle\langle k|_{K_A}\otimes|k\rangle\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E} \end{array}\right\|_1 \tag{15.1.232}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}X''^{KR}BE}\right) - |k\rangle\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E}\right\|_1. \tag{15.1.233}$$

Now let us define the state

$$\omega^{k',k}_{X''^{KR}E} := \frac{1}{q(k'|k)}\left(\frac{1}{R}\sum_{r,r'=1}^{R}\mathrm{Tr}_{X'^{KR}B}[\Lambda^{k',r'}_{X'^{KR}B}\rho^{k,r}_{X'^{KR}X''^{KR}BE}]\right), \tag{15.1.234}$$

$$q(k'|k) := \frac{1}{R}\sum_{r,r'=1}^{R}\mathrm{Tr}[\Lambda^{k',r'}_{X'^{KR}B}\rho^{k,r}_{X'^{KR}X''^{KR}BE}]. \tag{15.1.235}$$

Consider that

$$\sum_{k'=1}^{K}q(k'|k)\omega^{k',k}_{X''^{KR}E} = \frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E}. \tag{15.1.236}$$

Then we can write

$$\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}X''^{KR}BE}\right) = \sum_{k'=1}^{K}q(k'|k)|k'\rangle\langle k'|_{K_B}\otimes\omega^{k',k}_{X''^{KR}E}, \tag{15.1.237}$$

so that

$$\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}B}\right) = \sum_{k'=1}^{K}q(k'|k)|k'\rangle\langle k'|_{K_B}. \tag{15.1.238}$$

Using these observations, we can finally write

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'^{KR}X''^{KR}BE}\right) - |k\rangle\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''^{KR}E}\right\|_1$$

$$= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\left\|\sum_{k'=1}^{K} q(k'|k)|k'\rangle\!\langle k'|_{K_B}\otimes\omega^{k',k}_{X''KRE} - |k\rangle\!\langle k|_{K_B}\otimes\sum_{k'=1}^{K} q(k'|k)\omega^{k',k}_{X''KRE}\right\|_1 \tag{15.1.239}$$

$$\leq \frac{1}{K}\sum_{k=1}^{K}\sum_{k'=1}^{K} q(k'|k)\left(\frac{1}{2}\left\||k'\rangle\!\langle k'|_{K_B}\otimes\omega^{k',k}_{X''KRE} - |k\rangle\!\langle k|_{K_B}\otimes\omega^{k',k}_{X''KRE}\right\|_1\right) \tag{15.1.240}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\sum_{k'=1}^{K} q(k'|k)\left(\frac{1}{2}\left\||k'\rangle\!\langle k'|_{K_B} - |k\rangle\!\langle k|_{K_B}\right\|_1\right) \tag{15.1.241}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\sum_{\substack{k'=1,\\ k'\neq k}}^{K} q(k'|k) \tag{15.1.242}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\left\|\mathcal{M}_{X'KRB\to K_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X'KRB}\right) - |k\rangle\!\langle k|_{K_B}\right\|_1 \tag{15.1.243}$$

$$\leq \varepsilon - \delta. \tag{15.1.244}$$

We thus conclude that

$$\frac{1}{2}\left\|\mathcal{M}_{X'KRB\to K_B}(\rho_{K_AX'KRX''KRBE}) - \frac{1}{K}\sum_{k=1}^{K}|k\rangle\!\langle k|_{K_A}\otimes|k\rangle\!\langle k|_{K_B}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''KRE}\right\|_1 \leq \varepsilon - \delta. \tag{15.1.245}$$

We now turn to the analysis of privacy. Starting from the overall global state (15.1.217), and fixing a value of k, the reduced state of Eve's systems is as follows:

$$\rho^{k}_{X''KRE} = \frac{1}{R}\sum_{r=1}^{R}\rho^{k,r}_{X''KRE} = \rho_{X''_{1,1}}\otimes\cdots\otimes\rho_{X''_{k-1,R}} \otimes\frac{1}{R}\sum_{r=1}^{R}\rho_{X''_{k,1}}\otimes\cdots\otimes\rho_{X''_{k,r-1}}\otimes\rho_{X''_{k,r}E}\otimes\rho_{X''_{k,r+1}}\otimes\cdots\otimes\rho_{X''_{k,R}} \otimes\rho_{X''_{k+1,1}}\otimes\cdots\otimes\rho_{X''_{K,R}}. \tag{15.1.246}$$

Our goal is to show that

$$\frac{1}{2}\left\|\rho^{k}_{X''KRE} - \rho_{X''KR}\otimes\widetilde{\rho}_E\right\|_1 \leq \delta, \tag{15.1.247}$$

for some state $\widetilde{\rho}_E$. By the invariance of the trace distance with respect to tensor-product states, i.e., $\|\sigma\otimes\tau-\omega\otimes\tau\|_1=\|\sigma-\omega\|_1$, we find that

$$\frac{1}{2}\left\|\rho^k_{X''^{KR}E}-\rho_{X''^{KR}}\otimes\widetilde{\rho}_E\right\|_1 \tag{15.1.248}$$
$$=\frac{1}{2}\left\|\rho^k_{X''_{k,1}\cdots X''_{k,R}E}-\rho_{X''_{k,1}\cdots X''_{k,R}}\otimes\widetilde{\rho}_E\right\|_1 \tag{15.1.249}$$
$$=\frac{1}{2}\left\|\frac{1}{R}\sum_{r=1}^{R}\rho_{X''_{k,1}}\otimes\cdots\otimes\rho_{X''_{k,r-1}}\otimes\left(\rho_{X''_{k,r}E}-\rho_{X''_{k,r}}\otimes\widetilde{\rho}_E\right)\otimes\rho_{X''_{k,r+1}}\otimes\cdots\otimes\rho_{X''_{k,R}}\right\|_1. \tag{15.1.250}$$

By invoking the smooth convex-split lemma (Lemma 15.22) and the inequality relating normalized trace distance and sine distance (see (6.2.88)), we find that if we pick R such that

$$\log_2 R=\overline{I}^{\delta-\zeta}_{\max}(E;X)_\rho+\log_2\!\left(\frac{2}{\zeta^2}\right), \tag{15.1.251}$$

then we are guaranteed that

$$\frac{1}{2}\left\|\rho^k_{X''^{KR}E}-\rho_{X''^{KR}}\otimes\widetilde{\rho}_E\right\|_1\le\delta, \tag{15.1.252}$$

where $\widetilde{\rho}_E$ is some state such that $P(\widetilde{\rho}_E,\rho_E)\le\delta-\zeta$. Now combining (15.1.245) and (15.1.252) with the triangle inequality, we conclude the desired statement:

$$\frac{1}{2}\left\|\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE})-\overline{\Phi}_{K_AK_B}\otimes\rho_{X''^{KR}}\otimes\widetilde{\rho}_E\right\|_1\le\varepsilon. \tag{15.1.253}$$

We finally conclude that

$$1-F(\mathcal{M}_{X'^{KR}B\to K_B}(\rho_{K_AX'^{KR}X''^{KR}BE}),\overline{\Phi}_{K_AK_B}\otimes\rho_{X''^{KR}}\otimes\widetilde{\rho}_E)\le\varepsilon\,(2-\varepsilon) \tag{15.1.254}$$

by exploiting the inequality in (6.2.88) relating fidelity and trace distance. Now using the fact that the inverse function of $\varepsilon(2-\varepsilon)$, with domain and range given by $[0,1]$, is $1-\sqrt{1-\varepsilon}$ and reassigning $\varepsilon\,(2-\varepsilon)$ as ε, we conclude the desired statement in (15.1.213). ■

The result of Theorem 15.21 applies to the model of secret key distillation outlined in the paragraph containing (15.1.195)–(15.1.198). To extend it to the main model considered in this chapter (and outlined in Section 15.1), we can allow Alice and Bob to perform an LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ to obtain the following state:

$$\rho_{XB'EZ} := \mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}(\psi_{ABE}), \tag{15.1.255}$$

where ψ_{ABE} is a purification of the state ρ_{AB} of interest and $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ is an LOPC channel with classical output system X' and quantum output system B'. Then we obtain the following by applying Theorem 15.21:

Corollary 15.23

Let ρ_{AB} be a bipartite state, with system X held by Alice, B by Bob, and E by Eve. For all $\varepsilon \in (0,1]$, $\varepsilon' := 1-\sqrt{1-\varepsilon}$, $\delta \in (0,\varepsilon')$, $\eta \in (0,\varepsilon'-\delta)$, and $\zeta \in (0,\delta)$, there exists a (K,ε) one-way key distillation protocol for ρ_{AB} with

$$\log_2 K = \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B')_\rho - \overline{I}_{\max}^{\delta-\zeta}(EZ;X)_\rho - \log_2\!\left(\frac{4(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\!\left(\frac{2}{\zeta^2}\right), \tag{15.1.256}$$

where the hypothesis testing mutual information $\overline{I}_H^{\varepsilon'-\delta-\eta}(X;B')_\rho$ is defined in (7.11.88), the smooth max-mutual information $\overline{I}_{\max}^{\delta-\zeta}(EZ;X)_\rho$ is defined in (15.1.211), and these quantities are evaluated with respect to the state in (15.1.255), with $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ an LOPC channel with classical output system X and quantum output system B'. Consequently, for the one-shot distillable key of ρ_{AB}, we have

$$K_D^\varepsilon(A;B)_\rho \geq \sup_{\substack{\mathcal{L}^{\leftrightarrow},\delta\in(0,\varepsilon'),\\ \eta\in(0,\varepsilon'-\delta),\zeta\in(0,\delta)}} \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B')_\rho - \overline{I}_{\max}^{\delta-\zeta}(EZ;X)_\rho - \log_2\!\left(\frac{4(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\!\left(\frac{2}{\zeta^2}\right), \tag{15.1.257}$$

where $\varepsilon' := 1-\sqrt{1-\varepsilon}$ and the optimization is over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$.

Now combining Corollary 15.23 with Propositions 7.64 and 7.72, we conclude the following lower bound on one-shot distillable key:

Corollary 15.24

Let ρ_{AB} be a bipartite state with purification ψ_{ABE}. For all $\varepsilon \in (0,1)$, $\varepsilon' = 1 - \sqrt{1-\varepsilon}$, $\delta \in (0, \varepsilon')$, $\eta \in (0, \varepsilon' - \delta)$, $\zeta \in (0, \delta)$, $\nu \in (0, \delta - \zeta)$, $\alpha \in (0,1)$, and $\beta > 1$, there exists a (K, ε) one-way key distillation protocol for ρ_{AB} satisfying

$$\log_2 K \geq \overline{I}_\alpha(X; B')_\rho - \widetilde{I}'_\beta(X; EZ)_\rho - f(\varepsilon', \delta, \eta, \nu, \zeta, \alpha, \beta)$$

where

$$\rho_{XB'EZ} \coloneqq \mathcal{L}^{\leftrightarrow}_{AB \to XB'Z}(\psi_{ABE}), \tag{15.1.258}$$

$\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ is an LOPC channel with classical output system X and quantum output system B',

$$\widetilde{I}'_\beta(X; EZ)_\rho \coloneqq \widetilde{D}_\beta(\rho_{XEZ} \| \rho_X \otimes \rho_{EZ}), \tag{15.1.259}$$

and

$$\begin{aligned} f(\varepsilon', \delta, \eta, \nu, \zeta, \alpha, \beta) \coloneqq \frac{\alpha}{1-\alpha} \log_2\!\left(\frac{1}{\varepsilon' - \delta - \eta}\right) + \log_2\!\left(\frac{8}{\nu^2}\right) \\ + \frac{1}{\beta - 1} \log_2\!\left(\frac{1}{(\delta - \zeta - \nu)^2}\right) + \log_2\!\left(\frac{1}{1 - (\delta - \zeta - \nu)^2}\right) \\ + \log_2\!\left(\frac{4(\varepsilon' - \delta)}{\eta^2}\right) + \log_2\!\left(\frac{2}{\zeta^2}\right). \end{aligned} \tag{15.1.260}$$

Proof: The main idea here is to convert the smooth mutual information quantities $\overline{I}_H^{\varepsilon'-\delta-\eta}(X; B')_\rho$ and $\overline{I}_{\max}^{\delta-\zeta}(EZ; X)_\rho$ from Corollary 15.23 to Rényi mutual information quantities with correction terms related to the smoothing parameters. Let us first invoke Proposition 7.72 to conclude the following lower bound:

$$\overline{I}_H^{\varepsilon'-\delta-\eta}(X; B')_\rho \geq \overline{I}_\alpha(X; B')_\rho - \frac{\alpha}{1-\alpha} \log_2\left(\frac{1}{\varepsilon' - \delta - \eta}\right). \tag{15.1.261}$$

Next, we invoke Lemma 15.25 below to conclude that

$$\overline{I}_{\max}^{\delta-\zeta}(EZ; X)_\rho \leq D_{\max}^{\delta-\zeta-\nu}(\rho_{XEZ} \| \rho_X \otimes \rho_{EZ}) + \log_2\!\left(\frac{8}{\nu^2}\right), \tag{15.1.262}$$

where $\nu \in (0, \delta - \zeta)$. Then we invoke Proposition 7.64 to conclude that

$$D_{\max}^{\delta-\zeta-\nu}(\rho_{XEZ}\|\rho_X\otimes\rho_{EZ}) \leq \widetilde{D}_\beta(\rho_{XEZ}\|\rho_X\otimes\rho_{EZ}) + \frac{1}{\beta-1}\log_2\!\left(\frac{1}{(\delta-\zeta-\nu)^2}\right) + \log_2\!\left(\frac{1}{1-(\delta-\zeta-\nu)^2}\right). \quad (15.1.263)$$

Considering that

$$\widetilde{D}_\beta(\rho_{XEZ}\|\rho_X\otimes\rho_{EZ}) = \widetilde{I}'_\beta(X;EZ)_\rho. \quad (15.1.264)$$

Putting all of the above together with Corollary 15.23, we conclude the proof. ■

Lemma 15.25

Let ρ_{AE} be a bipartite state, and let $\varepsilon,\delta>0$ be such that $\varepsilon+\delta<1$. Then

$$\overline{I}_{\max}^{\varepsilon+\delta}(E;A)_\rho \leq D_{\max}^{\varepsilon}(\rho_{AE}\|\rho_A\otimes\rho_E) + \log_2\!\left(\frac{8}{\delta^2}\right), \quad (15.1.265)$$

where $\overline{I}_{\max}^{\varepsilon+\delta}(E;A)_\rho$ is defined in (15.1.211) and $D_{\max}^{\varepsilon}(\rho_{AE}\|\rho_A\otimes\rho_E)$ in (7.8.42).

PROOF: See Appendix 15.B. ■

15.2 Distillable Key of a Quantum State

Having found upper and lower bounds on the one-shot distillable key $K_D^\varepsilon(A;B)_\rho$ of a bipartite state ρ_{AB}, let us now move on to the asymptotic setting. In this setting, we allow Alice and Bob to make use of an arbitrarily large number n of copies of the state ρ_{AB} in order to obtain a secret-key state. A *secret key distillation protocol for n copies of* ρ_{AB} is defined by the triple $(n,K,\mathcal{L}^{\leftrightarrow}_{A^nB^n\to K_AK_BZ})$, consisting of the number n of copies of ρ_{AB}, an integer $K\in\mathbb{N}$, and an LOPC channel $\mathcal{L}^{\leftrightarrow}_{A^nB^n\to K_AK_BZ}$ with $d_{K_A}=d_{K_B}=K$. Observe that a secret-key distillation protocol for n copies of ρ_{AB} is equivalent to a one-shot secret-key distillation protocol for the state $\rho_{AB}^{\otimes n}$. All of the results of Section 15.1 thus carry over to the asymptotic setting simply by replacing ρ_{AB} with $\rho_{AB}^{\otimes n}$. In particular, the error probability for a secret-key distillation protocol for ρ_{AB} defined by $(n,K,\mathcal{L}^{\leftrightarrow}_{A^nB^n\to K_AK_BZ})$ is equal to

$$p_{\text{err}}(\mathcal{L}^{\leftrightarrow};\rho_{AB}^{\otimes n}) = \inf_{\gamma_{K_AK_BEZ}}\left(1-F(\gamma_{K_AK_BEZ},\mathcal{L}^{\leftrightarrow}_{A^nB^n\to K_AK_BZ}(\psi_{ABE}^{\otimes n}))\right), \quad (15.2.1)$$

where the infimum is with respect to every ideal tripartite key state $\gamma_{K_A K_B E Z}$ and ψ_{ABE} is a purification of ρ_{AB}. The definition in (15.2.1) is thus the same as that in (15.1.6), but for the tensor-power state $\rho_{AB}^{\otimes n}$.

Definition 15.26 (n, K, ε) Secret-Key Distillation Protocol

A secret-key distillation protocol $(n, K, \mathcal{L}^{\leftrightarrow}_{A^n B^n \to K_A K_B Z})$ for n copies of ρ_{AB}, with $d_{K_A} = d_{K_B} = K$, is called an (n, K, ε) *protocol*, with $\varepsilon \in [0,1]$, if $p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}^{\otimes n}) \leq \varepsilon$.

Based on the discussion above, we note here that an (n, K, ε) secret-key distillation protocol for ρ_{AB} is a (K, ε) secret-key distillation protocol for $\rho_{AB}^{\otimes n}$.

The rate $R(n, K)$ of an (n, K, ε) secret-key distillation protocol for n copies of a given state is

$$R(n, K) := \frac{\log_2 K}{n}, \tag{15.2.2}$$

which can be thought of as the number of ε-approximate secret-key bits contained in the final state of the protocol, per copy of the given initial state. Given a state ρ_{AB} and $\varepsilon \in [0,1]$, the maximum rate of secret key distillation among all (n, K, ε) secret-key distillation protocols for ρ_{AB} is

$$K_D^{n,\varepsilon}(\rho_{AB}) \equiv K_D^{n,\varepsilon}(A;B)_\rho := \frac{1}{n} K_D^{\varepsilon}(\rho_{AB}^{\otimes n}) \tag{15.2.3}$$

$$= \sup_{(K, \mathcal{L}^{\leftrightarrow})} \left\{ \frac{\log_2 K}{n} : p_{\text{err}}(\mathcal{L}^{\leftrightarrow}; \rho_{AB}^{\otimes n}) \leq \varepsilon \right\}, \tag{15.2.4}$$

where the optimization is with respect to all $K \in \mathbb{N}$ and every LOPC channel $\mathcal{L}^{\leftrightarrow}_{A^n B^n \to K_A K_B Z}$ with $d_{K_A} = d_{K_B} = K$.

Definition 15.27 Achievable Rate for Secret Key Distillation

Given a bipartite state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called an *achievable rate for secret key distillation for* ρ_{AB} if for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ secret-key distillation protocol for ρ_{AB}.

As we prove in Appendix A,

$$R \text{ achievable rate} \quad \Longleftrightarrow \quad \lim_{n\to\infty} \varepsilon_D(2^{n(R-\delta)}; \rho_{AB}^{\otimes n}) = 0 \qquad \forall \delta > 0. \tag{15.2.5}$$

In other words, a rate R is achievable if the optimal error probability for a sequence of protocols with rate $R - \delta$ vanishes as the number n of copies of ρ_{AB} increases.

Definition 15.28 Distillable Key of a Quantum State

The *distillable key of a bipartite state* ρ_{AB}, denoted by $K_D(A;B)_\rho$, is defined to be the supremum of all achievable rates for secret key distillation for ρ_{AB}, i.e.,

$$K_D(A;B)_\rho := \sup\{R : R \text{ is an achievable rate for } \rho_{AB}\}. \tag{15.2.6}$$

The distillable key can also be written as

$$K_D(A;B)_\rho = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} K_D^\varepsilon(\rho_{AB}^{\otimes n}). \tag{15.2.7}$$

See Appendix A for a proof.

Definition 15.29 Weak Converse Rate for Secret Key Distillation

Given a bipartite state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called a *weak converse rate for secret key distillation for* ρ_{AB} if every $R' > R$ is not an achievable rate for ρ_{AB}.

As we show in Appendix A,

$$R \text{ weak converse rate} \iff \lim_{n\to\infty} \varepsilon_D(2^{n(R-\delta)}; \rho_{AB}^{\otimes n}) > 0 \quad \forall \delta > 0. \tag{15.2.8}$$

Definition 15.30 Strong Converse Rate for Secret Key Distillation

Given a bipartite state ρ_{AB}, a rate $R \in \mathbb{R}^+$ is called a *strong converse rate for secret key distillation for* ρ_{AB} if for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ secret key distillation protocol for ρ_{AB}.

We show in Appendix A that

$$R \text{ strong converse rate} \iff \lim_{n\to\infty} \varepsilon_D(2^{n(R-\delta)}; \rho_{AB}^{\otimes n}) = 1 \quad \forall \delta > 0. \tag{15.2.9}$$

Definition 15.31 Strong Converse Distillable Entanglement of a Quantum State

The *strong converse distillable key* of a bipartite state ρ_{AB}, denoted by $\widetilde{K}_D(A;B)_\rho$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{K}_D(A;B)_\rho \coloneqq \inf\{R : R \text{ is a strong converse rate for } \rho_{AB}\}. \quad (15.2.10)$$

Note that

$$K_D(A;B)_\rho \le \widetilde{K}_D(A;B)_\rho \quad (15.2.11)$$

for every bipartite state ρ_{AB}. We can also write the strong converse distillable key as

$$\widetilde{K}_D(A;B)_\rho = \sup_{\varepsilon\in[0,1)} \limsup_{n\to\infty} \frac{1}{n} K_D^\varepsilon(\rho_{AB}^{\otimes n}). \quad (15.2.12)$$

See Appendix A for a proof.

We are now ready to present a general expression for the distillable key of a bipartite state, as well as two upper bounds on it.

Theorem 15.32 Distillable Key of a Bipartite State

The distillable key of a bipartite state ρ_{AB} is given by

$$K_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{(n)}} \left(I(X;B')_{\mathcal{L}^{(n)}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}^{(n)}(\psi^{\otimes n})} \right), \quad (15.2.13)$$

where the optimization is with respect to every two-way LOPC channel $\mathcal{L}^{(n)}_{A^nB^n\to XB'Z}$ with classical output system X and quantum output system B'. The information quantities are evaluated with respect to the state $\mathcal{L}^{(n)}_{A^nB^n\to XB'Z}(\psi_{ABE}^{\otimes n})$, where ψ_{ABE} is a purification of ρ_{AB}. Furthermore, the relative entropy of entanglement $E_R(A;B)_\rho$ from (9.2.2) is a strong converse rate for distillable key, in the sense that

$$\widetilde{K}_D(A;B)_\rho \le E_R(A;B)_\rho, \quad (15.2.14)$$

and the squashed entanglement from (9.4.1) is a weak converse rate, in the sense that

$$K_D(A;B)_\rho \le E_{\mathrm{sq}}(A;B)_\rho. \quad (15.2.15)$$

If we define

$$D_K^{\leftrightarrow}(\rho_{AB}) \equiv D_K^{\leftrightarrow}(A;B)_\rho := \sup_{\mathcal{L}} \left(I(X;B')_{\mathcal{L}(\psi)} - I(X;EZ)_{\mathcal{L}(\psi)} \right), \quad (15.2.16)$$

where the entropic quantities are evaluated with respect to $\mathcal{L}_{A^nB^n\to XB'Z}(\psi_{ABE})$, with ψ_{ABE} a purification of ρ_{AB}, then we can write (15.2.13) as

$$K_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} D_K^{\leftrightarrow}(\rho_{AB}^{\otimes n}) =: D_{\text{reg},K}^{\leftrightarrow}(\rho_{AB}). \quad (15.2.17)$$

Thus, the distillable key can be viewed as the regularization of $D_K^{\leftrightarrow}$, similar to what we found in (13.2.17) in Chapter 13 for distillable entanglement.

Let us make the following observations about Theorem 15.32.

- The private information is an achievable rate for secret key distillation, i.e.,

$$K_D(A;B)_\rho \geq \max\{I(X;B)_\tau - I(X;E)_\tau, I(A;Y)_\omega - I(Y;E)_\omega\}, \quad (15.2.18)$$

  where

$$\tau_{XBE} := \sum_x |x\rangle\langle x|_X \otimes \text{Tr}_A[\Lambda_A^x \psi_{ABE}], \quad (15.2.19)$$

$$\omega_{YAE} := \sum_y |y\rangle\langle y|_Y \otimes \text{Tr}_B[\Gamma_B^y \psi_{ABE}], \quad (15.2.20)$$

  ψ_{ABE} is a purification of the bipartite state ρ_{AB}, and $\{\Lambda_A^x\}_x$ and $\{\Gamma_B^y\}_y$ are POVMs. The idea behind the first achievable rate $I(X;B)_\tau - I(X;E)_\tau$ is that Alice performs the measurement $\{\Lambda_A^x\}_x$ on her system A, and this produces the classical–quantum–quantum state τ_{XBE}. Alice and Bob then execute the protocol from Theorem 15.21 on many copies of the state τ_{XBE}. Alternatively, the idea behind the second achievable rate $I(A;Y)_\omega - I(Y;E)_\omega$ is similar, but with the roles of Alice and Bob swapped and distilling a key from many copies of the state ω_{YAE}.

  We can also consider these conclusions to be immediate consequences of (15.2.13), which follow from dropping the optimization over two-way LOPC channels and from the fact that the unoptimized private informations in (15.2.18) are additive for product states.

- In order to obtain a higher key distillation rate than the private informations in (15.2.18), one strategy is to use $n \geq 2$ copies of ψ_{ABE} along with a two-way LOPC channel $\mathcal{L}_{A^n B^n \to XB'Z}$, in order to obtain a state $\omega_{A'B'E^nZ} := \mathcal{L}_{A^n B^n \to XB'Z}(\psi_{ABE}^{\otimes n})$. The normalized private informations of this latter state are potentially higher than that of τ_{XBE} and ω_{YAE} in (15.2.19)–(15.2.20). The overall rate of this strategy is then

$$\frac{1}{n}\left(I(X;B')_{\mathcal{L}^{(n)}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}^{(n)}(\psi^{\otimes n})}\right), \tag{15.2.21}$$

and Theorem 15.32 tells us that such a strategy is optimal in the large n limit. With increasingly more copies of ψ_{ABE} to start with, it might be possible to obtain a better rate, which is why we need to regularize in general.

As with other previous capacity theorem proofs in this book, we prove Theorem 15.32 in two steps:

1. *Achievability*: We show that the right-hand side of (15.2.13) is an achievable rate for secret key distillation for ρ_{AB}. Doing so involves exhibiting an explicit secret key distillation protocol. The protocol we use is based on the one we introduced in Section 15.1.4 to obtain a lower bound on the one-shot distillable secret key.

   The achievability part of the proof establishes that

$$K_D(A;B)_\rho \geq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{(n)}} \left(I(X;B')_{\mathcal{L}^{(n)}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}^{(n)}(\psi^{\otimes n})}\right). \tag{15.2.22}$$

2. *Weak converse*: We show that the right-hand side of (15.2.13) is a weak converse rate for secret key distillation for ρ_{AB}, from which it follows that

$$K_D(A;B)_\rho \leq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{(n)}} \left(I(X;B')_{\mathcal{L}^{(n)}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}^{(n)}(\psi^{\otimes n})}\right). \tag{15.2.23}$$

   In order to show this, we use the one-shot upper bounds from Section 15.1.3 to prove that every achievable rate R satisfies

$$R \leq \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{(n)}} \left(I(X;B')_{\mathcal{L}^{(n)}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}^{(n)}(\psi^{\otimes n})}\right). \tag{15.2.24}$$

We go through the achievability part of the proof of Theorem 15.32 in Section 15.2.1. We then proceed with the weak converse part in Section 15.2.2.

The expression in (15.2.13) for the distillable key involves both a limit over an unbounded number of copies of the state ρ_{AB}, as well as an optimization over all two-way LOPC channels. Computing the distillable key is therefore intractable in general. After establishing a proof of (15.2.13), we proceed to establish upper bounds on distillable entanglement that depend only on the given state ρ_{AB}. Specifically, in Section 15.2.3, we use the one-shot results in Section 15.1.3.2 to show that the relative entropy of entanglement is a strong converse rate for secret key distillation. We also show that the squashed entanglement is a weak converse rate for secret key distillation.

15.2.1 Proof of Achievability

As the first step in proving the achievability part of Theorem 15.32, let us recall Corollary 15.24: given a bipartite state ρ_{AB} with purification ψ_{ABE}, for all $\varepsilon \in (0,1)$, $\varepsilon' = 1-\sqrt{1-\varepsilon}$, $\delta \in (0,\varepsilon')$, $\eta \in (0,\varepsilon'-\delta)$, $\zeta \in (0,\delta)$, $\nu \in (0,\delta-\zeta)$, $\alpha \in (0,1)$, and $\beta > 1$, there exists a (K,ε) one-way key distillation protocol for ρ_{AB} satisfying

$$\log_2 K \geq \overline{I}_\alpha(X;B')_\rho - \widetilde{I}'_\beta(X;EZ)_\rho - f(\varepsilon',\delta,\eta,\nu,\zeta,\alpha,\beta) \tag{15.2.25}$$

where

$$\rho_{XB'EZ} \coloneqq \mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}(\psi_{ABE}), \tag{15.2.26}$$

$\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$ is an LOPC channel with classical output system X and quantum output system B', the Rényi mutual information $\widetilde{I}'_\beta(X;EZ)_\rho$ is defined in (15.1.259), and the function $f(\varepsilon',\delta,\eta,\nu,\zeta,\alpha,\beta)$ in (15.1.260). Applying this inequality to the state $\rho_{AB}^{\otimes n}$ for all $n \in \mathbb{N}$ leads to the following:

Proposition 15.33

For every state ρ_{AB} and $\varepsilon \in (0,1)$, there exists an (n,K,ε) key distillation protocol for ρ_{AB} such that the rate $\frac{\log_2 K}{n}$ satisfies

$$\frac{\log_2 K}{n} \geq \overline{I}_\alpha(X;B')_\rho - \widetilde{I}'_\beta(X;EZ)_\rho - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right), \tag{15.2.27}$$

for all $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\beta > 1$, where the information quantities are with respect to the state in (15.2.26). More generally, we have the following lower bound on the finite-length distillable key:

$$K_D^{n,\varepsilon}(A;B)_\rho \geq \frac{1}{n}\sup_{\mathcal{L}^{\leftrightarrow}}\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^nZ)_\tau\right) - \frac{1}{n}f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right), \quad (15.2.28)$$

for all $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\beta > 1$, where the optimization is over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{A^nB^n\to XB'Z}$ and the information quantities are with respect to the following state:

$$\tau_{XB'E^nZ} := \mathcal{L}^{\leftrightarrow}_{A^nB^n\to XB'Z}(\psi^{\otimes n}_{ABE}). \quad (15.2.29)$$

PROOF: Let ψ_{ABE} be a purification of ρ_{AB}, and use the tensor-product purification $\psi^{\otimes n}_{ABE}$ for $\rho^{\otimes n}_{AB}$. Also, let $\delta = \frac{\varepsilon'}{2}$, $\eta = \frac{\varepsilon'}{4}$, $\nu = \frac{\varepsilon'}{4}$, and $\zeta = \frac{\varepsilon'}{2}$. Substituting all of this into the inequality in (15.2.25) and simplifying leads to the inequality

$$\frac{\log_2 K}{n} \geq \frac{1}{n}\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^nZ)_\tau - f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right)\right), \quad (15.2.30)$$

where $\tau_{XB'E^nZ} := \mathcal{L}^{\leftrightarrow}_{A^nB^n\to XB'Z}(\psi^{\otimes n}_{ABE})$. Then, optimizing over every LOPC channel $\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}$, and using the definition of $K_D^{n,\varepsilon}(A;B)_\rho$ in (15.2.3), we obtain (15.2.28).

By restricting the state $\tau_{XB'E^nZ}$ to have the form $\tau^{\otimes n}_{XB'EZ} = \left(\mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}(\psi_{ABE})\right)^{\otimes n}$ (i.e., using a tensor-power LOPC strategy) and employing additivity of $\overline{I}_\alpha(X^n;B'^n)_{\tau^{\otimes n}}$ and $\widetilde{I}'_\beta(X^n;E^nZ^n)_{\tau^{\otimes n}}$, we conclude that

$$\overline{I}_\alpha(X^n;B'^n)_{\tau^{\otimes n}} - \widetilde{I}'_\beta(X^n;E^nZ^n)_{\tau^{\otimes n}} = n\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;EZ)_\tau\right). \quad (15.2.31)$$

The lower bound in (15.2.27) then follows. ■

Using the inequality in (15.2.27), we can prove the following lower bound on distillable key:

Theorem 15.34 Achievability of Private Information for Secret Key Distillation

The private information $I(X;B')_\tau - I(X;EZ)_\tau$ is an achievable rate for secret key distillation for ρ_{AB}, where the private information is evaluated with respect to the following state:

$$\tau_{XB'EZ} \coloneqq \mathcal{L}^{\leftrightarrow}_{AB\to XB'Z}(\psi_{ABE}), \tag{15.2.32}$$

and ψ_{ABE} is a purification of ρ_{AB}. In other words,

$$K_D(A;B)_\rho \geq I(X;B')_\tau - I(X;EZ)_\tau \tag{15.2.33}$$

for every bipartite state ρ_{AB}.

PROOF: Let ψ_{ABE} be a purification of ρ_{AB}. Fix $\varepsilon \in (0,1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that $\delta = \delta_1 + \delta_2$. Set $\alpha \in (0,1)$ and $\beta > 1$ such that

$$\delta_1 \geq I(X;B')_\tau - I(X;EZ)_\tau - \left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;EZ)_\tau\right). \tag{15.2.34}$$

Note that this is possible because $\overline{I}_\alpha(X;B')_\tau$ increases monotonically with increasing $\alpha \in (0,1)$ (see Proposition 7.23) and $\widetilde{I}'_\beta(X;EZ)_\tau$ decreases monotonically with decreasing β (see Proposition 7.31), so that

$$\lim_{\alpha\to 1^-} \overline{I}_\alpha(X;B')_\tau = \sup_{\alpha\in(0,1)} \overline{I}_\alpha(X;B')_\tau, \tag{15.2.35}$$

$$\lim_{\beta\to 1^+} \widetilde{I}'_\beta(X;EZ)_\tau = \inf_{\beta\in(1,\infty)} \widetilde{I}'_\beta(X;EZ)_\tau. \tag{15.2.36}$$

Also,

$$I(X;B')_\tau = \lim_{\alpha\to 1^-} \overline{I}_\alpha(X;B')_\tau, \tag{15.2.37}$$

$$I(X;EZ)_\tau = \lim_{\beta\to 1^+} \widetilde{I}'_\beta(X;EZ)_\tau. \tag{15.2.38}$$

With α and β chosen such that (15.2.34) holds, take n large enough so that

$$\delta_2 \geq \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{15.2.39}$$

Now, we use the fact that for the n and ε chosen above, there exists an (n, K, ε) protocol such that

$$\frac{\log_2 K}{n} \geq \overline{I}_\alpha(X;B')_\rho - \widetilde{I}'_\beta(X;EZ)_\rho - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{15.2.40}$$

(This follows from Proposition 15.33 above.) Rearranging the right-hand side of this inequality, and using (15.2.34), (15.2.39), and (15.2.40), we find that

$$\frac{\log_2 K}{n} \geq I(X;B')_\tau - I(X;EZ)_\tau - \left(\begin{array}{c} I(X;B')_\tau - I(X;EZ)_\tau - \left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;EZ)_\tau\right) \\ +\frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right) \end{array} \right) \tag{15.2.41}$$

$$\geq I(X;B')_\tau - I(X;EZ)_\tau - (\delta_1 + \delta_2) \tag{15.2.42}$$

$$= I(X;B')_\tau - I(X;EZ)_\tau - \delta. \tag{15.2.43}$$

We thus have shown that there exists an (n, K, ε) secret key distillation protocol with rate $\frac{\log_2 K}{n} \geq I(X;B')_\tau - I(X;EZ)_\tau - \delta$. Therefore, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ secret-key distillation protocol with $R = I(X;B')_\tau - I(X;EZ)_\tau$ for all sufficiently large n such that (15.2.39) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(I(X;B')_\tau - I(X;EZ)_\tau - \delta)}, \varepsilon)$ secret key distillation protocol. This means that, by definition, $I(X;B')_\tau - I(X;EZ)_\tau$ is an achievable rate. ■

Proof of the Achievability Part of Theorem 15.32

Let $\mathcal{L}^{\leftrightarrow}_{A^kB^k\to XB'Z}$ be an arbitrary LOPC channel with $k \in \mathbb{N}$, let

$$\tau_{XB'E^kZ} := \mathcal{L}^{\leftrightarrow}_{A^kB^k\to XB'Z}(\psi^{\otimes k}_{ABE}), \tag{15.2.44}$$

where ψ_{ABE} is a purification of ρ_{AB}. Fix $\varepsilon \in (0, 1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that $\delta = \delta_1 + \delta_2$. Set $\alpha \in (0, 1)$ and $\beta \in (1, \infty)$ such that

$$\delta_1 \geq \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) - \frac{1}{k}\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^kZ)_\tau\right), \tag{15.2.45}$$

which is possible based on the arguments given in the proof of Theorem 15.34 above. Then, with this choice of α and β, take n large enough so that

$$\delta_2 \geq \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{15.2.46}$$

Now, we use the fact that, for the chosen n and ε, there exists an (n, K, ε) secret-key distillation protocol such that (15.2.27) holds, i.e.,

$$\frac{\log_2 K}{n} \geq \overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^kZ)_\tau - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{15.2.47}$$

Dividing both sides by k gives

$$\frac{\log_2 K}{kn} \geq \frac{1}{k}\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^kZ)_\tau\right) - \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{15.2.48}$$

Rearranging the right-hand side of this inequality, and using (15.2.45)–(15.2.48), we find that

$$\begin{aligned}\frac{\log_2 K}{kn} &\geq \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right)\\ &\quad - \left(\frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) - \frac{1}{k}\left(\overline{I}_\alpha(X;B')_\tau - \widetilde{I}'_\beta(X;E^kZ)_\tau\right)\right)\\ &\quad - \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right) &(15.2.49)\\ &\geq \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) - (\delta_1 + \delta_2) &(15.2.50)\\ &= \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) - \delta. &(15.2.51)\end{aligned}$$

Thus, there exists a (kn, K, ε) secret-key distillation protocol with rate $\frac{\log_2 K}{kn} \geq \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) - \delta$. Therefore, letting $n' \equiv kn$, we conclude that there exists an $(n', 2^{n'(R-\delta)}, \varepsilon)$ secret-key distillation protocol with

$$R = \frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) \tag{15.2.52}$$

for all sufficiently large n such that (15.2.46) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(\frac{1}{k}(I(X;B')_\tau - I(X;E^kZ)_\tau)-\delta)}, \varepsilon)$ secret key distillation protocol. This means that $\frac{1}{k}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right)$ is an achievable rate.

Now, since in the arguments above the LOPC channel $\mathcal{L}^{\leftrightarrow}_{A^kB^k \to XB'Z}$ is arbitrary, we conclude that

$$\frac{1}{k}\sup_{\mathcal{L}^{\leftrightarrow}}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) \tag{15.2.53}$$

is an achievable rate. Finally, since the number k of copies of ρ_{AB} is arbitrary, we conclude that

$$\lim_{k\to\infty}\frac{1}{k}\sup_{\mathcal{L}^{\leftrightarrow}}\left(I(X;B')_\tau - I(X;E^kZ)_\tau\right) \tag{15.2.54}$$

is an achievable rate.

15.2.2 Proof of the Weak Converse

In order to prove the weak converse part of Theorem 15.32, we make use of Corollary 15.17, specifically (15.1.144): given a bipartite state ρ_{AB}, for every (K, ε) secret key distillation protocol for ρ_{AB}, with $\varepsilon \in [0, 1)$, the following bound holds

$$\left(1 - 2\sqrt{\varepsilon} - \delta\right) \log_2 K \leq \sup_{\mathcal{L}} \left(I(X; B')_{\mathcal{L}(\psi)} - I(X; EZ)_{\mathcal{L}(\psi)}\right) + h_2(\sqrt{\varepsilon} + \delta) + \left(1 - \sqrt{\varepsilon} - \delta\right) \log_2\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon}), \quad (15.2.55)$$

where $\delta \in \left(0, 1 - \sqrt{\varepsilon}\right)$, ψ_{ABE} is a purification of ρ_{AB}, and the information quantities are evaluated on the state $\mathcal{L}_{AB \to XB'Z}(\psi_{ABE})$. Applying this inequality to the state $\rho_{AB}^{\otimes n}$ leads to the following.

Proposition 15.35

Let ρ_{AB} be a bipartite state, let $n \in \mathbb{N}$, $\varepsilon \in [0, 1)$, and $\delta \in (0, 1 - \sqrt{\varepsilon})$. For an (n, K, ε) secret-key distillation protocol for ρ_{AB} with corresponding LOPC channel $\mathcal{L}_{A^n B^n \to XB'Z}$, with classical systems X and Z, the rate $\frac{\log_2 K}{n}$ satisfies

$$\left(1 - 2\sqrt{\varepsilon} - \delta\right) \frac{\log_2 K}{n} \leq \frac{1}{n} \sup_{\mathcal{L}} \left(I(X; B')_{\mathcal{L}(\psi^{\otimes n})} - I(X; EZ)_{\mathcal{L}(\psi^{\otimes n})}\right) + \frac{1}{n}\left(h_2(\sqrt{\varepsilon} + \delta) + \left(1 - \sqrt{\varepsilon} - \delta\right) \log_2\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon})\right). \quad (15.2.56)$$

Consequently,

$$\left(1 - 2\sqrt{\varepsilon} - \delta\right) K_D^{n,\varepsilon}(A; B)_\rho \leq \frac{1}{n} \sup_{\mathcal{L}} \left(I(X; B')_{\mathcal{L}(\psi^{\otimes n})} - I(X; EZ)_{\mathcal{L}(\psi^{\otimes n})}\right) + \frac{1}{n}\left(h_2(\sqrt{\varepsilon} + \delta) + \left(1 - \sqrt{\varepsilon} - \delta\right) \log_2\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon})\right), \quad (15.2.57)$$

where the optimization is over every LOCC channel $\mathcal{L}_{A^n B^n \to XB'Z}$.

Proof of the Weak Converse Part of Theorem 15.32

Suppose that R is an achievable rate for secret key distillation for the bipartite state ρ_{AB}. Then, by definition, for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ secret-key distillation protocol for ρ_{AB}. For all such protocols, the inequality in (15.2.56) holds, so that

$$\left(1 - 2\sqrt{\varepsilon} - \delta'\right)(R-\delta) \leq \frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right) + \frac{1}{n}\left(h_2(\sqrt{\varepsilon}+\delta') + \left(1-\sqrt{\varepsilon}-\delta'\right)\log_2\!\left(\frac{1}{\delta'}\right) + 2g_2(\sqrt{\varepsilon})\right). \quad (15.2.58)$$

Since the inequality holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$\begin{aligned}
&\left(1 - 2\sqrt{\varepsilon} - \delta'\right)(R-\delta) \\
&\leq \lim_{n\to\infty}\left(\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right)\right. \\
&\qquad \left. + \frac{1}{n}\left(h_2(\sqrt{\varepsilon}+\delta') + \left(1-\sqrt{\varepsilon}-\delta'\right)\log_2\!\left(\frac{1}{\delta}\right) + 2g_2(\sqrt{\varepsilon})\right)\right) \quad (15.2.59)\\
&= \lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right). \quad (15.2.60)
\end{aligned}$$

Then since this inequality holds for all $\varepsilon \in (0,1)$, $\delta > 0$, it holds in particular for $\delta' = \sqrt{\varepsilon}$, $\varepsilon \in (0, \frac{1}{9})$, which gives

$$R \leq \frac{1}{\left(1-3\sqrt{\varepsilon}\right)}\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right) + \delta, \quad (15.2.61)$$

and we thus conclude that

$$\begin{aligned}
R &\leq \lim_{\varepsilon,\delta\to 0}\frac{1}{\left(1-3\sqrt{\varepsilon}\right)}\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right) + \delta \quad (15.2.62)\\
&= \lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right). \quad (15.2.63)
\end{aligned}$$

We have thus shown that the quantity $\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{L}}\left(I(X;B')_{\mathcal{L}(\psi^{\otimes n})} - I(X;EZ)_{\mathcal{L}(\psi^{\otimes n})}\right)$ is a weak converse rate for secret key distillation for ρ_{AB}.

15.2.3 Relative Entropy of Entanglement Strong Converse Upper Bound

As indicated previously, the expression in (15.2.13) for distillable key involves both a limit over an unbounded number of copies of the initial state ρ_{AB}, as well as an optimization over all two-way LOPC channels. Computing the distillable key is therefore intractable in general. In this section, we use the one-shot upper bound established in Section 15.1.3.2 to show that the relative entropy of entanglement is a strong converse upper bound on the distillable key of a bipartite state ρ_{AB}.

We start by recalling the upper bound in (15.1.145), which tells us that

$$\log_2 K \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \qquad \forall \alpha > 1, \tag{15.2.64}$$

for an arbitrary (K,ε) secret-key distillation protocol, where $\varepsilon \in (0,1)$. Recall that

$$\widetilde{E}_\alpha(A;B)_\rho = \inf_{\sigma_{AB}\in \mathrm{SEP}(A:B)} \widetilde{D}_\alpha(\rho_{AB}\|\sigma_{AB}). \tag{15.2.65}$$

Recall that the upper bound above is a consequence of the fact that separable states are useless for secret key distillation.

Applying the upper bound in (15.2.64) to the state $\rho_{AB}^{\otimes n}$ leads to the following result:

Corollary 15.36

Let ρ_{AB} be a bipartite state, let $n \in \mathbb{N}$, $\varepsilon \in [0,1)$, and $\alpha > 1$. For an (n,K,ε) secret-key distillation protocol, the following bound holds

$$\frac{\log_2 K}{n} \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{n\,(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{15.2.66}$$

Consequently,

$$K_D^{n,\varepsilon}(A;B)_\rho \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{n\,(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{15.2.67}$$

Proof: An (n,K,ε) secret-key distillation protocol for ρ_{AB} is a (K,ε) secret-key distillation protocol for $\rho_{AB}^{\otimes n}$. Therefore, applying the inequality in (15.2.64) to the state $\rho_{AB}^{\otimes n}$ and dividing both sides by n leads to

$$\frac{\log_2 K}{n} \leq \frac{1}{n}\widetilde{E}_\alpha(A^n;B^n)_{\rho^{\otimes n}} + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{15.2.68}$$

Now, by subadditivity of the sandwiched Rényi relative entropy of entanglement (see (9.2.10)), we have that

$$\widetilde{E}_\alpha(A^n;B^n)_{\rho^{\otimes n}} \leq n\widetilde{E}_\alpha(A;B)_\rho. \tag{15.2.69}$$

Therefore,

$$\frac{\log_2 K}{n} \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \tag{15.2.70}$$

as required. Since this inequality holds for all (n, K, ε) protocols, we obtain (15.2.67) by optimizing over all key distillation protocols. ∎

Given an $\varepsilon \in (0,1)$, the inequality in (15.2.66) gives us a bound on the rate of an arbitrary (n, K, ε) secret-key distillation protocol for a state ρ_{AB}. If we instead fix the rate to be r, so that $K = 2^{nr}$, then the inequality in (15.2.66) is as follows:

$$r \leq \widetilde{E}_\alpha(A;B)_\rho + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right) \tag{15.2.71}$$

for all $\alpha > 1$. Rearranging this inequality gives us the following lower bound on ε:

$$\varepsilon \geq 1 - 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(r-\widetilde{E}_\alpha(A;B)_\rho\right)} \tag{15.2.72}$$

for all $\alpha > 1$.

Theorem 15.37 Strong Converse Upper Bound on Distillable Key

Let ρ_{AB} be a bipartite state. The relative entropy of entanglement $E_R(A;B)_\rho$ is a strong converse rate for secret key distillation for ρ_{AB}, i.e.,

$$\widetilde{K}_D(A;B)_\rho \leq E_R(A;B)_\rho, \tag{15.2.73}$$

where we recall that $E_R(A;B)_\rho$ is defined as

$$\inf_{\sigma_{AB}\in\mathrm{SEP}(A:B)} D(\rho_{AB}\|\sigma_{AB}). \tag{15.2.74}$$

PROOF: The proof is identical that given for Theorem 13.24, except we make use of (15.2.66). ■

Given that the relative entropy of entanglement is a strong converse rate for distillable key, by following arguments analogous to those in the referenced proof, we conclude that $\frac{1}{k}E_R(A^k;B^k)_{\rho^{\otimes k}}$ is a strong converse rate for all $k \in \mathbb{N}$. Therefore, the regularized quantity

$$E_R^{\text{reg}}(A;B)_\rho := \lim_{n\to\infty} \frac{1}{n} E_R(A^n;B^n)_{\rho^{\otimes n}} \tag{15.2.75}$$

is a strong converse rate for secret key distillation for ρ_{AB}, so that

$$\widetilde{K}_D(A;B)_\rho \leq E_R(A;B)_\rho. \tag{15.2.76}$$

By the subaddivity of relative entropy of entanglement (see (9.2.10)),

$$E_R^{\text{reg}}(A;B)_\rho \leq E_R(A;B)_\rho, \tag{15.2.77}$$

so that the regularized quantity in general gives a tighter upper bound on distillable key.

15.2.4 Squashed Entanglement Weak Converse Upper Bound

In this section, we establish the squashed entanglement of a bipartite state as a weak converse upper bound on its distillable key. The main idea is to apply the one-shot bound from Theorem 15.20 and the additivity of the squashed entanglement (Proposition 9.32) in order to arrive at this conclusion.

Corollary 15.38

Let ρ_{AB} be a bipartite state, let $n \in \mathbb{N}$, and let $\varepsilon \in [0,1)$. For an (n,K,ε) secret-key distillation protocol, the following bound holds

$$\left(1 - 2\sqrt{\varepsilon}\right) \frac{1}{n} \log_2 K \leq E_{\text{sq}}(A;B)_\rho + \frac{2}{n} g_2(\sqrt{\varepsilon}). \tag{15.2.78}$$

PROOF: An (n,K,ε) secret-key distillation protocol for ρ_{AB} is a (K,ε) secret-key distillation protocol for $\rho_{AB}^{\otimes n}$. Therefore, applying the inequality in (15.1.190) to the state $\rho_{AB}^{\otimes n}$ and dividing both sides by n leads to

$$\left(1 - 2\sqrt{\varepsilon}\right) \frac{1}{n} \log_2 K \leq \frac{1}{n} E_{\text{sq}}(A^n; B^n)_{\rho^{\otimes n}} + \frac{2}{n} g_2(\sqrt{\varepsilon}). \tag{15.2.79}$$

Now, by additivity of the squashed entanglement (Proposition 9.32), we have that

$$E_{\text{sq}}(A^n; B^n)_{\rho^{\otimes n}} = n E_{\text{sq}}(A; B)_\rho. \tag{15.2.80}$$

This concludes the proof. ■

We now provide a proof of (15.2.15), the statement that the squashed entanglement is a weak converse rate for secret key distillation. Suppose that R is an achievable rate for secret key distillation for the bipartite state ρ_{AB}. Then, by definition, for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ secret-key distillation protocol for ρ_{AB}. For all such protocols, the inequality in (15.2.78) holds, so that

$$\left(1 - 2\sqrt{\varepsilon}\right) (R - \delta) \leq E_{\text{sq}}(A; B)_\rho + \frac{2}{n} g_2(\sqrt{\varepsilon}). \tag{15.2.81}$$

Since the inequality holds for all sufficiently large n, it holds in the limit $n \to \infty$, so that

$$\left(1 - 2\sqrt{\varepsilon}\right) (R - \delta) \leq \lim_{n\to\infty} \left(E_{\text{sq}}(A; B)_\rho + \frac{2}{n} g_2(\sqrt{\varepsilon}) \right) \tag{15.2.82}$$

$$= E_{\text{sq}}(A; B)_\rho. \tag{15.2.83}$$

Then, since this inequality holds for all $\varepsilon \in (0, 1]$ and $\delta > 0$, it holds in particular for all $\varepsilon \in (0, \frac{1}{4})$ and $\delta > 0$, implying that

$$R \leq \frac{1}{1 - 2\sqrt{\varepsilon}} E_{\text{sq}}(A; B)_\rho + \delta, \tag{15.2.84}$$

and furthermore, that

$$R \leq \lim_{\varepsilon,\delta\to 0} \left(\frac{1}{1 - 2\sqrt{\varepsilon}} E_{\text{sq}}(A; B)_\rho + \delta \right) \tag{15.2.85}$$

$$= E_{\text{sq}}(A; B)_\rho. \tag{15.2.86}$$

We have thus shown that the squashed entanglement is a weak converse rate for secret key distillation.

15.3 One-Way Secret Key Distillation

In Section 15.1.4, we considered a one-way secret-key distillation protocol to derive a lower bound on the one-shot distillable key of a bipartite state. In the asymptotic setting, this leads to the private information lower bound on the distillable key of a bipartite state ρ_{AB}, i.e.,

$$K_D(A;B)_\rho \geq I(X;B)_\tau - I(X;E)_\tau, \tag{15.3.1}$$

where

$$\tau_{XBE} \coloneqq \sum_x |x\rangle\langle x|_X \otimes \mathrm{Tr}_A[\Lambda_A^x \psi_{ABE}], \tag{15.3.2}$$

ψ_{ABE} is a purification of ρ_{AB}, and $\{\Lambda_A^x\}_x$ is a POVM. By reversing the roles of Alice and Bob in the protocol, we find that

$$K_D(A;B)_\rho \geq I(A;Y)_\omega - I(Y;E)_\omega, \tag{15.3.3}$$

where

$$\omega_{YAE} \coloneqq \sum_y |y\rangle\langle y|_Y \otimes \mathrm{Tr}_B[\Gamma_B^y \psi_{ABE}], \tag{15.3.4}$$

where $\{\Gamma_B^y\}_y$ is a POVM. Then, in general, we have the following lower bound on distillable key:

$$K_D(A;B)_\rho \geq \max\{I(X;B)_\tau - I(X;E)_\tau, I(A;Y)_\omega - I(Y;E)_\omega\}. \tag{15.3.5}$$

This private information lower bound can be improved by first applying a two-way LOPC channel to n copies of the given state, and then performing a one-way secret-key distillation protocol at the private information rate. This leads to

$$K_D(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{\leftrightarrow}} I(X;B')_{\mathcal{L}^{\leftrightarrow}(\psi^{\otimes n})} - I(X;E^n Z)_{\mathcal{L}^{\leftrightarrow}(\psi^{\otimes n})}, \tag{15.3.6}$$

where the information quantities are evaluated on the state $\mathcal{L}^{\leftrightarrow}_{A^nB^n\to XB'Z}(\psi_{ABE}^{\otimes n})$, $\mathcal{L}^{\leftrightarrow}_{A^nB^n\to XB'Z}$ is an LOPC channel with classical systems X and Z, and ψ_{ABE} is a purification of ρ_{AB}.

If we restrict the optimization in (15.3.6) above to one-way LOPC channels of the form $\mathcal{L}^{\rightarrow}_{A^nB^n\to XB'Z}$, then we obtain what is called the one-way distillable key of ρ_{AB}, denoted by $K_D^{\rightarrow}(A;B)_\rho$, and defined operationally in a similar way to the distillable key $K_D(A;B)_\rho$, but with the free operations allowed restricted to one-way LOPC. A key result is the following equality:

$$K_D^{\rightarrow}(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\mathcal{L}^{\rightarrow}} I(X;B')_{\mathcal{L}^{\rightarrow}(\psi^{\otimes n})} - I(X;E^n Z)_{\mathcal{L}^{\rightarrow}(\psi^{\otimes n})} \tag{15.3.7}$$

$$= \lim_{n\to\infty} \frac{1}{n} D_K^{\rightarrow}(\rho_{AB}^{\otimes n}), \tag{15.3.8}$$

where

$$D_K^{\rightarrow}(\rho_{AB}) \coloneqq \sup_{\mathcal{L}^{\rightarrow}} I(X;B')_{\mathcal{L}^{\rightarrow}(\psi)} - I(X;EZ)_{\mathcal{L}^{\rightarrow}(\psi)}. \tag{15.3.9}$$

Like the distillable key, the one-way distillable key is an operational quantity of interest. Furthermore, the equality in (15.3.7) can be proved similarly to how we proved (15.2.13).

In what follows, we show that this expression for one-way distillable key can be simplified.

Theorem 15.39 One-Way Distillable Key of a Bipartite State

The one-way distillable key of a bipartite state ρ_{AB} is given by

$$K_D^{\rightarrow}(A;B)_\rho = \lim_{n\to\infty} \frac{1}{n} \sup_{\{\Lambda_{A^n}^{x,z}\}_{x\in\mathcal{X},z\in\mathcal{Z}}} I(X;B^n|Z)_\tau - I(X;E^n|Z)_\tau, \tag{15.3.10}$$

where

$$\tau_{XZB^nE^n} \coloneqq \sum_{x\in\mathcal{X},z\in\mathcal{Z}} |x\rangle\!\langle x|_X \otimes |z\rangle\!\langle z|_Z \otimes \operatorname{Tr}_{A^n}[\Lambda_{A^n}^{x,z}\psi_{ABE}^{\otimes n}], \tag{15.3.11}$$

and the optimization is over every POVM $\{\Lambda_{A^n}^{x,z}\}_{x\in\mathcal{X},z\in\mathcal{Z}}$ with output alphabets $\mathcal{X}$ and $\mathcal{Z}$.

This theorem tells us that, to determine the one-way distillable key of a bipartite state, it suffices to optimize over one-way LOPC channels that consist of a POVM conducted on Alice's systems.

Proof: Let us start by recalling from Definition 4.22 and the discussion around

(15.1.2) that every one-way LOPC channel $\mathcal{L}^{\rightarrow}_{A^nB^n\to XB'Z}$ can be expressed as

$$\omega_{XB'Z} \coloneqq \mathcal{L}^{\rightarrow}_{A^nB^n\to XB'Z}(\xi_{A^nB^n}) \tag{15.3.12}$$

$$= \sum_{z\in\mathcal{Z}} (\mathcal{E}^z_{A^n\to X}\otimes\mathcal{D}^z_{B^n\to B'})(\xi_{A^nB^n})\otimes|z\rangle\!\langle z|_Z, \tag{15.3.13}$$

$$= (\mathcal{D}_{Z_BB^n\to B'}\circ\mathcal{C}_{Z_A\to Z_BZ}\circ\mathcal{E}_{A^n\to XZ_A})(\xi_{A^nB^n}), \tag{15.3.14}$$

where $\mathcal{Z}$ is some finite alphabet, $\{\mathcal{E}^z_{A^n\to X}\}_{z\in\mathcal{Z}}$ is a set of completely positive maps such that $\sum_{z\in\mathcal{Z}}\mathcal{E}^z_{A^n\to X}$ is trace preserving, and $\{\mathcal{D}^z_{B^n\to B'}\}_{z\in\mathcal{Z}}$ is a set of channels. Furthermore,

$$\mathcal{E}_{A^n\to XZ_A}(\xi_{A^nB^n}) = \sum_{z\in\mathcal{Z}}\mathcal{E}^z_{A^n\to X}(\xi_{A^nB^n})\otimes|z\rangle\!\langle z|_Z, \tag{15.3.15}$$

$$\mathcal{D}_{Z_BB^n\to B'}(|z\rangle\!\langle z|_{Z_B}\otimes\xi_{A^nB^n}) = \mathcal{D}^z_{B^n\to B'}(\xi_{A^nB^n}), \tag{15.3.16}$$

and since the map $\mathcal{E}^z_{A^n\to X}$ has a classical output X, it can be written as

$$\mathcal{E}^z_{A^n\to X}(\xi_{A^nB^n}) = \sum_{x\in\mathcal{X}}\mathrm{Tr}_{A^n}[\Lambda^{x,z}_{A^n}\xi_{A^nB^n}]|x\rangle\!\langle x|_X, \tag{15.3.17}$$

where $\{\Lambda^{x,z}_{A^n}\}_{x\in\mathcal{X},z\in\mathcal{Z}}$ is a POVM.

For every $n\in\mathbb{N}$, if we restrict the optimization in (15.3.7) to $\mathcal{D}^z_{B^n\to B'} = \mathrm{id}_{B^n}$ for all $z\in\mathcal{Z}$ and $\mathcal{E}^z_{A^n\to X}(\cdot) = \sum_{x\in\mathcal{X}}\mathrm{Tr}_{A^n}[\Lambda^{x,z}_{A^n}(\cdot)]$ for all $z\in\mathcal{Z}$, then the LOPC channel $\mathcal{L}^{\rightarrow}_{A^nB^n\to XB'Z}$ reduces to

$$\mathcal{L}^{\rightarrow}_{A^nB^n\to XB^nZ}(\xi_{A^nB^n}) = \sum_{x\in\mathcal{X},z\in\mathcal{Z}}\mathrm{Tr}_{A^n}[\Lambda^{x,z}_{A^n}\xi_{A^nB^n}]\otimes|x\rangle\!\langle x|_X\otimes|z\rangle\!\langle z|_Z \tag{15.3.18}$$

for every input state $\xi_{A^nB^n}$. We thus conclude that

$$K^{\rightarrow}_D(A;B)_\rho \geq \lim_{n\to\infty}\frac{1}{n}\sup_{\{\Lambda^{x,z}_{A^n}\}_{x\in\mathcal{X},z\in\mathcal{Z}}} I(X;B^n|Z)_\tau - I(X;E^n|Z)_\tau. \tag{15.3.19}$$

The rest of the proof is devoted to proving the reverse inequality. Let $\mathcal{L}^{\rightarrow}_{A^nB^n\to XB'Z}$ be an arbitrary LOPC channel of the form in (15.3.12)–(15.3.17). Consider that

$$I(X;B')_{\mathcal{L}^{\rightarrow}(\psi^{\otimes n})} - I(X;E^nZ)_{\mathcal{L}^{\rightarrow}(\psi^{\otimes n})}$$

$$\leq I(X;B^nZ)_{\mathcal{L}'(\psi^{\otimes n})} - I(X;E^nZ)_{\mathcal{L}'(\psi^{\otimes n})} \tag{15.3.20}$$

$$= I(X;Z)_{\mathcal{L}'(\psi^{\otimes n})} + I(X;B^n|Z)_{\mathcal{L}'(\psi^{\otimes n})} - I(X;Z)_{\mathcal{L}'(\psi^{\otimes n})} - I(X;E^n|Z)_{\mathcal{L}'(\psi^{\otimes n})} \tag{15.3.21}$$

$$= I(X;B^n|Z)_{\mathcal{L}'(\psi^{\otimes n})} - I(X;E^n|Z)_{\mathcal{L}'(\psi^{\otimes n})} \tag{15.3.22}$$

where

$$\mathcal{L}'_{A^nB^n\to XB^nZ}(\xi_{A^nB^n}) := \sum_{x\in\mathcal{X},z\in\mathcal{Z}} \mathrm{Tr}_{A^n}[\Lambda^{x,z}_{A^n}\xi_{A^nB^n}] \otimes |x\rangle\langle x|_X \otimes |z\rangle\langle z|_Z. \tag{15.3.23}$$

The inequality follows from data-processing with respect to the decoding channel $\mathcal{D}_{Z_BB^n\to B'}$ of Bob. This concludes the proof. ■

Lemma 15.40

For every bipartite state ρ_{AB}, the optimized private information lower bound on distillable key is non-negative, i.e., $D_K^{\rightarrow}(\rho_{AB}) \geq 0$.

Proof: Let ψ_{ABE} be a purification of ρ_{AB}, and consider the following Schmidt decomposition of ψ_{ABE}:

$$|\psi\rangle_{ABE} = \sum_{k=0}^{r-1} \sqrt{\lambda_k}|\phi_k\rangle_A \otimes |\varphi_k\rangle_{BE}. \tag{15.3.24}$$

Then let $\Lambda_A^{x,z} = |\phi_x\rangle\langle\phi_x|\delta_{z,x}$ so that the POVM measures in the local Schmidt basis of A and broadcasts the measurement result through x and z. It is then straightforward to show that the private information $I(X;B|Z) - I(X;E|Z) = 0$. Since the POVM we chose is a particular choice in the optimization for $D_K^{\rightarrow}(\rho_{AB}) \geq 0$, we conclude that $D_K^{\rightarrow}(\rho_{AB}) \geq I(X;B|Z) - I(X;E|Z) = 0$. ■

Lemma 15.41

For every bipartite state ρ_{AB}, the optimized private information lower bound on distillable key is not smaller than the coherent information of ρ_{AB}, i.e., $D_K^{\rightarrow}(\rho_{AB}) \geq I(A\rangle B)_\rho$. Thus, the coherent information is a lower bound for one-way distillable key:

$$K_D^{\rightarrow}(\rho_{AB}) \geq I(A\rangle B)_\rho. \tag{15.3.25}$$

PROOF: Let $\Lambda_A^{x,z} = |\varphi_x\rangle\langle\varphi_x|_A$ be a rank-one POVM for which there is no output z. Let the state after the measurement be as follows:

$$\tau_{XBE} \coloneqq \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \mathrm{Tr}_A[|\varphi_x\rangle\langle\varphi_x|_A \psi_{ABE}] \tag{15.3.26}$$

$$= \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \psi_{BE}^x, \tag{15.3.27}$$

where

$$p(x) \coloneqq \mathrm{Tr}[|\varphi_x\rangle\langle\varphi_x|_A \psi_{ABE}], \tag{15.3.28}$$

$$\psi_{BE}^x \coloneqq \frac{1}{p(x)} \mathrm{Tr}_A[|\varphi_x\rangle\langle\varphi_x|_A \psi_{ABE}]. \tag{15.3.29}$$

Note that each ψ_{BE}^x is a pure state. Then it follows that

$$I(X;B|Z)_\tau - I(X;E|Z)_\tau = I(X;B)_\tau - I(X;E)_\tau \tag{15.3.30}$$

$$= H(B)_\tau - H(B|X)_\tau - H(E)_\tau + H(E|X)_\tau \tag{15.3.31}$$

$$= H(B)_\tau - H(E)_\tau \tag{15.3.32}$$

$$= H(B)_\rho - H(E)_\rho \tag{15.3.33}$$

$$= I(A\rangle B)_\rho. \tag{15.3.34}$$

The first equality follows because the Z system is trivial. The third equality follows because $H(B|X)_\tau = H(E|X)_\tau$, which in turn follows because each state ψ_{BE}^x is pure. ■

15.4 Examples

We now consider classes of bipartite states and evaluate the upper and lower bounds on their distillable key that we have established in this chapter. In some cases, the distillable key can be determined exactly because the upper and lower bounds coincide.

15.4.1 Pure States

The simplest example for which distillable key can be determined exactly is the class of pure bipartite states. In this case, the coherent information lower bound from Lemma 15.41 and the relative entropy of entanglement upper bound from Theorem 15.32 coincide and are equal to the entropy of the reduced state. Thus, applying this same reasoning from Section 13.3.1, we conclude the following:

Theorem 15.42 Distillable Key for Pure States

The distillable key of a pure bipartite state ψ_{AB} is equal to the entropy of the reduced state on A, i.e.,

$$K_D(A;B)_\psi = H(A)_\psi. \tag{15.4.1}$$

15.4.2 Degradable and Anti-Degradable States

In Section 13.3.2, we defined degradable and anti-degradable states, and we proved that the one-way distillable entanglement of a degradable state is equal to its coherent information. Also, we proved that the one-way distillable entanglement of an anti-degradable state vanishes. It turns out that the same results hold for one-way distillable key.

Theorem 15.43 One-Way Distillable Key for Anti-Degradable States

For an anti-degradable state ρ_{AB}, the one-way distillable key is equal to zero, i.e., $K_D^{\rightarrow}(A;B)_\rho = 0$.

PROOF: This is a direct consequence of the definition of an anti-degradable state and the result in Theorem 15.39. Indeed, for an anti-degradable state ρ_{AB} with purification ψ_{ABE}, there exists an anti-degrading channel $\mathcal{A}_{E\to B}$ such that $\rho_{AB} = \mathcal{A}_{E\to B}(\psi_{AE})$. A similar statement holds for $\rho_{AB}^{\otimes n}$, i.e., $\rho_{AB}^{\otimes n} = [\mathcal{A}_{E\to B}(\psi_{AE})]^{\otimes n}$. Applying this fact and the data-processing inequality to the expression $I(X;B^n|Z)_\tau - I(X;E^n|Z)_\tau$ from Theorem 15.39, we conclude that

$$I(X;B^n|Z)_\tau - I(X;E^n|Z)_\tau \leq 0. \tag{15.4.2}$$

So we conclude that $K_D^{\rightarrow}(A;B)_\rho \leq 0$. Combined with the general lower bound from Lemma 15.40, we conclude that $K_D^{\rightarrow}(A;B)_\rho = 0$ for an anti-degradable state ρ_{AB}. ■

Theorem 15.44 One-Way Distillable Key for Degradable States

For a degradable state ρ_{AB}, we have

$$D_K^{\rightarrow}(\rho_{AB}) = I(A\rangle B)_\rho. \tag{15.4.3}$$

Consequently, $D_K^{\rightarrow}(\rho_{AB}^{\otimes n}) = nD_K^{\rightarrow}(\rho_{AB})$, and thus the one-way distillable key of a degradable state ρ_{AB} is equal to its coherent information:

$$K_D^{\rightarrow}(A;B)_\rho = I(A\rangle B)_\rho. \tag{15.4.4}$$

Proof: It suffices to prove the upper bound $D_K^{\rightarrow}(\rho_{AB}) \leq I(A\rangle B)_\rho$ because Lemma 15.41 established the lower bound $D_K^{\rightarrow}(\rho_{AB}) \geq I(A\rangle B)_\rho$ in general. Recall that the defining property of a degradable state ρ_{AB} with purification ψ_{ABE} is that there exists a degrading channel $\mathcal{D}_{B\to E}$ such that $\psi_{AE} = \mathcal{D}_{B\to E}(\rho_{AB})$. Then the same is true for the tensor-power states, i.e., $\psi_{AE}^{\otimes n} = (\mathcal{D}_{B\to E}(\rho_{AB}))^{\otimes n}$. Then consider that

$$I(X;B^n|Z)_\tau - I(X;E^n|Z)_\tau$$
$$= I(XZ;B^n)_\tau - I(Z;B^n)_\tau - [I(XZ;E^n)_\tau - I(Z;E^n)_\tau] \tag{15.4.5}$$
$$= I(XZ;B^n)_\tau - I(XZ;E^n)_\tau - [I(Z;B^n)_\tau - I(Z;E^n)_\tau] \tag{15.4.6}$$
$$\leq I(XZ;B^n)_\tau - I(XZ;E^n)_\tau, \tag{15.4.7}$$

where

$$\tau_{XZB^nE^n} = \sum_{x\in\mathcal{X},z\in\mathcal{Z}} |x\rangle\langle x|_X \otimes |z\rangle\langle z|_Z \otimes \mathrm{Tr}_{A^n}[\Lambda_{A^n}^{x,z}\psi_{ABE}^{\otimes n}]. \tag{15.4.8}$$

The sole inequality above follows from the data-processing inequality for mutual information and the fact that there is a degrading channel from B^n to E^n. Now let $\Lambda_{A^n}^{x,z} = \sum_y |\varphi^{x,y,z}\rangle\langle\varphi^{x,y,z}|_{A^n}$ be a rank-one decomposition of the POVM $\{\Lambda_{A^n}^{x,z}\}_{x,z}$ and define the following extension of the state $\tau_{XZB^nE^n}$:

$$\tau_{XZYB^nE^n} = \sum_{x\in\mathcal{X},z\in\mathcal{Z}} |x\rangle\langle x|_X \otimes |z\rangle\langle z|_Z \otimes |y\rangle\langle y|_Y \otimes \mathrm{Tr}_{A^n}[|\varphi^{x,y,z}\rangle\langle\varphi^{x,y,z}|_{A^n}\psi_{ABE}^{\otimes n}]. \tag{15.4.9}$$

Then consider that

$$I(XZ;B^n)_\tau - I(XZ;E^n)_\tau$$

$$= I(XZY; B^n)_\tau - I(Y; B^n|XZ)_\tau - [I(XZY; E^n)_\tau - I(Y; E^n|XZ)_\tau] \quad (15.4.10)$$
$$= I(XZY; B^n)_\tau - I(XZY; E^n)_\tau - [I(Y; B^n|XZ)_\tau - I(Y; E^n|XZ)_\tau] \quad (15.4.11)$$
$$\leq I(XZY; B^n)_\tau - I(XZY; E^n)_\tau \quad (15.4.12)$$
$$= H(B^n)_\tau - H(E^n)_\tau - [H(B^n|XZY)_\tau - H(E^n|XZY)_\tau] \quad (15.4.13)$$
$$= H(B^n)_\tau - H(E^n)_\tau \quad (15.4.14)$$
$$= H(B^n)_\psi - H(E^n)_\psi \quad (15.4.15)$$
$$= n\left(H(B)_\psi - H(E)_\psi\right) \quad (15.4.16)$$
$$= nI(A\rangle B)_\rho. \quad (15.4.17)$$

The sole inequality above follows from the data-processing inequality for conditional mutual information and the fact that there is a degrading channel from B^n to E^n. This concludes the proof. ■

15.5 Summary

In this chapter, we considered the task of secret key distillation, in which the goal is for Alice and Bob to convert a bipartite state to an approximate tripartite key state with as many secret key bits as possible. In doing so, they are allowed to perform local operations and public classical communication, in which an eavesdropper obtains a copy of all of the classical communication exchanged. The highest rate at which this can be accomplished is called the distillable key of the state. We began with the one-shot setting, in which we allow some error in the distillation protocol, and we determined lower and upper bounds on the number of approximate secret key bits that can be distilled. In the asymptotic setting, we proved that the private information of the state is an achievable rate, and we proved that the squashed entanglement and the relative entropy of entanglement are upper bounds. These latter quantities are the best known upper bounds on distillable key.

By performing secret key distillation and then the one-time pad protocol (described in the introduction of this chapter), Alice can transmit a classical message privately to Bob. This process thus induces an ideal private classical channel from Alice to Bob. If Alice and Bob are connected by a quantum channel, then they can use it to share a bipartite state, from which they can induce a private classical channel in the aforementioned manner. This is one way to communicate privately over a quantum channel. In the next chapter, we discuss other, more direct approaches for private communication, which give an optimal private communication strategy for some quantum channels.

15.6 Bibliographic Notes

The task of secret key distillation, like many tasks in quantum information theory, has its roots in classical information theory. Maurer (1993) and Ahlswede and Csiszár (1993) developed the theory of secret key distillation in the classical case. There, the assumption is that Alice, Bob, and Eve share a tripartite distribution p_{XYZ}, from which they are trying to extract an approximation of an ideal secret key by means of local operations and public classical communication. The quantum case considered here is thus a generalization of this scenario, with a tripartite pure state ψ_{ABE} replacing the classical distribution p_{XYZ}. Recall that the eavesdropper sharing the purifying system of a purification of Alice and Bob's state gives Eve more power, because she can realize any possible extension of Alice and Bob's state by acting on the purifying system.

The one-time pad protocol traces its roots much further back. It was invented by Vernam (1926), and its security was established by Shannon (1949). As discussed in the introduction of this chapter, the main application of secret key distillation is to distill a secret key that can be used in conjunction with the one-time pad protocol in order to transmit a message privately.

Much of the technical work on secret key distillation was motivated by the development of quantum key distribution (Bennett and Brassard, 1984; Ekert, 1991). An early paper on the topic is about privacy amplification (Bennett et al., 1995), which is a component of a key distillation protocol. Secret key distillation from a bipartite quantum state was then studied by a number of researchers, including (Devetak and Winter, 2005; Horodecki et al., 2005a; Christandl, 2006; Horodecki et al., 2008a, 2009a; Christandl et al., 2007, 2012).

The one-shot setting of secret key distillation was studied by Renes and Renner (2012) and Khatri et al. (2019). We follow the approach of Khatri et al. (2019) closely in this chapter.

The connection between the tripartite picture of secret key distillation and the bipartite picture of private state distillation was identified by Horodecki et al. (2005a, 2009a). This work led to the understanding of the difference between entanglement and secret key, and it allowed for using the tools of entanglement theory (such as entanglement measures) in the context of secret key distillation. In the context of asymptotic secret key distillation, the relative entropy of entanglement upper bound on distillable key was established by Horodecki et al. (2005a, 2009a) and the squashed entanglement upper bound by Christandl (2006); Christandl et al. (2007, 2012) (see also (Wilde, 2016) in this context).

The privacy test was defined by Horodecki et al. (2008b,a), and its use in establishing one-shot converse bound was established by Wilde et al. (2017). Lemma 15.13 is due to Wilde et al. (2017). Lemma 15.14 is implicit in the work of Horodecki et al. (2009a) and was explicitly proved by Wilde et al. (2017). Proposition 15.15 and Theorem 15.16 were established by Wilde et al. (2017). Lemma 15.18, Proposition 15.19, and Theorem 15.20 were established by Wilde (2016).

Theorem 15.21 was established by Khatri et al. (2019). The convex split method was introduced by Anshu et al. (2017), and the smooth variant in Lemma 15.22 is due to Khatri et al. (2019), making use of methods in the appendix of Liu and Winter (2019). Lemma 15.25 is a variant of a result in Anshu et al. (2019) and was established by Khatri et al. (2019) (see also Wilde (2017b)).

The expression for distillable key in (15.2.13) of Theorem 15.32 is due to Devetak and Winter (2005). Eq. (15.2.66) is due to Wilde et al. (2017), and Eq. (15.2.78) to Wilde (2016). One-way secret key distillation was also considered by Devetak and Winter (2005). Theorems 15.43 and 15.44 were established by Leditzky (2019).

Appendix 15.A Proof of Smooth Convex Split Lemma

In this appendix, we prove Lemma 15.22.

Let $\widetilde{\rho}_{AE}$ be an arbitrary state satisfying $P(\widetilde{\rho}_{AE}, \rho_{AE}) \leq \sqrt{\varepsilon} - \eta$ and such that

$$\rho_A \otimes \widetilde{\rho}_E = p\widetilde{\rho}_{AE} + (1-p)\,\omega_{AE}, \tag{15.A.1}$$

for some $p \in (0,1)$ and ω_{AE} some state. We define the following state, which we think of as an approximation to $\tau_{A_1 \cdots A_R E}$:

$$\widetilde{\tau}_{A_1 \cdots A_R E} := \frac{1}{R}\sum_{r=1}^{R} \rho_{A_1} \otimes \cdots \otimes \rho_{A_{r-1}} \otimes \widetilde{\rho}_{A_r E} \otimes \rho_{A_{r+1}} \otimes \cdots \otimes \rho_{A_R}. \tag{15.A.2}$$

It is a good approximation if $\sqrt{\varepsilon} - \eta$ is small, because

$$\begin{aligned}
&\sqrt{F}(\tau_{A_1\cdots A_R E}, \widetilde{\tau}_{A_1\cdots A_R E}) \\
&\geq \frac{1}{R}\sum_{r=1}^{R} \sqrt{F}(\rho_A^{\otimes r-1} \otimes \rho_{A_r E} \otimes \rho_A^{\otimes R-r}, \rho_A^{\otimes r-1} \otimes \widetilde{\rho}_{A_r E} \otimes \rho_A^{\otimes R-r}) \qquad (15.A.3)\\
&= \frac{1}{R}\sum_{r=1}^{R} \sqrt{F}(\rho_{A_r E}, \widetilde{\rho}_{A_r E}) \qquad (15.A.4)\\
&= \sqrt{F}(\rho_{AE}, \widetilde{\rho}_{AE}), \qquad (15.A.5)
\end{aligned}$$

where the inequality follows from the concavity of the root fidelity (Theorem 6.11). This in turn implies that

$$\sqrt{F}(\tau_{A_1\cdots A_R E}, \widetilde{\tau}_{A_1\cdots A_R E}) \geq \sqrt{F}(\rho_{AE}, \widetilde{\rho}_{AE}). \qquad (15.A.6)$$

So the inequality in (15.A.6), the definition of the sine distance (Definition 6.16), and the fact that $P(\widetilde{\rho}_{AE}, \rho_{AE}) \leq \sqrt{\varepsilon} - \eta$, imply that

$$P(\tau_{A_1\cdots A_R E}, \widetilde{\tau}_{A_1\cdots A_R E}) \leq \sqrt{\varepsilon} - \eta. \qquad (15.A.7)$$

Now, let us define the following states:

$$\begin{aligned}
\beta_{AE} &:= \rho_A \otimes \widetilde{\rho}_E, \qquad (15.A.8)\\
\alpha_{AE} &:= \widetilde{\rho}_{AE}, \qquad (15.A.9)\\
\widetilde{\tau}_{A^R E^R} &:= \frac{1}{R}\sum_{r=1}^{R} \beta_{A_1 E_1} \otimes \cdots \otimes \beta_{A_{r-1}E_{r-1}} \otimes \alpha_{A_r E_r} \otimes \beta_{A_{r+1}E_{r+1}} \otimes \cdots \otimes \beta_{A_R E_R}, \qquad (15.A.10)
\end{aligned}$$

and observe that

$$\begin{aligned}
\mathrm{Tr}_{E_2^R}[(\beta_{AE})^{\otimes R}] &= \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E, \qquad (15.A.11)\\
\mathrm{Tr}_{E_2^R}[\widetilde{\tau}_{A^R E^R}] &= \widetilde{\tau}_{A_1\cdots A_R E}. \qquad (15.A.12)
\end{aligned}$$

Thus, it follows from the data-processing inequality for the sine distance that

$$P(\widetilde{\tau}_{A_1\cdots A_R E}, \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E) \leq P(\widetilde{\tau}_{A^R E^R}, (\beta_{AE})^{\otimes R}). \qquad (15.A.13)$$

Now consider that

$$(\beta_{AE})^{\otimes R} = (p\widetilde{\rho}_{AE} + (1-p)\,\omega_{AE})^{\otimes R} \qquad (15.A.14)$$

$$= \sum_{S \subset [R]} p^{|S|} (1-p)^{R-|S|} \widetilde{\rho}_{AE}^{\otimes S} \otimes \omega_{AE}^{\otimes [R] \setminus S} \tag{15.A.15}$$

$$= \sum_{k=0}^{R} \binom{R}{k} p^k (1-p)^{n-k} \theta_k \tag{15.A.16}$$

where θ_k is the following state:

$$\theta_k := \frac{1}{\binom{R}{k}} \sum_{|S|=k} \widetilde{\rho}_{AE}^{\otimes S} \otimes \omega_{AE}^{\otimes [R] \setminus S}. \tag{15.A.17}$$

Also, consider that

$$\widetilde{\tau}_{A^R E^R} = \frac{1}{R} \sum_{r=1}^{R} \beta_{A_1 E_1} \otimes \cdots \otimes \beta_{A_{r-1} E_{r-1}} \otimes \alpha_{A_r E_r}$$

$$\otimes \beta_{A_{r+1} E_{r+1}} \otimes \cdots \otimes \beta_{A_R E_R} \tag{15.A.18}$$

$$= \sum_{\emptyset \neq S \subset [R]} p^{|S|-1} (1-p)^{R-|S|} \widetilde{\rho}_{AE}^{\otimes S} \otimes \omega_{AE}^{\otimes [R] \setminus S} \tag{15.A.19}$$

$$= \sum_{k=1}^{R} \binom{R-1}{k-1} p^{k-1} (1-p)^{R-k} \theta_k \tag{15.A.20}$$

$$= \sum_{k=0}^{R} \frac{k}{Rp} \binom{R}{k} p^k (1-p)^{R-k} \theta_k. \tag{15.A.21}$$

In the last line, we used the identity $\binom{R-1}{k-1} = \frac{k}{Rp}\binom{R}{k}$. Defining the following classical–quantum states:

$$\beta_{A^R E^R K} := \sum_{k=0}^{R} \binom{R}{k} p^k (1-p)^{n-k} \theta_k \otimes |k\rangle\langle k|_K, \tag{15.A.22}$$

$$\widetilde{\tau}_{A^R E^R K} := \sum_{k=0}^{R} \frac{k}{Rp} \binom{R}{k} p^k (1-p)^{R-k} \theta_k \otimes |k\rangle\langle k|_K, \tag{15.A.23}$$

consider that

$$\sqrt{F}((\beta_{AE})^{\otimes R}, \widetilde{\tau}_{A^R E^R})$$

$$\geq \sqrt{F}(\beta_{A^R E^R K}, \widetilde{\tau}_{A^R E^R K}) \tag{15.A.24}$$

$$= \sum_{k=0}^{R} \sqrt{\binom{R}{k} p^k (1-p)^{n-k}} \sqrt{\frac{k}{Rp} \binom{R}{k} p^k (1-p)^{R-k}} \sqrt{F}(\theta_k, \theta_k) \tag{15.A.25}$$

$$= \sum_{k=0}^{R} \binom{R}{k} p^k (1-p)^{R-k} \sqrt{\frac{k}{Rp}} \tag{15.A.26}$$

$$= \sqrt{\frac{1}{Rp}} \sum_{k=0}^{R} \binom{R}{k} p^k (1-p)^{R-k} \sqrt{k} \tag{15.A.27}$$

$$= \sqrt{\frac{1}{Rp}} \mathbb{E}_K \left[\sqrt{K} \right], \tag{15.A.28}$$

where $\mathbb{E}_K$ denotes the expectation with respect to the binomial random variable K. The first inequality follows from the data-processing inequality for fidelity with respect to partial trace. The other steps follow by direct evaluation. Let $\mu = Rp$ (i.e., the mean of a binomial random variable). Consider that the following inequality holds for all $k \geq 0$ and $\mu > 0$:

$$\sqrt{k} \geq \sqrt{\mu} + \frac{k-\mu}{2\sqrt{\mu}} - \frac{(k-\mu)^2}{2\mu^{3/2}}. \tag{15.A.29}$$

Then we find that

$$\sqrt{\frac{1}{Rp}} \mathbb{E}_K \left[\sqrt{K} \right] \geq \sqrt{\frac{1}{Rp}} \mathbb{E}_K \left[\sqrt{\mu} + \frac{K-\mu}{2\sqrt{\mu}} - \frac{(K-\mu)^2}{2\mu^{3/2}} \right] \tag{15.A.30}$$

$$= \sqrt{\frac{1}{Rp}} \left(\sqrt{\mu} - \frac{\mathrm{Var}(K)}{2\mu^{3/2}} \right) \tag{15.A.31}$$

$$= \sqrt{\frac{1}{Rp}} \left(\sqrt{Rp} - \frac{\mathrm{Var}(K)}{2\,(Rp)^{3/2}} \right) \tag{15.A.32}$$

$$= 1 - \frac{Rp\,(1-p)}{(Rp)^2} \tag{15.A.33}$$

$$= 1 - \frac{(1-p)}{Rp} \tag{15.A.34}$$

$$\geq 1 - \frac{1}{Rp}. \tag{15.A.35}$$

Thus it follows that

$$\sqrt{F}((\beta_{AE})^{\otimes R}, \widetilde{\tau}_{A^R E^R}) \geq 1 - \frac{\eta^2}{2} \tag{15.A.36}$$

if

$$\log_2 R \geq \log_2(1/p) + \log_2\left(\frac{2}{\eta^2}\right). \tag{15.A.37}$$

This implies that

$$P((\beta_{AE})^{\otimes R}, \widetilde{\tau}_{A^R E^R}) \leq \eta. \tag{15.A.38}$$

For the same choice of R, it follows from (15.A.13) that

$$P(\widetilde{\tau}_{A_1 \cdots A_R E}, \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E) \leq \eta. \tag{15.A.39}$$

Applying the triangle inequality to (15.A.7) and (15.A.39), we find that

$$P(\tau_{A_1 \cdots A_R E}, \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E) \leq \sqrt{\varepsilon}. \tag{15.A.40}$$

The whole argument above holds for an arbitrary state $\widetilde{\rho}_{AE}$ satisfying $P(\widetilde{\rho}_{AE}, \rho_{AE}) \leq \sqrt{\varepsilon} - \eta$ and (15.A.1), and so taking an infimum of $\log_2(1/p)$ over p and all states satisfying these conditions, and applying the definition in (15.1.211), as well as Lemma 7.59, we find that

$$P(\tau_{A_1 \cdots A_R E}, \rho_{A_1} \otimes \cdots \otimes \rho_{A_R} \otimes \widetilde{\rho}_E) \leq \sqrt{\varepsilon} \tag{15.A.41}$$

if

$$\log_2 R \geq \overline{I}_{\max}^{\sqrt{\varepsilon}-\eta}(E;A)_\rho + \log_2\left(\frac{2}{\eta^2}\right). \tag{15.A.42}$$

This concludes the proof.

Appendix 15.B Relating Two Variants of Smooth-Max Mutual Information

In this appendix, we prove Lemma 15.25. The steps consist of constructing a state $\widehat{\rho}_{AE}$ such that

$$D_{\max}(\widehat{\rho}_{AE} \| \rho_A \otimes \widehat{\rho}_E) \leq D_{\max}(\widetilde{\rho}_{AE} \| \rho_A \otimes \rho_E) + \log_2\left(\frac{8}{\delta^2}\right) \tag{15.B.1}$$

and $P(\widehat{\rho}_{AE}, \rho_{AE}) \leq \varepsilon + \delta$. Using these inequalities, we can apply the definition of $\overline{I}_{\max}^{\varepsilon+\delta}(E;A)_\rho$ to conclude the desired inequality in (15.1.265). We begin by showing the first inequality, and after that, we establish the second one.

We begin by establishing some preparatory facts. Let $\widetilde{\rho}_{AE}$ be a state satisfying $P(\widetilde{\rho}_{AE}, \rho_{AE}) \leq \varepsilon$. Let $\gamma = \delta^2/8$, and set Π_E^γ to be the projection onto the positive eigenspace of $\frac{1}{\gamma}\widetilde{\rho}_E - \rho_E$. Then it follows that

$$\Pi_E^\gamma \left(\frac{1}{\gamma}\widetilde{\rho}_E - \rho_E\right) \Pi_E^\gamma \geq 0 \quad \Rightarrow \quad \Pi_E^\gamma \rho_E \Pi_E^\gamma \leq \frac{1}{\gamma}\Pi_E^\gamma \widetilde{\rho}_E \Pi_E^\gamma = \frac{8}{\delta^2}\Pi_E^\gamma \widetilde{\rho}_E \Pi_E^\gamma, \tag{15.B.2}$$

and

$$\left(I - \Pi_E^\gamma\right)\left(\frac{1}{\gamma}\widetilde{\rho}_E - \rho_E\right)\left(I - \Pi_E^\gamma\right) \leq 0$$
$$\Rightarrow \operatorname{Tr}[\left(I - \Pi_E^\gamma\right)\widetilde{\rho}_E] \leq \gamma \operatorname{Tr}[\left(I - \Pi_E^\gamma\right)\rho_E] \leq \gamma = \frac{\delta^2}{8}, \tag{15.B.3}$$

where the last inequality follows because $\operatorname{Tr}[\left(I - \Pi_E^\gamma\right)\rho_E] \leq 1$. The inequality in (15.B.3) can be rewritten as

$$\operatorname{Tr}[\Pi_E^\gamma \widetilde{\rho}_E] \geq 1 - \frac{\delta^2}{8}. \tag{15.B.4}$$

We now establish (15.B.1). Let us define the following states:

$$\overline{\rho}_{AEX} := \Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma \otimes |0\rangle\langle 0|_X + \left(I - \Pi_E^\gamma\right)\widetilde{\rho}_{AE}\left(I - \Pi_E^\gamma\right) \otimes |1\rangle\langle 1|_X, \tag{15.B.5}$$

$$\widehat{\rho}_{AEX} := \left(\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2}\left(I - \Pi_E^\gamma\right)\widetilde{\rho}_E^{1/2}\right) \otimes |0\rangle\langle 0|_X, \tag{15.B.6}$$

so that

$$\widehat{\rho}_{AE} = \operatorname{Tr}_X[\widehat{\rho}_{AEX}] \tag{15.B.7}$$
$$= \Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2}\left(I - \Pi_E^\gamma\right)\widetilde{\rho}_E^{1/2}. \tag{15.B.8}$$

Then, using the inequality $\widetilde{\rho}_{AE} \leq \mu \rho_A \otimes \rho_E$, with

$$\mu := 2^{D_{\max}(\widetilde{\rho}_{AE} \| \rho_A \otimes \rho_E)}, \tag{15.B.9}$$

and the fact that $\mu \frac{8}{\delta^2} \geq 1$ (which holds because $D_{\max}(\widetilde{\rho}_{AE} \| \rho_A \otimes \rho_E) \geq 0$ and $8 \geq \delta^2$), we find that

$$\widehat{\rho}_{AE} \leq \mu \rho_A \otimes \Pi_E^\gamma \rho_E \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2}\left(I - \Pi_E^\gamma\right)\widetilde{\rho}_E^{1/2} \tag{15.B.10}$$

$$\leq \mu \frac{8}{\delta^2} \rho_A \otimes \Pi_E^\gamma \widetilde{\rho}_E \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2} \left(I - \Pi_E^\gamma\right) \widetilde{\rho}_E^{1/2} \tag{15.B.11}$$

$$\leq \mu \frac{8}{\delta^2} \left[\rho_A \otimes \Pi_E^\gamma \widetilde{\rho}_E \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2} \left(I - \Pi_E^\gamma\right) \widetilde{\rho}_E^{1/2}\right] \tag{15.B.12}$$

$$= \mu \frac{8}{\delta^2} \rho_A \otimes \left[\Pi_E^\gamma \widetilde{\rho}_E \Pi_E^\gamma + \widetilde{\rho}_E^{1/2} \left(I - \Pi_E^\gamma\right) \widetilde{\rho}_E^{1/2}\right] \tag{15.B.13}$$

$$= \mu \frac{8}{\delta^2} \rho_A \otimes \widehat{\rho}_E. \tag{15.B.14}$$

The second inequality above follows from (15.B.2). Applying the definition of $D_{\max}(\widehat{\rho}_{AE} \| \rho_A \otimes \widehat{\rho}_E)$, we conclude that

$$D_{\max}(\widehat{\rho}_{AE} \| \rho_A \otimes \widehat{\rho}_E) \leq D_{\max}(\widetilde{\rho}_{AE} \| \rho_A \otimes \rho_E) + \log_2\!\left(\frac{8}{\delta^2}\right). \tag{15.B.15}$$

We can conclude the statement of the lemma if $P(\widehat{\rho}_{AE}, \rho_{AE}) \leq \varepsilon + \delta$, and so it is our aim to show this now. Consider that

$$P(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX}) = \sqrt{1 - F(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX})}. \tag{15.B.16}$$

The following chain of inequalities holds

$$\sqrt{F(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX})}$$

$$= \operatorname{Tr}\!\left[\left(\sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma} \widehat{\rho}_{AE} \sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma}\right)^{1/2}\right] \tag{15.B.17}$$

$$\geq \operatorname{Tr}\!\left[\left(\sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma} \left(\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma\right) \sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma}\right)^{1/2}\right] \tag{15.B.18}$$

$$= \operatorname{Tr}\left[\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma\right] \tag{15.B.19}$$

$$= \operatorname{Tr}[\Pi_E^\gamma \widetilde{\rho}_E] \tag{15.B.20}$$

$$\geq 1 - \frac{\delta^2}{8}, \tag{15.B.21}$$

where the inequality follows from operator monotonicity of the square root and the fact that

$$\widehat{\rho}_{AE} = \Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma + \rho_A \otimes \widetilde{\rho}_E^{1/2} \left(I - \Pi_E^\gamma\right) \widetilde{\rho}_E^{1/2} \tag{15.B.22}$$

$$\geq \Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma \tag{15.B.23}$$

From the above and (15.B.4), we conclude that $F(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX}) \geq 1 - \frac{\delta^2}{4}$, which implies that

$$P(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX}) \leq \frac{\delta}{2}. \tag{15.B.24}$$

Now consider that

$$\begin{aligned}
&P(\overline{\rho}_{AEX}, \rho_{AE} \otimes |0\rangle\!\langle 0|_X) \\
&\leq P(\overline{\rho}_{AEX}, \widetilde{\rho}_{AE} \otimes |0\rangle\!\langle 0|_X) \\
&\qquad + P(\widetilde{\rho}_{AE} \otimes |0\rangle\!\langle 0|_X, \rho_{AE} \otimes |0\rangle\!\langle 0|_X) &(15.B.25)\\
&= \sqrt{1 - F(\overline{\rho}_{AEX}, \widetilde{\rho}_{AE} \otimes |0\rangle\!\langle 0|_X)} + P(\widetilde{\rho}_{AE}, \rho_{AE}) &(15.B.26)\\
&= \sqrt{1 - \left\| \sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_E^\gamma} \sqrt{\widetilde{\rho}_{AE}} \right\|_1^2} + P(\widetilde{\rho}_{AE}, \rho_{AE}) &(15.B.27)\\
&= \sqrt{1 - \left\| \sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_A^\gamma} \sqrt{\Pi_E^\gamma \widetilde{\rho}_{AE} \Pi_A^\gamma} \right\|_1^2} + P(\widetilde{\rho}_{AE}, \rho_{AE}) &(15.B.28)\\
&= \sqrt{1 - \left(\mathrm{Tr}[\Pi_E^\gamma \widetilde{\rho}_{AE}]\right)^2} + P(\widetilde{\rho}_{AE}, \rho_{AE}) &(15.B.29)\\
&\leq \frac{\delta}{2} + \varepsilon, &(15.B.30)
\end{aligned}$$

where we applied the triangle inequality of the sine distance (Lemma 6.17) for the first inequality and the fact that $\left\| \sqrt{\Pi \omega \Pi} \sqrt{\tau} \right\|_1 = \left\| \sqrt{\Pi \omega \Pi} \sqrt{\Pi \tau \Pi} \right\|_1$ for a projector Π and states ω and τ. Combining this with (15.B.24), we find that

$$\begin{aligned}
P(\widehat{\rho}_{AE}, \rho_{AE}) &= P(\widehat{\rho}_{AEX}, \rho_{AE} \otimes |0\rangle\!\langle 0|_X) &(15.B.31)\\
&\leq P(\widehat{\rho}_{AEX}, \overline{\rho}_{AEX}) + P(\overline{\rho}_{AEX}, \rho_{AE} \otimes |0\rangle\!\langle 0|_X) &(15.B.32)\\
&= \varepsilon + \delta. &(15.B.33)
\end{aligned}$$

Since we have found a state $\widehat{\rho}_{AE}$ satisfying $P(\widehat{\rho}_{AE}, \rho_{AE}) \leq \varepsilon + \delta$ and (15.B.15), we conclude that

$$\widetilde{I}_{\max}^{\varepsilon+\delta}(E;A)_\rho \leq D_{\max}(\widetilde{\rho}_{AE} \| \rho_A \otimes \rho_E) + \log_2\!\left(\frac{8}{\delta^2}\right). \tag{15.B.34}$$

Since this inequality has been shown for all states $\widetilde{\rho}_{AE}$ satisfying $P(\widetilde{\rho}_{AE}, \rho_{AE}) \leq \varepsilon$, we conclude the statement of the lemma.

Chapter 16

Private Communication

This chapter focuses on the task of private communication, in which the goal is for a sender to communicate classical information privately over a quantum channel to a receiver, such that the environment of the channel gains essentially no information about the message transmitted. There are connections between this task and secret key distillation from Chapter 15, as well as with quantum communication from Chapter 14. Private communication can be considered a dynamic version of the general problem of establishing secret correlations between two parties, whereas secret key distillation is a static version of the same problem. Indeed, the resource shared between the two parties in the former task is a quantum channel (a dynamic resource), whereas the resource shared in the latter is a bipartite quantum state (a static resource). The cryptographic models are similar as well: in key distillation, we assumed that an eavesdropper possesses the purifying system of a purification of the shared state, whereas, in this chapter, we assume that an eavesdropper possesses the purifying system of a purification of the channel connecting the sender to receiver (i.e., the eavesdropper possesses the environment of the channel). The connection of private communication to quantum communication is as follows: if two parties can communicate some amount of quantum information with some error, then the amount of private information that they can communicate is related to this amount by an inequality. This inequality in turn implies that the private capacity of a quantum channel is not smaller than its quantum capacity.

As with other communication tasks that we have considered in previous chapters, there are multiple ways to define how communication can be private, based on various error criteria. In this chapter, we define two such criteria that lead to two different but related communication tasks, one that we call secret-key transmission and another that we call private communication. The criterion for the former task is most similar to an average error criterion, in which the goal is for the sender to use the channel to transmit one share of a secret key to the receiver, and the criterion for the latter task is a maximal infidelity criterion, in which all messages transmitted over the channel are required to meet a particular error criterion, which captures both the decoding error probability of the receiver, as well as the security of the message transmitted.

As usual by now, we begin our development in the one-shot setting, with the goal of establishing lower and upper bounds on the one-shot private capacity. We find several upper bounds on the one-shot private capacity, in terms of the one-shot private information of the channel, the hypothesis testing relative entropy of entanglement, and the squashed entanglement. The lower bound that we establish is related to a different variation of the one-shot private information of the channel (not the same quantity as in the upper bound), and we juxtapose the methods of position-based coding and convex splitting to prove the achievability of this one-shot private information. Some of the mathematical steps in the proof of the lower bound are similar to those that we used in the previous chapter, in which we established a lower bound on the one-shot distillable key. Moving on to the asymptotic setting, we prove that the private capacity of a quantum channel is equal to its regularized private information. This quantity is difficult to compute in general, and so we then establish some upper bounds on it in terms of the relative entropy of entanglement and squashed entanglement.

16.1 One-Shot Setting

Let $\mathcal{N}_{A\to B}$ be a quantum channel connecting a sender Alice to a receiver Bob, and let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending $\mathcal{N}_{A\to B}$, in the sense that $\mathcal{N}_{A\to B} = \operatorname{Tr}_E \circ \mathcal{U}^{\mathcal{N}}_{A\to BE}$. The goal of a private communication protocol is for Alice to communicate a classical message to Bob reliably, in the sense that Bob can decode it with high probability, and such that it is secure from anyone who possesses the environment system E (we personify the environment as the eavesdropper Eve). A private communication protocol in the one-shot setting is illustrated in Figure [REF]. It is defined by the three elements $(\mathcal{M}, \mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\hat{M}})$, in which $\mathcal{M}$ is a message set, $\mathcal{E}_{M'\to A}$ is an encoding channel, and $\mathcal{D}_{B\to\hat{M}}$ is a decoding channel. The pair $(\mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\hat{M}})$, consisting of the encoding and decoding channels, is called a private communication code or, more simply, a code. The encoding channel is a classical–quantum channel, and the decoding channel is a quantum–classical or measurement channel.

The steps of the protocol proceed similarly to those of the classical communication protocol discussed in Section 12.1, with the key difference that the message transmitted should be kept private from Eve. Let us employ notation similar to that discussed in (12.1.1)–(12.1.9), in which Alice's probability for selecting message $m \in \mathcal{M}$ is denoted by $p(m)$, the initial state is denoted by

$$\overline{\Phi}^p_{MM'} \coloneqq \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'}, \tag{16.1.1}$$

the state after the encoding channel by

$$\rho^p_{MA} \coloneqq \mathcal{E}_{M'\to A}(\overline{\Phi}^p_{MM'}) \tag{16.1.2}$$

$$= \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes \rho^m_A, \tag{16.1.3}$$

where we have defined

$$\rho^m_A \coloneqq \mathcal{E}_{M'\to A}(|m\rangle\langle m|_{M'}), \tag{16.1.4}$$

the state before the decoding channel by

$$\mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^p_{MA}), \tag{16.1.5}$$

and the final state of the protocol by

$$\omega^p_{M\hat{M}E} \coloneqq (\mathcal{D}_{B\to\hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}^p_{MM'}) \tag{16.1.6}$$

$$= \sum_{m,\hat{m}\in\mathcal{M}} p(m)|m\rangle\langle m|_M \otimes |\hat{m}\rangle\langle\hat{m}|_{\hat{M}} \otimes \mathrm{Tr}_B[\Lambda^{\hat{m}}_B \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^m_A)], \tag{16.1.7}$$

where we have used the fact that the decoding channel is a measurement channel and thus can be written in terms of a POVM $\{\Lambda^m_B\}_{m\in\mathcal{M}}$ as

$$\mathcal{D}_{B\to\hat{M}}(\tau_B) \coloneqq \sum_{\hat{m}\in\mathcal{M}} \mathrm{Tr}[\Lambda^{\hat{m}}_B \tau_B]\ |\hat{m}\rangle\langle\hat{m}|_{\hat{M}}. \tag{16.1.8}$$

If we define the following states

$$\omega^{m,\hat{m}}_E \coloneqq \frac{\mathrm{Tr}_B[\Lambda^{\hat{m}}_B \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^m_A)]}{q(\hat{m}|m)}, \tag{16.1.9}$$

$$q(\hat{m}|m) := \text{Tr}[\Lambda_B^{\hat{m}} \mathcal{U}_{A\to BE}^{\mathcal{N}}(\rho_A^m)] = \text{Tr}[\Lambda_B^{\hat{m}} \mathcal{N}_{A\to B}(\rho_A^m)], \tag{16.1.10}$$

then we can write the final state of the protocol alternatively as follows:

$$\omega_{M\hat{M}E}^p = \sum_{m,\hat{m}\in\mathcal{M}} p(m)q(\hat{m}|m)|m\rangle\langle m|_M \otimes |\hat{m}\rangle\langle\hat{m}|_{\hat{M}} \otimes \omega_E^{m,\hat{m}}. \tag{16.1.11}$$

The difference between a protocol for (public) communication, as discussed in Section 12.1, and one for private communication is the metric used for characterizing performance. Here, we demand that the message remain private from the eavesdropper in addition to being decodable by the receiver. We combine these constraints into a single metric, which we define in terms of the infidelity. We delineate two different cases, similar to how we did in Section 12.1:

1. The average infidelity of the code is given by

$$\begin{aligned}
&\overline{p}_{\text{err}}(\mathcal{E},\mathcal{D};p,\mathcal{N}) \\
&:= \inf_{\sigma_E}\left(1 - F(\overline{\Phi}_{M\hat{M}}^p \otimes \sigma_E, \mathcal{P}_{M'\to\hat{M}E}(\overline{\Phi}_{MM'}^p))\right) &(16.1.12)\\
&= \inf_{\sigma_E}\left(1 - \left(\sum_{m\in\mathcal{M}} p(m)\sqrt{F}(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E, \mathcal{P}_{M'\to\hat{M}E}(|m\rangle\langle m|_{M'}))\right)^2\right), &(16.1.13)
\end{aligned}$$

   where

$$\mathcal{P}_{M'\to\hat{M}E} := \mathcal{D}_{B\to\hat{M}} \circ \mathcal{U}_{A\to BE}^{\mathcal{N}} \circ \mathcal{E}_{M'\to A} \tag{16.1.14}$$

   and the infimum is taken over every state σ_E of the eavesdropper's system E. Also, we employed Proposition 7.31 with $\alpha = \frac{1}{2}$ in the last line above. If the prior probability distribution $p(m)$ is the uniform distribution (i.e., $p(m) = 1/|\mathcal{M}|$), then the communication task is called *secret-key transmission*, because the goal is for Alice to transmit one share of a secret key to the receiver Bob.

2. An alternative error criterion is the maximal infidelity of the code, defined as

$$p_{\text{err}}^*(\mathcal{E},\mathcal{D};\mathcal{N}) := \inf_{\sigma_E}\max_{m\in\mathcal{M}}\left(1 - F(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E, \mathcal{P}_{M'\to\hat{M}E}(|m\rangle\langle m|_{M'}))\right). \tag{16.1.15}$$

   When the communication task employs this error criterion, we refer to it as *private communication*.

The interpretation of the average infidelity obeying the inequality

$$\overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N}) \leq \varepsilon \tag{16.1.16}$$

is that there exists a state σ_E of the eavesdropper's system E such that the state of systems $\hat{M}$ and E is close to the product state $|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E$, on average. This means that not only can Bob can decode well, but also, that the state of Eve's system is close to the constant state σ_E, such that her system is not useful for figuring out the message transmitted (on average). Indeed, by applying the data-processing inequality with respect to partial trace of system E and letting σ_E be the state that achieves $\overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N})$, we conclude that

$$\begin{aligned}
\varepsilon &\geq \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N}) && (16.1.17)\\
&= 1 - \left(\sum_{m \in \mathcal{M}} p(m) \sqrt{F}(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E, \mathcal{P}_{M' \to \hat{M}E}(|m\rangle\langle m|_{M'})) \right)^2 && (16.1.18)\\
&\geq 1 - \sum_{m \in \mathcal{M}} p(m) F(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E, \mathcal{P}_{M' \to \hat{M}E}(|m\rangle\langle m|_{M'})) && (16.1.19)\\
&\geq \sum_{m \in \mathcal{M}} p(m) \left(1 - F(|m\rangle\langle m|_{\hat{M}}, (\mathcal{D}_{B \to \hat{M}} \circ \mathcal{N}_{A \to B} \circ \mathcal{E}_{M' \to A})(|m\rangle\langle m|_{M'}))\right) && (16.1.20)\\
&= \sum_{m \in \mathcal{M}} p(m) \left(1 - \langle m|_{\hat{M}} (\mathcal{D}_{B \to \hat{M}} \circ \mathcal{N}_{A \to B} \circ \mathcal{E}_{M' \to A})(|m\rangle\langle m|_{M'}) |m\rangle_{\hat{M}}\right) && (16.1.21)\\
&= \sum_{m \in \mathcal{M}} p(m) \left(1 - \operatorname{Tr}[\Lambda_B^m \mathcal{N}_{A \to B}(\rho_A^m)]\right). && (16.1.22)
\end{aligned}$$

The second inequality follows from convexity of the square function and the third from the data-processing inequality for fidelity. The latter expression is the same as the average error probability from (12.1.13). Now applying the data-processing inequality with respect to partial trace over $\hat{M}$, we conclude that

$$\begin{aligned}
\varepsilon &\geq \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N}) && (16.1.23)\\
&= 1 - F(\overline{\Phi}^p_{M\hat{M}} \otimes \sigma_E, (\mathcal{D}_{B \to \hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A \to BE} \circ \mathcal{E}_{M' \to A})(\overline{\Phi}^p_{MM'})) && (16.1.24)\\
&\geq 1 - F(\operatorname{Tr}_{\hat{M}}[\overline{\Phi}^p_{M\hat{M}} \otimes \sigma_E], \operatorname{Tr}_{\hat{M}}[(\mathcal{D}_{B \to \hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A \to BE} \circ \mathcal{E}_{M' \to A})(\overline{\Phi}^p_{MM'})]) && (16.1.25)\\
&= 1 - F(\pi^p_M \otimes \sigma_E, (\mathcal{N}^c_{A \to E} \circ \mathcal{E}_{M' \to A})(\overline{\Phi}^p_{MM'})) && (16.1.26)
\end{aligned}$$

$$= 1 - \left(\sum_{m \in \mathcal{M}} p(m) \sqrt{F}(\sigma_E, \mathcal{N}^c_{A\to E}(\rho^m_A)) \right)^2, \tag{16.1.27}$$

which indicates that the state of Eve's system E is close to the constant state σ_E on average. In the above, $\mathcal{N}^c_{A\to E}$ is a complementary channel of $\mathcal{N}_{A\to B}$, as defined in Section 4.3.2, and is given by $\mathcal{N}^c_{A\to E} = \mathrm{Tr}_B \circ \mathcal{U}^{\mathcal{N}}_{A\to BE}$. Also, in the last line above, we employed Proposition 7.31 with $\alpha = \frac{1}{2}$.

The interpretation of the maximum infidelity obeying the constraint

$$p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon \tag{16.1.28}$$

is similar. If this condition holds, then there exists a state σ_E of the eavesdropper's system E such that the state of systems $\hat{M}$ and E is close to the product state $|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E$, for every message $m \in \mathcal{M}$. So this is a much stronger constraint in general and the one we aim to achieve for private communication. By applying the data-processing inequality to $p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ and letting σ_E be the state that achieves $p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N})$, we conclude by similar reasoning as given above that

$$\varepsilon \geq p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \geq \max_{m \in \mathcal{M}} \left(1 - \mathrm{Tr}[\Lambda^m_B \mathcal{N}_{A\to B}(\rho^m_A)]\right), \tag{16.1.29}$$

and

$$\varepsilon \geq p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \geq \max_{m \in \mathcal{M}} \left(1 - F(\sigma_E, \mathcal{N}^c_{A\to E}(\rho^m_A))\right). \tag{16.1.30}$$

Thus, if $p^*_{\mathrm{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ holds, then Bob can reliably decode every message $m \in \mathcal{M}$, in the sense that

$$\mathrm{Tr}[\Lambda^m_B \mathcal{N}_{A\to B}(\rho^m_A)] \geq 1 - \varepsilon \qquad \forall m \in \mathcal{M}, \tag{16.1.31}$$

and Eve's system E is not useful for determining any of the messages, in the sense that

$$F(\sigma_E, \mathcal{N}^c_{A\to E}(\rho^m_A)) \geq 1 - \varepsilon \qquad \forall m \in \mathcal{M}. \tag{16.1.32}$$

These two different infidelity criteria can be used to assess the performance of a protocol, i.e., how well Bob can decode the message and how secure it is from Eve.

Definition 16.1 $(|\mathcal{M}|, \varepsilon)$ Private Communication Protocol

A private communication protocol $(\mathcal{M}, \mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\hat{M}})$ over the channel $\mathcal{N}_{A\to B}$ is called an $(|\mathcal{M}|, \varepsilon)$ protocol, with $\varepsilon \in [0,1]$, if $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$.

Similar to the case of entanglement-assisted and unassisted classical communication, the infidelity criterion $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon$ is equivalent to

$$\inf_{\sigma_E} \max_{p:\mathcal{M}\to[0,1]} \left(1 - F(\overline{\Phi}^p_{M\hat{M}} \otimes \sigma_E, (\mathcal{D}_{B\to\hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}^p_{MM'}))\right) \leq \varepsilon, \tag{16.1.33}$$

where the optimization is over every probability distribution $p(m)$ for the messages in $\mathcal{M}$. This follows because

$$F(\overline{\Phi}^p_{M\hat{M}} \otimes \sigma_E, (\mathcal{D}_{B\to\hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}^p_{MM'})) = \left[\sum_{m\in\mathcal{M}} p(m)\sqrt{F}(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_E, (\mathcal{D}_{B\to\hat{M}} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(|m\rangle\langle m|_{M'})))\right]^2, \tag{16.1.34}$$

as a consequence of Proposition 7.31 with $\alpha = \frac{1}{2}$, and then one can employ arguments similar to those in (11.1.25)–(11.1.34) to conclude (16.1.33).

The one-shot private capacity of the channel $\mathcal{N}$ is equal to the maximum number of private bits that can be transmitted for a fixed infidelity threshold ε:

Definition 16.2 One-Shot Private Capacity of a Quantum Channel

Given a quantum channel $\mathcal{N}_{A\to B}$ and $\varepsilon \in [0,1]$, the one-shot ε-error private capacity of $\mathcal{N}$, denoted by $P^\varepsilon(\mathcal{N})$, is defined to be the maximum number $\log_2|\mathcal{M}|$ of private bits among all $(|\mathcal{M}|, \varepsilon)$ private communication protocols over $\mathcal{N}$. In other words,

$$P^\varepsilon(\mathcal{N}) := \sup_{(\mathcal{M},\mathcal{E},\mathcal{D})} \left\{\log_2|\mathcal{M}| : p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \leq \varepsilon\right\}, \tag{16.1.35}$$

where the optimization is over all protocols $(\mathcal{M}, \mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\hat{M}})$ satisfying $d_{M'} = d_{\hat{M}} = |\mathcal{M}|$.

16.1.1 Private Communication and Quantum Communication

This subsection establishes that a quantum communication protocol can always be converted to one for private communication, such that there is negligible loss with respect to code parameters. This result then implies an inequality relating the one-shot quantum capacity to the one-shot private capacity.

Proposition 16.3

The existence of an (M, ε) quantum communication protocol for a quantum channel $\mathcal{N}_{A\to B}$ implies the existence of an $(\lfloor M/2 \rfloor, \min\{1, 2\varepsilon\})$ private communication protocol for $\mathcal{N}_{A\to B}$.

PROOF: Starting from an (M, ε) quantum communication protocol, we can use it to transmit one share of a maximally entangled state

$$\Phi_{RS} := \frac{1}{M} \sum_{m,m'=1}^{M} |m\rangle\langle m'|_R \otimes |m\rangle\langle m'|_S \tag{16.1.36}$$

of Schmidt rank M faithfully, by definition (see Definition 14.1):

$$F(\Phi_{RS}, (\mathcal{D}_{B\to S} \circ \mathcal{N}_{A\to B} \circ \mathcal{E}_{S'\to A})(\Phi_{RS'})) \geq 1 - \varepsilon. \tag{16.1.37}$$

Consider that the state

$$\sigma_{RSE} := (\mathcal{D}_{B\to S} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{S'\to A})(\Phi_{RS'}) \tag{16.1.38}$$

extends the state output from the actual protocol. By Uhlmann's theorem (Theorem 6.8), there exists an extension of Φ_{RS} such that the fidelity between this extension and the state σ_{RSE} is equal to the fidelity in (16.1.37). However, the maximally entangled state Φ_{RS} is unextendible in the sense that the only possible extension is a tensor-product state $\Phi_{RS} \otimes \omega_E$ for some state ω_E. So, putting these statements together, we find that

$$F(\Phi_{RS} \otimes \omega_E, (\mathcal{D}_{B\to S} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{S'\to A})(\Phi_{RS'})) \geq 1 - \varepsilon. \tag{16.1.39}$$

Furthermore, measuring the R and S systems locally in the Schmidt basis of Φ_{RS} only increases the fidelity, so that

$$F(\overline{\Phi}_{RS} \otimes \omega_{E^n}, (\overline{\mathcal{D}}_{B\to S} \circ \mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{S'\to A})(\overline{\Phi}_{RS})) \geq 1 - \varepsilon, \tag{16.1.40}$$

where $\overline{\mathcal{D}}_{B\to S}$ denotes the concatenation of the original decoder $\mathcal{D}_{B\to S}$ followed by the local measurement:

$$\overline{\mathcal{D}}_{B\to S}(\cdot) \coloneqq \sum_m |m\rangle\langle m|\mathcal{D}_{B\to S}(\cdot)|m\rangle\langle m| \tag{16.1.41}$$

$$= \sum_m \operatorname{Tr}[(\mathcal{D}_{B\to S})^\dagger[|m\rangle\langle m|](\cdot)]|m\rangle\langle m|_S. \tag{16.1.42}$$

Observe that $\{(\mathcal{D}_{B\to S})^\dagger[|m\rangle\langle m|]\}_m$ is a valid POVM. Using the direct-sum property of the fidelity (Proposition 7.31 with $\alpha = \frac{1}{2}$) and defining $\rho_A^m \coloneqq \mathcal{E}_{S'\to A}(|m\rangle\langle m|_{S'})$, we can then rewrite this as

$$\left(\frac{1}{M}\sum_{m=1}^{M}\sqrt{F}(|m\rangle\langle m|_S \otimes \omega_E, (\overline{\mathcal{D}}_{B\to S}\circ \mathcal{U}^{\mathcal{N}}_{A\to BE})(\rho_A^m))\right)^2 \geq 1-\varepsilon. \tag{16.1.43}$$

We can in turn rewrite this inequality as

$$\frac{1}{M}\sum_{m=1}^{M}\sqrt{F}(|m\rangle\langle m|_S \otimes \omega_E, (\overline{\mathcal{D}}_{B\to S}\circ \mathcal{U}^{\mathcal{N}}_{A\to BE})(\rho_A^m)) \geq \sqrt{1-\varepsilon} \tag{16.1.44}$$

and again as

$$\frac{1}{M}\sum_{m=1}^{M}\left(1-\sqrt{F}(|m\rangle\langle m|_S \otimes \omega_E, (\overline{\mathcal{D}}_{B\to S}\circ \mathcal{U}^{\mathcal{N}}_{A\to BE})(\rho_A^m))\right) \leq 1-\sqrt{1-\varepsilon} \tag{16.1.45}$$

Markov's inequality then guarantees that there exists a subset $\mathcal{M}'$ of the set $\{1,\ldots,M\}$ of size $\lfloor M/2\rfloor$ such that the following condition holds for all $m\in\mathcal{M}'$:

$$1-\sqrt{F}(|m\rangle\langle m|_S \otimes \omega_E, (\overline{\mathcal{D}}_{B\to S}\circ \mathcal{U}^{\mathcal{N}}_{A\to BE})(\rho_A^m)) \leq 2\left(1-\sqrt{1-\varepsilon}\right). \tag{16.1.46}$$

We can rewrite this condition as

$$F(|m\rangle\langle m|_S \otimes \omega_E, (\overline{\mathcal{D}}_{B\to S}\circ \mathcal{U}^{\mathcal{N}}_{A\to BE})(\rho_A^m)) \geq \left(1-2\left(1-\sqrt{1-\varepsilon}\right)\right)^2 \tag{16.1.47}$$

$$= \left(1-2\sqrt{1-\varepsilon}\right)^2 \tag{16.1.48}$$

$$\geq 1-2\varepsilon. \tag{16.1.49}$$

We now define the private communication protocol to consist of codewords $\{\rho_A^m := \mathcal{E}_{S\to A}(|m\rangle\langle m|_S)\}_{m\in\mathcal{M}'}$ and the decoding POVM to be

$$\{\Lambda_B^m \equiv (\mathcal{D}_{B\to S})^\dagger(|m\rangle\langle m|)\}_{m\in\mathcal{M}'} \cup \left\{\Lambda_{B^n}^0 := (\mathcal{D}_{B\to S})^\dagger\left(\sum_{m\notin\mathcal{M}'} |m\rangle\langle m|\right)\right\}. \quad (16.1.50)$$

Thus, we have shown that from an (M,ε) quantum communication protocol, one can realize an $(\lfloor M/2\rfloor, 2\varepsilon)$ protocol for private communication. ■

Proposition 16.3 then implies the following for the one-shot capacities:

Theorem 16.4

For a quantum channel $\mathcal{N}_{A\to B}$ and $\varepsilon \in (0,1)$, the following inequality relates the one-shot quantum capacity $Q^{\frac{\varepsilon}{2}}(\mathcal{N})$ to the one-shot private capacity $P^\varepsilon(\mathcal{N})$:

$$Q^{\frac{\varepsilon}{2}}(\mathcal{N}) \leq P^\varepsilon(\mathcal{N}) + 1. \quad (16.1.51)$$

PROOF: Given an arbitrary $(M,\varepsilon/2)$ quantum communication protocol, by Proposition 16.3, we can realize an arbitrary $(M/2,\varepsilon)$ private communication protocol. Letting the protocol be one that achieves the one-shot quantum capacity $Q^{\frac{\varepsilon}{2}}(\mathcal{N})$ (i.e., $\log_2 M = Q^{\frac{\varepsilon}{2}}(\mathcal{N})$), we conclude that there exists an $(M/2,\varepsilon)$ private communication protocol. Since this is a particular $(M/2,\varepsilon)$ private communication protocol, we conclude that

$$\log_2(M/2) \leq P^\varepsilon(\mathcal{N}), \quad (16.1.52)$$

which follows from the definition of the one-shot private capacity $P^\varepsilon(\mathcal{N})$. We finally use the fact that $\log_2(M/2) = Q^{\frac{\varepsilon}{2}}(\mathcal{N}) - 1$. ■

16.1.2 Secret-Key Transmission and Bipartite Private-State Transmission

In this section, we establish a connection between secret-key transmission and bipartite private-state transmission. Before doing so, we first define what is meant by a bipartite private-state transmission protocol. To do so, we follow the same spirit in Section 15.1.1, and we purify each step of a secret-key transmission protocol and trace out the system possessed by the eavesdropper Eve. In this case, it is only the environment E of the isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ that belongs to the eavesdropper, and tracing it out leads to the original channel $\mathcal{N}_{A\to B}$.

A bipartite private-state transmission protocol is defined by the triple

$$(\mathcal{M}, \mathcal{U}^{\mathcal{E}}_{M'\to AA'}, \mathcal{U}^{\mathcal{D}}_{B\to \hat{M}B'}), \tag{16.1.53}$$

where $\mathcal{M}$ is a message set, $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}$ is an isometric encoding channel, and $\mathcal{U}^{\mathcal{D}}_{B\to \hat{M}B'}$ is an isometric decoding channel. The protocol begins with Alice preparing a GHZ state $\Phi_{M''MM'}$ of the following form:

$$\Phi_{M''MM'} \coloneqq |\Phi\rangle\langle\Phi|_{M''MM'}, \tag{16.1.54}$$

where

$$|\Phi\rangle_{M''MM'} \coloneqq \frac{1}{\sqrt{|\mathcal{M}|}} \sum_{m\in\mathcal{M}} |m\rangle_{M''}|m\rangle_M|m\rangle_{M'}. \tag{16.1.55}$$

She transmits the M' system through the isometric encoding channel $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}$, leading to the state $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}(\Phi_{M''MM'})$. She transmits the A system through the channel $\mathcal{N}_{A\to B}$, leading to the state

$$\mathcal{N}_{A\to B}(\mathcal{U}^{\mathcal{E}}_{M'\to AA'}(\Phi_{M''MM'})). \tag{16.1.56}$$

Bob finally performs the isometric decoding channel $\mathcal{U}^{\mathcal{D}}_{B\to \hat{M}B'}$. The final state of the protocol is then as follows:

$$\omega_{M''MA'\hat{M}B'} \coloneqq (\mathcal{U}^{\mathcal{D}}_{B\to \hat{M}B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{U}^{\mathcal{E}}_{M'\to AA'})(\Phi_{M''MM'}), \tag{16.1.57}$$

where the systems $M''MA'$ are in possession of Alice and systems $\hat{M}B'$ are in possession of Bob.

Observe that each step of the protocol involves a purification of the steps in a secret-key transmission protocol, as outlined in Section 16.1. The initial GHZ state is a purification of the maximally classically correlated state $\overline{\Phi}_{MM'}$. The isometric encoding channel $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}$ purifies the encoding channel $\mathcal{E}_{M'\to A}$, and the isometric decoding channel $\mathcal{U}^{\mathcal{D}}_{B\to \hat{M}B'}$ purifies the decoding channel $\mathcal{D}_{B\to \hat{M}}$.

The infidelity of a bipartite private-state transmission protocol of the form above is then defined as follows:

$$p^{b}_{\text{err}}(\mathcal{U}^{\mathcal{E}}, \mathcal{U}^{\mathcal{D}}; \mathcal{N}) \coloneqq \inf_{\gamma_{M''MA'\hat{M}B'}} \left(1 - F(\gamma_{M''MA'\hat{M}B'}, \omega_{M''MA'\hat{M}B'})\right), \tag{16.1.58}$$

where the optimization is with respect to every ideal bipartite private state $\gamma_{M''MA'\hat{M}B'}$, with key system M held by Alice, shield systems $M''A'$ by Alice, key system $\hat{M}$ by Bob, and shield system B' by Bob (see Section 15.1.1).

Definition 16.5 ($(|\mathcal{M}|, \varepsilon)$ Private-State Transmission Protocol)

A bipartite private-state transmission protocol $(\mathcal{M}, \mathcal{U}^{\mathcal{E}}_{M' \to AA'}, \mathcal{U}^{\mathcal{D}}_{B \to \hat{M}B'})$ for the channel $\mathcal{N}_{A \to B}$ is called an $(|\mathcal{M}|, \varepsilon)$ protocol, with $\varepsilon \in [0, 1]$, if $p^b_{\text{err}}(\mathcal{U}^{\mathcal{E}}, \mathcal{U}^{\mathcal{D}}; \mathcal{N}) \leq \varepsilon$.

We now establish the main result of this section, which is the equivalence of secret-key transmission and bipartite private-state transmission:

Theorem 16.6

Let $\mathcal{M}$ be a message set, and let $\varepsilon \in [0, 1]$. Let $\mathcal{N}_{A \to B}$ be a quantum channel. There exists an $(|\mathcal{M}|, \varepsilon)$ secret-key transmission protocol for $\mathcal{N}_{A \to B}$ if and only if there exists an $(|\mathcal{M}|, \varepsilon)$ bipartite private-state transmission protocol for $\mathcal{N}_{A \to B}$.

Proof: We start by proving that there exists an $(|\mathcal{M}|, \varepsilon)$ bipartite private-state transmission protocol if there exists an $(|\mathcal{M}|, \varepsilon)$ secret-key transmission protocol. Let $\mathcal{U}^{\mathcal{N}}_{A \to BE}$ be an isometric channel extending $\mathcal{N}_{A \to B}$. Let $\mathcal{E}_{M' \to A}$ be the encoding channel, and let $\mathcal{D}_{B \to \hat{M}}$ be the decoding channel. The final state of the protocol is as follows:

$$\omega_{M\hat{M}E} := (\mathcal{D}_{B \to \hat{M}} \circ \mathcal{N}_{A \to B} \circ \mathcal{E}_{M' \to A})(\overline{\Phi}_{MM'}) \tag{16.1.59}$$

and satisfies the inequality

$$1 - F(\overline{\Phi}_{M\hat{M}} \otimes \sigma_E, \omega_{M\hat{M}E}) \leq \varepsilon. \tag{16.1.60}$$

Observe that $\overline{\Phi}_{M\hat{M}} \otimes \sigma_E$ is an ideal tripartite key state, according to Definition 15.1. Now let $\mathcal{U}^{\mathcal{E}}_{M' \to AA'}$ be an isometric channel that extends the encoding channel $\mathcal{E}_{M' \to A}$, and let $\mathcal{U}^{\mathcal{D}}_{B \to \hat{M}B'}$ be an isometric channel that extends the decoding channel $\mathcal{D}_{B \to \hat{M}}$. Let $\Phi_{M''MM'}$ be a GHZ state that purifies $\overline{\Phi}_{MM'}$. Then the following state is a purification of $\omega_{M\hat{M}E}$:

$$\omega_{MM''A'\hat{M}B'E} := (\mathcal{U}^{\mathcal{D}}_{B \to \hat{M}B'} \circ \mathcal{U}^{\mathcal{N}}_{A \to BE} \circ \mathcal{U}^{\mathcal{E}}_{M' \to AA'})(\Phi_{M''MM'}). \tag{16.1.61}$$

Applying Uhlmann's theorem, we conclude that there is a purification of $\overline{\Phi}_{M\hat{M}} \otimes \sigma_E$, call it $\gamma_{MM''A'\hat{M}B'E}$, such that

$$F(\overline{\Phi}_{M\hat{M}} \otimes \sigma_E, \omega_{M\hat{M}E}) = F(\gamma_{MM''A'\hat{M}B'E}, \omega_{MM''A'\hat{M}B'E}). \tag{16.1.62}$$

Tracing over the E system, we conclude from the data-processing inequality for fidelity that

$$1 - F(\gamma_{MM''A'\hat{M}B'}, \omega_{MM''A'\hat{M}B'}) \leq \varepsilon, \tag{16.1.63}$$

where

$$\omega_{MM''A'\hat{M}B'} = (\mathcal{U}^{\mathcal{D}}_{B\to\hat{M}B'} \circ \mathcal{N}_{A\to B} \circ \mathcal{U}^{\mathcal{E}}_{M'\to AA'})(\Phi_{M''MM'}). \tag{16.1.64}$$

Furthermore, by Definition 15.4, the state $\gamma_{MM''A'\hat{M}B'}$ is an ideal bipartite private state with key system M held by Alice, shield systems $M''A'$ held by Alice, key system $\hat{M}$ held by Bob, and shield system B' held by Bob. Thus, we have shown the first claim.

Now we establish the opposite implication (which follows essentially by running the argument above backwards). To this end, let $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}$ be an isometric encoding channel, and let $\mathcal{U}^{\mathcal{D}}_{B\to\hat{M}B'}$ be an isometric decoding channel for a bipartite private-state transmission protocol. The initial state of the protocol is the GHZ state $\Phi_{M''MM'}$, and the final state is $\omega_{MM''A'\hat{M}B'}$, as given in (16.1.64). For an $(|\mathcal{M}|, \varepsilon)$ bipartite private-state transmission protocol, the following inequality holds

$$1 - F(\gamma_{MM''A'\hat{M}B'}, \omega_{MM''A'\hat{M}B'}) \leq \varepsilon, \tag{16.1.65}$$

where $\gamma_{MM''A'\hat{M}B'}$ is an ideal bipartite private state. The state $\omega_{MM''A'\hat{M}B'E}$ in (16.1.61) is a purification of $\omega_{MM''A'\hat{M}B'}$, and by Uhlmann's theorem, there exists a purification $\gamma_{MM''A'\hat{M}B'E}$ of $\gamma_{MM''A'\hat{M}B'}$ such that

$$F(\gamma_{MM''A'\hat{M}B'}, \omega_{MM''A'\hat{M}B'}) = F(\gamma_{MM''A'\hat{M}B'E}, \omega_{MM''A'\hat{M}B'E}). \tag{16.1.66}$$

Tracing over the systems M'', A', and B', the following inequality holds

$$1 - F(\gamma_{M\hat{M}E}, \omega_{M\hat{M}E}) \leq \varepsilon. \tag{16.1.67}$$

By the definition of an ideal private state (see Definition 15.4) and since the state $\gamma_{MM''A'\hat{M}B'}$ is an ideal bipartite private state, it follows that $\gamma_{M\hat{M}E}$ is an ideal tripartite key state. Thus, we have proven the second claim. ■

16.1.3 Upper Bounds on the Number of Transmitted Private Bits

We now establish some general upper bounds on the number of private bits that can be communicated in an arbitrary private communication protocol. The results are stated in Proposition 16.7 and Theorems 16.9 and 16.11, and, like the upper bounds established in previous chapters, they hold independently of the encoding and decoding channels used in the protocol and depends only on the given communication channel $\mathcal{N}$. The first upper bound is in terms of the one-shot private information of the channel, and the others are in terms of the channel's ε-relative entropy of entanglement and squashed entanglement.

16.1.3.1 Private Information Upper Bound

Proposition 16.7 Upper Bound on One-Shot Private Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel. For every $(|\mathcal{M}|, \varepsilon)$ private communication protocol over $\mathcal{N}$, with $\varepsilon \in [0,1]$, the number of private bits transmitted over $\mathcal{N}$ is bounded from above by the one-shot private information of $\mathcal{N}$:

$$\log_2|\mathcal{M}| \leq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I_H^{\varepsilon}(X;B)_\rho - I_{\max}^{\sqrt{\varepsilon}}(X;E)_\rho \right), \tag{16.1.68}$$

where the optimization is over every ensemble $\{p(x),\rho_A^x\}_{x\in\mathcal{X}}$ and the state ρ_{XBE} is given by

$$\rho_{XBE} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}_{A\to BE}^{\mathcal{N}}(\rho_A^x), \tag{16.1.69}$$

with $\mathcal{U}_{A\to BE}^{\mathcal{N}}$ an isometric channel extending $\mathcal{N}_{A\to B}$. The hypothesis testing mutual information $I_H^{\varepsilon}(X;B)_\rho$ is defined in (7.11.88) and the smooth max-mutual information $I_{\max}^{\sqrt{\varepsilon}}(X;E)_\rho$ in (15.1.59). Therefore,

$$P^{\varepsilon}(\mathcal{N}) \leq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I_H^{\varepsilon}(X;B)_\rho - I_{\max}^{\sqrt{\varepsilon}}(X;E)_\rho \right). \tag{16.1.70}$$

Proof: The proof has some similarities with the proof of Lemma 15.10. Since

$p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}) \geq \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N})$ for every probability distribution $p(m)$ over the messages, it follows by definition that

$$P^{\varepsilon}(\mathcal{N}) \leq \sup_{(\mathcal{M},\mathcal{E},\mathcal{D})} \left\{\log_2|\mathcal{M}| : \overline{p}_{\text{err}}(\mathcal{E}, \mathcal{D}; p, \mathcal{N}) \leq \varepsilon\right\}, \tag{16.1.71}$$

with p set to the uniform distribution over messages. So we bound the right-hand side instead (note that it is equal to the one-shot secret-key transmission capacity). Let $(\mathcal{M}, \mathcal{E}_{M'\to A}, \mathcal{D}_{B\to\hat{M}})$ be an arbitrary private communication protocol. By the reasoning in (16.1.17)–(16.1.22), it follows that

$$\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \text{Tr}[\Lambda^m_B \mathcal{N}_{A\to B}(\rho^m_A)] \geq 1 - \varepsilon. \tag{16.1.72}$$

By the same reasoning given in the proof of Proposition 12.3, we conclude that

$$\log_2|\mathcal{M}| \leq I^{\varepsilon}_H(M; B)_\tau, \tag{16.1.73}$$

where the state τ_{MBE} is defined as

$$\tau_{MBE} \coloneqq \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^m_A). \tag{16.1.74}$$

Observe that

$$\tau_{MBE} = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes (\mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(|m\rangle\langle m|_{M'}) \tag{16.1.75}$$

$$= (\mathcal{U}^{\mathcal{N}}_{A\to BE} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}_{MM'}). \tag{16.1.76}$$

From (16.1.23)–(16.1.27), we know that there exists a state σ_E such that

$$\begin{aligned}\varepsilon &\geq 1 - F(\pi_M \otimes \sigma_E, (\widehat{\mathcal{N}}_{A\to E} \circ \mathcal{E}_{M'\to A})(\overline{\Phi}_{MM'})) \\ &= 1 - F(\tau_M \otimes \sigma_E, \tau_{ME}),\end{aligned}$$

which, by applying the same reasoning in (15.1.93)–(15.1.98), allows us to conclude that

$$I^{\sqrt{\varepsilon}}_{\max}(M; E)_\tau \leq 0. \tag{16.1.77}$$

Putting together (16.1.73) and (16.1.77) implies that

$$\log_2|\mathcal{M}| \leq I^{\varepsilon}_H(M; B)_\tau - I^{\sqrt{\varepsilon}}_{\max}(M; E)_\tau \tag{16.1.78}$$

$$\leq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I_H^{\varepsilon}(X;B)_\rho - I_{\max}^{\sqrt{\varepsilon}}(X;E)_\rho \right), \tag{16.1.79}$$

where the final inequality follows by noting that $\left\{\frac{1}{|\mathcal{M}|}, \rho_A^m\right\}_{m\in\mathcal{M}}$ is a particular input ensemble and the one-shot private information in the last line involves an optimization over all input ensembles. ■

As a consequence of the reasoning behind Proposition 16.7, along with (7.2.96), Proposition 7.70, and (15.1.155), we obtain the following:

Corollary 16.8

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(|\mathcal{M}|, \varepsilon)$ private communication protocols for $\mathcal{N}$, the following bound holds

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right)\log_2|\mathcal{M}| \leq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I(X;B)_\rho - I(X;E)_\rho\right) + h_2(\varepsilon) + 2g(\sqrt{\varepsilon}). \tag{16.1.80}$$

Consequently, the following bound holds for the one-shot private capacity of a channel $\mathcal{N}$:

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right) P^{\varepsilon}(\mathcal{N}) \leq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I(X;B)_\rho - I(X;E)_\rho\right) + h_2(\varepsilon) + 2g(\sqrt{\varepsilon}). \tag{16.1.81}$$

Proof: Employing the same reasoning that led to (16.1.73) and (16.1.77), consider that the following bounds hold for a given $(|\mathcal{M}|, \varepsilon)$ private communication protocol:

$$\log_2|\mathcal{M}| \leq I_H^{\varepsilon}(M;B)_\tau, \tag{16.1.82}$$

$$I_{\max}^{\sqrt{\varepsilon}}(M;E)_\tau \leq 0, \tag{16.1.83}$$

where the state τ_{MBE} is defined in (16.1.74). Now we apply (7.2.96) and Proposition 7.70 to conclude that

$$I_H^{\varepsilon}(M;B)_\tau \leq \frac{1}{1-\varepsilon}\left(I(M;B)_\tau + h_2(\varepsilon)\right), \tag{16.1.84}$$

which implies that

$$(1-\varepsilon)\log_2|\mathcal{M}| \le I(M;B)_\tau + h_2(\varepsilon), \tag{16.1.85}$$

and the same reasoning that led to (15.1.155) to conclude that

$$I_{\max}^{\sqrt{\varepsilon}}(M;E)_\tau \ge I(M;E)_\tau - \sqrt{\varepsilon}\log_2|\mathcal{M}| - 2g_2(\sqrt{\varepsilon}). \tag{16.1.86}$$

Combining (16.1.85) and (16.1.86), we conclude that

$$\begin{aligned}
&\left(1-\varepsilon-\sqrt{\varepsilon}\right)\log_2|\mathcal{M}| \\
&\le I(M;B)_\tau - I(M;E)_\tau + h_2(\varepsilon) + 2g(\sqrt{\varepsilon}) &(16.1.87)\\
&\le \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I(X;B)_\rho - I(X;E)_\rho\right) + h_2(\varepsilon) + 2g(\sqrt{\varepsilon}), &(16.1.88)
\end{aligned}$$

where the last inequality follows by optimizing over all input ensembles. ∎

16.1.3.2 Relative Entropy of Entanglement Upper Bound

We now consider an upper bound based on the channel's relative entropy of entanglement. In order to do so, we exploit the equivalence between secret-key transmission and bipartite private-state transmission established in Section 16.1.2. We also make use of Proposition 15.15, which gives an upper bound on the number $\log_2|\mathcal{M}|$ of private bits contained in the final state of a bipartite private-state transmission protocol.

In the previous chapter on secret key distillation, our approach to obtaining upper bounds on distillable key consisted of 1) establishing a connection between a tripartite key distillation protocol and a bipartite private state distillation protocol (see Section 15.1.2) and 2) comparing the state at the output of a bipartite private-state distillation protocol with one that is useless for this task. We considered the set of separable states as the useless set, and we proved that certain state entanglement measures are upper bounds on distillable key in the one-shot and asymptotic settings.

In Section 16.1.2, we established a similar correspondence between secret-key transmission, as defined in Section 16.1, and bipartite private-state transmission. Here, we observe that bipartite private-state transmission is similar to bipartite private-state distillation in the sense that, like private-state distillation, the error criterion for private-state transmission involves comparing the output state to an ideal private state (see Definition 16.5). This suggests that the state entanglement measures defined in Section 9.2 are relevant (in particular the result of Proposition 15.15). However, as was the case for quantum communication in Chapter 14, the main resource that we are considering in this chapter is a quantum channel and not a quantum state, and so we have an extra degree of freedom in the input state to the channel, which we can optimize. This suggests that the channel entanglement measures from Chapter 10 are relevant, and it is indeed what we find.

Theorem 16.9 Relative Entropy of Entanglement Upper Bound on One-Shot Private Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(|\mathcal{M}|, \varepsilon)$ private communication protocols for $\mathcal{N}$, the following bound holds

$$\log_2|\mathcal{M}| \leq E_R^\varepsilon(\mathcal{N}), \tag{16.1.89}$$

where $E_R^\varepsilon(\mathcal{N})$ is the ε-relative entropy of entanglement of $\mathcal{N}$, defined in (10.3.8) as

$$E_R^\varepsilon(\mathcal{N}) \coloneqq \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\mathrm{SEP}(S:B)} D_H^\varepsilon(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}). \tag{16.1.90}$$

Consequently, we have the following bound on the one-shot private capacity:

$$P^\varepsilon(\mathcal{N}) \leq E_R^\varepsilon(\mathcal{N}). \tag{16.1.91}$$

Proof: By definition, it follows that the one-shot secret-key transmission capacity is an upper bound on $P^\varepsilon(\mathcal{N})$. Applying Theorem 16.6, we conclude that the one-shot bipartite private-state transmission capacity is equal to the one-shot secret-key transmission capacity. So let us bound the one-shot bipartite private-state transmission capacity. Consider an arbitrary (M, ε) bipartite private-state transmission protocol. The final state of such a protocol satisfies the condition in Definition 16.5, which means that there is an ideal private state $\gamma_{M''MA'\hat{M}B'}$ such that

$$1 - F(\gamma_{M''MA'\hat{M}B'}, \omega_{M''MA'\hat{M}B'}) \leq \varepsilon. \tag{16.1.92}$$

As such, Proposition 15.15 applies, and we conclude that

$$\log_2|\mathcal{M}| \leq E_R^\varepsilon(M''MA';\hat{M}B')_\omega \tag{16.1.93}$$

$$\leq E_R^\varepsilon(M''MA';B)_\rho \tag{16.1.94}$$

$$\leq E_R^{\varepsilon}(\mathcal{N}). \tag{16.1.95}$$

The second inequality follows from the data-processing inequality for D_H^{ε} under the action of the isometric decoding channel $\mathcal{U}^{\mathcal{D}}_{B\to\hat{M}B'}$ and where the state $\rho_{M''MA'B}$ is defined as

$$\rho_{M''MA'B} := (\mathcal{N}_{A\to B} \circ \mathcal{U}^{\mathcal{E}}_{M'\to AA'})(\Phi_{M''MM'}). \tag{16.1.96}$$

The systems $M''MA'$ extend the system A of the state $\mathcal{U}^{\mathcal{E}}_{M'\to AA'}(\Phi_{M''MM'})$, with A being the input to the channel $\mathcal{N}_{A\to B}$. As such, we can optimize over all such input states, and then conclude the final inequality above (here, we need to apply the remark after Definition 10.1 as well). ■

We then have the following bound as a direct application of Proposition 7.71:

Corollary 16.10

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(|\mathcal{M}|, \varepsilon)$ private communication protocols for $\mathcal{N}$, the following bound holds for all $\alpha > 1$:

$$\log_2|\mathcal{M}| \leq \widetilde{E}_{\alpha}(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{16.1.97}$$

where $\widetilde{E}_{\alpha}(\mathcal{N})$ is the sandwiched Renyi relative entropy of entanglement of $\mathcal{N}$, defined in (10.3.10) as

$$\widetilde{E}_{\alpha}(\mathcal{N}) := \sup_{\psi_{SA}} \inf_{\sigma_{SB}\in\mathrm{SEP}(S:B)} \widetilde{D}_{\alpha}(\mathcal{N}_{A\to B}(\psi_{SA})\|\sigma_{SB}). \tag{16.1.98}$$

16.1.3.3 Squashed Entanglement Upper Bound

We now turn to squashed entanglement and establish it as an upper bound on one-shot private capacity. The reasoning behind this result is very similar to that given in the proof of Proposition 16.9, except that we employ Proposition 15.19 instead:

Theorem 16.11 Squashed Entanglement Upper Bound on One-Shot Private Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(|\mathcal{M}|, \varepsilon)$ private communication protocols for $\mathcal{N}$, the following bound holds

$$\left(1 - 2\sqrt{\varepsilon}\right) \log_2|\mathcal{M}| \leq E_{\text{sq}}(\mathcal{N}) + 2g_2(\sqrt{\varepsilon}), \tag{16.1.99}$$

where $E_{\text{sq}}(\mathcal{N})$ is the squashed entanglement of the channel $\mathcal{N}$, given in Definition 10.14 as

$$E_{\text{sq}}(\mathcal{N}) := \sup_{\psi_{SA}} E_{\text{sq}}(S;B)_\tau, \tag{16.1.100}$$

and $\tau_{SB} := \mathcal{N}_{A\to B}(\psi_{SA})$. Consequently, we have the following bound on the one-shot private capacity:

$$\left(1 - 2\sqrt{\varepsilon}\right) P^\varepsilon(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N}) + 2g_2(\sqrt{\varepsilon}). \tag{16.1.101}$$

PROOF: As indicated above, the argument is precisely the same as in the proof of Proposition 16.9, except that we apply the following bound from Proposition 15.19 instead:

$$\left(1 - 2\sqrt{\varepsilon}\right) \log_2|\mathcal{M}| \leq E_{\text{sq}}(M''MA'; \hat{M}B')_\omega + 2g_2(\sqrt{\varepsilon}). \tag{16.1.102}$$

After this step, we apply the data-processing inequality for E_{sq} and optimize over channel input states. ■

16.1.4 Lower Bound on the Number of Transmitted Private Bits via Position-Based Coding and Convex Splitting

Having derived upper bounds on the number of private bits that can be transmitted in an arbitrary private communication protocol, let us now determine a lower bound. Here we use the methods of position-based coding and convex splitting to derive an explicit $(|\mathcal{M}|, \varepsilon)$ protocol for all $\varepsilon \in (0,1)$.

To derive this lower bound, let us consider a slightly different model of communication, in which there is a one-input, two-output classical–quantum channel connecting the sender Alice to the legitimate receiver Bob and the eavesdropper

Eve:

$$x \to \rho_{BE}^{x}, \tag{16.1.103}$$

where $x \in \mathcal{X}$ is the classical input symbol and ρ_{BE}^{x} is the bipartite quantum state that appears at the output when x is input. Bob has access to the system B of the output and Eve to E. The channel can alternatively be written as a quantum channel as follows:

$$\mathcal{N}_{X\to BE}(\omega) \coloneqq \sum_{x\in\mathcal{X}} \langle x|_X \omega |x\rangle_X \rho_{BE}^{x}, \tag{16.1.104}$$

where $\{|x\rangle_X\}_{x\in\mathcal{X}}$ is an orthonormal basis. In this way, a private communication protocol for $\mathcal{N}_{X\to BE}$ is defined exactly as we did in Section 16.1, with $\mathcal{N}_{X\to BE}$ replacing the isometric channel $\mathcal{U}_{A\to BE}$ therein. Furthermore, the notions of code infidelity, an $(|\mathcal{M}|, \varepsilon)$ private communication protocol, and one-shot private capacity are defined in the same way, but with $\mathcal{N}_{X\to BE}$ replacing the isometric channel $\mathcal{U}_{A\to BE}$.

The main result of this section is the following lower bound on the one-shot private capacity $P^{\varepsilon}(\mathcal{N})$ of a classical–quantum wiretap channel $\mathcal{N}_{X\to BE}$:

$$P^{\varepsilon}(\mathcal{N}) \geq \overline{I}_{H}^{\varepsilon'-\delta-\eta}(X;B)_{\rho} - \overline{I}_{\max}^{\delta-\zeta}(E;X)_{\rho} \\ - \log_2\!\left(\frac{8\,(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\!\left(\frac{2}{\zeta^2}\right), \tag{16.1.105}$$

where $\varepsilon' = 1 - \sqrt{1-\varepsilon/2}$, $\delta \in (0,\varepsilon')$, $\eta \in (0,\varepsilon'-\delta)$, and $\zeta \in (0,\delta)$, and the information measures are evaluated with respect to the state

$$\rho_{XBE} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_{BE}^{x}. \tag{16.1.106}$$

In the above, $P^{\varepsilon}(\mathcal{N})$ represents the maximum number of bits that can be sent from Alice to Bob, using a classical–quantum wiretap channel once, such that the infidelity does not exceed $\varepsilon \in (0,1)$. The quantities on the right-hand side of the inequality in (16.1.105) are particular one-shot generalizations of the Holevo information to Bob and Eve, which are defined in (7.11.88) and (15.1.211), respectively.

To prove the one-shot bound in (16.1.105), we employ position-based coding (Section 11.1.3) and convex splitting (Section 15.1.4). The main idea of position-based coding is conceptually simple and we review it briefly here. To communicate a classical message from Alice to Bob, we allow them to share a quantum state $\rho_{RA}^{\otimes M}$ before communication begins, where M is the number of messages, Bob possesses the R systems, and Alice the A systems. If Alice wishes to communicate message m, then she sends the mth A system through the channel. The reduced state of Bob's systems is then

$$\rho_{R_1} \otimes \cdots \otimes \rho_{R_{m-1}} \otimes \rho_{R_m B} \otimes \rho_{R_{m+1}} \otimes \cdots \otimes \rho_{R_M}, \tag{16.1.107}$$

where $\rho_{R_m B} = \mathcal{N}_{A_m \to B}(\rho_{R_m A_m})$ and $\mathcal{N}_{A_m \to B}$ is the quantum channel. For all $m' \neq m$, the reduced state for systems $R_{m'}$ and B is the product state $\rho_{R_{m'}} \otimes \rho_B$. However, the reduced state of systems $R_m B$ is the (generally) correlated state $\rho_{R_m B}$. So if Bob has a binary measurement that can distinguish the joint state ρ_{RB} from the product state $\rho_R \otimes \rho_B$ sufficiently well, he can base a decoding strategy off of this, and the scheme is reliable as long as the number of bits $\log_2 M$ to be communicated is chosen to be roughly equal to the hypothesis testing mutual information. This is exactly what is used in position-based coding, thus forging a transparent and intuitive link between quantum hypothesis testing and communication for the case of entanglement-assisted communication.

Convex splitting is rather intuitive as well and can be thought of as dual to the coding scenario mentioned above. Suppose instead that Alice and Bob have a means of generating the state in (16.1.107), perhaps by the strategy mentioned above. But now suppose that Alice chooses the variable m uniformly at random, so that the state, from the perspective of someone ignorant of the choice of m, is the following mixture:

$$\frac{1}{M} \sum_{m=1}^{M} \rho_{R_1} \otimes \cdots \otimes \rho_{R_{m-1}} \otimes \rho_{R_m B} \otimes \rho_{R_{m+1}} \otimes \cdots \otimes \rho_{R_M}. \tag{16.1.108}$$

The convex-split lemma (Lemma 15.22) guarantees that as long as $\log_2 M$ is roughly equal to the smooth max-mutual information in (15.1.211), then the state above is nearly indistinguishable from the product state $\rho_R^{\otimes M} \otimes \rho_B$.

Here we use the approaches of position-based coding and convex splitting in conjunction to construct codes for the classical–quantum wiretap channel. The main underlying idea is to have a message variable $m \in \{1, \ldots, M\}$ and a local randomness variable $r \in \{1, \ldots, R\}$, the latter of which is selected uniformly at random and used to confuse the eavesdropper Eve. Before communication begins, Alice, Bob, and Eve are allowed share to MR copies of the common randomness state

$$\rho_{XX'X''} \coloneqq \sum_{x \in \mathcal{X}} p_X(x) |xxx\rangle\langle xxx|_{XX'X''}. \tag{16.1.109}$$

We can think of the MR copies of $\rho_{XX'X''}$ as being partitioned into M blocks, each of which contain R copies of the state $\rho_{XX'X''}$. If Alice wishes to send message m, then she picks r uniformly at random and sends the X_A system labeled by (m, r) through the classical–quantum wiretap channel in (16.1.103). As long as $\log_2 MR$ is roughly equal to the hypothesis testing mutual information $\overline{I}_H^{\varepsilon}(X; B)$, then Bob can use a position-based decoder to figure out both m and r. As long as $\log_2 R$ is roughly equal to the smooth max-mutual information $\overline{I}_{\max}^{\varepsilon}(E; X)$, then the convex-split lemma guarantees that the overall state of Eve's systems, regardless of which message m was chosen, is nearly indistinguishable from the product state $\rho_{X_E}^{\otimes MR} \otimes \rho_E$. Thus, in such a scheme, Bob can figure out m while Eve cannot figure out anything about m. This is the intuition behind the coding scheme and gives a sense of why $\log_2 M = \log_2 MR - \log_2 R \approx \overline{I}_H^{\varepsilon}(X; B) - \overline{I}_{\max}^{\varepsilon}(E; X)$ is an achievable number of bits that can be sent privately from Alice to Bob. The main purpose of this section is to develop the details of this argument and furthermore to show how the scheme can be derandomized, so that the MR copies of the common randomness state $\rho_{XX'X''}$ are in fact not necessary.

There are strong connections between the approach for establishing a lower bound on one-shot distillable key detailed in Section 15.1.4, and the approach we have outlined above and detail below. In fact, there is a point in the analysis below at which it becomes precisely the same, and at that point, we simply invoke the proof of Theorem 15.21 to complete the analysis.

We now state the main theorem of this section:

Theorem 16.12

Let $\mathcal{N}_{X\to BE} : x \to \rho_{BE}^{x}$ be a classical–quantum wiretap channel, in which Alice has access to the input, Bob to the output system B, and Eve to the output system E. For all $\varepsilon \in (0, 1]$, $\varepsilon' = 1 - \sqrt{1 - \varepsilon/2}$, $\delta \in (0, \varepsilon')$, $\eta \in (0, \varepsilon' - \delta)$, and $\zeta \in (0, \delta)$, there exists an $(|\mathcal{M}|, \varepsilon)$ private communication protocol for $\mathcal{N}_{X\to BE}$, such that

$$\log_2|\mathcal{M}| = \overline{I}_H^{\varepsilon'-\delta-\eta}(X; B)_\rho - \overline{I}_{\max}^{\delta-\zeta}(E; X)_\rho - \log_2\!\left(\frac{8(\varepsilon' - \delta)}{\eta^2}\right) - \log_2\!\left(\frac{2}{\zeta^2}\right).$$

where the hypothesis testing mutual information $\overline{I}_H^{\varepsilon'-\delta-\eta}(X; B)_\rho$ is defined in

(7.11.88) and the smooth max-mutual information $\overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho$ in (15.1.211), and they are evaluated with respect to the state ρ_{XBE} in (16.1.106).

Proof: We first exhibit a public shared randomness assisted protocol for private communication and then show later how to derandomize it. The protocol proceeds exactly as discussed above. We suppose that Alice, Bob, and Eve share the state $\rho_{XX'X''}^{\otimes MR}$ before communication begins, where $M = |\mathcal{M}|$. If Alice wants to send the message m, she picks r uniformly at random from $\{1, \ldots, R\}$ and transmits a classical copy X''' of the X system labeled by (m, r) through the channel $\mathcal{N}_{X\to BE}$. The resulting state of Alice, Bob, and Eve, for fixed m and r, is then as follows:

$$\rho^{m,r}_{X^{MR}X'^{MR}X''^{MR}BE} \coloneqq \rho_{X_{1,1}X'_{1,1}X''_{1,1}} \otimes \cdots \otimes \rho_{X_{m,r-1}X'_{m,r-1}X''_{m,r-1}} \otimes \\ \rho_{X_{m,r}X'_{m,r}X''_{m,r}BE} \otimes \rho_{X_{m,r+1}X'_{m,r+1}X''_{m,r+1}} \otimes \cdots \otimes \rho_{X_{M,R}X'_{M,R}X''_{M,R}}, \quad (16.1.110)$$

where

$$\rho_{X_{1,1}X'_{1,1}X''_{1,1}} = \cdots = \rho_{X_{m,r-1}X'_{m,r-1}X''_{m,r-1}} \quad (16.1.111)$$
$$= \rho_{X_{m,r+1}X'_{m,r+1}X''_{m,r+1}} = \cdots = \rho_{X_{M,R}X'_{M,R}X''_{M,R}} \quad (16.1.112)$$
$$= \sum_{x\in\mathcal{X}} p_X(x)|xxx\rangle\langle xxx|_{XX'X''}, \quad (16.1.113)$$

and

$$\rho_{X_{m,r}X'_{m,r}X''_{m,r}BE} = \sum_{x\in\mathcal{X}} p_X(x)|xxx\rangle\langle xxx|_{XX'X''} \otimes \mathcal{N}_{X'''\to BE}(|x\rangle\langle x|_{X'''}) \quad (16.1.114)$$
$$= \sum_{x\in\mathcal{X}} p_X(x)|xxx\rangle\langle xxx|_{XX'X''} \otimes \rho^x_{BE}. \quad (16.1.115)$$

At this point, the state here is precisely the same as that given in (15.1.200), and the goal from here is the same as well. Thus, we can apply the same reasoning given there to conclude that the following infidelity condition holds

$$1 - F(\mathcal{M}_{X'^{MR}B\to M_B}(\rho_{M_A X'^{MR}X''^{MR}BE}), \overline{\Phi}_{M_A M_B} \otimes \rho_{X''^{MR}} \otimes \widetilde{\rho}_E) \leq \varepsilon \quad (16.1.116)$$

if

$$\log_2|\mathcal{M}| = \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\tau - \overline{I}_{\max}^{\delta-\zeta}(E;X)_\tau$$

$$- \log_2\left(\frac{4\,(\varepsilon' - \delta)}{\eta^2}\right) - \log_2\left(\frac{2}{\zeta^2}\right), \quad (16.1.117)$$

where

$$\tau_{XBE} := \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|_X \otimes \rho^x_{BE} \quad (16.1.118)$$

and $\rho_{M_A X'^{MR} X''^{MR} BE}$ is the reduction of the following state:

$$\rho_{M_A R_A X^{MR} X'^{MR} X''^{MR} BE} := \frac{1}{MR} \sum_{m=1}^{M} \sum_{r=1}^{R} |m\rangle\langle m|_{M_A} \otimes |r\rangle\langle r|_{R_A} \otimes \rho^{m,r}_{X^{MR} X'^{MR} X''^{MR} BE}. \quad (16.1.119)$$

That is,

$$\rho_{M_A X'^{MR} X''^{MR} BE} = \operatorname{Tr}_{R_A X^{MR}}[\rho_{M_A R_A X'^{MR} X''^{MR} BE}]. \quad (16.1.120)$$

Furthermore, $\mathcal{M}_{X'^{MR} B \to M_B}$ is a measurement channel of the form in (15.1.222), and $\widetilde{\rho}_E$ is a state satisfying

$$P(\rho_E, \widetilde{\rho}_E) \leq \delta - \zeta. \quad (16.1.121)$$

Thus, Bob's strategy is to decode both m and r (as before), and he can do so as long as $\log_2 MR = \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\tau - \log_2\left(\frac{4(\varepsilon'-\delta)}{\eta^2}\right)$. At the same time, the message m should be private from Eve, and this is possible as long as $\log_2 R = \overline{I}_{\max}^{\delta-\zeta}(E;X)_\tau + \log_2\left(\frac{2}{\zeta^2}\right)$. Calculating $\log_2 M = \log_2 MR - \log_2 R$ gives (16.1.117). Then the analysis in the proof of Theorem 15.21 guarantees that the condition in (16.1.116) holds.

We now discuss how to derandomize the protocol. First, let us define the following measurement channels

$$\mathcal{M}^{x_{1,1},\ldots,x_{M,R}}_{B \to M_B}(\omega_B) := \mathcal{M}_{X'^{MR} B \to M_B}(|x_{1,1},\ldots,x_{M,R}\rangle\langle x_{1,1},\ldots,x_{M,R}|_{X'^{MR}} \otimes \omega_B), \quad (16.1.122)$$

where ω_B is an input state. Also, consider that the state $\rho_{M_A X'^{MR} X''^{MR} BE}$ can be written as

$$\rho_{M_A X'^{MR} X''^{MR} BE} = \frac{1}{M} \sum_{m=1}^{M} |m\rangle\langle m|_{M_A} \otimes$$

$$\sum_{x_{1,1},\ldots,x_{M,R}} p(x_{1,1})\cdots p(x_{M,R})|x_{1,1},\ldots,x_{M,R}\rangle\langle x_{1,1},\ldots,x_{M,R}|_{X'^{MR}}\otimes$$

$$|x_{1,1},\ldots,x_{M,R}\rangle\langle x_{1,1},\ldots,x_{M,R}|_{X''^{MR}}\otimes\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}. \quad (16.1.123)$$

With this in mind, the state $\mathcal{M}_{X'^{MR}B\to M_B}(\rho_{M_AX'^{MR}X''^{MR}BE})$ can be written as follows:

$$\mathcal{M}_{X'^{MR}B\to M_B}(\rho_{M_AX'^{MR}X''^{MR}BE}) =$$
$$\frac{1}{M}\sum_{m=1}^{M}\sum_{x_{1,1},\ldots,x_{M,R}} p(x_{1,1})\cdots p(x_{M,R})|m\rangle\langle m|_{M_A}\otimes$$
$$|x_{1,1},\ldots,x_{M,R}\rangle\langle x_{1,1},\ldots,x_{M,R}|_{X''^{MR}}\otimes\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right), \quad (16.1.124)$$

and the condition in (16.1.116) as

$$1-\varepsilon$$
$$\leq F(\mathcal{M}_{X'^{MR}B\to M_B}(\rho_{M_AX'^{MR}X''^{MR}BE}),\overline{\Phi}_{M_AM_B}\otimes\rho_{X''^{MR}}\otimes\widetilde{\rho}_E) \quad (16.1.125)$$
$$=\left[\begin{array}{c}\frac{1}{M}\sum_{m=1}^{M}\sum_{x_{1,1},\ldots,x_{M,R}} p(x_{1,1})\cdots p(x_{M,R})\times \\ \sqrt{F}\left(\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\end{array}\right]^2, \quad (16.1.126)$$

which is the same as

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{x_{1,1},\ldots,x_{M,R}} p(x_{1,1})\cdots p(x_{M,R})\times$$
$$\sqrt{F}\left(\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)$$
$$\geq\sqrt{1-\varepsilon}. \quad (16.1.127)$$

We can now exploit the "Shannon trick" of exchanging the sum over the messages m and the sum over the codewords to rewrite this inequality as

$$\sum_{x_{1,1},\ldots,x_{M,R}} p(x_{1,1})\cdots p(x_{M,R})\times$$

$$\left(\frac{1}{M}\sum_{m=1}^{M}\sqrt{F}\left(\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\right)$$
$$\geq\sqrt{1-\varepsilon}. \quad (16.1.128)$$

Since the average does not exceed the maximum, we conclude that there exists some choice of codewords $x_{1,1},\ldots,x_{M,R}$ such that the following inequality holds

$$\frac{1}{M}\sum_{m=1}^{M}\sqrt{F}\left(\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\geq\sqrt{1-\varepsilon}. \quad (16.1.129)$$

Let us then use the shorthand $\mathcal{M}_{B\to M_B}\equiv\mathcal{M}_{B\to M_B}^{x_{1,1},\ldots,x_{M,R}}$, so that we can rewrite the above as

$$\frac{1}{M}\sum_{m=1}^{M}\sqrt{F}\left(\mathcal{M}_{B\to M_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\geq\sqrt{1-\varepsilon}. \quad (16.1.130)$$

This completes the *derandomization* part of the proof.

Finally, we are interested in a code that satisfies the maximal infidelity criterion $p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N})\leq\varepsilon$. To find such a code, we can apply *expurgation* to the code found above. Since the square root function is concave, after bringing the average inside the square root and squaring both sides of the inequality, we conclude that

$$\frac{1}{M}\sum_{m=1}^{M}F\left(\mathcal{M}_{B\to M_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\geq 1-\varepsilon, \quad (16.1.131)$$

which we can rewrite one more time as

$$\frac{1}{M}\sum_{m=1}^{M}\left(1-F\left(\mathcal{M}_{B\to M_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\right)\leq\varepsilon. \quad (16.1.132)$$

Now applying Markov's inequality, we conclude that at least half of the messages are such that the following inequality holds

$$1-F\left(\mathcal{M}_{B\to M_B}\left(\frac{1}{R}\sum_{r=1}^{R}\rho_{BE}^{x_{m,r}}\right),|m\rangle\langle m|_{M_B}\otimes\widetilde{\rho}_E\right)\leq 2\varepsilon. \quad (16.1.133)$$

Thus, these messages and the corresponding codewords are retained as the final code. To be clear, suppose without loss of generality, that messages $1,\ldots,\lfloor M/2\rfloor$ are retained and messages $\lfloor M/2 \rfloor + 1, \ldots, M$ are expurgated. Then this means that the corresponding codewords retained are $x_{1,1}, \ldots, x_{1,R}, x_{2,1}, \ldots, x_{2,R}, \ldots, x_{\lfloor M/2 \rfloor,1}, \ldots, x_{\lfloor M/2 \rfloor,R}$, and the ones discarded are $x_{\lfloor M/2 \rfloor+1,1}, \ldots, x_{\lfloor M/2 \rfloor+1,R}, x_{\lfloor M/2 \rfloor+2,1}, \ldots, x_{\lfloor M/2 \rfloor+2,R}, \ldots, x_{M,1}, \ldots, x_{M,R}$. After the expurgation, the rate of the code is given by

$$\log_2 |\mathcal{M}| / 2 = \log_2 |\mathcal{M}| - \log_2(2) \tag{16.1.134}$$

$$= \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho - \overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho - \log_2\left(\frac{4(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\left(\frac{2}{\zeta^2}\right) - \log_2(2) \tag{16.1.135}$$

$$= \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho - \overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho - \log_2\left(\frac{8(\varepsilon'-\delta)}{\eta^2}\right) - \log_2\left(\frac{2}{\zeta^2}\right). \tag{16.1.136}$$

By a final substitution of 2ε with ε and rewriting, we arrive at the claim of the theorem. ■

We can induce a classical–quantum wiretap channel from an isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ extending $\mathcal{N}_{A\to B}$ by the following pre-processing:

$$x \to \rho_A^x \to \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho_A^x). \tag{16.1.137}$$

That is, based on the value of a letter x, Alice inputs the state ρ_A^x into the isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$. Optimizing over all such preprocessings and applying Theorem 16.12, we arrive at the following lower bound on the one-shot private capacity of a quantum channel $\mathcal{N}_{A\to B}$ (according to the definition given in Section 16.1):

Corollary 16.13

Let $\mathcal{N}_{A\to B}$ be a quantum channel that is extended by the isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$. For all $\varepsilon \in (0,1]$, $\varepsilon' = 1 - \sqrt{1-\varepsilon/2}$, $\delta \in (0,\varepsilon')$, $\eta \in (0, \varepsilon'-\delta)$, and $\zeta \in (0,\delta)$, there exists an $(|\mathcal{M}|, \varepsilon)$ private communication protocol for $\mathcal{N}_{A\to B}$, such that

$$\log_2 |\mathcal{M}| = \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho - \overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho$$

$$- \log_2\left(\frac{8(\varepsilon' - \delta)}{\eta^2}\right) - \log_2\left(\frac{2}{\zeta^2}\right).$$

where the hypothesis testing mutual information $\overline{I}_H^{\varepsilon'-\delta-\eta}(X;B)_\rho$ is defined in (7.11.88) and the smooth max-mutual information $\overline{I}_{\max}^{\delta-\zeta}(E;X)_\rho$ in (15.1.211), and the information quantities are evaluated with respect to the following state:

$$\rho_{XBE} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^x_A). \tag{16.1.138}$$

Now applying Propositions 7.72 and 7.64, we conclude the following bound:

Corollary 16.14

Let $\mathcal{N}_{A\to B}$ be a quantum channel that is extended by the isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$. For all $\varepsilon \in (0,1]$, $\varepsilon' = 1 - \sqrt{1-\varepsilon/2}$, $\delta \in (0,\varepsilon')$, $\eta \in (0, \varepsilon' - \delta)$, $\zeta \in (0,\delta)$, $\nu \in (0, \delta - \zeta)$, $\alpha \in (0,1)$, and $\beta > 1$, there exists an $(|\mathcal{M}|, \varepsilon)$ private communication protocol for $\mathcal{N}_{A\to B}$, such that

$$\log_2|\mathcal{M}| \geq \sup_{\{p(x),\rho^x_A\}_{x\in\mathcal{X}}} \left(\overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho\right) - f(\varepsilon',\delta,\eta,\nu,\zeta,\alpha,\beta). \tag{16.1.139}$$

where the Petz–Renyi mutual information $\overline{I}_\alpha(X;B)_\rho$ is defined in (11.1.136), the sandwiched Renyi mutual information $\widetilde{I}'_\beta(X;E)_\rho$ as

$$\widetilde{I}'_\beta(X;E)_\rho := \widetilde{D}_\beta(\rho_{XE}\|\rho_X \otimes \rho_E), \tag{16.1.140}$$

and the information quantities are evaluated with respect to the following state:

$$\rho_{XBE} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^x_A). \tag{16.1.141}$$

Furthermore,

$$f(\varepsilon',\delta,\eta,\nu,\zeta,\alpha,\beta) := \frac{\alpha}{1-\alpha}\log_2\left(\frac{1}{\varepsilon'-\delta-\eta}\right) + \log_2\left(\frac{8}{\nu^2}\right) + \frac{1}{\beta-1}\log_2\left(\frac{1}{(\delta-\zeta-\nu)^2}\right) + \log_2\left(\frac{1}{1-(\delta-\zeta-\nu)^2}\right)$$

$$+\log_2\left(\frac{8(\varepsilon'-\delta)}{\eta^2}\right)+\log_2\left(\frac{2}{\zeta^2}\right). \quad (16.1.142)$$

PROOF: The reasoning here is precisely the same as that given in the proof of Corollary 15.24. The only difference is that we optimize over every ensemble $\{p(x), \rho_A^x\}_{x\in\mathcal{X}}$. ■

16.2 Private Capacity of a Quantum Channel

We now consider the asymptotic setting. In this scenario, depicted in Figure [REF], the classical message system M' to be transmitted to Bob is encoded into n copies $A_1, \ldots, A_n$ of a quantum system A, for $n \in \mathbb{N}$. Each of the systems is then sent through an independent use of the isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$, which extends the point-to-point channel $\mathcal{N}_{A\to B}$. As before, this is the asymptotic setting because n can be arbitrarily large.

Due to the fact that n independent uses of the channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ is no different from one use of the tensor-power channel $(\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n}$, the information theory underlying the asymptotic setting is no different from that in the one-shot setting, and the main task we accomplish here is to analyze performance of the protocols in the large n limit. Indeed, the only change that we make here is to replace $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ with $(\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n}$ and define the encoding and decoding channels as acting on n systems instead of one. If Alice transmits message m, then the final state of the protocol is

$$(\mathcal{D}_{B^n\to\hat{M}} \circ (\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n} \circ \mathcal{E}_{M'\to A^n})(|m\rangle\langle m|_{M'}), \quad (16.2.1)$$

where $\mathcal{E}_{M'\to A^n}$ is the encoding channel and $\mathcal{D}_{B^n\to\hat{M}}$ the decoding channel. Just as in the one-shot setting, we define the maximal infidelity of the code as

$$p^*_{\text{err}}(\mathcal{E},\mathcal{D};\mathcal{N}^{\otimes n}) = \\ \inf_{\sigma_{E^n}} \max_{m\in\mathcal{M}} (1 - F(|m\rangle\langle m|_{\hat{M}} \otimes \sigma_{E^n}, (\mathcal{D}_{B^n\to\hat{M}} \circ (\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n} \circ \mathcal{E}_{M'\to A^n})(|m\rangle\langle m|_{M'}))), \quad (16.2.2)$$

where the infimum is with respect to every state σ_{E^n} of the eavesdropper's system E.

Definition 16.15 $(n, |\mathcal{M}|, \varepsilon)$ Private Communication Protocol

Let $(\mathcal{M}, \mathcal{E}_{M' \to A^n}, \mathcal{D}_{B^n \to \hat{M}})$ be the elements of a private communication protocol for n independent uses of the channel $\mathcal{N}_{A\to B}$, where $d_{M'} = d_{\hat{M}} = |\mathcal{M}|$. The protocol is called an $(n, |\mathcal{M}|, \varepsilon)$ protocol, with $\varepsilon \in [0, 1]$, if $p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) \leq \varepsilon$.

The rate of an $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol is defined as the number of private bits transmitted per channel use, i.e.,

$$R(n, |\mathcal{M}|) := \frac{\log_2|\mathcal{M}|}{n}. \tag{16.2.3}$$

The rate depends only on the size $|\mathcal{M}|$ of the message set $\mathcal{M}$ and on the number of channel uses. In particular, it does not depend on the communication channel nor on the encoding and decoding channels. For a given $\varepsilon \in [0, 1]$ and $n \in \mathbb{N}$, the highest rate among all $(n, |\mathcal{M}|, \varepsilon)$ protocols is denoted by $P^{n,\varepsilon}(\mathcal{N})$, and it is defined as

$$P^{n,\varepsilon}(\mathcal{N}) := \frac{1}{n} P^{\varepsilon}(\mathcal{N}^{\otimes n}) = \sup_{(\mathcal{M},\mathcal{E},\mathcal{D})} \left\{ \frac{\log_2|\mathcal{M}|}{n} : p^*_{\text{err}}(\mathcal{E}, \mathcal{D}; \mathcal{N}^{\otimes n}) \leq \varepsilon \right\}, \tag{16.2.4}$$

where, in the second equality, we use the definition of the one-shot private capacity P^{ε} given in (16.1.35), and the supremum is over every message set $\mathcal{M}$, encoding channel $\mathcal{E}$ with input dimension $|\mathcal{M}|$, and decoding channel $\mathcal{D}$ with output dimension $|\mathcal{M}|$.

We now provide several definitions related to private capacity and its associated concepts.

Definition 16.16 Achievable Rate for Private Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an achievable rate for private communication over $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ private communication protocol for $\mathcal{N}$.

Definition 16.17 Private Capacity of a Quantum Channel

The private capacity of a quantum channel $\mathcal{N}$, denoted by $P(\mathcal{N})$, is defined to be the supremum of all achievable rates, i.e.,

$$P(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{16.2.5}$$

An equivalent definition of private capacity is

$$P(\mathcal{N}) = \inf_{\varepsilon\in(0,1]} \liminf_{n\to\infty} \frac{1}{n} P^{\varepsilon}(\mathcal{N}^{\otimes n}), \tag{16.2.6}$$

which we prove in Appendix A.

Definition 16.18 Weak Converse Rate for Private Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a weak converse rate for private communication over $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

Definition 16.19 Strong Converse Rate for Private Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a strong converse rate for private communication over $\mathcal{N}$ if for all $\varepsilon \in [0,1)$, $\delta > 0$, and sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ private communication protocol for $\mathcal{N}$.

Definition 16.20 Strong Converse Private Capacity of a Quantum Channel

The strong converse private capacity of a quantum channel $\mathcal{N}$, denoted by $\widetilde{P}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{P}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{16.2.7}$$

We can also write the strong converse private capacity as

$$\widetilde{P}(\mathcal{N}) = \sup_{\varepsilon\in[0,1)} \limsup_{n\to\infty} \frac{1}{n} P^{\varepsilon}(\mathcal{N}^{\otimes n}). \tag{16.2.8}$$

See Appendix A for a proof. We also show in Appendix A that

$$P(\mathcal{N}) \leq \widetilde{P}(\mathcal{N}) \tag{16.2.9}$$

for every quantum channel $\mathcal{N}$.

We now state one of the main theorems of this chapter, which gives an experssion for the private capacity of a quantum channel.

Theorem 16.21 Private Capacity

The private capacity of a quantum channel $\mathcal{N}_{A\to B}$ is equal to the regularized private information $P_{\text{reg}}(\mathcal{N})$ of $\mathcal{N}$, i.e.,

$$P(\mathcal{N}) = I^p_{\text{reg}}(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} I^p(\mathcal{N}^{\otimes n}), \tag{16.2.10}$$

where the private information of a channel is defined as

$$I^p(\mathcal{N}) := \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} I(X;B)_\rho - I(X;E)_\rho, \tag{16.2.11}$$

and the information quantities are evaluated with respect to the state

$$\rho_{XBE} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho_A^x), \tag{16.2.12}$$

with $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ an isometric channel extending $\mathcal{N}_{A\to B}$.

Observe that the expression in (16.2.10) for the private capacity involves a regularization of the private information. Thus, in general, it is difficult to compute because the optimization is over an arbitrarily large number of channel uses.

By following an argument similar to that given in Section 14.2.3, it follows that the private information is always superadditive, meaning that $I^p(\mathcal{N}^{\otimes n}) \geq nI^p(\mathcal{N})$ for every $n \in \mathbb{N}$ and channel $\mathcal{N}$. This means that the private information is always a lower bound on the private capacity of a channel $\mathcal{N}$:

$$P(\mathcal{N}) \geq I^p(\mathcal{N}) \text{ for every channel } \mathcal{N}. \tag{16.2.13}$$

If the private information happens to be additive for a particular channel, then the regularization in (16.2.10) is not required. For example, the private information is known to be additive for all degradable and anti-degradable channels (see Definition 4.6). Furthermore, for degradable channels, the private information is equal to the coherent information and so there is no difference between the quantum capacity and the private capacity for these channels. That is, for degradable channels, we have that

$$P(\mathcal{N}) = Q(\mathcal{N}) = I^c(\mathcal{N}), \tag{16.2.14}$$

where the coherent information $I^c(\mathcal{N})$ of a channel $\mathcal{N}$ is defined in (7.11.107). For anti-degradable channels, the private information is equal to zero, which is what we prove in Section 16.3.2.

Theorem 16.21 only makes a statement about the private capacity and not about the strong converse private capacity. Later on, we prove that a channel's relative entropy of entanglement is a strong converse rate for private communication, and for some channels, it coincides with the private information, thus leading to the strong converse property holding for these channels. More generally, however, the best statement we can make is that the regularized private information is a weak converse rate for all quantum channels.

There are two ingredients to the proof of Theorem 16.21:

1. *Achievability*: We prove that $I^p_{\text{reg}}(\mathcal{N})$ is an achievable rate, which involves explicitly constructing a private communication protocol. The developments in Section 16.1.4 on a lower bound for one-shot private capacity can be used, via the substitution $\mathcal{N} \to \mathcal{N}^{\otimes n}$, to argue for the existence of a private communication protocol for $\mathcal{N}$ in the asymptotic setting at the rate $I^p_{\text{reg}}(\mathcal{N})$.

   The achievability part of the proof establishes that $P(\mathcal{N}) \geq I^p_{\text{reg}}(\mathcal{N})$.

2. *Weak Converse*: We prove that $I^p_{\text{reg}}(\mathcal{N})$ is a weak converse rate, from which it follows that $P(\mathcal{N}) \leq I^p_{\text{reg}}(\mathcal{N})$. To show that $I^p_{\text{reg}}(\mathcal{N})$ is a weak converse rate, we use the upper bounds on one-shot private capacity from Section 16.1.3 to conclude that every achievable rate R satisfies $R \leq I^p_{\text{reg}}(\mathcal{N})$.

We first establish in Section 16.2.1 that the quantity $I^p_{\text{reg}}(\mathcal{N})$ is an achievable rate for private communication over $\mathcal{N}$. Then, in Section 16.2.2, we prove that $I^p_{\text{reg}}(\mathcal{N})$ is a weak converse rate.

Before proceeding, we establish a relationship between the private information of a quantum channel and its coherent information, which mirrors the operational relationship established in Section 16.1.1. This relationship gives another way to arrive at the conclusion that the private capacity of a quantum channel is not smaller than its quantum capacity.

Theorem 16.22

For a quantum channel $\mathcal{N}_{A\to B}$, its private information is not smaller than its coherent information:

$$I^c(\mathcal{N}) \leq I^p(\mathcal{N}), \tag{16.2.15}$$

where the coherent information is defined in (7.11.107) and the private information in (16.2.11). As a consequence, the private capacity is not smaller than the quantum capacity:

$$Q(\mathcal{N}) \leq P(\mathcal{N}). \tag{16.2.16}$$

PROOF: Picking a pure-state ensemble in (16.2.11), i.e., $\{p(x), \psi_A^x\}_{x\in\mathcal{X}}$, and setting

$$\rho_{XBE} \coloneqq \sum_{x\in\mathcal{X}} p(x)|x\rangle\!\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\psi_A^x), \tag{16.2.17}$$

with $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ an isometric channel extending $\mathcal{N}_{A\to B}$, we find that

$$I^p(\mathcal{N}) \geq I(X;B)_\rho - I(X;E)_\rho \tag{16.2.18}$$
$$= H(B)_\rho - H(B|X)_\rho - \big(H(E)_\rho - H(E|X)_\rho\big) \tag{16.2.19}$$
$$= H(B)_\rho - H(E)_\rho. \tag{16.2.20}$$

The first equality follows from rewriting the mutual information, and the second follows because the conditional entropies can be written as

$$H(B|X)_\rho = \sum_{x\in\mathcal{X}} p(x) H(\operatorname{Tr}_E[\mathcal{U}^{\mathcal{N}}_{A\to BE}(\psi_A^x)]), \tag{16.2.21}$$
$$H(E|X)_\rho = \sum_{x\in\mathcal{X}} p(x) H(\operatorname{Tr}_B[\mathcal{U}^{\mathcal{N}}_{A\to BE}(\psi_A^x)]). \tag{16.2.22}$$

They are equal because the entropies of the marginal states of a pure bipartite state are equal. Now consider that the reduced state of the BE systems is

$$\rho_{BE} = \sum_{x\in\mathcal{X}} p(x)\mathcal{U}^{\mathcal{N}}_{A\to BE}(\psi_A^x) = \mathcal{U}^{\mathcal{N}}_{A\to BE}\left(\sum_{x\in\mathcal{X}} p(x)\psi_A^x\right). \tag{16.2.23}$$

Since we can realize an arbitrary input density operator by taking convex combinations of pure states, and by applying (7.11.113), we conclude the claim in (16.2.15).

The conclusion in (16.2.16) follows by applying (16.2.15) and Theorems 14.16 and 16.21. ■

16.2.1 Proof of Achievability

In this section, we prove that $I^{p}_{\text{reg}}(\mathcal{N})$ is an achievable rate for private communication over a channel $\mathcal{N}$.

First, recall Corollary 16.14. Applying it, we find the following:

Corollary 16.23 Lower Bound for Private Communication in the Asymptotic Setting

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending it. For all $n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $\alpha \in (0,1)$, and $\beta > 1$, there exists an $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol for $\mathcal{N}_{A\to B}$ such that the rate $\frac{\log_2|\mathcal{M}|}{n}$ satisfies

$$\frac{\log_2|\mathcal{M}|}{n} \geq \sup_{\{p(x),\rho^x_A\}_{x\in\mathcal{X}}} \overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho \\ - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right), \quad (16.2.24)$$

where the information quantities are evaluated with respect to the state

$$\rho_{XBE} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho^x_A), \quad (16.2.25)$$

and the function f is defined in (16.1.142).

PROOF: We evaluate the quantities in Corollary 16.14 with respect to the tensor-power isometric channel $(\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n}$ and choose the input ensemble to be a tensor-power ensemble $\{p(x_1)\cdots p(x_n), \rho^{x_1}_{A_1} \otimes \cdots \otimes \rho^{x_n}_{A_n}\}_{x_1,\ldots,x_n\in\mathcal{X}^{\times n}}$. This implies that the state being evaluated for the Renyi information quantities is the tensor-power state $\rho^{\otimes n}_{XBE}$, where ρ_{XBE} is defined in (16.2.25). Let $\delta = \frac{\varepsilon'}{2}$, $\eta = \frac{\varepsilon'}{4}$, $\nu = \frac{\varepsilon'}{4}$, and $\zeta = \frac{\varepsilon'}{2}$. Exploiting the additivity of $\overline{I}_\alpha$ and $\widetilde{I}'_\beta$, substituting into the inequality in

(16.1.139), and simplifying leads to the inequality

$$\frac{\log_2|\mathcal{M}|}{n} \geq \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(\overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho\right) - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \quad (16.2.26)$$

This concludes the proof. ■

Using the inequality in (16.2.24), we conclude the following lower bound on the private capacity of a quantum channel:

Theorem 16.24 Achievability of Private Information for Private Communication

The private information $I^p(\mathcal{N})$ of a quantum channel $\mathcal{N}$, as defined in (16.2.11), is an achievable rate for private communication over $\mathcal{N}$. In other words,

$$P(\mathcal{N}) \geq I^p(\mathcal{N}), \quad (16.2.27)$$

for every quantum channel $\mathcal{N}$.

Proof: Let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending the channel $\mathcal{N}_{A\to B}$ of interest. Fix $\varepsilon \in (0,1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that $\delta = \delta_1 + \delta_2$. Set $\alpha \in (0,1)$ and $\beta > 1$ such that

$$\delta_1 \geq I(X;B)_\rho - I(X;E)_\rho - \left(\overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho\right), \quad (16.2.28)$$

where the information quantities are evaluated with respect to the state ρ_{XBE} in (16.2.25). Note that this is possible because $\overline{I}_\alpha(X;B)_\rho$ increases monotonically with increasing $\alpha \in (0,1)$ (see Proposition 7.23) and $\widetilde{I}'_\beta(X;E)_\rho$ decreases monotonically with decreasing β (see Proposition 7.31), so that

$$\lim_{\alpha\to 1^-} \overline{I}_\alpha(X;B)_\rho = \sup_{\alpha\in(0,1)} \overline{I}_\alpha(X;B)_\rho, \quad (16.2.29)$$

$$\lim_{\beta\to 1^+} \widetilde{I}'_\beta(X;E)_\rho = \inf_{\beta\in(1,\infty)} \widetilde{I}'_\beta(X;E)_\rho. \quad (16.2.30)$$

Also,

$$I(X;B)_\rho = \lim_{\alpha\to 1^-} \overline{I}_\alpha(X;B)_\rho, \quad (16.2.31)$$

$$I(X;E)_\rho = \lim_{\beta\to 1^+} \widetilde{I}'_\beta(X;E)_\rho. \tag{16.2.32}$$

With α and β chosen such that (16.2.28) holds, take n large enough so that

$$\delta_2 \geq \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{16.2.33}$$

Now, we use the fact that for the n and ε chosen above, there exists an $(n, |\mathcal{M}|, \varepsilon)$ protocol such that

$$\frac{\log_2|\mathcal{M}|}{n} \geq \overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right), \tag{16.2.34}$$

which follows from Corollary 16.23 above. Rearranging the right-hand side of this inequality, and using (16.2.28), (16.2.33), and (16.2.34), we find that

$$\begin{aligned}
\frac{\log_2|\mathcal{M}|}{n} &\geq I(X;B)_\rho - I(X;E)_\rho \\
&\quad - \begin{pmatrix} I(X;B)_\rho - I(X;E)_\rho - \left(\overline{I}_\alpha(X;B)_\rho - \widetilde{I}'_\beta(X;E)_\rho\right) \\ +\frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right) \end{pmatrix} && (16.2.35) \\
&\geq I(X;B)_\rho - I(X;E)_\rho - (\delta_1 + \delta_2) && (16.2.36) \\
&= I(X;B)_\rho - I(X;E)_\rho - \delta. && (16.2.37)
\end{aligned}$$

We thus have shown that there exists an $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol with rate $\frac{\log_2|\mathcal{M}|}{n} \geq I(X;B)_\rho - I(X;E)_\rho - \delta$. Therefore, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ private communication protocol with $R = I(X;B)_\rho - I(X;E)_\rho$ for all sufficiently large n such that (16.2.33) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0, 1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ private communication protocol. This means that, by definition, $I(X;B)_\rho - I(X;E)_\rho$ is an achievable rate. Since this is true for all input ensembles, we can finally take a supremum over all input ensembles to arrive at the conclusion in (16.2.27). ■

16.2.1.1 Proof of the Achievability Part of Theorem 16.21

Let $\{p(x), \rho^x_{A^k}\}_{x\in\mathcal{X}}$ be an arbitrary ensemble over k channel input systems, with $k \in \mathbb{N}$. Let

$$\tau_{XB^kE^k} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes (\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes k}(\rho^x_{A^k}). \tag{16.2.38}$$

Fix $\varepsilon \in (0,1]$ and $\delta > 0$. Let $\delta_1, \delta_2 > 0$ be such that $\delta = \delta_1 + \delta_2$. Set $\alpha \in (0,1)$ and $\beta > 1$ such that

$$\delta_1 \geq \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) - \frac{1}{k}\left(\overline{I}_\alpha(X;B^k)_\tau - \widetilde{I}'_\beta(X;E^k)_\tau\right), \tag{16.2.39}$$

which is possible based on the arguments given in the proof of Theorem 16.24 above. Then, with this choice of α and β, take n large enough so that

$$\delta_2 \geq \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{16.2.40}$$

Now, we use the fact that, for the chosen n and ε, there exists an $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol such that (16.2.24) holds, i.e.,

$$\frac{\log_2|\mathcal{M}|}{n} \geq \overline{I}_\alpha(X;B^k)_\tau - \widetilde{I}'_\beta(X;E^k)_\tau - \frac{1}{n} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{16.2.41}$$

Dividing both sides by k gives

$$\frac{\log_2|\mathcal{M}|}{kn} \geq \frac{1}{k}\left(\overline{I}_\alpha(X;B^k)_\tau - \widetilde{I}'_\beta(X;E^k)_\tau\right) - \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right). \tag{16.2.42}$$

Rearranging the right-hand side of this inequality, and using (16.2.39)–(16.2.42), we find that

$$\begin{aligned}
\frac{\log_2|\mathcal{M}|}{kn} &\geq \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) \\
&\quad - \left(\frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) - \overline{I}_\alpha(X;B^k)_\tau - \widetilde{I}'_\beta(X;E^k)_\tau\right) \\
&\quad - \frac{1}{kn} f\left(\varepsilon', \frac{\varepsilon'}{2}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{4}, \frac{\varepsilon'}{2}, \alpha, \beta\right) && (16.2.43) \\
&\geq \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) - (\delta_1 + \delta_2) && (16.2.44) \\
&= \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) - \delta. && (16.2.45)
\end{aligned}$$

Thus, there exists a $(kn, |\mathcal{M}|, \varepsilon)$ private communication protocol with rate $\frac{\log_2|\mathcal{M}|}{kn} \geq \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) - \delta$. Therefore, letting $n' \equiv kn$, we conclude that there exists an $(n', 2^{n'(R-\delta)}, \varepsilon)$ private communication protocol with

$$R = \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) \tag{16.2.46}$$

for all sufficiently large n such that (16.2.40) holds. Since ε and δ are arbitrary, we conclude that for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n\left(\frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right)-\delta\right)}, \varepsilon)$ private communication protocol. This means that $\frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right)$ is an achievable rate.

Now, since the input ensemble is arbitrary in the arguments above, we conclude that

$$\frac{1}{k} I^p(\mathcal{N}^{\otimes k}) = \sup_{\{p(x),\rho^x_{A^k}\}_{x\in\mathcal{X}}} \frac{1}{k}\left(I(X;B^k)_\tau - I(X;E^k)_\tau\right) \tag{16.2.47}$$

is an achievable rate. Finally, since the number k of instances of the channel $\mathcal{N}$ is arbitrary, we conclude that the regularized private information $\lim_{k\to\infty} \frac{1}{k} I^p(\mathcal{N}^{\otimes k})$ is an achievable rate.

16.2.2 Proof of the Weak Converse

In order to prove the weak converse part of Theorem 16.21, we make use of Corollary 16.8, specifically (16.1.80). Applying this inequality to the tensor-power channel $\mathcal{N}^{\otimes n}_{A\to B}$ leads to the following:

Proposition 16.25

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending it. Let $n \in \mathbb{N}$ and $\varepsilon \in [0,1)$. For an $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol for $\mathcal{N}_{A\to B}$, the rate $\frac{\log_2|\mathcal{M}|}{n}$ satisfies

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right)\frac{\log_2|\mathcal{M}|}{n} \le \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \frac{1}{n}\left(h_2(\varepsilon) + 2g(\sqrt{\varepsilon})\right), \tag{16.2.48}$$

where the information quantities are evaluated with respect to the state

$$\rho_{XB^nE^n} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes (\mathcal{U}^{\mathcal{N}}_{A\to BE})^{\otimes n}(\rho^x_{A^n}), \tag{16.2.49}$$

with $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ an isometric channel extending $\mathcal{N}_{A\to B}$. Consequently,

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right) P^{n,\varepsilon}(\mathcal{N}) \leq \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \frac{1}{n}\left(h_2(\varepsilon)+2g(\sqrt{\varepsilon})\right). \quad (16.2.50)$$

16.2.2.1 Proof of the Weak Converse Part of Theorem 16.21

Suppose that R is an achievable rate for private communication over the channel $\mathcal{N}_{A\to B}$. Then, by definition, for all $\varepsilon \in (0,1]$, $\delta > 0$, and sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ private communication protocol for $\mathcal{N}_{A\to B}$. For all such protocols, the inequality in (16.2.48) holds, so that

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right)(R-\delta) \leq \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \frac{1}{n}\left(h_2(\varepsilon)+2g(\sqrt{\varepsilon})\right). \quad (16.2.51)$$

Since the inequality holds for all sufficiently large n, it holds in the limit $n\to\infty$, so that

$$\left(1-\varepsilon-\sqrt{\varepsilon}\right)(R-\delta) \leq \lim_{n\to\infty}\left(\sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \frac{1}{n}\left(h_2(\varepsilon)+2g(\sqrt{\varepsilon})\right)\right) \quad (16.2.52)$$

$$= \lim_{n\to\infty} \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right). \quad (16.2.53)$$

Then since this inequality holds for all $\varepsilon \in (0,1)$, $\delta > 0$, it holds in particular for ε satisfying $\varepsilon + \sqrt{\varepsilon} < 1$, which gives

$$R \leq \frac{1}{1-\varepsilon-\sqrt{\varepsilon}} \lim_{n\to\infty} \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \delta, \quad (16.2.54)$$

and we thus conclude that

$$R \leq \lim_{\varepsilon,\delta\to 0} \frac{1}{1-\varepsilon-\sqrt{\varepsilon}} \lim_{n\to\infty} \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) + \delta \quad (16.2.55)$$

$$= \lim_{n\to\infty} \sup_{\{p(x),\rho^x_{A^n}\}_{x\in\mathcal{X}}} \frac{1}{n}\left(I(X;B^n)_\rho - I(X;E^n)_\rho\right) \tag{16.2.56}$$

$$= I^p_{\text{reg}}(\mathcal{N}). \tag{16.2.57}$$

We have thus shown that the quantity $I^p_{\text{reg}}(\mathcal{N})$ is a weak converse rate for private communication over $\mathcal{N}$.

16.2.3 Relative Entropy of Entanglement Strong Converse Bound

Except for channels for which the private information is known to be additive (such as the class of degradable channels; see Section 16.3.1 below), the private capacity of a channel is difficult to compute. This prompts us to find upper bounds on the private capacity. In this section, we do so in terms of the channel's relative entropy of entanglement, and in terms of the channel's squashed entanglement in the next section.

We begin by recalling the bound from (16.1.97), which holds for all $(|\mathcal{M}|, \varepsilon)$ private communication protocols and for all $\alpha > 1$:

$$P^\varepsilon(\mathcal{N}) \leq \widetilde{E}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{16.2.58}$$

For n channel uses, the bound in (16.1.97) becomes

$$\frac{\log_2|\mathcal{M}|}{n} \leq \frac{1}{n}\widetilde{E}_\alpha(\mathcal{N}^{\otimes n}) + \frac{\alpha}{n(\alpha-1)}\log_2\left(\frac{1}{1-\varepsilon}\right), \tag{16.2.59}$$

which holds for all $\alpha > 1$ and for all $(n, |\mathcal{M}|, \varepsilon)$ private communication protocols, with $n \in \mathbb{N}$ and $\varepsilon \in [0,1)$. We can simplify this inequality by making use of the following fact:

Proposition 16.26 Weak Subadditivity of a Channel's Renyi Relative Entropy of Entanglement

Let $\mathcal{N}_{A\to B}$ be a quantum channel, with d_A the dimension of the input system A. For all $\alpha > 1$ and $n \in \mathbb{N}$, we have

$$\widetilde{E}_\alpha(\mathcal{N}^{\otimes n}) \leq n\widetilde{E}_\alpha(\mathcal{N}) + \frac{\alpha\left(d_A^2-1\right)}{\alpha-1}\log_2(n+1). \tag{16.2.60}$$

PROOF: The proof is identical to the proof of Proposition 14.21, but making use of Proposition 10.9 at the beginning instead of Proposition 10.12. ■

Combining (16.2.60) with (16.2.59), we conclude the following upper bound on the rate of an arbitrary $(n, |\mathcal{M}|, \varepsilon)$ private communication protocol:

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \widetilde{E}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha - 1)} \log_2\!\left(\frac{(n+1)^{d_A^2 - 1}}{1 - \varepsilon}\right), \tag{16.2.61}$$

which holds for all $\alpha > 1$. Consequently, the following upper bound holds for the n-shot private capacity:

$$P^{n,\varepsilon}(\mathcal{N}) \leq \widetilde{E}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha - 1)} \log_2\!\left(\frac{(n+1)^{d_A^2 - 1}}{1 - \varepsilon}\right), \tag{16.2.62}$$

for all $\alpha > 1$.

With this bound, we are now ready to state the main result of this section, which is that a channel's relative entropy of entanglement is an upper bound on the strong converse private capacity of an arbitrary quantum channel $\mathcal{N}$.

Theorem 16.27 Strong Converse Upper Bound on Private Capacity

A channel's relative entropy of entanglement, denoted by $E_R(\mathcal{N})$, is a strong converse rate for private communication over $\mathcal{N}$. In other words, $\widetilde{P}(\mathcal{N}) \leq E_R(\mathcal{N})$ for every quantum channel $\mathcal{N}$.

Recall from (10.3.6) that

$$E_R(\mathcal{N}) := \sup_{\psi_{SA}} \inf_{\sigma_{SB} \in \mathrm{SEP}(S:B)} D(\mathcal{N}_{A\to B}(\psi_{SA}) \| \sigma_{SB}), \tag{16.2.63}$$

where the supremum is with respect to every pure state ψ_{SA} with $d_S = d_A$.

PROOF: The proof here is identical to that given for Theorem 14.22, but using the relative entropy of entanglement $E_R(\mathcal{N})$ instead of the Rains information $R(\mathcal{N})$. ■

16.2.4 Squashed Entanglement Weak Converse Bound

We showed earlier in Section 16.1.3.3 that the squashed entanglement gives an upper bound on the one-shot private capacity. Here we extend these results to the n-shot setting, establishing a bound on the n-shot private capacity and we conclude from it that the squashed entanglement of a quantum channel is a weak converse rate for private communication over it. Later on in the book, in Chapter 20, we prove that the squashed entanglement of a channel is an upper bound on its secret-key-agreement capacity, which generally can be much larger than its private capacity. Thus, the squashed entanglement bound is generally a loose upper bound on its (unassisted) private capacity.

Theorem 16.28 Squashed Entanglement Upper Bound on n-Shot Private Capacity

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0, 1)$. For all $(n, |\mathcal{M}|, \varepsilon)$ private communication protocols for $\mathcal{N}$, the following bound holds

$$\left(1 - 2\sqrt{\varepsilon}\right) \frac{\log_2|\mathcal{M}|}{n} \leq E_{\text{sq}}(\mathcal{N}) + \frac{2}{n} g_2(\sqrt{\varepsilon}), \tag{16.2.64}$$

where $E_{\text{sq}}(\mathcal{N})$ is the squashed entanglement of the channel $\mathcal{N}$, defined in Section 10.5.

Proof: Plugging the tensor-power channel $\mathcal{N}^{\otimes n}$ into the bound from Theorem 16.11, we conclude the following bound

$$\left(1 - 2\sqrt{\varepsilon}\right) \frac{\log_2|\mathcal{M}|}{n} \leq \frac{1}{n} E_{\text{sq}}(\mathcal{N}^{\otimes n}) + \frac{2}{n} g_2(\sqrt{\varepsilon}). \tag{16.2.65}$$

The desired statement then follows from the additivity of squashed entanglement of a channel (Corollary 10.21), which implies that $\frac{1}{n} E_{\text{sq}}(\mathcal{N}^{\otimes n}) = E_{\text{sq}}(\mathcal{N})$. ■

Theorem 16.29 Weak Converse Upper Bound on Private Capacity

The squashed entanglement $E_{\text{sq}}(\mathcal{N})$ of a quantum channel $\mathcal{N}_{A\to B}$ is a weak converse rate for private communication over $\mathcal{N}$. In other words, $P(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N})$ for every quantum channel $\mathcal{N}$.

Proof: We exploit the bound from Theorem 16.28 and an argument similar to that from Section 15.2.4 to conclude the desired statement. ■

16.3 Examples

We now consider the private capacity for particular classes of quantum channels. As we indicated earlier, computing the private capacity of an arbitrary channel is a difficult task. This task is made more difficult by the fact that, in some cases, the private information is known to be strictly superadditive in the following sense:

$$I^p(\mathcal{N}^{\otimes n}) \geq nI^p(\mathcal{N}). \tag{16.3.1}$$

This fact confirms that regularization of the private information is really needed in general in order to compute the private capacity, and that additivity of private information does not hold for all channels. Please consult the Bibliographic Notes in Section 16.5 for more information about strict superadditivity of private information for certain quantum channels.

Before starting the development below, recall that the private information of a channel $\mathcal{N}_{A\to B}$ is defined as

$$I^p(\mathcal{N}) = \sup_{\{p(x),\rho_A^x\}_{x\in\mathcal{X}}} \left(I(X;B)_\rho - I(X;E)_\rho\right), \tag{16.3.2}$$

where the state ρ_{XBE} is defined as

$$\rho_{XBE} := \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho_A^x), \tag{16.3.3}$$

with $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ an isometric channel extending $\mathcal{N}_{A\to B}$ and the optimization over every ensemble $\{p(x),\rho_A^x\}_{x\in\mathcal{X}}$.

16.3.1 Degradable Channels

Recall from Definition 4.6 that a channel $\mathcal{N}_{A\to B}$ is degradable if there exists a degrading channel $\mathcal{D}_{B\to E}$ such that

$$\mathcal{N}^c = \mathcal{D}\circ\mathcal{N}, \tag{16.3.4}$$

where $\mathcal{N}^c$ is a channel complementary to $\mathcal{N}$ (see Definition 4.5) and $d_E \geq \text{rank}(\Gamma^{\mathcal{N}}_{AB})$. In particular, if $V_{A\to BE}$ is an isometric extension of $\mathcal{N}$, so that

$$\mathcal{N}(\rho) = \text{Tr}_E[V\rho V^\dagger] \tag{16.3.5}$$

for every state ρ, then

$$\mathcal{N}^c(\rho) = \mathrm{Tr}_B[V\rho V^\dagger]. \tag{16.3.6}$$

We now show that the private information is equal to the coherent information for every degradable channel, i.e.,

$$I^p(\mathcal{N}) = I^c(\mathcal{N}), \tag{16.3.7}$$

where $I^c(\mathcal{N})$ is defined in (7.11.107). As a consequence of this observation and the fact that a tensor product of degradable channels is also degradable, it follows that the private capacity of a degradable channel is equal to its coherent information, and there is no difference between the private capacity and the quantum capacity in this case, i.e.,

$$Q(\mathcal{N}) = P(\mathcal{N}) = I^p(\mathcal{N}) = I^c(\mathcal{N}) \text{ for every degradable channel } \mathcal{N}. \tag{16.3.8}$$

Proposition 16.30 Private Information of Degradable Channels

Let $\mathcal{N}_{A\to B}$ be a degradable channel. Then its private information is equal to its coherent information:

$$I^p(\mathcal{N}) = I^c(\mathcal{N}), \tag{16.3.9}$$

where the channel's private information $I^p(\mathcal{N})$ is defined in (16.3.2) and its coherent information in (7.11.107). As a consequence, (16.3.8) holds.

PROOF: By Theorem 16.22, we only need to prove the inequality $I^p(\mathcal{N}) \leq I^c(\mathcal{N})$ for the case of a degradable channel. Let $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ be an isometric channel extending $\mathcal{N}_{A\to B}$. Let $\rho^x_A = \sum_y p(y|x)\psi^{x,y}_A$ be a spectral decomposition of the input state ρ^x_A, and define the following extension of the state ρ_{XBE} in (16.3.3):

$$\rho_{XYBE} = \sum_{x,y} p(x)p(y|x)|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\psi^{x,y}_A). \tag{16.3.10}$$

Consider that

$$\begin{aligned}
&I(X;B)_\rho - I(X;E)_\rho \\
&= I(XY;B)_\rho - I(Y;B|X)_\rho - \big(I(XY;E)_\rho - I(Y;E|X)_\rho\big) && (16.3.11)\\
&= I(XY;B)_\rho - I(XY;E)_\rho - \big(I(Y;B|X)_\rho - I(Y;E|X)_\rho\big) && (16.3.12)\\
&\leq I(XY;B)_\rho - I(XY;E)_\rho && (16.3.13)
\end{aligned}$$

$$= H(B)_\rho - H(B|XY)_\rho - \big(H(E)_\rho - H(E|XY)_\rho\big) \quad (16.3.14)$$

$$= H(B)_\rho - H(E)_\rho \quad (16.3.15)$$

$$\leq I^c(\mathcal{N}). \quad (16.3.16)$$

The first equality follows by applying the chain rule for conditional mutual information. The first inequality follows by applying the data-processing inequality for conditional mutual information and the fact that there is a degrading channel $\mathcal{D}_{B\to E}$ such that $\rho_{XYE} = \mathcal{D}_{B\to E}(\rho_{XYB})$. The last few steps follow the same reasoning given in the proof of Theorem 16.22. ■

16.3.1.1 Generalized Dephasing Channels

Recall the definition of a generalized dephasing channel from the discussion surrounding (14.3.30). Similar to what was found for generalized dephasing channels in Section 14.3.1.1, we can also consider the question of whether the strong converse property holds for the private capacity of these channels. Indeed, it is the case, and the reasoning is essentially the same as that given in the proof of Theorem 14.28, except that we use the strong converse bound on private capacity given by the relative entropy of entanglement. The main observation to make while examining the proof of Theorem 14.28 is that the state in (14.3.48) is a separable state.

Theorem 16.31 Private Capacity of Generalized Dephasing Channels

For every generalized dephasing channel $\mathcal{N}$ (defined by the isometric extension in (14.3.30)), the following equalities hold

$$P(\mathcal{N}) = \widetilde{P}(\mathcal{N}) = E_R(\mathcal{N}) = I^p(\mathcal{N}) \quad (16.3.17)$$

$$= Q(\mathcal{N}) = \widetilde{Q}(\mathcal{N}) = R(\mathcal{N}) = I^c(\mathcal{N}). \quad (16.3.18)$$

Proof: The following inequalities hold in general

$$I^c(\mathcal{N}) \leq Q(\mathcal{N}) \leq P(\mathcal{N}) \leq \widetilde{P}(\mathcal{N}) \leq E_R(\mathcal{N}), \quad (16.3.19)$$

$$I^c(\mathcal{N}) \leq I^p(\mathcal{N}) \leq P(\mathcal{N}), \quad (16.3.20)$$

$$I^c(\mathcal{N}) \leq Q(\mathcal{N}) \leq \widetilde{Q}(\mathcal{N}) \leq R(\mathcal{N}) \leq E_R(\mathcal{N}), \quad (16.3.21)$$

and the reasoning given above establishes that $I^c(\mathcal{N}) = E_R(\mathcal{N})$ for generalizing dephasing channels. ■

16.3.2 Anti-Degradable Channels

Let us consider the private capacity for anti-degradable channels. Recall from Definition 4.6 that a channel $\mathcal{N}_{A\to B}$ is anti-degradable if there exists an anti-degrading channel $\mathcal{A}_{E\to B}$ such that

$$\mathcal{N}_{A\to B} = \mathcal{A}_{E\to B} \circ \mathcal{N}^c_{A\to E}, \tag{16.3.22}$$

where $\mathcal{N}^c_{A\to E}$ is a channel complementary to $\mathcal{N}_{A\to B}$ and $d_E \geq \operatorname{rank}(\Gamma^{\mathcal{N}}_{AB})$.

Proposition 16.32 Private Information for Anti-Degradable Channels

The private information vanishes for all anti-degradable channels, i.e., $I^p(\mathcal{N}) = 0$ for every anti-degradable channel $\mathcal{N}$. Therefore, the private capacity of an anti-degradable channel is equal to zero, i.e., $P(\mathcal{N}) = 0$ for every anti-degradable channel $\mathcal{N}$.

Proof: The first claim is a direct consequence of the definition of the private information in (16.3.2), the fact that there is an anti-degrading channel $\mathcal{A}_{E\to B}$ such that $\rho_{XB} = \mathcal{A}_{E\to B}(\rho_{XE})$, where the state ρ_{XBE} is defined in (16.3.3), and the data-processing inequality for mutual information. The second claim follows from the regularized expression for private capacity from Theorem 16.21 and the fact that a tensor product of anti-degradable channels is anti-degradable. ■

16.4 Summary

In this chapter, we studied private communication over a quantum channel $\mathcal{N}_{A\to B}$. The communication model that we employed is that a sender has access to the input system A, a legitimate receiver has access to the output system B, and an eavesdropper has access to the system E of an isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ extending $\mathcal{N}_{A\to B}$. This model gives the most power to the eavesdropper, subject to the constraints that the systems A and B are physically secure in the laboratories of the sender and receiver, respectively. The goal of a private communication protocol is for the sender to transmit a classical message such that the receiver can decode it with high probability and the eavesdropper cannot determine which message was transmitted (i.e., her system E should be essentially useless for figuring out the transmitted message). The private capacity is defined as the largest rate at which private communication is possible, such that the decoding error probability tends to zero and the eavesdropper's system becomes decoupled with the message system. In our definitions, we combined these requirements into a single constraint. We found that the private information $I^p(\mathcal{N})$ of quantum channel $\mathcal{N}$ is a lower bound on its private capacity, and that, in general, computing the exact value of the private capacity involves a regularization, i.e., $P(\mathcal{N}) = I^p_{\text{reg}}(\mathcal{N})$.

Following the same course as in previous chapters, we began with the one-shot setting for private communication, in which only one use of the channel is allowed, along with some non-zero error. We then determined upper and lower bounds on the number of private bits that can be transmitted. We established three upper bounds on the one-shot private capacity, involving the one-shot private information, the hypothesis testing relative entropy of entanglement, as well as the squashed entanglement. These in turn led to upper bounds on the asymptotic private capacity. To obtain a lower bound on the one-shot private capacity, we employed the methods of position-based coding and convex splitting, similar to how we did in the previous chapter on secret key distillation (Chapter 15). This lower bound is optimal when employed in the asymptotic setting because it leads to the regularized private information as an achievable rate for private communication, and this matches the upper bound. For degradable channels, there is no difference between the private information and the coherent information, and this implies that there is no difference between the private capacity and quantum capacity for these channels. We also proved that the private capacity of anti-degradable channels is equal to zero.

Since the regularized private information is difficult to compute, we established other upper bounds on private capacity, in terms of relative entropy of entanglement (strong converse upper bound) and squashed entanglement (weak converse upper bound). We then concluded that the strong converse property holds for all generalized dephasing channels and their private capacity is equal to their coherent information.

16.5 Bibliographic Notes

Shannon (1949) studied the information-theoretic security of communication systems. Some years later, the private capacity of a classical channel (also known as secrecy capacity) was introduced and studied by Wyner (1975) and some years later by Csiszár and Körner (1978), who established a general formula for the private capacity of a classical channel.

Bennett and Brassard (1984) devised the first protocol for sending private classical information over a quantum channel, which is known as quantum key distribution. The private capacity of a quantum channel was studied by Devetak (2005); Cai et al. (2004), who independently established the regularized expression for it in Theorem 16.21.

Private communication was studied from the one-shot perspective by Renes and Renner (2011); Wilde et al. (2017); Wilde (2017b); Radhakrishnan et al. (2017). Proposition 16.3 was established by Wilde and Qi (2018). The connection between secret-key transmission and bipartite private-state transmission is a direct consequence of the insights of Horodecki et al. (2005a, 2009a) and was discussed by Wilde et al. (2017). The upper bound in Proposition 16.7 is similar to that established by Qi et al. (2018a). The upper bound in Theorem 16.9 is due to Wilde et al. (2017) and the upper bound in Theorem 16.11 to Takeoka et al. (2014). The lower bound in Section 16.1.4 is due to Wilde (2017b).

As mentioned above, the asymptotic theory of private communication was developed by Devetak (2005); Cai et al. (2004). Devetak (2005) proved Theorem 16.22, relating coherent and private information and the private to quantum capacity. Strict superadditivity of the private information of a quantum channel was established by Smith et al. (2008), and this result was strengthened by Elkouss and Strelchuk (2015). The relative entropy of entanglement strong converse bound on private capacity in Theorem 16.27 was proven by Wilde et al. (2017). The squashed entanglement weak converse bound on private capacity in Theorem 16.29 was proven by Takeoka et al. (2014). The private capacity of degradable channels (i.e., Theorem 16.30 and (16.3.8)) was established by Smith (2008). The strong converse property for the private capacity of generalized dephasing channels was established by Wilde et al. (2017).

Part III

Quantum Communication Protocols With Feedback Assistance

We now delve into interactive quantum communication protocols. Such protocols involve interaction between the sender and receiver of a quantum channel, beyond the quantum channel that connects them, and this interaction can potentially increase a given communication capacity because it represents an additional resource that the sender and receiver have at their disposal. Such protocols are richer than the non-interactive protocols that we considered previously, and as such, their analysis is more involved.

One objective of the following chapters is to understand what role this interaction plays and whether it can increase capacity. For the most part, what we accomplish is the establishment of limitations on the ability of feedback to increase capacity. In some cases, such as the case presented in the first chapter, a surprising conclusion is that interaction does not increase capacity at all, so that the theory simplifies.

Chapter 17

Quantum-Feedback-Assisted Communication

In this chapter, we begin our foray into interactive quantum communication by analyzing communication protocols in which the goal is for the sender to communicate a classical message to the receiver, with the assistance of a free noiseless quantum feedback channel. By a quantum feedback channel, we mean a quantum channel from the receiver to the sender that is separate from the channel from the sender to the receiver being used to communicate the message. We thus call this communication scenario "quantum-feedback-assisted communication."

One simple (yet effective) way to make use of this free noiseless quantum feedback channel is for the receiver to transmit one share of a bipartite quantum state to the sender. By doing so, they can establish shared entanglement, and the rates of classical communication that are achievable with such a strategy are given by the limits on entanglement-assisted communication that we studied previously in Chapter 11.

Perhaps surprisingly, we show here that the same non-asymptotic converse bounds established in (11.2.61) and (11.2.92) apply to protocols assisted by noiseless quantum feedback. These non-asymptotic converse bounds imply that the quantum-feedback-assisted classical capacity of a channel is no larger than its entanglement-assisted capacity. Furthermore, the strong converse property holds for the quantum-feedback-assisted capacity, so that the strong converse capacity is equal to the mutual information of a quantum channel.

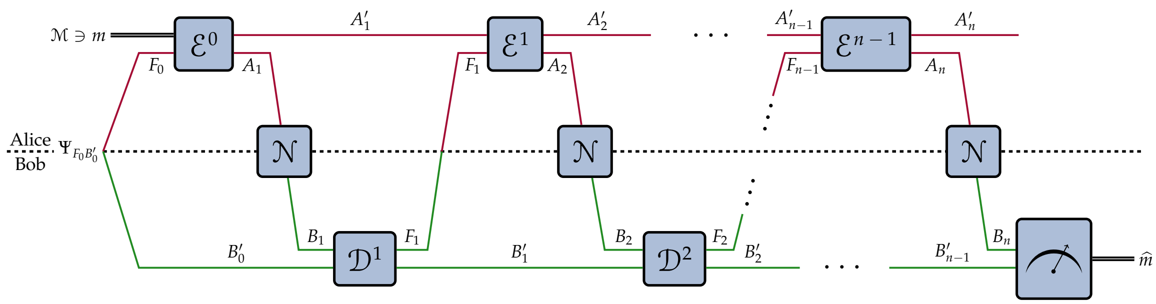


FIGURE 17.1: A general quantum-feedback-assisted communication protocol for the channel $\mathcal{N}$, which uses it n times.

This result demonstrates that the entanglement-assisted capacity of a quantum channel is a rather robust communication capacity. Not only is the mutual information of a channel equal to the strong converse entanglement-assisted capacity for all channels, but it is also equal to the strong converse quantum feedback-assisted capacity for all channels. Thus, the theory of entanglement-assisted and quantum-feedback-assisted communication simplifies immensely.

It is worth remarking that Shannon proved that a similar result holds for classical channels, and the strong converse property was later demonstrated as well. In this sense, the entanglement-assisted capacity of a quantum channel represents the fully quantum generalization of the classical capacity of a classical channel. Related, the quantum mutual information of a quantum channel represents the fully quantum generalization of the classical mutual information of a classical channel.

17.1 n-Shot Quantum Feedback-Assisted Communication Protocols

We begin by defining the most general form for an n-shot classical communication protocol assisted by a noiseless quantum feedback channel, where $n \in \mathbb{N}$. Such a protocol is depicted in Figure 17.1, and it is defined by the following elements:

$$(\mathcal{M}, \Psi_{F_0 B'_0}, \mathcal{E}^0_{M' F_0 \to A'_1 A_1}, \{\mathcal{E}^i_{A'_i F_i \to A'_{i+1} A_{i+1}}, \mathcal{D}^i_{B_i B'_{i-1} \to F_i B'_i}\}_{i=1}^{n-1}, \mathcal{D}^n_{B_n B'_{n-1} \to \widehat{M}}), \tag{17.1.1}$$

where $\mathcal{M}$ is the message set, $\Psi_{F_0 B'_0}$ denotes a bipartite quantum state, the objects denoted by $\mathcal{E}$ are encoding channels, and the objects denoted by $\mathcal{D}$ are decoding channels. Let $\mathcal{C}$ denote all of these elements, which together constitute the quantum-feedback-assisted code. The quantum systems labeled by F represent the feedback systems that Bob sends back to Alice. The primed systems A'_i and B'_i represent local quantum memory or "scratch" registers that Alice and Bob can exploit in the feedback-assisted protocol.

In such an n-round feedback-assisted protocol, the protocol proceeds as follows: let $p : \mathcal{M} \to [0, 1]$ be a probability distribution over the message set. Alice starts by preparing two classical registers M and M' in the following state:

$$\overline{\Phi}^p_{MM'} \coloneqq \sum_{m \in \mathcal{M}} p(m) |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'}. \tag{17.1.2}$$

Furthermore, Alice and Bob also initially share a quantum state $\Psi_{F_0 B'_0}$ on Alice's system F_0 and Bob's system B'_0. This state is prepared by Bob locally, and then he transmits the system F_0 to Alice via the noiseless quantum feedback channel. The initial global state shared between them is

$$\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0 B'_0}. \tag{17.1.3}$$

Alice then sends the M' and F_0 registers through the first encoding channel $\mathcal{E}^0_{M'F_0 \to A'_1 A_1}$. This encoding channel realizes a set $\{\mathcal{E}^{0,m}_{F_0 \to A'_1 A_1}\}_{m \in \mathcal{M}}$ of quantum channels as follows:

$$\mathcal{E}^{0,m}_{F_0 \to A'_1 A_1}(\tau_{F_0}) \coloneqq \mathcal{E}^0_{M'F_0 \to A'_1 A_1}(|m\rangle\langle m|_{M'} \otimes \tau_{F_0}), \tag{17.1.4}$$

for all input states τ_{F_0}. The global state after the first encoding channel is then as follows:

$$\mathcal{E}^0_{M'F_0 \to A'_1 A_1}(\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0 B'_0}). \tag{17.1.5}$$

Note that the scratch system A'_1 can contain a classical copy of the particular message m that is being communicated, and the same is true for all of the later scratch systems A'_i, for $i \in \{2, \ldots, n\}$. In fact, this is necessary in order for the communication protocol to be effective. Alice then transmits the A_1 system through the channel $\mathcal{N}_{A_1 \to B_1}$, leading to the state

$$\rho^1_{MA'_1 B_1 B'_0} \coloneqq (\mathcal{N}_{A_1 \to B_1} \circ \mathcal{E}^0_{M'F_0 \to A'_1 A_1})(\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0 B'_0}). \tag{17.1.6}$$

After receiving the B_1 system, Bob performs the decoding channel $\mathcal{D}^1_{B_1 B'_0 \to F_1 B'_1}$, such that the state is then

$$\mathcal{D}^1_{B_1 B'_0 \to F_1 B'_1}(\rho^1_{MA'_1 B_1 B'_0}), \tag{17.1.7}$$

with it being understood that the system B'_1 is Bob's new scratch register and the feedback system F_1 gets sent over the noiseless quantum feedback channel back to Alice.

In the next round, Alice processes the A'_1F_1 systems with the encoding channel $\mathcal{E}^1_{A'_1F_1\to A'_2A_2}$, and she sends system A_2 over the channel $\mathcal{N}_{A_2\to B_2}$, leading to the state

$$\rho^2_{MA'_2B_2B'_1} \coloneqq (\mathcal{N}_{A_2\to B_2} \circ \mathcal{E}^1_{A'_1F_1\to A'_2A_2} \circ \mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\rho^1_{MA'_1B_1B'_0}). \tag{17.1.8}$$

Bob then applies the second decoding channel $\mathcal{D}^2_{B_2B'_1\to F_2B'_2}$. This process then iterates $n-2$ more times, and the state after each use of the channel is as follows:

$$\rho^i_{MA'_iB_iB'_{i-1}} \coloneqq \\ (\mathcal{N}_{A_i\to B_i} \circ \mathcal{E}^{i-1}_{A'_{i-1}F_{i-1}\to A'_iA_i} \circ \mathcal{D}^{i-1}_{B_{i-1}B'_{i-2}\to F_{i-1}B'_{i-1}})(\rho^{i-1}_{MA'_{i-1}B_{i-1}B'_{i-2}}), \tag{17.1.9}$$

for $i \in \{3, \ldots, n\}$.

In the final round, Bob performs the decoding channel $\mathcal{D}^n_{B_nB'_{n-1}\to\widehat{M}}$, which is a quantum-to-classical channel that finally decodes the transmitted message. The final classical–classical state of the protocol is then as follows:

$$\omega^p_{M\widehat{M}} \coloneqq \mathcal{D}^n_{B_nB'_{n-1}\to\widehat{M}}(\mathrm{Tr}_{A'_n}[\rho^n_{MA'_nB_nB'_{n-1}}]). \tag{17.1.10}$$

Now, just as we did in Chapter 11 in the case of entanglement-assisted classical communication, we can define the message error probability, average error probability, and maximal error probability as in (11.1.13), (11.1.14), and (11.1.15), respectively. Using the alternative expression in (11.1.24) for the average error probability, we have that the average error probability for the quantum-feedback-assisted code $\mathcal{C}$ is given by

$$\overline{p}_{\mathrm{err}}(\mathcal{C}; p) = \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1. \tag{17.1.11}$$

Using the alternative expression in (11.1.36) for the maximal error probability, we have that the maximal error probability of the quantum-feedback-assisted code $\mathcal{C}$ is given by

$$p^*_{\mathrm{err}}(\mathcal{C}) = \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2}\left\|\overline{\Phi}^p_{MM'} - \omega^p_{M\widehat{M}}\right\|_1. \tag{17.1.12}$$

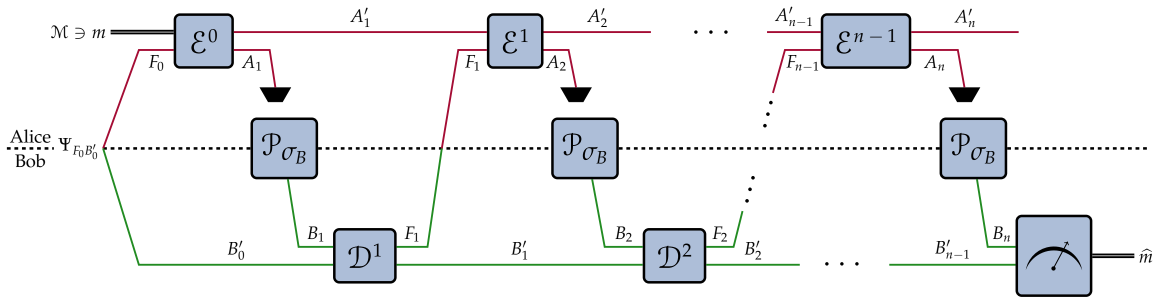


FIGURE 17.2: Depiction of a protocol that is useless for quantum-feedback-assisted classical communication. In each round, the encoded state is discarded and replaced with an arbitrary (but fixed) state σ_B.

Definition 17.1 **$(n, |\mathcal{M}|, \varepsilon)$ Quantum-Feedback-Assisted Classical Communication Protocol**

Let $(\mathcal{M}, \Psi_{F_0 B'_0}, \mathcal{E}^0_{M' F_0 \to A'_1 A_1}, \{\mathcal{E}^i_{A'_i F_i \to A'_{i+1} A_{i+1}}, \mathcal{D}^i_{B_i B'_{i-1} \to F_i B'_i}\}_{i=1}^{n-1}, \mathcal{D}^n_{B_n B'_{n-1} \to \widehat{M}})$ be the elements of an n-shot quantum-feedback-assisted classical communication protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an $(n, |\mathcal{M}|, \varepsilon)$ protocol, with $\varepsilon \in [0,1]$, if $p^*_{\text{err}}(\mathcal{C}) \leq \varepsilon$.

17.1.1 Protocol over a Useless Channel

As before, when determining converse bounds on the rate at which classical messages can be communicated reliably using such feedback-assisted protocols, it is helpful to consider a useless channel. Our plan is again to use relative entropy (or some generalized divergence) to compare the states at each time step of the actual protocol with those resulting from employing a useless channel instead of the actual channel. As before, a useless channel that conveys no information at all is one in which the input state is discarded and replaced with some state at the output:

$$\mathcal{R}_{A\to B} \coloneqq \mathcal{P}_{\sigma_B} \circ \text{Tr}_A, \tag{17.1.13}$$

where $\mathcal{P}_{\sigma_B}$ denotes a preparation channel that prepares the arbitrary (but fixed) state σ_B at the output.

We can modify the i^{th} step of the protocol discussed in the previous section, such that instead of the actual channel $\mathcal{N}_{A_i \to B_i}$ being applied, the replacement channel $\mathcal{R}_{A_i \to B_i}$ is applied; see Figure 17.2.

The state after the first round in this protocol over the useless channel is

$$\tau^1_{MA'_1B_1B'_0} := (\mathcal{R}_{A_1\to B_1} \circ \mathcal{E}^0_{M'F_0\to A'_1A_1})(\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0B'_0}) \tag{17.1.14}$$

$$= \operatorname{Tr}_{A_1}[\mathcal{E}^0_{M'F_0\to A'_1A_1}(\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0B'_0})] \otimes \sigma_{B_1}, \tag{17.1.15}$$

where we observe that

$$\tau^1_{MA'_1B_1B'_0} = \tau^1_{MA'_1B'_0} \otimes \sigma_{B_1}, \tag{17.1.16}$$

and furthermore that

$$\tau^1_{MB_1B'_0} = \operatorname{Tr}_{A'_1A_1}[\mathcal{E}^0_{M'F_0\to A'_1A_1}(\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0B'_0})] \otimes \sigma_{B_1} \tag{17.1.17}$$

$$= \operatorname{Tr}_{M'F_0}[\overline{\Phi}^p_{MM'} \otimes \Psi_{F_0B'_0}] \otimes \sigma_{B_1} \tag{17.1.18}$$

$$= \pi^p_M \otimes \Psi_{B'_0} \otimes \sigma_{B_1}, \tag{17.1.19}$$

where the second equality holds due to the fact that the first encoding channel $\mathcal{E}^0_{M'F_0\to A'_1A_1}$ is trace preserving, and where

$$\pi^p_M := \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M. \tag{17.1.20}$$

Thus, there is no correlation whatsoever between the message system M and Bob's systems $B_1B'_0$ after tracing over all of Alice's systems. Intuitively, this is a consequence of the fact that the "communication line has been cut" when employing the replacement channel.

The state after the second replacement channel $\mathcal{R}_{A_2\to B_2}$ is then given by

$$\tau^2_{MA'_2B_2B'_1}$$

$$:= (\mathcal{R}_{A_2\to B_2} \circ \mathcal{E}^1_{A'_1F_1\to A'_2A_2} \circ \mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\tau^1_{MA'_1B_1B'_0}) \tag{17.1.21}$$

$$= \operatorname{Tr}_{A_2}[(\mathcal{E}^1_{A'_1F_1\to A'_2A_2} \circ \mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\tau^1_{MA'_1B_1B'_0})] \otimes \sigma_{B_2} \tag{17.1.22}$$

$$= \operatorname{Tr}_{A_2}[(\mathcal{E}^1_{A'_1F_1\to A'_2A_2} \circ \mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\tau^1_{MA'_1B'_0} \otimes \sigma_{B_1})] \otimes \sigma_{B_2}, \tag{17.1.23}$$

where we used (17.1.16) to obtain the last line. If we take the partial trace over system A'_2, then the fact that the encoding channel $\mathcal{E}^1_{A'_1F_1\to A'_2A_2}$ is trace preserving implies that

$$\tau^2_{MB_2B'_1}$$

$$= \mathrm{Tr}_{A_2'A_2}[(\mathcal{E}^1_{A_1'F_1\to A_2'A_2} \circ \mathcal{D}^1_{B_1B_0'\to F_1B_1'})(\tau^1_{MA_1'B_0'} \otimes \sigma_{B_1})] \otimes \sigma_{B_2} \tag{17.1.24}$$

$$= \mathrm{Tr}_{A_1'F_1}[\mathcal{D}^1_{B_1B_0'\to F_1B_1'}(\tau^1_{MA_1'B_0'} \otimes \sigma_{B_1})] \otimes \sigma_{B_2} \tag{17.1.25}$$

$$= \mathrm{Tr}_{F_1}[\mathcal{D}^1_{B_1B_0'\to F_1B_1'}(\tau^1_{MB_0'} \otimes \sigma_{B_1})] \otimes \sigma_{B_2}. \tag{17.1.26}$$

Then, using (17.1.19), which implies that $\tau^1_{MB_0'} = \pi^p_M \otimes \Psi_{B_0'}$, we find that

$$\tau^2_{MB_2B_1'} = \mathrm{Tr}_{F_1}[\mathcal{D}^1_{B_1B_0'\to F_1B_1'}(\pi^p_M \otimes \Psi_{B_0'} \otimes \sigma_{B_1})] \otimes \sigma_{B_2} \tag{17.1.27}$$

$$= \pi^p_M \otimes \mathrm{Tr}_{F_1}[\mathcal{D}^1_{B_1B_0'\to F_1B_1'}(\Psi_{B_0'} \otimes \sigma_{B_1})] \otimes \sigma_{B_2} \tag{17.1.28}$$

$$= \pi^p_M \otimes \tau^2_{B_1'} \otimes \sigma_{B_2}. \tag{17.1.29}$$

Thus, we find again that there is no correlation whatsoever between the message system M and Bob's systems B_2B_1' after tracing over all of Alice's systems.

The states for the other rounds $i \in \{3, \dots, n\}$ are given by

$$\tau^i_{MA_i'B_iB_{i-1}'}$$

$$:= (\mathcal{R}_{A_i\to B_i} \circ \mathcal{E}^{i-1}_{A_{i-1}'F_{i-1}\to A_i'A_i} \circ \mathcal{D}^{i-1}_{B_{i-1}B_{i-2}'\to F_{i-1}B_{i-1}'})(\tau^{i-1}_{MA_{i-1}'B_{i-1}B_{i-2}'}) \tag{17.1.30}$$

$$= \mathrm{Tr}_{A_i}[\mathcal{E}^{i-1}_{A_{i-1}'F_{i-1}\to A_i'A_i} \circ \mathcal{D}^{i-1}_{B_{i-1}B_{i-2}'\to F_{i-1}B_{i-1}'})(\tau^{i-1}_{MA_{i-1}'B_{i-1}B_{i-2}'})] \otimes \sigma_{B_i} \tag{17.1.31}$$

$$= \mathrm{Tr}_{A_i}[\mathcal{E}^{i-1}_{A_{i-1}'F_{i-1}\to A_i'A_i} \circ \mathcal{D}^{i-1}_{B_{i-1}B_{i-2}'\to F_{i-1}B_{i-1}'})(\tau^{i-1}_{MA_{i-1}'B_{i-2}'} \otimes \sigma_{B_{i-1}})] \otimes \sigma_{B_i}. \tag{17.1.32}$$

Repeating a calculation similar to the above leads to a similar conclusion as above:

$$\tau^i_{MB_iB_{i-1}'} = \pi^p_M \otimes \tau^i_{B_{i-1}'} \otimes \sigma_{B_i}, \tag{17.1.33}$$

for all $i \in \{3, \dots, n\}$. That is, there is no correlation whatsoever between the message system M and Bob's systems B_iB_{i-1}' after tracing over all of Alice's systems. Again, this is intuitively a consequence of the fact that the "communication line has been cut" when employing the replacement channel.

Bob's final decoding channel $\mathcal{D}^n_{B_nB_{n-1}'\to\widehat{M}}$ therefore leads to the following classical–classical state:

$$\tau_{M\widehat{M}} := \pi^p_M \otimes \mathcal{D}^n_{B_nB_{n-1}'\to\widehat{M}}(\tau^2_{B_{n-1}'} \otimes \sigma_{B_n}) = \pi^p_M \otimes \tau_{\widehat{M}}, \tag{17.1.34}$$

where $\tau_{\widehat{M}} := \sum_{\widehat{m}\in\mathcal{M}} t(\widehat{m})|\widehat{m}\rangle\langle\widehat{m}|_{\widehat{M}}$ for some probability distribution $t : \mathcal{M} \to [0, 1]$, which corresponds to Bob's measurement.

17.1.2 Upper Bound on the Number of Transmitted Bits

We now give a general upper bound on the number transmitted bits in any quantum-feedback-assisted classical communication protocol. This result is stated in Theorem 17.2, and it holds independently of the encoding and decoding channels used in the protocol and depends only on the given communication channel $\mathcal{N}$. Recall from the previous section that $\log_2 |\mathcal{M}|$ represents the number of bits that are transmitted over the channel $\mathcal{N}$.

Theorem 17.2 ***n*-Shot Upper Bounds for Quantum-Feedback-Assisted Classical Communication**

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all $(n, |\mathcal{M}|, \varepsilon)$ quantum-feedback-assisted classical communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bounds hold,

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \frac{1}{1-\varepsilon}\left(I(\mathcal{N}) + \frac{1}{n}h_2(\varepsilon)\right), \tag{17.1.35}$$

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right) \qquad \forall\, \alpha > 1, \tag{17.1.36}$$

where $I(\mathcal{N})$ is the mutual information of $\mathcal{N}$, as defined in (7.11.102), and $\widetilde{I}_\alpha(\mathcal{N})$ is the sandwiched Rényi mutual information of $\mathcal{N}$, as defined in (7.11.91).

Proof: Let us start with an arbitrary $(n, |\mathcal{M}|, \varepsilon)$ quantum-feedback-assisted classical communication protocol over a channel $\mathcal{N}_{A\to B}$, corresponding to, as described earlier, a message set $\mathcal{M}$, a shared quantum state $\Psi_{F_0 B_0'}$, the encoding channels $\mathcal{E}^0_{M' F_0 \to A_1' A_1}$ and $\{\mathcal{E}^i_{A_i' F_i \to A_{i+1}' A_{i+1}}\}_{i=1}^{n-1}$, and the decoding channels $\{\mathcal{D}^i_{B_i B_{i-1}' \to F_i B_i'}\}_{i=1}^{n-1}$ and $\mathcal{D}^n_{B_n B_{n-1}' \to \widehat{M}}$. Recall that we refer to all of these objects collectively as the code $\mathcal{C}$. The error criterion $p^*_{\text{err}}(\mathcal{C}) \leq \varepsilon$ holds by the definition of an $(n, |\mathcal{M}|, \varepsilon)$ protocol, which implies that for all probability distributions $p : \mathcal{M} \to [0,1]$ on the message set $\mathcal{M}$:

$$\overline{p}_{\text{err}}(\mathcal{C}; p) \leq p^*_{\text{err}}(\mathcal{C}) \leq \varepsilon. \tag{17.1.37}$$

(The reasoning for this is analogous to that in (11.1.63)–(11.1.66).) In particular, the above inequality holds with p being the uniform distribution on $\mathcal{M}$, so that $p(m) = \frac{1}{|\mathcal{M}|}$ for all $m \in \mathcal{M}$.

Now, let $\overline{\Phi}_{M\widehat{M}}$ be the state defined in (17.1.2) with p the uniform distribution, and similarly let $\omega_{M\widehat{M}}$, defined in (17.1.10), be the state at the end of the protocol such that p is the uniform prior probability distribution. Observe that $\operatorname{Tr}[\omega_{M\widehat{M}}] = \pi_M$. Also, letting

$$\Pi_{M\widehat{M}} = \sum_{m\in\mathcal{M}} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\widehat{M}} \tag{17.1.38}$$

be the projection defining the comparator test, as in (11.1.37), observe that

$$1 - \operatorname{Tr}[\Pi_{M\widehat{M}}\omega_{M\widehat{M}}] = \frac{1}{2}\left\|\overline{\Phi}_{M\widehat{M}} - \omega_{M\widehat{M}}\right\|_1 \leq \varepsilon, \tag{17.1.39}$$

where the first equality follows by combining (11.1.24) with (11.1.41). This means that

$$\operatorname{Tr}[\Pi_{M\widehat{M}}\omega_{M\widehat{M}}] \geq 1 - \varepsilon. \tag{17.1.40}$$

We thus have all of the ingredients to apply Lemma 11.4. Doing so gives the following critical first bound:

$$\log_2|\mathcal{M}| \leq I_H^{\varepsilon}(M;\widehat{M})_{\omega}. \tag{17.1.41}$$

Invoking Proposition 7.70, the definition of $I_H^{\varepsilon}(M;\widehat{M})$ from (7.11.88), and the expression for mutual information from (7.2.97), we find that

$$I_H^{\varepsilon}(M;\widehat{M})_{\omega} \leq \frac{1}{1-\varepsilon}\left(I(M;\widehat{M})_{\omega} + h_2(\varepsilon)\right). \tag{17.1.42}$$

Now, using the data-processing inequality for the mutual information (see Proposition 7.19) with respect to the last decoding channel, $\mathcal{D}^n_{B_nB'_{n-1}\to\widehat{M}}$, we find that

$$I(M;\widehat{M})_{\omega} \leq I(M;B_nB'_{n-1})_{\rho^n}. \tag{17.1.43}$$

Then, using the chain rule for mutual information in (7.2.112), we obtain

$$\begin{aligned} I(M;B_nB'_{n-1})_{\rho^n} &= I(M;B_n|B'_{n-1})_{\rho^n} + I(M;B'_{n-1})_{\rho^n} && (17.1.44)\\ &\leq I(MB'_{n-1};B_n)_{\rho^n} + I(M;B'_{n-1})_{\rho^n}, && (17.1.45) \end{aligned}$$

where the second line is a consequence of the chain rule, as well as non-negativity of mutual information:

$$I(M;B_n|B'_{n-1})_{\rho^n} = I(MB'_{n-1};B_n)_{\rho^n} - I(B'_{n-1};B_n)_{\rho^n} \tag{17.1.46}$$

$$\leq I(MB'_{n-1}; B_n)_{\rho^n}. \tag{17.1.47}$$

Finally, observe that the state $\rho^n_{MB_nB'_{n-1}}$ has the following form:

$$\rho^n_{MB_nB'_{n-1}} = \mathcal{N}_{A_n\to B_n}(\zeta^n_{MB'_{n-1}A_n}), \tag{17.1.48}$$

where

$$\zeta^n_{MB'_{n-1}A_n} := \\ \operatorname{Tr}_{A'_n}[(\mathcal{E}^{n-1}_{A'_{n-1}F_{n-1}\to A'_nA_n} \circ \mathcal{D}^{n-1}_{B_{n-1}B'_{n-2}\to F_{n-1}B'_{n-1}})(\rho^{n-1}_{MA'_{n-1}B_{n-1}B'_{n-2}})]. \tag{17.1.49}$$

That is, the state $\zeta^n_{MB'_{n-1}A_n}$ is a particular state to consider in the optimization of the mutual information of a channel (with the channel input system being A_n and the external correlated systems being MB'_{n-1}), whereas the definition of the mutual information of a channel involves an optimization over all such states. This means that

$$I(MB'_{n-1}; B_n)_{\rho^n} \leq I(\mathcal{N}). \tag{17.1.50}$$

Putting together (17.1.43), (17.1.45), and (17.1.50), we find that

$$I(M;\widehat{M})_\omega \leq I(\mathcal{N}) + I(M; B'_{n-1})_{\rho^n}. \tag{17.1.51}$$

The quantity $I(M; B'_{n-1})_{\rho^n}$ can be bounded using steps analogous to the above. In particular, using the data-processing inequality for the mutual information with respect to the second-to-last decoding channel $\mathcal{D}^{n-1}_{B_{n-1}B'_{n-2}\to F_{n-1}B'_{n-1}}$, then employing the same steps as above, we conclude that

$$\begin{aligned} I(M; B'_{n-1})_{\rho^n} &\leq I(M; B_{n-1}B'_{n-2})_{\rho^{n-1}} && (17.1.52)\\ &= I(M; B_{n-1}|B'_{n-2})_{\rho^{n-1}} + I(M; B'_{n-2})_{\rho^{n-1}} && (17.1.53)\\ &\leq I(MB'_{n-2}; B_{n-1})_{\rho^{n-1}} + I(M; B'_{n-2})_{\rho^{n-1}} && (17.1.54)\\ &\leq I(\mathcal{N}) + I(M; B'_{n-2})_{\rho^{n-1}}, && (17.1.55) \end{aligned}$$

Overall, this leads to

$$I(M;\widehat{M})_\omega \leq 2I(\mathcal{N}) + I(M; B'_{n-2})_{\rho^{n-1}}. \tag{17.1.56}$$

Then, bounding $I(M; B'_{n-2})$ in the same manner as above, and continuing this process $n-3$ more times such that we completely "unwind" the protocol, we obtain

$$I(M;\widehat{M})_\omega \leq 2I(\mathcal{N}) + I(M; B'_{n-1})_{\rho^{n-1}} \tag{17.1.57}$$

$$\leq 3I(\mathcal{N}) + I(M; B'_{n-3})_{\rho^{n-2}} \tag{17.1.58}$$

$$\vdots \tag{17.1.59}$$

$$\leq nI(\mathcal{N}) + I(M; B'_0)_{\rho^1}. \tag{17.1.60}$$

However, from (17.1.6), we have that $\rho^1_{MB'_0} = \overline{\Phi}_M \otimes \Psi_{B'_0}$, which means that $I(M; B'_0)_{\rho^1} = 0$. Therefore, putting together (17.1.41), (17.1.42), and (17.1.57)–(17.1.60), we get

$$\log_2 |\mathcal{M}| \leq I^{\varepsilon}_H(M; \widehat{M})_{\omega} \tag{17.1.61}$$

$$\leq \frac{1}{1-\varepsilon}\left(I(M; \widehat{M})_{\omega} + h_2(\varepsilon)\right) \tag{17.1.62}$$

$$\leq \frac{1}{1-\varepsilon}\left(nI(\mathcal{N}) + h_2(\varepsilon)\right), \tag{17.1.63}$$

and the last line is equivalent to (17.1.35), as required.

We now establish the bound in (17.1.36). Combining (17.1.41) with Proposition 7.71, we conclude that the following bound holds for all $\alpha > 1$:

$$\log_2 |\mathcal{M}| \leq \widetilde{I}_{\alpha}(M; \widehat{M})_{\omega} + \frac{\alpha}{\alpha - 1}\log_2\left(\frac{1}{1-\varepsilon}\right). \tag{17.1.64}$$

Recall that the sandwiched Rényi mutual information $\widetilde{I}_{\alpha}(M; \widehat{M})_{\omega}$ is defined as

$$\widetilde{I}_{\alpha}(M; \widehat{M})_{\omega} = \inf_{\xi_{\widehat{M}}} \widetilde{D}_{\alpha}(\omega_{M\widehat{M}} \| \omega_M \otimes \xi_{\widehat{M}}) \tag{17.1.65}$$

$$= \inf_{\xi_{\widehat{M}}} \widetilde{D}_{\alpha}(\omega_{M\widehat{M}} \| \pi_M \otimes \xi_{\widehat{M}}). \tag{17.1.66}$$

Our goal now is to compare the actual protocol with one that results from employing a useless, replacement channel. To this end, let $\mathcal{R}^{\sigma_B}_{A\to B}$ be the replacement channel defined in (17.1.13), with σ_B an arbitrary (but fixed) state. Then as discussed in Section 17.1.1 (in particular, in (17.1.34)), the final state of the protocol conducted with the replacement channel is given by $\tau_{M\widehat{M}} = \pi_M \otimes \tau_{\widehat{M}}$. Then, we find that

$$\widetilde{I}_{\alpha}(M; \widehat{M})_{\omega} = \inf_{\xi_{\widehat{M}}} \widetilde{D}_{\alpha}(\omega_{M\widehat{M}} \| \pi_M \otimes \xi_{\widehat{M}}) \tag{17.1.67}$$

$$\leq \widetilde{D}_{\alpha}(\omega_{M\widehat{M}} \| \pi_M \otimes \tau_{\widehat{M}}) \tag{17.1.68}$$

$$= \widetilde{D}_{\alpha}(\omega_{M\widehat{M}} \| \tau_{M\widehat{M}}). \tag{17.1.69}$$

We now proceed with a similar method considered in the proof of the bound in (17.1.35), but using the sandwiched Rényi relative entropy as our main tool for analysis. By applying the data-processing inequality for the sandwiched Rényi relative entropy with respect to the last decoding channel, and using (17.1.33), we find that

$$\widetilde{D}_\alpha(\omega_{M\widehat{M}}\|\tau_{M\widehat{M}}) \leq \widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}}\|\tau^n_{MB_nB'_{n-1}}) \tag{17.1.70}$$

$$= \widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}}\|\pi_M \otimes \tau^n_{B'_{n-1}} \otimes \sigma_{B_n}) \tag{17.1.71}$$

$$= \frac{\alpha}{\alpha-1}\log_2 \widetilde{Q}_\alpha(\rho^n_{MB_nB'_{n-1}}\|\pi_M \otimes \tau^n_{B'_{n-1}} \otimes \sigma_{B_n})^{\frac{1}{\alpha}}, \tag{17.1.72}$$

where in the last line we used the definition in (7.5.2) of the sandwiched Rényi relative entropy. Now, recalling that $\rho^n_{MB_nB'_{n-1}} = \mathcal{N}_{A_n\to B_n}(\zeta^n_{MB'_{n-1}A_n})$ with the state $\zeta^{(n)}_{MB'_{n-1}A_n}$ defined in (17.1.49), and defining the positive semi-definite operator

$$X^{(\alpha)}_{MB'_{n-1}A_n} := \left(\pi_M \otimes \tau^n_{B'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \zeta^n_{MB'_{n-1}A_n} \left(\pi_M \otimes \tau^n_{B'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}, \tag{17.1.73}$$

as well as the completely positive map

$$\mathcal{S}^{(\alpha)}_{\sigma_B}(\cdot) := \sigma_{B_n}^{\frac{1-\alpha}{2\alpha}}(\cdot)\sigma_{B_n}^{\frac{1-\alpha}{2\alpha}}, \tag{17.1.74}$$

we use the definition of $\widetilde{Q}_\alpha$ in (7.5.3) to obtain

$$\widetilde{Q}_\alpha(\rho^n_{MB_nB'_{n-1}}\|\pi_M \otimes \tau^n_{B'_{n-1}} \otimes \sigma_{B_n})^{\frac{1}{\alpha}}$$

$$= \left\|\left(\pi_M \otimes \tau^n_{B'_{n-1}} \otimes \sigma_{B_n}\right)^{\frac{1-\alpha}{2\alpha}} \rho^n_{MB_nB'_{n-1}} \left(\pi_M \otimes \tau^n_{B'_{n-1}} \otimes \sigma_{B_n}\right)^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha \tag{17.1.75}$$

$$= \left\|\mathcal{S}^{(\alpha)}_{\sigma_B}\left(\left(\pi_M \otimes \tau^n_{B'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \mathcal{N}_{A_n\to B_n}(\zeta^n_{MB'_{n-1}A_n}) \left(\pi_M \otimes \tau^n_{B'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha \tag{17.1.76}$$

$$= \left\|(\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n\to B_n})(X^{(\alpha)}_{MB'_{n-1}A_n})\right\|_\alpha. \tag{17.1.77}$$

Multiplying and dividing by $\left\|X^{(\alpha)}_{MB'_{n-1}}\right\|_\alpha$ leads to

$$\left\|(\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n\to B_n})(X^{(\alpha)}_{MB'_{n-1}A_n})\right\|_\alpha$$

$$= \frac{\left\|(\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n\to B_n})(X^{(\alpha)}_{MB'_{n-1}A_n})\right\|_\alpha}{\left\|X^{(\alpha)}_{MB'_{n-1}}\right\|_\alpha}\left\|X^{(\alpha)}_{MB'_{n-1}}\right\|_\alpha \tag{17.1.78}$$

$$= \frac{\left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n \to B_n})(X^{(\alpha)}_{MB'_{n-1}A_n}) \right\|_\alpha}{\left\| X^{(\alpha)}_{MB'_{n-1}} \right\|_\alpha} \times \left\| \left(\pi_M^{\frac{1-\alpha}{2\alpha}} \otimes \left(\tau^n_{B'_{n-1}} \right)^{\frac{1-\alpha}{2\alpha}} \right) \zeta^n_{MB'_{n-1}} \left(\pi_M^{\frac{1-\alpha}{2\alpha}} \otimes \left(\tau^n_{B'_{n-1}} \right)^{\frac{1-\alpha}{2\alpha}} \right) \right\|_\alpha \tag{17.1.79}$$

$$= \frac{\left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n \to B_n})(X^{(\alpha)}_{MB'_{n-1}A_n}) \right\|_\alpha}{\left\| X^{(\alpha)}_{MB'_{n-1}} \right\|_\alpha} \cdot \widetilde{Q}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}})^{\frac{1}{\alpha}} \tag{17.1.80}$$

$$\leq \sup_{Y_{MB'_{n-1}A_n} \geq 0} \frac{\left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n \to B_n})(Y_{MB'_{n-1}A_n}) \right\|_\alpha}{\left\| Y_{MB'_{n-1}} \right\|_\alpha} \cdot \widetilde{Q}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}})^{\frac{1}{\alpha}} \tag{17.1.81}$$

$$= \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n \to B_n} \right\|_{\mathrm{CB},1\to\alpha} \cdot \widetilde{Q}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}})^{\frac{1}{\alpha}}, \tag{17.1.82}$$

where to obtain the inequality we performed an optimization with respect to all positive semi-definite operators $Y_{MB'_{n-1}A_n}$. To obtain the last line, we have used the norm $\|\cdot\|_{\mathrm{CB},1\to\alpha}$ defined in (11.2.68), which we show in Appendix 11.E can be written as

$$\|\mathcal{M}\|_{\mathrm{CB},1\to\alpha} = \sup_{Y_{RA} \geq 0} \frac{\|\mathcal{M}_{A\to B}(Y_{RA})\|_\alpha}{\|\mathrm{Tr}_A[Y_{RA}]\|_\alpha} \tag{17.1.83}$$

for any completely positive map $\mathcal{M}$. Plugging (17.1.82) back in to (17.1.72), we conclude the following bound:

$$\widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}} \| \tau^n_{MB_nB'_{n-1}}) \leq \frac{\alpha}{\alpha - 1} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_n \to B_n} \right\|_{\mathrm{CB},1\to\alpha} + \widetilde{D}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}}). \tag{17.1.84}$$

As in the proof of (17.1.35), we now iterate the above by successively bounding the sandwiched Rényi relative entropy terms $\widetilde{D}_\alpha(\zeta^i_{MB'_{i-1}} \| \pi_M \otimes \tau^i_{B'_{i-1}})$ for $i \in \{1, \dots, n\}$. Starting with the term $\widetilde{D}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}})$, we use the data-processing inequality for the sandwiched Rényi relative entropy under the second-to-last decoding channel $\mathcal{D}^{n-1}_{B_{n-1}B'_{n-2} \to F_{n-1}B'_{n-1}}$, then apply the same reasoning as in (17.1.75)–(17.1.82) to obtain

$$\widetilde{D}_\alpha(\zeta^n_{MB'_{n-1}} \| \pi_M \otimes \tau^n_{B'_{n-1}})$$

$$\leq \widetilde{D}_\alpha(\rho^{n-1}_{MB_{n-1}B'_{n-2}} \| \pi_M \otimes \tau^{n-1}_{B_{n-1}B'_{n-2}}) \tag{17.1.85}$$
$$= \widetilde{D}_\alpha(\rho^{n-1}_{MB_{n-1}B'_{n-2}} \| \pi_M \otimes \tau^{n-1}_{B'_{n-2}} \otimes \sigma_B) \tag{17.1.86}$$
$$\leq \frac{\alpha}{\alpha-1}\log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A_{n-1}\to B_{n-1}} \right\|_{\text{CB},1\to\alpha} + \widetilde{D}_\alpha(\zeta^{n-1}_{MB'_{n-2}} \| \pi_M \otimes \tau^{n-1}_{B'_{n-2}}). \tag{17.1.87}$$

Iterating this reasoning $n-2$ more times, we end up with the following bound:

$$\begin{aligned}
&\widetilde{D}_\alpha(\omega_{M\widehat{M}} \| \tau_{M\widehat{M}}) \\
&\leq n\frac{\alpha}{\alpha-1}\log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} + \widetilde{D}_\alpha(\rho^1_{MB'_0} \| \pi_M \otimes \Psi_{B'_0}) \\
&= n\frac{\alpha}{\alpha-1}\log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha},
\end{aligned} \tag{17.1.88}$$

where the equality holds because $\rho^1_{MB'_0} = \pi_M \otimes \Psi_{B'_0}$. Putting together (17.1.64), (17.1.69), and (17.1.88), we finally obtain

$$\log_2 |\mathcal{M}| \leq n\frac{\alpha}{\alpha-1}\log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{17.1.89}$$

Since we proved that this bound holds for any choice of the state σ_B, we conclude that

$$\begin{aligned}
&\log_2 |\mathcal{M}| \\
&\leq n\frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right) && (17.1.90)\\
&= n\widetilde{I}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right), && (17.1.91)
\end{aligned}$$

where the last equality follows from Lemma 11.20, and it implies (17.1.36). ■

17.1.3 The Amortized Perspective

In this section, we revisit the proofs above for Theorem 17.2 that establish bounds on non-asymptotic quantum feedback-assisted capacity. In particular, we adopt a different perspective, which we call the amortized perspective and which turns out to be useful in establishing bounds for all kinds of feedback-assisted protocols other than the ones considered in this chapter.

For the case of quantum feedback-assisted protocols, the amortized perspective consists of defining the amortized mutual information of a quantum channel as the largest net difference of the output and input mutual information that can be realized by the channel, when Alice and Bob are allowed to share an arbitrary state before the use of the channel. A key property of the amortized mutual information of a channel is that it is more readily seen to lead to an upper bound on the non-asymptotic quantum feedback-assisted capacity. Furthermore, another key property is that the amortized mutual information of a channel collapses to the usual mutual information of a channel, and so this leads to an alternative way of understanding the previous results. Furthermore, as indicated above, this perspective becomes quite useful in later chapters when we analyze LOCC-assisted quantum communication protocols and LOPC-assisted private communication protocols.

17.1.3.1 Quantum Mutual Information

We begin by defining the following key concept, the amortized mutual information of a quantum channel:

Definition 17.3 Amortized Mutual Information of a Channel

The *amortized mutual information of a quantum channel* $\mathcal{N}_{A\to B}$ is defined as

$$I^{\mathcal{A}}(\mathcal{N}) := \sup_{\rho_{A'AB'}} \left[I(A';BB')_\omega - I(A'A;B')_\rho \right], \tag{17.1.92}$$

where

$$\omega_{A'BB'} := \mathcal{N}_{A\to B}(\rho_{A'AB'}) \tag{17.1.93}$$

and the optimization is over states $\rho_{A'AB'}$.

Intuitively, the amortized mutual information is equal to the largest net mutual information that can be realized by the channel, if we allow Alice and Bob to share an arbitrary state before communication begins. As mentioned above, this concept turns out to be useful for understanding the feedback-assisted protocols presented previously.

We have the following simple relationship between mutual information and amortized mutual information:

Lemma 17.4

The mutual information of any channel $\mathcal{N}_{A\to B}$ does not exceed its amortized mutual information:

$$I(\mathcal{N}) \leq I^{\mathcal{A}}(\mathcal{N}). \tag{17.1.94}$$

Proof: Let us restrict the optimization in the definition of the amortized mutual information to states $\rho_{A'AB'}$ that have a trivial B' system. This means that $\rho_{A'AB'}$ is of the form $\rho_{A'AB'} = \rho_{A'A} \otimes |0\rangle\langle 0|_{B'}$. Therefore, $I(A'A;B')_\rho = 0$ and $I(A';BB')_\omega = I(A';B)_\omega$, where $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$, so that

$$I^{\mathcal{A}}(\mathcal{N}) = \sup_{\rho_{A'AB'}} [I(A';BB')_\omega - I(A'A;B)_\omega] \tag{17.1.95}$$

$$\geq \sup_{\rho_{A'A}} I(A';B)_\omega \tag{17.1.96}$$

$$= I(\mathcal{N}), \tag{17.1.97}$$

i.e., $I^{\mathcal{A}}(\mathcal{N}) \geq I(\mathcal{N})$, as required. ■

An important question is whether the opposite inequality, i.e., $I^{\mathcal{A}} \leq I(\mathcal{N})$, holds, which would allow us to conclude that $I^{\mathcal{A}} = I(\mathcal{N})$. In this case, we find that it does. Specifically, we have the following.

Proposition 17.5

Given an arbitrary quantum channel $\mathcal{N}$, amortization does not increase its mutual information:

$$I(\mathcal{N}) = I^{\mathcal{A}}(\mathcal{N}). \tag{17.1.98}$$

Proof: To see this, consider that for an arbitrary input state $\rho_{A'AB'}$, we can use the chain rule for mutual information in (7.2.112) twice to obtain

$$I(A';BB')_\omega - I(A'A;B')_\rho$$

$$= I(A';B)_\omega + I(A';B'|B)_\omega - I(A'A;B')_\rho \tag{17.1.99}$$

$$\leq I(A';B)_\omega + I(A'B;B')_\omega - I(A'A;B')_\rho. \tag{17.1.100}$$

In particular, to obtain the third line, note that (7.2.112) implies

$$I(A'B;B')_\omega = I(B;B')_\omega + I(A';B'|B)_\omega \geq I(A';B'|B)_\omega \tag{17.1.101}$$

since $I(B;B')_\omega \geq 0$. Continuing, we apply the data-processing inequality for the mutual information under the channel $\mathcal{N}$, which implies that $I(A'A;B)_\rho \geq I(A'B;B')_\omega$. Therefore,

$$I(A';BB')_\omega - I(A'A;B')_\rho \tag{17.1.102}$$
$$\leq I(A';B)_\omega + I(A'B;B')_\omega - I(A'B;B')_\omega \tag{17.1.103}$$
$$= I(A';B)_\omega \tag{17.1.104}$$
$$\leq I(\mathcal{N}), \tag{17.1.105}$$

where the last line follows because the state $\omega_{A'B} = \mathcal{N}_{A\to B}(\rho_{A'A})$ has the form of states that we consider when performing the optimization in the definition of the mutual information of a channel. Since the inequality

$$I(A';BB')_\omega - I(A'A;B')_\rho \leq I(\mathcal{N}) \tag{17.1.106}$$

holds for an arbitrary input state $\rho_{A'AB'}$, we conclude the bound in (17.1.98). ■

We note here that the equality in (17.1.98) is stronger than the additivity of mutual information shown in Chapter 11 (in particular, that shown in Theorem 11.19). Indeed, the equality in (17.1.98) actually implies the additivity relation discussed previously. To see this, consider that the equality in (17.1.98) implies that

$$I(A';BB')_\omega - I(A'A;B')_\rho \leq I(\mathcal{N}) \tag{17.1.107}$$

for an arbitrary input state $\rho_{A'AB'}$, where $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$. Now let $\rho_{A'AB'} = \mathcal{M}_{A''\to B'}(\sigma_{A'AA''})$ for some channel $\mathcal{M}_{A''\to B'}$ and some state $\sigma_{A'AA''}$. Then it follows that

$$\omega_{A'BB'} = (\mathcal{N}_{A\to B} \otimes \mathcal{M}_{A''\to B'})(\sigma_{A'AA''}), \tag{17.1.108}$$

and applying (17.1.107), we have that

$$I(A';BB')_\omega \leq I(\mathcal{N}) + I(A'A;B')_\rho \tag{17.1.109}$$
$$\leq I(\mathcal{N}) + \sup_{\sigma_{A'AA''}} I(A'A;B')_\rho \tag{17.1.110}$$
$$= I(\mathcal{N}) + I(\mathcal{M}), \tag{17.1.111}$$

where the inequality follows because the state $\sigma_{A'AA''}$ is a particular state to consider for the optimization in the definition of the mutual information of the channel

$\mathcal{M}_{A''\to B'}$. Since the inequality holds for all input states $\sigma_{A'AA''}$ to $\mathcal{N}_{A\to B}\otimes\mathcal{M}_{A''\to B'}$, we conclude that

$$I(\mathcal{N}\otimes\mathcal{M}) \leq I(\mathcal{N}) + I(\mathcal{M}), \tag{17.1.112}$$

which is the non-trivial inequality needed in the proof of the additivity of the mutual information of a channel (see the proof of Theorem 11.19).

How is the amortized mutual information relevant for analyzing a feedback-assisted protocol? Consider that the bound in (17.1.43) involves the mutual information $I(M;B_nB'_{n-1})_{\rho^n}$, so that

$$\begin{aligned}
&I(M;B_nB'_{n-1})_{\rho^n}\\
&= I(M;B_nB'_{n-1})_{\rho^n} - I(M;B'_0)_{\rho^1} &(17.1.113)\\
&= I(M;B_nB'_{n-1})_{\rho^n} - I(M;B'_0)_{\rho^1} + \sum_{i=1}^{n-1} I(M;B'_i)_{\rho^i} - I(M;B'_i)_{\rho^i} &(17.1.114)\\
&\leq I(M;B_nB'_{n-1})_{\rho^n} - I(M;B'_0)_{\rho^1} + \sum_{i=1}^{n-1} I(M;B_iB'_{i-1})_{\rho^i} - I(M;B'_i)_{\rho^i} &(17.1.115)\\
&= \sum_{i=1}^{n} I(M;B_iB'_{i-1})_{\rho^i} - I(M;B'_{i-1})_{\rho^i} &(17.1.116)\\
&\leq n\cdot \sup_{\rho_{A'AB'}} I(A';BB')_\omega - I(A'A;B')_\rho &(17.1.117)\\
&= n\cdot I^{\mathcal{A}}(\mathcal{N}) = n\cdot I(\mathcal{N}). &(17.1.118)
\end{aligned}$$

The first equality follows because the state $\rho^1_{MB'_0}$ is a product state. The second equality follows by adding and subtracting the mutual information of the state of the message system M and Bob's memory system B'_i. The inequality is a consequence of data processing under the action of the decoding channels. The third equality follows from collecting terms. The final inequality follows because the state $\rho^i_{MB_iB'_{i-1}}$ is a particular state to consider in the optimization of the amortized mutual information, and the final equality follows from the amortization collapse in Proposition 17.5.

Thus, we observe that the bound in (17.1.35), at a fundamental level, is a consequence of the amortization collapse from Proposition 17.5.

17.1.3.2 Sandwiched Rényi Mutual Information

We can also consider the concept of amortization for the sandwiched Rényi mutual information, and in this subsection, we revisit the bound in (17.1.36) to understand it from this perspective.

Definition 17.6 Amortized Sandwiched Mutual Information

The amortized sandwiched Rényi mutual information of a quantum channel $\mathcal{N}_{A\to B}$ is defined for $\alpha \in (0,1) \cup (1,\infty)$ as follows:

$$\widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}) := \sup_{\rho_{A'AB'}} \left[\widetilde{I}_\alpha(A';BB')_\omega - \widetilde{I}_\alpha(A'A;B')_\rho\right], \tag{17.1.119}$$

where $\omega_{A'BB'} := \mathcal{N}_{A\to B}(\rho_{A'AB'})$ and the optimization is over states $\rho_{A'AB'}$.

Just as with the mutual information of a channel, we find that for all $\alpha \in (0,1) \cup (1,\infty)$,

$$\widetilde{I}_\alpha(\mathcal{N}) \leq \widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}), \tag{17.1.120}$$

and the proof of this analogous to the proof of Lemma 17.4, which establishes the corresponding inequality for the mutual information. So the question is to determine whether the opposite inequality holds. Indeed, we find again that it is the case, at least for $\alpha > 1$.

Proposition 17.7

Amortization does not increase the sandwiched Rényi mutual information of a quantum channel $\mathcal{N}$ for all $\alpha > 1$:

$$\widetilde{I}_\alpha(\mathcal{N}) = \widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}). \tag{17.1.121}$$

Proof: Let $\rho_{A'AB'}$ be an arbitrary input state, and let σ_B and $\tau_{B'}$ be arbitrary states. Then, letting $\omega_{A'BB'} = \mathcal{N}_{A\to B}(\rho_{A'AB'})$, we find that

$$\widetilde{I}_\alpha(A';BB')_\omega$$
$$= \inf_{\xi_{BB'}} \widetilde{D}_\alpha(\omega_{A'BB'} \| \omega_{A'} \otimes \xi_{BB'}) \tag{17.1.122}$$
$$\leq \widetilde{D}_\alpha(\omega_{A'BB'} \| \omega_{A'} \otimes \sigma_B \otimes \tau_{B'}) \tag{17.1.123}$$

$$= \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{A'AB'}) \| \rho_{A'} \otimes \sigma_B \otimes \tau_{B'}) \tag{17.1.124}$$
$$= \frac{\alpha}{\alpha-1} \log_2 \left\| (\rho_{A'} \otimes \sigma_B \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} \mathcal{N}_{A\to B}(\rho_{A'AB'}) \, (\rho_{A'} \otimes \sigma_B \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha , \tag{17.1.125}$$

where to obtain the last equality we used the alternate expression in (7.5.3) for the sandwiched Rényi relative entropy. Defining

$$X^{(\alpha)}_{A'AB'} := (\rho_{A'} \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} \rho_{A'AB'} \, (\rho_{A'} \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} , \tag{17.1.126}$$

and making use of the completely positive map $\mathcal{S}^{(\alpha)}_{\sigma_B}$ from (17.1.74), we find that

$$\left\| (\rho_{A'} \otimes \sigma_B \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} \mathcal{N}_{A\to B}(\rho_{A'AB'}) \, (\rho_{A'} \otimes \sigma_B \otimes \tau_{B'})^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha$$
$$= \left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B})(X^{(\alpha)}_{A'AB'}) \right\|_\alpha \tag{17.1.127}$$
$$= \frac{\left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B})(X^{(\alpha)}_{A'AB'}) \right\|_\alpha}{\left\| X^{(\alpha)}_{A'B'} \right\|_\alpha} \left\| X^{(\alpha)}_{A'B'} \right\|_\alpha \tag{17.1.128}$$
$$\leq \sup_{Y_{A'AB'} \geq 0} \frac{\left\| (\mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B})(Y_{A'AB'}) \right\|_\alpha}{\| Y_{A'B'} \|_\alpha} \times \left\| [\rho_{A'} \otimes \tau_{B'}]^{\frac{1-\alpha}{2\alpha}} \rho_{A'B'} \, [\rho_{A'} \otimes \tau_{B'}]^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \tag{17.1.129}$$
$$= \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} \cdot \widetilde{Q}_\alpha(\rho_{A'B'} \| \rho_{A'} \otimes \tau_{B'})^{\frac{1}{\alpha}}, \tag{17.1.130}$$

where in the last line we have used the expression in (11.E.1) for the norm $\|\cdot\|_{\text{CB},1\to\alpha}$. Plugging (17.1.130) back into (17.1.125), we find that

$$\widetilde{I}_\alpha(A';BB')_\omega \leq \frac{\alpha}{\alpha-1} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} + \widetilde{D}_\alpha(\rho_{A'B'} \| \rho_{A'} \otimes \tau_{B'}). \tag{17.1.131}$$

Since the inequality holds for arbitrary states σ_B and $\tau_{B'}$, we conclude that

$$\widetilde{I}_\alpha(A';BB')_\omega$$
$$\leq \frac{\alpha}{\alpha-1} \inf_{\sigma_B} \log_2 \left\| \mathcal{S}^{(\alpha)}_{\sigma_B} \circ \mathcal{N}_{A\to B} \right\|_{\text{CB},1\to\alpha} + \inf_{\tau_{B'}} \widetilde{D}_\alpha(\rho_{A'B'} \| \rho_{A'} \otimes \tau_{B'}) \tag{17.1.132}$$
$$= \widetilde{I}_\alpha(\mathcal{N}) + \widetilde{I}_\alpha(A';B')_\rho \tag{17.1.133}$$

$$\leq \widetilde{I}_\alpha(\mathcal{N}) + \widetilde{I}_\alpha(A'A;B')_\rho, \tag{17.1.134}$$

where the equality follows from Lemma 11.20 and the final inequality from the data-processing inequality for the mutual information under the partial trace Tr_A. Since we have shown that the following inequality holds for an arbitrary input state $\rho_{A'AB'}$:

$$\widetilde{I}_\alpha(A';BB')_\omega - \widetilde{I}_\alpha(A'A;B')_\rho \leq \widetilde{I}_\alpha(\mathcal{N}), \tag{17.1.135}$$

we conclude that $\widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}) \leq \widetilde{I}_\alpha(\mathcal{N})$, which leads to $\widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}) = \widetilde{I}_\alpha(\mathcal{N})$ after combining with (17.1.120). ■

By following exactly the same steps in (17.1.107)–(17.1.112), but replacing I with $\widetilde{I}_\alpha$, we can conclude that the amortization collapse in Proposition 17.7 implies the additivity relation (Theorem 11.22) for sandwiched Rényi mutual information of quantum channels for all $\alpha > 1$:

$$\widetilde{I}_\alpha(\mathcal{N}) = \widetilde{I}_\alpha^{\mathcal{A}}(\mathcal{N}) \quad \forall\, \alpha > 1 \quad \Longrightarrow \quad \widetilde{I}_\alpha(\mathcal{N}\otimes\mathcal{M}) \leq \widetilde{I}_\alpha(\mathcal{N}) + \widetilde{I}_\alpha(\mathcal{M}), \tag{17.1.136}$$

where $\mathcal{N}$ and $\mathcal{M}$ are quantum channels. Furthermore, by following exactly the same steps as in (17.1.113)–(17.1.118), but replacing I with $\widetilde{I}_\alpha$, and employing Proposition 17.7, we conclude the following bound

$$\widetilde{I}_\alpha(M;B_nB'_{n-1})_{\rho^n} \leq n \cdot \widetilde{I}_\alpha(\mathcal{N}), \tag{17.1.137}$$

which in turn implies the bound in (17.1.36). Thus, we can alternatively analyze feedback-assisted protocols and arrive at the bound in (17.1.36) by utilizing the concept of amortization.

17.2 Quantum Feedback-Assisted Classical Capacity of a Quantum Channel

In this section, we analyze the asymptotic case, in which we allow for an arbitrarily large number of rounds of feedback. This task is now rather straightforward, given the bounds that we have established in the previous section. So we keep this section brief, only stating some definitions and then some theorems that follow as a direct consequence of definitions and developments in previous chapters.

Definition 17.8 Achievable Rate for Quantum-Feedback-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an achievable rate for quantum-feedback-assisted classical communication over $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, all $\delta > 0$, and all sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ quantum-feedback-assisted classical communication protocol.

Definition 17.9 Quantum-Feedback-Assisted Classical Capacity of a Quantum Channel

The quantum-feedback-assisted classical capacity of a quantum channel $\mathcal{N}$, denoted by $C_{\text{QFB}}(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$C_{\text{QFB}}(\mathcal{N}) \coloneqq \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{17.2.1}$$

Definition 17.10 Strong Converse Rate for Quantum-Feedback-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a strong converse rate for quantum-feedback-assisted classical communication over $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, all $\delta > 0$, and all sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ quantum-feedback-assisted classical communication protocol.

Definition 17.11 Strong Converse Quantum-Feedback-Assisted Classical Capacity of a Quantum Channel

The strong converse quantum-feedback-assisted classical capacity of a quantum channel $\mathcal{N}$, denoted by $\widetilde{C}_{\text{QFB}}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{C}_{\text{QFB}}(\mathcal{N}) \coloneqq \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{17.2.2}$$

The main result of this section is the following capacity theorem:

Theorem 17.12 Quantum-Feedback-Assisted Classical Capacity

For any quantum channel $\mathcal{N}$, its quantum-feedback-assisted classical capacity $C_{\text{QFB}}(\mathcal{N})$ and its strong converse quantum-feedback-assisted classical capacity are both equal to its mutual information $I(\mathcal{N})$, i.e.,

$$C_{\text{QFB}}(\mathcal{N}) = \widetilde{C}_{\text{QFB}}(\mathcal{N}) = I(\mathcal{N}), \tag{17.2.3}$$

where $I(\mathcal{N})$ is defined in (7.11.102).

Proof: By previous reasoning, we have that

$$C_{\text{QFB}}(\mathcal{N}) \leq \widetilde{C}_{\text{QFB}}(\mathcal{N}), \tag{17.2.4}$$

and by Theorem 11.16 and the fact that any entanglement-assisted classical communication protocol is a particular kind of quantum-feedback-assisted classical communication protocol, we have that

$$I(\mathcal{N}) \leq C_{\text{QFB}}(\mathcal{N}) \leq \widetilde{C}_{\text{QFB}}(\mathcal{N}). \tag{17.2.5}$$

The upper bound $\widetilde{C}_{\text{QFB}}(\mathcal{N}) \leq I(\mathcal{N})$ follows from (17.1.36) and the same reasoning given in the proof detailed in Section 11.2.3. ■

17.3 Bibliographic Notes

Shannon (1956) proved that feedback does not increase the classical capacity of a classical channel (his result is a weak-converse bound). The strong converse for the feedback-assisted classical capacity of a classical channel was established independently by Kemperman and Kesten. Kesten's proof appeared in (Wolfowitz, 1964, Chapter 4) and Kemperman's proof appeared later in (Kemperman, 1971). See (Ulrey, 1976) for a discussion of this history. Polyanskiy and Verdú (2010) employed a Rényi-entropic method to extend Shannon's result to a strong converse statement, i.e., that the mutual information of a classical channel is equal to the strong converse feedback-assisted classical capacity.

Bowen (2004) proved that a quantum feedback channel does not increase the entanglement-assisted classical capacity of a quantum channel (his result is a weak-converse bound). Bennett et al. (2014) proved that the mutual information of a quantum channel is equal to the strong converse quantum-feedback-assisted classical capacity. Their approach was to employ the quantum reverse Shannon theorem to do so. Cooney et al. (2016) used a Rényi-entropic method to prove this same result, and this is the approach that we have followed in this book. As far as we are aware, the concept of amortized mutual information of a quantum channel and the fact that it reduces to the mutual information of a quantum channel are original to this book.

Chapter 18

Classical-Feedback-Assisted Communication

In this chapter, we continue with our study of feedback-assisted capacities. The class of protocols that we consider in this chapter are very similar to those from the previous chapter (Chapter 17), with the exception that the feedback channel is a classical channel instead of a quantum channel. The resulting communication task is then called classical communication assisted by a classical channel (or classical-feedback-assisted communication for short).

Interestingly, this slight change has the effect of complicating the theory quite a bit: a general expression for the capacity is not known. It is only known for certain channels such as entanglement-breaking channels and erasure channels. Additionally, there are examples of channels for which classical feedback can increase the classical capacity significantly, due to the interplay between classical feedback and entanglement that can be generated by the channel. We do not discuss this example in this chapter and instead point to the Bibliographic Notes for details (Section 18.7). All of the above implies that the increase of capacity due to classical feedback is a truly quantum-mechanical phenomenon that separates the classical and quantum theories of communication. Indeed, it is necessary for a channel to have the ability to generate entanglement in order for classical feedback to give a boost to capacity.

Our main focus in this chapter is on establishing upper bounds on the classical-feedback-assisted capacity. First, we prove that classical feedback does not increase the capacity of entanglement-breaking channels. The main tools here are similar to those employed in Section 12.2.3.1. Next, we establish that the average output entropy of a channel is an upper bound on the feedback-assisted capacity. Finally, we establish that the Υ-information of a channel, introduced in Section 12.2.5.1, is actually an upper bound on the feedback-assisted capacity. We close out the chapter by discussing some example channels and summarizing the main concepts presented.

18.1 *n*-Shot Classical Feedback-Assisted Communication Protocols

In this section, we briefly summarize what is meant by an n-shot protocol for classical communication assisted by a classical feedback channel, where $n \in \mathbb{N}$. This section is brief because such a protocol is defined exactly as in Section 17.1, with the exception that every feedback channel acting on system F_i, sent from the receiver to the sender, for all $i \in \{0, \ldots, n\}$, is a classical feedback channel of the following form:

$$\overline{\Delta}_{F_i}(\rho_{F_i}) \coloneqq \sum_{j=0}^{d_{F_i}-1} |j\rangle\langle j|_{F_i} \rho_{F_i} |j\rangle\langle j|_{F_i}, \tag{18.1.1}$$

where $\{|j\rangle\}_{j=0}^{d_{F_i}-1}$ is some orthonormal basis known to both the sender and the receiver.

In short, every such protocol has the form given in Figure 17.1, with the aforementioned exception that every channel acting on F_i, for $i \in \{0, \ldots, n\}$, is a classical feedback channel as in (18.1.1). Every such protocol is defined by the following elements:

$$(\mathcal{M}, \Psi_{F_0 B_0'}, \mathcal{E}^0_{M' F_0 \to A_1' A_1}, \{\mathcal{E}^i_{A_i' F_i \to A_{i+1}' A_{i+1}}, \mathcal{D}^i_{B_i B_{i-1}' \to F_i B_i'}\}_{i=1}^{n-1}, \mathcal{D}^n_{B_n B_{n-1}' \to \hat{M}}), \tag{18.1.2}$$

where $\mathcal{M}$ is the message set, $\Psi_{F_0 B_0'}$ is a bipartite quantum state, the objects denoted by $\mathcal{E}$ are encoding channels, and those denoted by $\mathcal{D}$ decoding channels. Let $\mathcal{C}$ denote all of these elements, which together constitute the classical-feedback-assisted code. The systems labeled by F are feedback systems, the ith of which is sent by the receiver Bob to the sender Alice through the classical feedback channel in (18.1.1). The initial state of such a protocol, prepared by Alice, is $\overline{\Phi}^p_{MM'}$, as defined in (17.1.2). The states throughout the protocol are the same as defined in

Section 17.1, with the exception that every state with an F label is replaced by the same state succeeded by the completely dephasing channel $\overline{\Delta}_{F_i}$. That is, the initial state is

$$\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0 B'_0}), \tag{18.1.3}$$

and the other states are

$$\rho^1_{MA'_1 B_1 B'_0} \coloneqq (\mathcal{N}_{A_1 \to B_1} \circ \mathcal{E}^0_{M'F_0 \to A'_1 A_1})(\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0 B'_0})), \tag{18.1.4}$$

$$\rho^2_{MA'_2 B_2 B'_1} \coloneqq (\mathcal{N}_{A_2 \to B_2} \circ \mathcal{E}^1_{A'_1 F_1 \to A'_2 A_2} \circ \overline{\Delta}_{F_1} \circ \mathcal{D}^1_{B_1 B'_0 \to F_1 B'_1})(\rho^1_{MA'_1 B_1 B'_0}), \tag{18.1.5}$$

$$\rho^i_{MA'_i B_i B'_{i-1}} \coloneqq (\mathcal{N}_{A_i \to B_i} \circ \mathcal{E}^{i-1}_{A'_{i-1} F_{i-1} \to A'_i A_i} \circ \overline{\Delta}_{F_{i-1}} \circ \mathcal{D}^{i-1}_{B_{i-1} B'_{i-2} \to F_{i-1} B'_{i-1}})(\rho^2_{MA'_2 B_2 B'_1}), \tag{18.1.6}$$

where $i \in \{3, \ldots, n\}$. The final state of the protocol is then as follows:

$$\omega^p_{M\hat{M}} \coloneqq \mathcal{D}^n_{B_n B'_{n-1} \to \hat{M}}(\mathrm{Tr}_{A'_n}[\rho^n_{MA'_n B_n B'_{n-1}}]). \tag{18.1.7}$$

Consider that the initial state of the protocol, as given in (18.1.3), has the following form:

$$\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0 B'_0}) = \sum_{m \in \mathcal{M}} p(m)|m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'} \otimes \sum_{f_0} p(f_0)|f_0\rangle\langle f_0|_{F_0} \otimes \Psi^{f_0}_{B'_0}, \tag{18.1.8}$$

where $p(f_0)$ is a probability distribution over the possible classical values sent through the feedback channel and each $\Psi^{f_0}_{B'_0}$ is a state of the system B'_0. After the encoding channel $\mathcal{E}^0_{M'F_0 \to A'_1 A_1}$ acts, the state becomes as follows:

$$\mathcal{E}^0_{M'F_0 \to A'_1 A_1}(\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0 B'_0})) = \sum_{m \in \mathcal{M}} \sum_{f_0} p(m) p(f_0) |m\rangle\langle m|_M \otimes \varsigma^{0,m,f_0}_{A'_1 A_1} \otimes \Psi^{f_0}_{B'_0}, \tag{18.1.9}$$

where the state $\varsigma^{0,m,f_0}_{A'_1 A_1}$ is defined as

$$\varsigma^{0,m,f_0}_{A'_1 A_1} \coloneqq \mathcal{E}^0_{M'F_0 \to A'_1 A_1}(|m\rangle\langle m|_{M'} \otimes |f_0\rangle\langle f_0|_{F_0}). \tag{18.1.10}$$

Then one can proceed from here, defining states of the protocol conditioned on the value of the message and the classical feedback.

Just as we did in Chapter 11, we define the message error probability, average error probability, and maximal error probability, as in (11.1.13), (11.1.14), and (11.1.15), respectively. Using the expression in (11.1.24), the average error probability for the classical-feedback-assisted code $\mathcal{C}$ is given by

$$\overline{p}_{\text{err}}(\mathcal{C}; p) := \frac{1}{2} \left\| \overline{\Phi}^p_{M\hat{M}} - \omega^p_{M\hat{M}} \right\|_1, \tag{18.1.11}$$

and using the expression in (11.1.36), the maximal error probability for the classical-feedback-assisted code $\mathcal{C}$ is given by

$$p^*_{\text{err}}(\mathcal{C}) := \max_{p:\mathcal{M}\to[0,1]} \frac{1}{2} \left\| \overline{\Phi}^p_{M\hat{M}} - \omega^p_{M\hat{M}} \right\|_1, \tag{18.1.12}$$

where the maximization is over every probability distribution p.

Definition 18.1 **$(n, |\mathcal{M}|, \varepsilon)$ Classical-Feedback-Assisted Classical Communication Protocol**

Let $(\mathcal{M}, \Psi_{F_0B'_0}, \mathcal{E}^0_{M'F_0\to A'_1A_1}, \{\mathcal{E}^i_{A'_iF_i\to A'_{i+1}A_{i+1}}, \mathcal{D}^i_{B_iB'_{i-1}\to F_iB'_i}\}_{i=1}^n, \mathcal{D}^n_{B_nB'_{n-1}\to\hat{M}})$ be the elements of an n-shot classical-feedback-assisted classical communication protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an $(n, |\mathcal{M}|, \varepsilon)$ protocol, with $\varepsilon \in [0, 1]$, if $p^*_{\text{err}}(\mathcal{C}) \leq \varepsilon$.

18.2 Protocol over a Useless Channel

A common theme in this book has been that we can derive converse bounds by using a generalized divergence to compare the output of the actual protocol with one that is useless for the task. We did exactly this in Section 17.1.1 of the previous chapter, and the only change that we make here, as in the previous section, is to replace every quantum feedback channel with a classical one. Figure 17.2 applies again and we briefly define the steps and states exactly as before, except with this key difference for both the figure and the states involved. A useless channel is one that traces out the input and replaces it with some state at the output:

$$\mathcal{R}_{A\to B} := \mathcal{P}_{\sigma_B} \circ \text{Tr}_A, \tag{18.2.1}$$

where $\mathcal{P}_{\sigma_B}$ denotes a preparation channel that prepares the state σ_B at the output. The initial state of this protocol is

$$\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0 B'_0}), \tag{18.2.2}$$

and the others are as follows:

$$\tau^1_{MA'_1B_1B'_0} := (\mathcal{R}_{A_1\to B_1} \circ \mathcal{E}^0_{M'F_0\to A'_1A_1})(\overline{\Phi}^p_{MM'} \otimes \overline{\Delta}_{F_0}(\Psi_{F_0B'_0})), \tag{18.2.3}$$

$$\tau^2_{MA'_2B_2B'_1} := (\mathcal{R}_{A_2\to B_2} \circ \mathcal{E}^1_{A'_1F_1\to A'_2A_2} \circ \overline{\Delta}_{F_1} \circ \mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\rho^1_{MA'_1B_1B'_0}), \tag{18.2.4}$$

$$\tau^i_{MA'_iB_iB'_{i-1}} := (\mathcal{R}_{A_i\to B_i} \circ \mathcal{E}^{i-1}_{A'_{i-1}F_{i-1}\to A'_iA_i} \circ \overline{\Delta}_{F_{i-1}} \circ \mathcal{D}^{i-1}_{B_{i-1}B'_{i-2}\to F_{i-1}B'_{i-1}})(\rho^2_{MA'_2B_2B'_1}), \tag{18.2.5}$$

where $i \in \{3,\ldots,n\}$. The final state of the protocol is then as follows:

$$\omega^p_{M\hat{M}} := \mathcal{D}^n_{B_nB'_{n-1}\to\hat{M}}(\mathrm{Tr}_{A'_n}[\rho^n_{MA'_nB_nB'_{n-1}}]). \tag{18.2.6}$$

Going through calculations similar to those in (17.1.15)–(17.1.33), we arrive at the following conclusions:

$$\tau^1_{MA'_1B_1B'_0} = \tau^1_{MA'_1B'_0} \otimes \sigma_{B_1}, \tag{18.2.7}$$

$$\tau^1_{MB_1B'_0} = \pi^p_M \otimes \Psi_{B'_0} \otimes \sigma_{B_1}, \tag{18.2.8}$$

$$\tau^2_{MB_2B'_1} = \pi^p_M \otimes \tau^2_{B'_1} \otimes \sigma_{B_2}, \tag{18.2.9}$$

$$\tau^i_{MB_iB'_{i-1}} = \pi^p_M \otimes \tau^i_{B'_{i-1}} \otimes \sigma_{B_i}, \tag{18.2.10}$$

where $i \in \{3,\ldots,n\}$ and

$$\pi^p_M := \sum_{m\in\mathcal{M}} p(m)|m\rangle\langle m|_M. \tag{18.2.11}$$

Thus, there is no correlation whatsoever between the message system M and Bob's systems $B_iB'_{i-1}$, for each $i \in \{1,\ldots,n\}$, after tracing over Alice's systems. As before, this is a consequence of the fact that the "communication line has been cut" when employing the replacement channel.

Bob's final decoding channel $\mathcal{D}^n_{B_nB'_{n-1}\to\hat{M}}$ thus leads to the following classical–classical state:

$$\tau_{M\hat{M}} := \pi^p_M \otimes \tau_{\hat{M}}, \tag{18.2.12}$$

where $\tau_{\hat{M}} := \sum_{\hat{m}\in\mathcal{M}} t(\hat{m})|\hat{m}\rangle\langle\hat{m}|_{\hat{M}}$, for some probability distribution $t : \mathcal{M} \to [0,1]$, which corresponds to Bob's measurement.

18.3 Upper Bounds on the Number of Transmitted Bits

We now provide several upper bounds on the number of transmitted bits for classical communication protocols assisted by classical feedback. The first bounds that we discuss apply only to entanglement-breaking channels (recall Definition 4.12), and they imply that classical feedback increases neither the asymptotic classical capacity of entanglement-breaking channels nor their strong converse classical capacity. Then the next two bounds apply to all quantum channels, and they are known as the entropy bound and the geometric Υ-information bound.

18.3.1 Upper Bounds for Entanglement-Breaking Channels

Before stating the main theorem of this section, we discuss particular aspects of a classical-feedback-assisted protocol for classical communication over an entanglement-breaking channel. Indeed, suppose that $\mathcal{N}_{A\to B}$ is an entanglement-breaking channel. We begin our analysis by inspecting the state in (18.1.9). This state is fully separable with respect to the cut $M : A_1'A_1 : B_0'$. That is, it can be written as follows:

$$\sum_z q(z)\tau_M^z \otimes \sigma_{A_1'A_1}^z \otimes \omega_{B_0}^z, \tag{18.3.1}$$

for q a probability distribution and $\{\tau_M^z\}_z$, $\{\sigma_{A_1'A_1}^z\}_z$, and $\{\omega_{B_0}^z\}_z$ sets of states. Since the channel $\mathcal{N}_{A\to B}$ is entanglement breaking, when it acts on system A_1 of the state $\varsigma_{A_1'A_1}^{0,m,f_0}$ in (18.1.9), the resulting state is a separable state of the following form:

$$\mathcal{N}_{A_1\to B_1}(\varsigma_{A_1'A_1}^{0,m,f_0}) = \sum_y p(y|m,f_0)\varsigma_{A_1'}^{y,m,f_0} \otimes \varsigma_{B_1}^{y,m,f_0}. \tag{18.3.2}$$

So this implies that the state $\rho_{MA_1'B_1B_0'}^1$, as defined in (18.1.4) and with $\mathcal{N}_{A_1\to B_1}$ entanglement breaking, is fully separable across all systems (i.e., with respect to the cut $M : A_1' : B_1 : B_0'$).

Bob then applies the decoding channel $\overline{\Delta}_{F_1} \circ \mathcal{D}_{B_1B_0'\to F_1B_1'}^1$ and the state at this point is as follows:

$$(\overline{\Delta}_{F_1} \circ \mathcal{D}_{B_1B_0'\to F_1B_1'}^1)(\rho_{MA_1'B_1B_0'}^1) =$$

$$\sum_{m\in\mathcal{M}}\sum_{f_0,y} p(y|m,f_0)p(m)p(f_0)|m\rangle\langle m|_M\otimes$$
$$\varsigma^{y,m,f_0}_{A'_1}\otimes(\overline{\Delta}_{F_1}\circ\mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\varsigma^{y,m,f_0}_{B_1}\otimes\Psi^{f_0}_{B'_0}). \quad (18.3.3)$$

Since the F_1 system is classical, the state $(\overline{\Delta}_{F_1}\circ\mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\varsigma^{y,m,f_0}_{B_1}\otimes\Psi^{f_0}_{B'_0})$ can be written as

$$(\overline{\Delta}_{F_1}\circ\mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\varsigma^{y,m,f_0}_{B_1}\otimes\Psi^{f_0}_{B'_0})=\sum_{f_1}p(f_1|y,m,f_0)|f_1\rangle\langle f_1|_{F_1}\otimes\varsigma^{f_1,y,m,f_0}_{B'_1}. \quad (18.3.4)$$

This means that the state $(\overline{\Delta}_{F_1}\circ\mathcal{D}^1_{B_1B'_0\to F_1B'_1})(\rho^1_{MA'_1B_1B'_0})$ is fully separable with respect to the cut $M : A'_1 : F_1 : B'_1$.

This process continues, and since the channel $\mathcal{N}_{A\to B}$ is entanglement breaking, by following an analysis similar to that given above, we observe that the state of the message system M, Alice's, and Bob's is always fully separable throughout the protocol. This is the key reason that we obtain the bounds given in the following theorem:

Theorem 18.2 ***n*-Shot Upper Bounds for Classical Feedback Assisted Classical Communication over Entanglement Breaking Channels**

Let $\mathcal{N}_{A\to B}$ be an entanglement-breaking channel, and let $\varepsilon\in[0,1)$. For all $(n,|\mathcal{M}|,\varepsilon)$ classical-feedback-assisted classical communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bounds hold

$$\frac{\log_2|\mathcal{M}|}{n}\leq\frac{1}{1-\varepsilon}\left(\chi(\mathcal{N})+\frac{1}{n}h_2(\varepsilon)\right), \quad (18.3.5)$$
$$\frac{\log_2|\mathcal{M}|}{n}\leq\widetilde{\chi}_\alpha(\mathcal{N})+\frac{\alpha}{n(\alpha-1)}\log_2\!\left(\frac{1}{1-\varepsilon}\right)\qquad\forall\alpha>1, \quad (18.3.6)$$

where $\chi(\mathcal{N})$ is the Holevo information of $\mathcal{N}_{A\to B}$, as defined in (7.11.106), and $\widetilde{\chi}_\alpha(\mathcal{N})$ is the sandwiched Rényi Holevo information of $\mathcal{N}_{A\to B}$, as defined in (7.11.95).

Proof: Applying precisely the same reasoning as in the beginning of the proof of

Theorem 17.2, we conclude the following bound:

$$\log_2|\mathcal{M}| \leq I_H^{\varepsilon}(M;\hat{M})_{\omega}, \tag{18.3.7}$$

where $\omega_{M\hat{M}}$ is the final state of the protocol when the distribution p is set to the uniform distribution.

Invoking Proposition 7.70, the definition of $I_H^{\varepsilon}(M;\hat{M})$ from (7.11.88), and the expression for the mutual information from (7.2.97), we find that

$$I_H^{\varepsilon}(M;\hat{M})_{\omega} \leq \frac{1}{1-\varepsilon}\left(I(M;\hat{M})_{\omega} + h_2(\varepsilon)\right). \tag{18.3.8}$$

Now employing the data-processing inequality for the mutual information with respect to the last decoding channel $\mathcal{D}^n_{B_nB'_{n-1}\to\hat{M}}$, we find that

$$I(M;\hat{M})_{\omega} \leq I(M;B_nB'_{n-1})_{\rho^n}. \tag{18.3.9}$$

Then using the chain for the mutual information in (7.2.112), we obtain

$$I(M;B_nB'_{n-1})_{\rho^n} = I(M;B_n|B'_{n-1})_{\rho^n} + I(M;B'_{n-1})_{\rho^n} \tag{18.3.10}$$

$$\leq I(MB'_{n-1};B_n)_{\rho^n} + I(M;B'_{n-1})_{\rho^n}. \tag{18.3.11}$$

As mentioned above, the state shared between Alice and Bob, at any point during the protocol, is a separable state. Thus, the global state before the nth channel use can be written as follows:

$$\rho^n_{MA'_nA_nB'_{n-1}} = \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}|m\rangle\langle m|_M \otimes \sum_y p(y|m)\varsigma^{m,y}_{A'_nA_n} \otimes \varsigma^{m,y}_{B'_{n-1}}. \tag{18.3.12}$$

Then the state after the nth channel acts is as follows:

$$\rho^n_{MA'_nB_nB'_{n-1}} = \mathcal{N}_{A_n\to B_n}(\rho^n_{MA'_nA_nB'_{n-1}}) \tag{18.3.13}$$

$$= \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}|m\rangle\langle m|_M \otimes \sum_y p(y|m)\mathcal{N}_{A_n\to B_n}(\varsigma^{m,y}_{A'_nA_n}) \otimes \varsigma^{m,y}_{B'_{n-1}}. \tag{18.3.14}$$

An extension of the state above is as follows:

$$\rho^n_{MYA'_nB_nB'_{n-1}} =$$

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y} p(y|m) |m\rangle\langle m|_M \otimes |y\rangle\langle y|_Y \otimes \mathcal{N}_{A_n \to B_n}(\varsigma^{m,y}_{A'_n A_n}) \otimes \varsigma^{m,y}_{B'_{n-1}}, \quad (18.3.15)$$

and tracing over the system A'_n leads to the following state:

$$\begin{aligned} &\rho^n_{MYB_nB'_{n-1}} \\ &= \mathrm{Tr}_{A'_n}[\rho^n_{MYA'_nB_nB'_{n-1}}] && (18.3.16) \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y} p(y|m) |m\rangle\langle m|_M \otimes |y\rangle\langle y|_Y \otimes \mathcal{N}_{A_n \to B_n}(\varsigma^{m,y}_{A_n}) \otimes \varsigma^{m,y}_{B'_{n-1}}. && (18.3.17) \end{aligned}$$

Then consider that

$$\begin{aligned} I(MB'_{n-1}; B_n)_{\rho^n} &\leq I(MYB'_{n-1}; B_n)_{\rho^n} && (18.3.18) \\ &= I(MY; B_n)_{\rho^n} + I(B'_{n-1}; B_n|MY)_{\rho^n} && (18.3.19) \\ &= I(MY; B_n)_{\rho^n} && (18.3.20) \\ &\leq \chi(\mathcal{N}). && (18.3.21) \end{aligned}$$

The first inequality follows from the data-processing inequality for mutual information. The first equality follows from the chain rule. The second equality follows because the state in (18.3.17) is product when conditioning on the systems M and Y. The last inequality follows because the state $\rho^n_{MYB_n}$ is a classical–quantum state of the following form:

$$\begin{aligned} \rho^n_{MYB_n} &= \mathrm{Tr}_{B'_{n-1}}[\rho^n_{MYB_nB'_{n-1}}] && (18.3.22) \\ &= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{y} p(y|m) |m\rangle\langle m|_M \otimes |y\rangle\langle y|_Y \otimes \mathcal{N}_{A_n \to B_n}(\varsigma^{m,y}_{A_n}). && (18.3.23) \end{aligned}$$

Thus, the definition of the Holevo information in (7.11.106) implies the last inequality in (18.3.21). Putting together (18.3.9), (18.3.10)–(18.3.11), and (18.3.18)–(18.3.21), we conclude that

$$\begin{aligned} I(M; \hat{M})_\omega &\leq \chi(\mathcal{N}) + I(M; B'_{n-1})_{\rho^n} && (18.3.24) \\ &\leq \chi(\mathcal{N}) + I(M; B_{n-1}B'_{n-2})_{\rho^{n-1}}, && (18.3.25) \end{aligned}$$

where the last inequality follows from the data-processing inequality for mutual information.

Now, we recognize the term $I(M;B_{n-1}B'_{n-2})_{\rho^{n-1}}$ as being of the same form as $I(M;B_nB'_{n-1})_{\rho^n}$ in (18.3.10). Thus, we iterate through the same sequence of arguments to conclude that

$$I(M;B_{n-1}B'_{n-2})_{\rho^{n-1}} \leq \chi(\mathcal{N}) + I(M;B_{n-2}B'_{n-3})_{\rho^{n-2}}, \tag{18.3.26}$$

which in turn implies that

$$I(M;\hat{M})_\omega \leq 2\chi(\mathcal{N}) + I(M;B_{n-2}B'_{n-3})_{\rho^{n-2}}. \tag{18.3.27}$$

Continuing all the way back to the first channel use, we find that

$$I(M;\hat{M})_\omega \leq n\chi(\mathcal{N}) \tag{18.3.28}$$

because $I(M;B'_0) = 0$ (the systems M and B'_0 are in a product state at the start of the protocol). Putting together (18.3.7), (18.3.8), and (18.3.28), we conclude that

$$\log_2|\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(n\chi(\mathcal{N}) + h_2(\varepsilon)\right), \tag{18.3.29}$$

which implies the claim in (18.3.5).

We now prove the inequality in (18.3.6). Our starting point is again (18.3.7), but from there, we instead apply Proposition 7.71, the definition of $I_H^\varepsilon(M;\hat{M})_\omega$ from (7.11.88), and the expression for the sandwiched Rényi mutual information from (7.11.92) to find that the following holds for all $\alpha > 1$:

$$I_H^\varepsilon(M;\hat{M})_\omega \leq \widetilde{I}_\alpha(M;\hat{M})_\omega + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{18.3.30}$$

Recall that the sandwiched Rényi mutual information $\widetilde{I}_\alpha(M;\hat{M})_\omega$ is defined as

$$\widetilde{I}_\alpha(M;\hat{M})_\omega = \inf_{\xi_{\hat{M}}} \widetilde{D}_\alpha(\omega_{M\hat{M}}\|\omega_M \otimes \xi_{\hat{M}}) \tag{18.3.31}$$

$$= \inf_{\xi_{\hat{M}}} \widetilde{D}_\alpha(\omega_{M\hat{M}}\|\pi_M \otimes \xi_{\hat{M}}). \tag{18.3.32}$$

We adopt a similar approach to that given for the proof of (17.1.36). Our goal is thus to compare the actual protocol with one that results from employing a useless, replacement channel (of the form discussed in Section 18.2). To this end, let $\mathcal{R}^{\sigma_B}_{A\to B}$ be the replacement channel defined in (18.2.1), with σ_B an arbitrary state. Then as discussed in Section 18.2 (in particular, in (18.2.12)), the final state of the protocol conducted with the replacement channel is given by $\tau_{M\hat{M}} = \pi_M \otimes \tau_{\hat{M}}$. Then, we find that

$$\widetilde{I}_\alpha(M;\hat{M})_\omega = \inf_{\xi_{\hat{M}}} \widetilde{D}_\alpha(\omega_{M\hat{M}} \| \pi_M \otimes \xi_{\hat{M}}) \tag{18.3.33}$$

$$\leq \widetilde{D}_\alpha(\omega_{M\hat{M}} \| \pi_M \otimes \tau_{\hat{M}}) \tag{18.3.34}$$

$$= \widetilde{D}_\alpha(\omega_{M\hat{M}} \| \tau_{M\hat{M}}). \tag{18.3.35}$$

By applying the data-processing inequality for the sandwiched Rényi relative entropy with respect to the last decoding channel, and using (18.2.10), we find that

$$\widetilde{D}_\alpha(\omega_{M\hat{M}} \| \tau_{M\hat{M}}) \leq \widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}} \| \tau^n_{MB_nB'_{n-1}}) \tag{18.3.36}$$

$$= \widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}} \| \tau^n_{MB'_{n-1}} \otimes \sigma_{B_n}), \tag{18.3.37}$$

It is our goal to bound this last term. To do so, consider that

$$\widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}} \| \tau^n_{MB'_{n-1}} \otimes \sigma_{B_n}) = \frac{\alpha}{\alpha-1} \log_2 \left\| \left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n \to B_n}\right) \left(\left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \rho^n_{MA_nB'_{n-1}} \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \right) \right\|_\alpha, \tag{18.3.38}$$

where we define the completely positive map Θ_X by

$$\Theta_X(\rho) \coloneqq X^{\frac{1}{2}} \rho X^{\frac{1}{2}}. \tag{18.3.39}$$

We now employ the key observation from before: if the channel $\mathcal{N}$ is entanglement breaking, then Alice and Bob's systems are always separable throughout the protocol. Thus, the state $\rho^n_{MA_nB'_{n-1}}$ is fully separable with respect to the cut $M : A_n : B'_{n-1}$. It is in turn separable with respect to the bipartite cut $A_n : MB'_{n-1}$ and can be written as

$$\rho^n_{MA_nB'_{n-1}} = \sum_j p(j) \rho^j_{A_n} \otimes \rho^j_{MB'_{n-1}}, \tag{18.3.40}$$

which implies that

$$\left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \rho^n_{MA_nB'_{n-1}} \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}$$

$$= \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \left(\sum_j p(j) \rho^j_{A_n} \otimes \rho^j_{MB'_{n-1}} \right) \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \tag{18.3.41}$$

$$= \sum_j p(j)\rho^j_{A_n} \otimes \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \rho^j_{MB'_{n-1}} \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}. \tag{18.3.42}$$

Since conjugation by a positive semi-definite operator is a completely positive map, we can apply Lemma 12.18 to conclude that

$$\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n\to B_n}\right)\left(\left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \rho^n_{MA_nB'_{n-1}} \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha$$
$$\leq \nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n\to B_n}\right) \cdot \left\|\left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}} \rho^n_{MB'_{n-1}} \left(\tau^n_{MB'_{n-1}}\right)^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha, \tag{18.3.43}$$

where $\nu_\alpha(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n\to B_n})$ is defined from (12.2.82) and we have identified MB'_{n-1} with system R of P_{RA} in (12.2.81) and A_n with system A of P_{RA}. We then have the following chain of inequalities:

$$\widetilde{D}_\alpha(\rho^n_{MB_nB'_{n-1}}\|\tau^n_{MB_nB'_{n-1}})$$
$$\leq \frac{\alpha}{\alpha-1}\log_2 \nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n\to B_n}\right) + \widetilde{D}_\alpha(\rho^n_{MB'_{n-1}}\|\tau^n_{MB'_{n-1}}) \tag{18.3.44}$$
$$\leq \frac{\alpha}{\alpha-1}\log_2 \nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A_n\to B_n}\right) + \widetilde{D}_\alpha(\rho^n_{MB_{n-1}B'_{n-2}}\|\tau^n_{MB_{n-1}B'_{n-2}}) \tag{18.3.45}$$
$$\leq n\frac{\alpha}{\alpha-1}\log_2 \nu_\alpha\left(\Theta_{\sigma_{B}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right) + \widetilde{D}_\alpha(\rho^n_{MB'_0}\|\tau^n_{MB'_0}) \tag{18.3.46}$$
$$= n\frac{\alpha}{\alpha-1}\log_2 \nu_\alpha\left(\Theta_{\sigma_{B}^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right) \tag{18.3.47}$$

The first inequality follows by combining (18.3.36)–(18.3.37) and (18.3.43). The second inequality follows from the data-processing inequality for the sandwiched Rényi relative entropy, with respect to the channel $\operatorname{Tr}_{F_{n-1}} \circ \overline{\Delta}_{F_{n-1}} \circ \mathcal{D}^{n-1}_{B_{n-1}B'_{n-2}\to F_{n-1}B'_{n-1}}$. The third inequality follows by recognizing that

$$\widetilde{D}_\alpha(\rho^n_{MB_{n-1}B'_{n-2}}\|\tau^n_{MB_{n-1}B'_{n-2}}) \tag{18.3.48}$$

is the sandwiched Rényi relative entropy at round $n-1$ of the protocol, which allows us to apply the argument inductively. The first equality follows because $\rho^n_{MB'_0} = \tau^n_{MB'_0}$ because no channels have been applied at this point in the protocol. Putting together (18.3.7), (18.3.30), (18.3.33)–(18.3.35), (18.3.36), and (18.3.44)–(18.3.47), we conclude that

$$\log_2|\mathcal{M}| \leq n\frac{\alpha}{\alpha-1}\log_2 \nu_\alpha\left(\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). \qquad (18.3.49)$$

Since this upper bound holds for every state σ_B, we can take an infimum over all such states and conclude that

$$\begin{aligned}
&\log_2|\mathcal{M}| \\
&\leq n\frac{\alpha}{\alpha-1}\inf_{\sigma_B}\log_2 \nu_\alpha\left(\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right) && (18.3.50)\\
&= n\widetilde{K}_\alpha(\mathcal{N}_{A\to B}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right) && (18.3.51)\\
&= n\widetilde{\chi}_\alpha(\mathcal{N}_{A\to B}) + \frac{\alpha}{\alpha-1}\log_2\left(\frac{1}{1-\varepsilon}\right). && (18.3.52)
\end{aligned}$$

The first equality follows from the definition of ν_α in (12.2.82) and the definition of $\widetilde{K}_\alpha$ in (12.2.58). The last equality follows from Lemma 12.17. ■

18.3.2 Entropy Upper Bound on the Number of Transmitted Bits

We now establish an upper bound that holds for an arbitrary quantum channel. It is equal to the maximum output entropy of the channel (Theorem 18.5). A refinement of this upper bound leads to an upper bound equal to the maximum expected output entropy of the channel (Theorem 18.6), by writing it as a convex combination of other channels.

We begin by establishing the first upper bound. The main idea for doing so is to consider a protocol that simulates the general protocol detailed in Section 18.1. The simulation is a purified protocol, in which every step of the original protocol is purified. Each state of the purified protocol, when conditioned on the message being transmitted and the values of the classical feedback, is in a pure state. We now detail the form of this purified protocol. In order to simplify notation, we let $\hat{A}$ denote a joint system throughout, referring to both the original system A' and a purifying reference system, and we take the same convention when using the notation $\hat{B}$. By inspecting (18.1.8), the initial state of Bob in the purified protocol is as follows:

$$\sigma_{F_0F_0'\hat{B}_0} \coloneqq \sum_{f_0} p(f_0)|f_0\rangle\langle f_0|_{F_0} \otimes |f_0\rangle\langle f_0|_{F_0'} \otimes \psi^{f_0}_{\hat{B}_0}, \tag{18.3.53}$$

where the state $\psi^{f_0}_{\hat{B}_0}$ purifies Bob's state $\Psi^{f_0}_{B_0'}$, such that tracing over a subsystem of $\psi^{f_0}_{\hat{B}_0}$ gives $\Psi^{f_0}_{B_0'}$. Additionally, Bob keeps an extra copy F_0' of the classical data transmitted over the classical feedback channel. Let $\mathcal{U}^0_{M'F_0\to\hat{A}_1A_1}$ denote an isometric channel extending the encoding channel $\mathcal{E}^0_{M'F_0\to A_1'A_1}$. After $\mathcal{U}^0_{M'F_0\to\hat{A}_1A_1}$ acts, the global state is as follows:

$$\omega^1_{M\hat{A}_1A_1F_0'\hat{B}_0} \coloneqq \sum_{m\in\mathcal{M}}\sum_{f_0} p(m)p(f_0)|m\rangle\langle m|_M \otimes \varphi^{0,m,f_0}_{\hat{A}_1A_1} \otimes |f_0\rangle\langle f_0|_{F_0'} \otimes \psi^{f_0}_{\hat{B}_0}, \tag{18.3.54}$$

where

$$\mathcal{U}^0_{M'F_0\to\hat{A}_1A_1}(|m\rangle\langle m|_{M'} \otimes |f_0\rangle\langle f_0|_{F_0}). \tag{18.3.55}$$

We perform this purification for each step of the protocol. Let $\mathcal{U}^i_{A_i'F_i\to\hat{A}_{i+1}A_{i+1}}$ denote an isometric channel extending the encoding channel $\mathcal{E}^i_{A_i'F_i\to A_{i+1}'A_{i+1}}$ for each $i \in \{1,\ldots,n-1\}$. Since the system F_i is classical, for each $i \in \{1,\ldots,n-1\}$, the decoding channel $\overline{\Delta}_{F_i} \circ \mathcal{D}^i_{B_iB_{i-1}'\to F_iB_i'}$ can be written explicitly as

$$\overline{\Delta}_{F_i} \circ \mathcal{D}^i_{B_iB_{i-1}'\to F_iB_i'} = \sum_{f_i} \mathcal{D}^{i,f_i}_{B_iB_{i-1}'\to B_i'} \otimes |f_i\rangle\langle f_i|_{F_i}, \tag{18.3.56}$$

where $\{\mathcal{D}^{i,f_i}_{B_iB_{i-1}'\to B_i'}\}_{f_i}$ is a collection of completely positive maps such that the sum map $\sum_{f_i} \mathcal{D}^{i,f_i}_{B_iB_{i-1}'\to B_i'}$ is trace preserving. Let $V^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i}$ be a linear map such that tracing over a subsystem of $V^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i}(\cdot)(V^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i})^\dagger$ gives the original map $\mathcal{D}^{i,f_i}_{B_iB_{i-1}'\to B_i'}$, and define the map

$$\mathcal{V}^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i}(\tau_{B_iB_{i-1}'}) \coloneqq V^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i}\tau_{B_iB_{i-1}'}(V^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i})^\dagger. \tag{18.3.57}$$

Then we define the extended decoding channel $\mathcal{V}^i_{B_iB_{i-1}'\to F_i\hat{B}_iF_i'}$ for each $i \in \{1,\ldots,n-1\}$ as

$$\mathcal{V}^i_{B_iB_{i-1}'\to F_i\hat{B}_iF_i'}(\tau_{B_iB_{i-1}'}) \coloneqq \sum_{f_i} \mathcal{V}^{i,f_i}_{B_iB_{i-1}'\to\hat{B}_i}(\tau_{B_iB_{i-1}'}) \otimes |f_i\rangle\langle f_i|_{F_i} \otimes |f_i\rangle\langle f_i|_{F_i'}. \tag{18.3.58}$$

This extended decoding channel keeps an extra copy of the classical feedback value f_i for Bob in the classical register F_i'. The final decoding channel in the original protocol is a measurement channel and thus can be written as

$$\mathcal{D}^n_{B_nB'_{n-1}\to\hat{M}}(\tau_{B_nB'_{n-1}}) = \sum_{m\in\mathcal{M}} \text{Tr}[\Lambda^m_{B_nB'_{n-1}}\tau_{B_nB'_{n-1}}]|m\rangle\langle m|_{\hat{M}}, \tag{18.3.59}$$

where $\{\Lambda^m_{B_nB'_{n-1}}\}_{m\in\mathcal{M}}$ is a POVM. We enlarge it as follows in the simulation protocol:

$$\mathcal{V}^n_{B_nB'_{n-1}\to\hat{B}_n\hat{M}}(\tau_{B_nB'_{n-1}}) \coloneqq \sum_{m\in\mathcal{M}} \sqrt{\Lambda^m_{B_nB'_{n-1}}}\tau_{B_nB'_{n-1}}\sqrt{\Lambda^m_{B_nB'_{n-1}}} \otimes |m\rangle\langle m|_{\hat{M}}, \tag{18.3.60}$$

where the system $\hat{B}_n$ is isomorphic to the systems $B_nB'_{n-1}$. In the simulation protocol, we also consider an isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ that simulates the original channel $\mathcal{N}_{A\to B}$ as follows: $\mathcal{N}_{A\to B} = \text{Tr}_E \circ \mathcal{U}^{\mathcal{N}}_{A\to BE}$.

Thus, the various states involved in the purified protocol are as follows. The global initial state is $\overline{\Phi}^p_{MM'} \otimes \sigma_{F_0F_0'\hat{B}_0}$. Alice performs the extended encoding channel $\mathcal{U}^0_{M'F_0\to\hat{A}_1A_1}$ and the state becomes as follows:

$$\omega^1_{M\hat{A}_1A_1F_0'\hat{B}_0} = \mathcal{U}^0_{M'F_0\to\hat{A}_1A_1}(\overline{\Phi}^p_{MM'} \otimes \sigma_{F_0F_0'\hat{B}_0}). \tag{18.3.61}$$

Alice transmits system A_1 through the first use of the extended channel $\mathcal{U}^{\mathcal{N}}_{A_1\to B_1E_1}$, resulting in the following state:

$$\rho^1_{M\hat{A}_1B_1E_1F_0'\hat{B}_0} \coloneqq \mathcal{U}^{\mathcal{N}}_{A_1\to B_1E_1}(\omega^1_{M\hat{A}_1A_1F_0'\hat{B}_0}). \tag{18.3.62}$$

Bob processes his systems B_1 and B_0' with the extended decoding channel $\mathcal{V}^1_{B_1B_0'\to F_1\hat{B}_1F_1'}$, and Alice acts with the extended encoding channel $\mathcal{U}^1_{A_1'F_1\to\hat{A}_2A_2}$, resulting in the state

$$\omega^2_{M\hat{A}_2A_2\hat{B}_1E_1F_0'F_1'} \coloneqq (\mathcal{U}^1_{A_1'F_1\to\hat{A}_2A_2} \circ \mathcal{V}^1_{B_1B_0'\to F_1\hat{B}_1F_1'})(\rho^1_{M\hat{A}_1B_1E_1F_0'\hat{B}_0}). \tag{18.3.63}$$

This process iterates $n-2$ more times, resulting in the following states:

$$\rho^i_{M\hat{A}_iB_i\hat{B}_{i-1}E_1^i[F_0^{i-1}]'} \coloneqq \mathcal{U}^{\mathcal{N}}_{A_i\to B_iE_i}(\omega^i_{M\hat{A}_iA_i\hat{B}_{i-1}E_1^{i-1}[F_0^{i-1}]'}), \tag{18.3.64}$$

$$\omega^{i+1}_{M\hat{A}_{i+1}A_{i+1}\hat{B}_iE_1^i[F_0^i]'} \coloneqq (\mathcal{U}^i_{A_i'F_i\to\hat{A}_{i+1}A_{i+1}} \circ \mathcal{V}^i_{B_iB'_{i-1}\to F_i\hat{B}_iF_i'})(\rho^i_{M\hat{A}_iB_i\hat{B}_{i-1}E_1^i[F_0^{i-1}]'}), \tag{18.3.65}$$

for $i \in \{2, \ldots, n-1\}$. The final extended decoding channel results in the following state:

$$\rho^n_{M\hat{A}_n\tilde{B}\hat{M}E^n_1[F^{n-1}_0]'} := \mathcal{V}^n_{B_nB'_{n-1}\to\hat{B}_n\hat{M}}(\rho^n_{M\hat{A}_nB_n\hat{B}_{n-1}E^n_1[F^{n-1}_0]'}), \tag{18.3.66}$$

where the $\tilde{B}$ system encompasses all systems in Bob's possession at the end. Note that we recover each state of the original protocol described in Section 18.1 by performing particular partial traces.

Before stating the main theorem of this section, we prove two lemmas that play an important role in its proof. Both lemmas involve the following information measure:

$$I(X;CY)_\tau + H(C|XY)_\tau, \tag{18.3.67}$$

where the information quantities are evaluated with respect to the following classical–quantum state:

$$\tau_{XYC} := \sum_{x,y} p(x,y)|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \tau^{x,y}_C. \tag{18.3.68}$$

In the above, $p(x,y)$ is a probability distribution and $\tau^{x,y}_C$ is a quantum state for all x and y.

Lemma 18.3

Let τ_{XYAB} be a classical–quantum state, with classical systems XY and quantum systems AB pure when conditioned on XY. Let $\mathcal{L}_{AB\to A'B'Z}$ be a one-way LOCC channel of the following form:

$$\mathcal{L}_{AB\to A'B'Z} := \sum_z \mathcal{U}^z_{A\to A'} \otimes \mathcal{V}^z_{B\to B'} \otimes |z\rangle\langle z|_Z, \tag{18.3.69}$$

where $\{\mathcal{V}^z_{B\to B'}\}_z$ is a collection of completely positive trace-non-increasing maps with $\mathcal{V}^z_{B\to B'}(\cdot) := V^z_{B\to B'}(\cdot)(V^z_{B\to B'})^\dagger$ and $\{\mathcal{U}^z_{A\to A'}\}_z$ is a collection of isometric channels. Then the following inequality holds

$$I(X;BY)_\tau + H(B|XY)_\tau \geq I(X;B'YZ)_\omega + H(B'|XYZ)_\omega, \tag{18.3.70}$$

where $\omega_{XYZA'B'} := \mathcal{L}_{AB\to A'B'Z'}(\tau_{XYAB})$.

PROOF: The inequality $I(X;BY)_\tau \geq I(X;B'YZ)_\omega$ follows from the data-processing inequality for mutual information. In more detail, consider that $\omega_{XYZB'}$ is equal to

$$\omega_{XYZB'} = \operatorname{Tr}_{A'}[\omega_{XYZA'B'}] \tag{18.3.71}$$

$$= \operatorname{Tr}_{A'}\left[\sum_z (\mathcal{U}^z_{A\to A'} \otimes \mathcal{V}^z_{B\to B'})(\tau_{XYAB}) \otimes |z\rangle\langle z|_Z\right] \tag{18.3.72}$$

$$= \sum_z ((\operatorname{Tr}_{A'} \circ \mathcal{U}^z_{A\to A'}) \otimes \mathcal{V}^z_{B\to B'})(\tau_{XYAB}) \otimes |z\rangle\langle z|_Z \tag{18.3.73}$$

$$= \sum_z \mathcal{V}^z_{B\to B'}(\operatorname{Tr}_A[\tau_{XYAB}]) \otimes |z\rangle\langle z|_Z \tag{18.3.74}$$

$$= \sum_z \mathcal{V}^z_{B\to B'}(\tau_{XYB}) \otimes |z\rangle\langle z|_Z. \tag{18.3.75}$$

The fourth equality follows because $\mathcal{U}^z_{A\to A'}$ is an isometric channel for all z. Thus, the state $\omega_{XYZB'}$ can be understood as arising from the action of the quantum instrument $\sum_z \mathcal{V}^z_{B\to B'} \otimes |z\rangle\langle z|_Z$ on the state τ_{XYB}, and since this a channel taking system B to $B'Z$, the data-processing inequality for mutual information applies. The inequality $H(B|XY)_\tau \geq H(B'|XYZ)_\omega$ is a consequence of the LOCC monotonicity of the entanglement of formation (see Proposition 9.6). Indeed, consider that

$$H(B|XY)_\tau = E_F(A;BXY)_\tau, \tag{18.3.76}$$

$$H(B'|XYZ)_\omega = E_F(A';B'XYZ)_\omega, \tag{18.3.77}$$

which follows from the direct-sum property of the entanglement of formation (see the proof of Proposition 9.6) and its reduction to entropy of entanglement for pure states (see (9.1.40)). Thus, we apply these equalities and the LOCC monotonicity of entanglement of formation (i.e., $E_F(A;BXY)_\tau \geq E_F(A';B'XYZ)_\omega$). ■

The following lemma places an entropic upper bound on the amount by which the information quantity in (18.3.67) can increase by the action of a channel $\mathcal{N}_{A\to B}$:

Lemma 18.4

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\tau_{XYAB'}$ be a classical–quantum state of the following form:

$$\tau_{XYAB'} := \sum_{x,y} p(x,y)|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \tau^{x,y}_{AB'}. \tag{18.3.78}$$

Then

$$I(X;BB'Y)_\omega + H(BB'|XY)_\omega - [I(X;B'Y)_\tau + H(B'|XY)_\tau] \leq H(B)_\omega, \tag{18.3.79}$$

where $\omega_{XYBB'} := \mathcal{N}_{A\to B}(\tau_{XYAB'})$.

PROOF: Consider that

$$\begin{aligned}
&I(X;BB'Y)_\omega + H(BB'|XY)_\omega - [I(X;B'Y)_\tau + H(B'|XY)_\tau] \\
&= I(X;BB'Y)_\omega + H(BB'|XY)_\omega - [I(X;B'Y)_\omega + H(B'|XY)_\omega] && (18.3.80)\\
&= I(X;B|B'Y)_\omega + H(B|B'XY)_\omega && (18.3.81)\\
&= H(B|B'Y)_\omega - H(B|B'XY)_\omega + H(B|B'XY)_\omega && (18.3.82)\\
&= H(B|B'Y)_\omega && (18.3.83)\\
&\leq H(B)_\omega. && (18.3.84)
\end{aligned}$$

All equalities follow from applying definitions and chain rules for mutual information and entropy. The final inequality follows because conditioning does not increase entropy. ■

The key properties of the information quantity in (18.3.67) is that it does not increase under the action of a one-way LOCC channel from Bob to Alice (i.e., the decoding channel of Bob, the classical feedback channel, and the encoding channel of Alice) and it cannot increase by more than the output entropy of a channel under its action. We can use these properties to establish the following entropy bound on the number of bits that can be transmitted by a feedback-assisted communication protocol:

Theorem 18.5

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For an $(n, |\mathcal{M}|, \varepsilon)$ protocol for classical communication over a quantum channel $\mathcal{N}_{A\to B}$ assisted by classical feedback, as described in Section 18.1, the following bound holds

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \frac{1}{1-\varepsilon}\left(\sup_{\rho_A} H(\mathcal{N}_{A\to B}(\rho_A)) + \frac{h_2(\varepsilon)}{n}\right). \tag{18.3.85}$$

PROOF: Our starting point is the general bounds in (18.3.7)–(18.3.8), which imply

that

$$\log_2 |\mathcal{M}| \leq \frac{1}{1-\varepsilon}\left(I(M;\hat{M})_\omega + h_2(\varepsilon)\right), \tag{18.3.86}$$

where $\omega_{M\hat{M}}$ is the final state of the protocol, as given in (18.1.7), with p therein set to the uniform distribution over the set $\mathcal{M}$ of messages. Continuing, and considering the purified protocol outlined above, we find that

$$I(M;\hat{M})_\omega$$

$$\leq I(M;B_n\hat{B}_{n-1}[F_0^{n-1}]')_{\rho^n} + H(B_n\hat{B}_{n-1}|[F_0^{n-1}]'M)_{\rho^n} \tag{18.3.87}$$

$$= I(M;B_n\hat{B}_{n-1}[F_0^{n-1}]')_{\rho^n} + H(B_n\hat{B}_{n-1}|[F_0^{n-1}]'M)_{\rho^n} - \left[I(M;\hat{B}_0F_0')_{\omega^1} + H(\hat{B}_0|F_0'M)_{\omega^1}\right] \tag{18.3.88}$$

$$\begin{aligned}= {} & I(M;B_n\hat{B}_{n-1}[F_0^{n-1}]')_{\rho^n} + H(B_n\hat{B}_{n-1}|[F_0^{n-1}]'M)_{\rho^n} \\ & - \left[I(M;\hat{B}_0F_0')_{\omega^1} + H(\hat{B}_0|F_0'M)_{\omega^1}\right] \\ & + \sum_{i=2}^{n} I(M;\hat{B}_{i-1}[F_0^{i-1}]')_{\omega^i} + H(\hat{B}_{i-1}|[F_0^{i-1}]'M)_{\omega^i} \\ & \qquad - \left[I(M;\hat{B}_{i-1}[F_0^{i-1}]')_{\omega^i} + H(\hat{B}_{i-1}|[F_0^{i-1}]'M)_{\omega^i}\right].\end{aligned} \tag{18.3.89}$$

The first inequality follows from non-negativity of quantum entropy and data processing under the action of the final decoding channel. The first equality follows because $I(M;\hat{B}_0F_0')_{\omega^1} + H(\hat{B}_0|F_0'M)_{\omega^1} = 0$ for the initial state $\omega^1_{M\hat{A}_1A_1F_0'\hat{B}_0}$ (indeed, the systems M and $F_0'\hat{B}_0$ of the reduced state $\omega^1_{MF_0'\hat{B}_0}$ are product, and the state on system $\hat{B}_0$ is pure when conditioned on $F_0'M$). The last equality follows by adding and subtracting the same term. Continuing, we find that the quantity in the last line above is bounded as

$$\begin{aligned}\leq {} & I(M;B_n\hat{B}_{n-1}[F_0^{n-1}]')_{\rho^n} + H(B_n\hat{B}_{n-1}|[F_0^{n-1}]'M)_{\rho^n} \\ & - \left[I(M;\hat{B}_0F_0')_{\omega^1} + H(\hat{B}_0|F_0'M)_{\omega^1}\right] \\ & + \sum_{i=2}^{n} I(M;B_{i-1}\hat{B}_{i-2}[F_0^{i-2}]')_{\rho^{i-1}} + H(B_{i-1}\hat{B}_{i-2}|[F_0^{i-2}]'M)_{\rho^{i-1}} \\ & \qquad - \left[I(M;\hat{B}_{i-1}[F_0^{i-1}]')_{\omega^i} + H(\hat{B}_{i-1}|[F_0^{i-1}]'M)_{\omega^i}\right]\end{aligned} \tag{18.3.90}$$

$$\begin{aligned}= {} & \sum_{i=1}^{n} I(M;B_i\hat{B}_{i-1}[F_0^{i-1}]')_{\rho^i} + H(B_i\hat{B}_{i-1}|[F_0^{i-1}]'M)_{\rho^i} \\ & - \left[I(M;\hat{B}_{i-1}[F_0^{i-1}]')_{\omega^i} + H(\hat{B}_{i-1}|[F_0^{i-1}]'M)_{\omega^i}\right]\end{aligned} \tag{18.3.91}$$

$$\leq \sum_{i=1}^{n} H(B_i)_{\rho^i} \tag{18.3.92}$$

$$\leq n \sup_{\rho_A} H(\mathcal{N}_{A\to B}(\rho_A)) \tag{18.3.93}$$

The first inequality follows from Lemma 18.3 and the second from Lemma 18.4. So we conclude that

$$I(M;\hat{M})_\omega \leq n \sup_{\rho_A} H(\mathcal{N}_{A\to B}(\rho_A)). \tag{18.3.94}$$

Putting together this inequality and (18.3.86), we conclude the inequality in (18.3.85). ■

18.3.2.1 Maximum Average Output Entropy Upper Bound for Probabilistic Mixtures of Channels

In this section, we provide a brief proof of the following theorem, which generalizes Theorem 18.5 to the maximum average output entropy of a quantum channel:

Theorem 18.6

Let $\mathcal{N}_{A\to B} = \sum_x p_X(x)\mathcal{N}^x_{A\to B}$, where p_X is a probability distribution and $\{\mathcal{N}^x_{A\to B}\}_x$ is a set of channels. For an $(n, |\mathcal{M}|, \varepsilon)$ protocol for classical communication over the channel $\mathcal{N}_{A\to B}$ assisted by classical feedback, of the form described in Section 18.1, the following bound applies

$$(1-\varepsilon)\log_2|\mathcal{M}| \leq n \cdot \sup_{\rho_A} \sum_x p_X(x) H(\mathcal{N}^x_{A\to B}(\rho_A)) + h_2(\varepsilon).$$

Proof: The main idea behind the proof is to observe that an arbitrary feedback-assisted protocol of the form discussed in Section 18.1, which is for communication over a probabilistic mixture channel $\mathcal{N}_{A\to B} = \sum_z p_Z(z)\mathcal{N}^z_{A\to B}$, has a simulation of the following form:

1. Before the ith use of the channel $\mathcal{N}_{A\to B}$ in the feedback-assisted protocol, Bob selects a random variable Z_i independently according to the distribution p_Z. He transmits Z_i over the classical feedback channel to Alice.

2. Each channel use $\mathcal{N}_{A\to B}$ from the original protocol is replaced by a simulation in terms of another channel $\mathcal{M}_{AZ'\to B}$, which accepts a quantum input on system A and a classical input on system Z'. Conditioned on the value z in system Z', the channel $\mathcal{M}_{AZ'\to B}$ applies $\mathcal{N}^z_{A\to B}$ to the quantum system A. Thus, if the random variable $Z \sim p_Z$ is fed into the input system Z' of $\mathcal{M}_{AZ'\to B}$, then the channel $\mathcal{M}_{AZ'\to B}$ is indistinguishable from the original channel $\mathcal{N}_{A\to B}$.

3. Alice feeds a copy of the classical random variable Z_i into the ith use of the channel $\mathcal{M}_{AZ'\to B}$.

4. All other aspects of the protocol are executed in the same way as before. Namely, even though it would be an advantage to Alice to modify her encodings and Bob to modify later decodings based on the realizations of Z_i, they do not do so, and they instead blindly operate all other aspects of the simulation protocol as they are in the original protocol.

Our goal now is to establish the inequality in Theorem 18.6, relating the n, $|\mathcal{M}|$, and ε parameters of the original $(n, |\mathcal{M}|, \varepsilon)$ protocol by using the above simulation.

The main observation to make from here is that the same proof from Lemma 18.4 gives the following bound:

$$I(X; BB'YZ)_\omega + H(BB'|XYZ)_\omega - [I(X; B'YZ)_\tau + H(B'|XYZ)_\tau] \leq H(B|Z)_\omega, \quad (18.3.95)$$

where $\omega_{XYZBB'}$ is the following state:

$$\omega_{XYZBB'} := \mathcal{M}_{AZ'\to B}(\tau_{XYZZ''}), \quad (18.3.96)$$

$$\tau_{XYZZ''AB'} := \sum_{x,y,z} p(x, y, z)|x, y, z, z\rangle\langle x, y, z, z|_{X,Y,Z,Z'} \otimes \tau^{x,y,z}_{AB'}. \quad (18.3.97)$$

This follows by grouping Z with Y, but then discarding only Y and B' at the end of the proof. We then apply this bound, and the same reasoning in the proof of Theorem 18.5, except that the variables $Z_0, \ldots, Z_i$ are grouped together with the feedback variables $[F_0^{i-1}]'$ and then the same reasoning in (18.3.87)–(18.3.93) applies. At this point, we invoke (18.3.95) and find that

$$(1 - \varepsilon) \log_2 |\mathcal{M}| \leq \sum_{i=1}^{n} H(B_i|Z_i)_{\rho^{(i)}} + h_2(\varepsilon). \quad (18.3.98)$$

We can then bound the sum over entropies as follows:

$$\sum_{i=1}^{n} H(B_i|Z_i)_{\rho^{(i)}} \leq nH(B|Z)_{\overline{\rho}} \tag{18.3.99}$$

$$= n \sum_{z} p_Z(z) H(\mathcal{N}^z(\overline{\omega})) \tag{18.3.100}$$

$$\leq n \sup_{\rho} \sum_{z} p_Z(z) H(\mathcal{N}^z(\rho)). \tag{18.3.101}$$

The first inequality is by concavity of conditional entropy, and the conditional entropy is defined on the averaged channel output state over n uses, defined as

$$\overline{\rho}_{BZ} \coloneqq \sum_{z} p_Z(z) |z\rangle\langle z| \otimes \mathcal{N}^z(\overline{\omega}), \tag{18.3.102}$$

$$\overline{\omega}_A \coloneqq \frac{1}{n} \sum_{i=1}^{n} \omega_{A_i}^{(i)}. \tag{18.3.103}$$

The second equality follows from the definition of conditional entropy. The third inequality follows from optimizing over all states. ■

18.3.3 Geometric Υ-Information Upper Bound on the Number of Transmitted Bits

In this section, we prove that the Υ-information bound from Section 12.2.5.1 is actually an upper bound on the classical capacity assisted by classical feedback. The main idea behind the approach detailed in this section is to establish a correlation measure for bipartite channels, which is non-increasing under the action of one-way LOCC channels and measures the forward classical communication that can be generated by the bipartite channel for which it is evaluated. Such a measure is relevant in the context of a feedback-assisted protocol because, in such a protocol, Alice and Bob employ a one-way LOCC channel from Bob to Alice. In particular, local channels are allowed for free, as well as the use of a classical feedback channel. Both of these actions can be considered as particular kinds of bipartite channels and both of them fall into the class of bipartite channels that are non-signaling from Alice to Bob and C-PPT-P (call this class $\mathrm{NS}_{A\not\to B} \cap \mathrm{PPT}$). Recall the definition of non-signaling channels from Section 4.6.4 and C-PPT-P channels from Section 4.6.3. As such, if we employ a measure of bipartite channels that involves a comparison between a bipartite channel of interest to all bipartite channels in $\mathrm{NS}_{A\not\to B} \cap \mathrm{PPT}$, then the two kinds of free channels would have zero value and the measure would indicate how different the channel of interest is from this set (i.e., how different it is from a channel that has no ability to send quantum information and no ability to signal from Alice to Bob). This is the main idea behind the measure that we define below in Definition 18.7, but one should keep in mind that the measure below does not follow this reasoning precisely.

In Definition 18.7, although we motivated the measure for bipartite channels, we define it more generally for completely positive bipartite maps, as it turns out to be useful to do so when we define other measures later.

Definition 18.7 **β-Measure of Classical Communication for Bipartite Channels**

Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Then we define

$$C_\beta(\mathcal{M}_{AB\to A'B'}) := \log_2 \beta(\mathcal{M}_{AB\to A'B'}), \tag{18.3.104}$$

$$\beta(\mathcal{M}_{AB\to A'B'}) := \inf_{\substack{S_{AA'BB'},\\ V_{AA'BB'}\in\mathrm{Herm}}} \left\{ \begin{array}{c} \|\mathrm{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty : \\ T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0, \\ S_{AA'BB'} \pm V_{AA'BB'} \geq 0, \\ \mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}, \tag{18.3.105}$$

where Herm denotes the set of Hermitian operators and $\Gamma^{\mathcal{M}}_{AA'BB'}$ is the Choi operator of $\mathcal{M}_{AB\to A'B'}$:

$$\Gamma^{\mathcal{M}}_{AA'BB'} := \mathcal{M}_{\hat{A}\hat{B}\to A'B'}(\Gamma_{A\hat{A}} \otimes \Gamma_{B\hat{B}}). \tag{18.3.106}$$

In the above, $\hat{A}$ is isomorphic to A, system $\hat{B}$ is isomorphic to B,

$$\Gamma_{A\hat{A}} := \sum_{i,j=0}^{d_A-1} |i\rangle\!\langle j|_A \otimes |i\rangle\!\langle j|_{\hat{A}}, \qquad \Gamma_{B\hat{B}} := \sum_{i,j=0}^{d_B-1} |i\rangle\!\langle j|_B \otimes |i\rangle\!\langle j|_{\hat{B}}, \tag{18.3.107}$$

and $\pi_A := I_A/d_A$.

Since $S_{AA'BB'} \pm V_{AA'BB'} \geq 0$ implies that $S_{AA'BB'} \geq 0$, we can also express

$\beta(\mathcal{M}_{AB\to A'B'})$ as follows:

$$\inf_{\substack{S_{AA'BB'},\\ V_{AA'BB'}\in\text{Herm}}} \left\{ \begin{array}{c} \lambda : \\ \text{Tr}_{A'B'}[S_{AA'BB'}] \leq \lambda I_{AB} \\ T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0, \\ S_{AA'BB'} \pm V_{AA'BB'} \geq 0, \\ \text{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \text{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\} \tag{18.3.108}$$

By exploiting the equality constraint $\text{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \text{Tr}_{AA'}[S_{AA'BB'}]$, we find that

$$\|\text{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty = \|\text{Tr}_{B'}[\text{Tr}_{A'}[S_{AA'BB'}]]\|_\infty \tag{18.3.109}$$

$$= \|\text{Tr}_{B'}[\pi_A \otimes \text{Tr}_{AA'}[S_{AA'BB'}]]\|_\infty \tag{18.3.110}$$

$$= \|\pi_A \otimes \text{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty \tag{18.3.111}$$

$$= \frac{1}{d_A} \|\text{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty . \tag{18.3.112}$$

Then we find that

$$\beta(\mathcal{M}_{AB\to A'B'}) \coloneqq \inf_{\substack{S_{AA'BB'},\\ V_{AA'BB'}\in\text{Herm}}} \left\{ \begin{array}{c} \frac{1}{d_A} \|\text{Tr}_{AA'B'}[S_{AA'BB'}]\|_\infty : \\ T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0, \\ S_{AA'BB'} \pm V_{AA'BB'} \geq 0, \\ \text{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \text{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}. \tag{18.3.113}$$

Since $S_{AA'BB'} \pm V_{AA'BB'} \geq 0$ implies that $S_{AA'BB'} \geq 0$, we can also rewrite $\beta(\mathcal{M}_{AB\to A'B'})$ as

$$\beta(\mathcal{M}_{AB\to A'B'}) \coloneqq \inf_{\substack{\lambda, S_{AA'BB'}\geq 0,\\ V_{AA'BB'}\in\text{Herm}}} \left\{ \begin{array}{c} \lambda : \\ \frac{1}{d_A} \text{Tr}_{AA'B'}[S_{AA'BB'}] \leq \lambda I_B, \\ T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0, \\ S_{AA'BB'} \pm V_{AA'BB'} \geq 0, \\ \text{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \text{Tr}_{AA'}[S_{AA'BB'}] \end{array} \right\}. \tag{18.3.114}$$

18.3.3.1 Properties of the basic measure

We now establish several properties of $C_\beta(\mathcal{N}_{AB\to A'B'})$, which are basic properties that we might expect of a measure of forward classical communication for a bipartite channel. These include the following:

1. non-negativity (Proposition 18.8),
2. stability under tensoring with identity channels (Proposition 18.9),
3. zero value for classical feedback channels (Proposition 18.10),
4. zero value for a tensor product of local channels (Proposition 18.11),
5. subadditivity under serial composition (Proposition 18.12),
6. data processing under pre- and post-processing by local channels (Corollary 18.13),
7. invariance under local unitary channels (Corollary 18.14),
8. convexity of β (Proposition 18.15).

All of the properties above hold for bipartite channels, while the second and fifth through eighth hold more generally for completely positive bipartite maps.

Proposition 18.8 Non-Negativity

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$C_\beta(\mathcal{N}_{AB\to A'B'}) \geq 0. \tag{18.3.115}$$

PROOF: We prove the equivalent statement $\beta(\mathcal{N}_{AB\to A'B'}) \geq 1$. Let λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (18.3.114). Then consider that

$$\lambda d_B = \lambda \operatorname{Tr}_B[I_B] \tag{18.3.116}$$
$$\geq \frac{1}{d_A} \operatorname{Tr}_{AA'BB'}[S_{AA'BB'}] \tag{18.3.117}$$
$$\geq \frac{1}{d_A} \operatorname{Tr}_{AA'BB'}[V_{AA'BB'}] \tag{18.3.118}$$
$$= \frac{1}{d_A} \operatorname{Tr}_{AA'BB'}[T_{BB'}(V_{AA'BB'})] \tag{18.3.119}$$
$$\geq \operatorname{Tr}_{AA'BB'}[\Gamma^{\mathcal{N}}_{AA'BB'}] \tag{18.3.120}$$
$$= \frac{1}{d_A} \operatorname{Tr}_{AB}[I_{AB}] \tag{18.3.121}$$

$$= d_B. \tag{18.3.122}$$

This implies that $\lambda \geq 1$. Since the inequality holds for all λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ satisfying the constraints in (18.3.114), we conclude the statement above. ■

Proposition 18.9 Stability

Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Then

$$C_\beta(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}) = C_\beta(\mathcal{M}_{AB\to A'B'}). \tag{18.3.123}$$

Proof: Let $S_{AA'BB'}$ and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (18.3.105) for $\mathcal{M}_{AB\to A'B'}$. The Choi operator of $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$ is given by

$$\Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}. \tag{18.3.124}$$

Let us show that $\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}$ and $\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}$ satisfy the constraints in (18.3.105) for $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$. Consider that

$$T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0 \tag{18.3.125}$$

$$\iff \quad T_{BB'}(\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq 0 \tag{18.3.126}$$

$$\iff \quad T_{BB'\bar{B}\tilde{B}}(\Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq 0 \tag{18.3.127}$$

$$S_{AA'BB'} \pm V_{AA'BB'} \geq 0 \tag{18.3.128}$$

$$\iff \quad \Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes V_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}} \geq 0 \tag{18.3.129}$$

$$\mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'}],$$

the latter equivalent to

$$\begin{aligned} &\mathrm{Tr}_{A'\tilde{A}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}] \\ &= I_{\bar{A}} \otimes \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}] \qquad (18.3.130) \\ &= \pi_{\bar{A}} \otimes \pi_A \otimes \mathrm{Tr}_{AA'\bar{A}\tilde{A}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}]. \qquad (18.3.131) \end{aligned}$$

Also, consider that

$$\frac{1}{d_A d_{\bar{A}}} \left\| \mathrm{Tr}_{AA'\bar{A}\tilde{A}B'\tilde{B}}[\Gamma_{\bar{A}\tilde{A}} \otimes S_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}] \right\|_\infty$$

$$= \frac{1}{d_A d_{\bar{A}}} \left\| d_{\bar{A}} \operatorname{Tr}_{AA'B'}[S_{AA'BB'} \otimes I_{\bar{B}}] \right\|_\infty \tag{18.3.132}$$

$$= \frac{1}{d_A} \left\| \operatorname{Tr}_{AA'B'}[S_{AA'BB'}] \otimes I_{\bar{B}} \right\|_\infty \tag{18.3.133}$$

$$= \frac{1}{d_A} \left\| \operatorname{Tr}_{AA'B'}[S_{AA'BB'}] \right\|_\infty . \tag{18.3.134}$$

Thus, it follows that

$$\beta(\mathcal{M}_{AB\to A'B'}) \geq \beta(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}). \tag{18.3.135}$$

Now let us show the opposite inequality. Let $S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}$ and $V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}$ be arbitrary Hermitian operators satisfying the constraints in (18.3.105) for $\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}$. Set

$$S'_{AA'BB'} \coloneqq \frac{1}{d_{\bar{A}} d_{\bar{B}}} \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}], \tag{18.3.136}$$

$$V'_{AA'BB'} \coloneqq \frac{1}{d_{\bar{A}} d_{\bar{B}}} \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]. \tag{18.3.137}$$

Consider that

$$\Gamma^{\mathrm{id}\otimes\mathcal{N}\otimes\mathrm{id}}_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} = \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}. \tag{18.3.138}$$

Then

$$T_{BB'\bar{B}\tilde{B}}(V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}}) \geq 0 \tag{18.3.139}$$

$$\Longrightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[T_{BB'\bar{B}\tilde{B}}(V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm \Gamma_{\bar{A}\tilde{A}} \otimes \Gamma^{\mathcal{M}}_{AA'BB'} \otimes \Gamma_{\bar{B}\tilde{B}})] \geq 0 \tag{18.3.140}$$

$$\Longleftrightarrow \quad T_{BB'}(V_{AA'BB'} \pm d_{\bar{A}} d_{\bar{B}} \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0 \tag{18.3.141}$$

$$\Longleftrightarrow \quad T_{BB'}(V'_{AA'BB'} \pm \Gamma^{\mathcal{M}}_{AA'BB'}) \geq 0. \tag{18.3.142}$$

Also

$$S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \geq 0 \tag{18.3.143}$$

$$\Longrightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}} \pm V_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \geq 0 \tag{18.3.144}$$

$$\Longleftrightarrow \quad S'_{AA'BB'} \pm V'_{AA'BB'} \geq 0, \tag{18.3.145}$$

and

$$\operatorname{Tr}_{\tilde{A}A'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] = \pi_{\bar{A}A} \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \tag{18.3.146}$$

$$\Longrightarrow \quad \operatorname{Tr}_{\bar{A}\tilde{A}A'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] = \operatorname{Tr}_{\bar{A}\bar{B}\tilde{B}}[\pi_{\bar{A}A} \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]] \tag{18.3.147}$$

$$= \pi_A \otimes \operatorname{Tr}_{\bar{A}\tilde{A}AA'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \tag{18.3.148}$$

$$\Longleftrightarrow \quad \operatorname{Tr}_{A'}[S'_{AA'BB'}] = \pi_A \otimes \operatorname{Tr}_{AA'}[S'_{AA'BB'}]. \tag{18.3.149}$$

Finally, let λ be such that

$$\frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \leq \lambda I_{B\bar{B}}. \tag{18.3.150}$$

Then it follows that

$$\operatorname{Tr}_{\bar{B}}\left[\frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}]\right] \leq \operatorname{Tr}_{\bar{B}}[\lambda I_{B\bar{B}}] \tag{18.3.151}$$

$$\Longleftrightarrow \quad \frac{1}{d_A d_{\bar{A}}} \operatorname{Tr}_{\bar{A}\tilde{A}AA'B'\bar{B}\tilde{B}}[S_{\bar{A}\tilde{A}AA'BB'\bar{B}\tilde{B}}] \leq d_{\bar{B}} \lambda I_B \tag{18.3.152}$$

$$\Longleftrightarrow \quad \frac{1}{d_A} \operatorname{Tr}_{AA'B'}[S'_{AA'BB'}] \leq \lambda I_B. \tag{18.3.153}$$

Thus, we conclude that

$$\beta(\mathcal{M}_{AB\to A'B'}) \leq \beta(\operatorname{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \operatorname{id}_{\bar{B}\to\tilde{B}}). \tag{18.3.154}$$

This concludes the proof. ■

Proposition 18.10 Zero on Classical Feedback Channels

Let $\overline{\Delta}_{B\to A'}$ be a classical feedback channel:

$$\overline{\Delta}_{B\to A'}(\cdot) \coloneqq \sum_{i=0}^{d-1} |i\rangle_{A'}\langle i|_B(\cdot)|i\rangle_B\langle i|_{A'}, \tag{18.3.155}$$

where system A' is isomorphic to B and $d = d_{A'} = d_B$. Then

$$C_\beta(\overline{\Delta}_{B\to A'}) = 0. \tag{18.3.156}$$

Proof: We prove the equivalent statement that $\beta(\overline{\Delta}_{B\to A'}) = 1$. In this case, the A and B' systems are trivial, so that $d_A = 1$, and the Choi operator of $\overline{\Delta}_{B\to A'}$ is given by

$$\Gamma^{\overline{\Delta}}_{BA'} = \overline{\Gamma}_{BA'}, \tag{18.3.157}$$

where

$$\overline{\Gamma}_{BA'} \coloneqq \sum_{i=0}^{d_B-1} |i\rangle\!\langle i|_B \otimes |i\rangle\!\langle i|_{A'}. \tag{18.3.158}$$

Pick $S_{BA'} = V_{BA'} = \overline{\Gamma}_{BA'}$. Then we need to check that the constraints in (18.3.105) are satisfied for these choices. Consider that

$$T_B(V_{BA'} \pm \Gamma^{\overline{\Delta}}_{BA'}) \geq 0 \tag{18.3.159}$$

$$\iff \quad T_B(\overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'}) \geq 0 \tag{18.3.160}$$

$$\iff \quad \overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'} \geq 0, \tag{18.3.161}$$

and the last inequality is trivially satisfied. Also,

$$S_{BA'} \pm V_{BA'} \geq 0 \tag{18.3.162}$$

$$\iff \quad \overline{\Gamma}_{BA'} \pm \overline{\Gamma}_{BA'} \geq 0, \tag{18.3.163}$$

and the no-signaling condition $\mathrm{Tr}_{A'}[S_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S_{AA'BB'}]$ is trivially satisfied because the A system is trivial, having dimension equal to one. Finally, let us evaluate the objective function for these choices:

$$\frac{1}{d_A}\left\|\mathrm{Tr}_{AA'B'}[S_{AA'BB'}]\right\|_\infty = \left\|\mathrm{Tr}_{A'}[S_{A'B}]\right\|_\infty \tag{18.3.164}$$

$$= \left\|\mathrm{Tr}_{A'}[\overline{\Gamma}_{BA'}]\right\|_\infty \tag{18.3.165}$$

$$= \|I_B\|_\infty \tag{18.3.166}$$

$$= 1. \tag{18.3.167}$$

Combined with the general lower bound from Proposition 18.8, we conclude (18.3.156). ■

Proposition 18.11 Zero on Tensor Products of Local Channels

Let $\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$ be quantum channels. Then

$$C_\beta(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 0. \tag{18.3.168}$$

PROOF: We prove the equivalent statement that $\beta(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 1$. Set $S_{AA'BB'} = V_{AA'BB'} = \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}$, where $\Gamma^{\mathcal{E}}_{AA'}$ and $\Gamma^{\mathcal{F}}_{BB'}$ are the Choi operators of

$\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$, respectively. We need to check that the constraints in (18.3.105) are satisfied for these choices. Consider that

$$T_{BB'}(V_{AA'BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}) \geq 0 \tag{18.3.169}$$

$$\iff \quad T_{BB'}(\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}) \geq 0 \tag{18.3.170}$$

$$\iff \quad \Gamma^{\mathcal{E}}_{AA'} \otimes T_{BB'}(\Gamma^{\mathcal{F}}_{BB'}) \pm \Gamma^{\mathcal{E}}_{AA'} \otimes T_{BB'}(\Gamma^{\mathcal{F}}_{BB'}) \geq 0, \tag{18.3.171}$$

and the last inequality trivially holds because $T_{BB'}$ acts as a positive map on $\Gamma^{\mathcal{F}}_{BB'}$. Also,

$$S_{AA'BB'} \pm V_{AA'BB'} \geq 0 \tag{18.3.172}$$

$$\iff \quad \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \pm \Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'} \geq 0, \tag{18.3.173}$$

and

$$\operatorname{Tr}_{A'}[S_{AA'BB'}] = \operatorname{Tr}_{A'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}] \tag{18.3.174}$$

$$= I_A \otimes \Gamma^{\mathcal{F}}_{BB'} \tag{18.3.175}$$

$$= \pi_A \otimes \operatorname{Tr}_{AA'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}] \tag{18.3.176}$$

$$= \pi_A \otimes \operatorname{Tr}_{AA'}[S_{AA'BB'}]. \tag{18.3.177}$$

Finally, consider that the objective function evaluates to

$$\|\operatorname{Tr}_{A'B'}[S_{AA'BB'}]\|_\infty = \left\|\operatorname{Tr}_{A'B'}[\Gamma^{\mathcal{E}}_{AA'} \otimes \Gamma^{\mathcal{F}}_{BB'}]\right\|_\infty \tag{18.3.178}$$

$$= \|I_{AB}\|_\infty \tag{18.3.179}$$

$$= 1. \tag{18.3.180}$$

Combined with the general lower bound from Proposition 18.8, we conclude (18.3.168). ■

Proposition 18.12 Subadditivity under Composition

Let $\mathcal{M}^1_{AB\to A'B'}, \mathcal{M}^2_{A'B'\to A''B''}$ be completely positive bipartite maps, and define

$$\mathcal{M}^3_{AB\to A''B''} := \mathcal{M}^2_{A'B'\to A''B''} \circ \mathcal{M}^1_{AB\to A'B'}. \tag{18.3.181}$$

Then

$$C_\beta(\mathcal{M}^3_{AB\to A''B''}) \leq C_\beta(\mathcal{M}^2_{A'B'\to A''B''}) + C_\beta(\mathcal{M}^1_{AB\to A'B'}). \tag{18.3.182}$$

PROOF: We prove the equivalent statement that

$$\beta(\mathcal{M}^3_{AB\to A''B''}) \leq \beta(\mathcal{M}^2_{A'B'\to A''B''}) \cdot \beta(\mathcal{M}^1_{AB\to A'B'}). \tag{18.3.183}$$

Let $S^1_{AA'BB'}$ and $V^1_{AA'BB'}$ satisfy

$$T_{BB'}(V^1_{AA'BB'} \pm \Gamma^{\mathcal{M}^1}_{AA'BB'}) \geq 0, \tag{18.3.184}$$
$$S^1_{AA'BB'} \pm V^1_{AA'BB'} \geq 0, \tag{18.3.185}$$
$$\mathrm{Tr}_{A'}[S^1_{AA'BB'}] = \pi_A \otimes \mathrm{Tr}_{AA'}[S^1_{AA'BB'}], \tag{18.3.186}$$

and let $S^2_{A'A''B'B''}$ and $V^2_{A'A''B'B''}$ satisfy

$$T_{B'B''}(V^2_{A'A''B'B''} \pm \Gamma^{\mathcal{M}^2}_{A'A''B'B''}) \geq 0, \tag{18.3.187}$$
$$S^2_{A'A''B'B''} \pm V^2_{A'A''B'B''} \geq 0, \tag{18.3.188}$$
$$\mathrm{Tr}_{A''}[S^2_{A'A''B'B''}] = \pi_{A'} \otimes \mathrm{Tr}_{A'A''}[S^2_{A'A''B'B''}]. \tag{18.3.189}$$

Then it follows that

$$T_{BB'B'B''}(V^1_{AA'BB'} \otimes V^2_{A'A''B'B''} \pm \Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''}) \geq 0, \tag{18.3.190}$$
$$S^1_{AA'BB'} \otimes S^2_{A'A''B'B''} \pm V^1_{AA'BB'} \otimes V^2_{A'A''B'B''} \geq 0. \tag{18.3.191}$$

This latter statement is a consequence of the general fact that if A, B, C, and D are Hermitian operators satisfying $A \pm B \geq 0$ and $C \pm D \geq 0$, then $A \otimes C \pm B \otimes D \geq 0$. To see this, consider that the original four operator inequalities imply the four operator inequalities $(A \pm B) \otimes (C \pm D) \geq 0$, and then summing these four different operator inequalities in various ways leads to $A \otimes C \pm B \otimes D \geq 0$.

Now apply the following positive map to (18.3.190)–(18.3.191):

$$(\cdot) \to (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(\cdot)(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{18.3.192}$$

where

$$|\Gamma\rangle_{A'A'} \coloneqq \sum_i |i\rangle_{A'}|i\rangle_{A'}, \tag{18.3.193}$$
$$|\Gamma\rangle_{B'B'} \coloneqq \sum_i |i\rangle_{B'}|i\rangle_{B'}. \tag{18.3.194}$$

This gives

$$T_{BB''}(V^3_{AA''BB''} \pm \Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''}) \geq 0, \tag{18.3.195}$$

$$S^3_{AA''BB''} \pm V^3_{AA''BB''} \geq 0, \tag{18.3.196}$$

where

$$V^3_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(V^1_{AA'BB'} \otimes V^2_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{18.3.197}$$

$$\Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(\Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{18.3.198}$$

$$S^3_{AA''BB''} \coloneqq (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes S^2_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}), \tag{18.3.199}$$

and we applied (4.2.20) to conclude that

$$(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(\Gamma^{\mathcal{M}^1}_{AA'BB'} \otimes \Gamma^{\mathcal{M}^2}_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) = \Gamma^{\mathcal{M}^2\circ\mathcal{M}^1}_{AA''BB''}. \tag{18.3.200}$$

Also, consider that

$$\operatorname{Tr}_{A''}[S^3_{AA''BB''}]$$

$$= \operatorname{Tr}_{A''}[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes S^2_{A'A''B'B''})(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})] \tag{18.3.201}$$

$$= (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes \operatorname{Tr}_{A''}[S^2_{A'A''B'B''}])(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) \tag{18.3.202}$$

$$= (\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes \pi_{A'} \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) \tag{18.3.203}$$

$$= \frac{1}{d_{A'}}(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'})(S^1_{AA'BB'} \otimes I_{A'} \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])(|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'}) \tag{18.3.204}$$

$$= \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\operatorname{Tr}_{A'}[S^1_{AA'BB'}] \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'} \tag{18.3.205}$$

$$= \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\pi_A \otimes \operatorname{Tr}_{AA'}[S^1_{AA'BB'}] \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'} \tag{18.3.206}$$

$$= \pi_A \otimes \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\operatorname{Tr}_{AA'}[S^1_{AA'BB'}] \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'}. \tag{18.3.207}$$

Now consider that

$$\operatorname{Tr}_{AA''}[S^3_{AA''BB''}] = \frac{1}{d_{A'}}\langle\Gamma|_{B'B'}(\operatorname{Tr}_{AA'}[S^1_{AA'BB'}] \otimes \operatorname{Tr}_{A'A''}[S^2_{A'A''B'B''}])|\Gamma\rangle_{B'B'}. \tag{18.3.208}$$

So we conclude that

$$\mathrm{Tr}_{A''}[S^3_{AA''BB''}] = \pi_A \otimes \mathrm{Tr}_{AA''}[S^3_{AA''BB''}]. \tag{18.3.209}$$

Finally, consider that

$$\begin{aligned}
&\left\|\mathrm{Tr}_{A''B''}[S^3_{AA''BB''}]\right\|_\infty \\
&= \left\|\mathrm{Tr}_{A''B''}[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \left(S^1_{AA'BB'} \otimes S^2_{A'A''B'B''}\right) (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})]\right\|_\infty && (18.3.210)\\
&= \left\|[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \left(S^1_{AA'BB'} \otimes \mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right) (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})]\right\|_\infty && (18.3.211)\\
&\leq \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \cdot \\
&\qquad \left\|[(\langle\Gamma|_{A'A'} \otimes \langle\Gamma|_{B'B'}) \left(S^1_{AA'BB'} \otimes I_{A'B'}\right) (|\Gamma\rangle_{A'A'} \otimes |\Gamma\rangle_{B'B'})]\right\|_\infty && (18.3.212)\\
&= \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \left\|\mathrm{Tr}_{A'B'}[S^1_{AA'BB'}]\right\|_\infty . && (18.3.213)
\end{aligned}$$

Since $S^3_{AA''BB''}$ and $V^3_{AA''BB''}$ are particular choices that satisfy the constraints in (18.3.195)–(18.3.209), we conclude that

$$\beta(\mathcal{M}^3_{AB\to A''B''}) \leq \left\|\mathrm{Tr}_{A''B''}[S^2_{A'A''B'B''}]\right\|_\infty \left\|\mathrm{Tr}_{A'B'}[S^1_{AA'BB'}]\right\|_\infty . \tag{18.3.214}$$

Since $S^1_{AA'BB'}$ and $V^1_{AA'BB'}$ are arbitrary Hermitian operators satisfying the constraints in (18.3.184)–(18.3.186) and $S^2_{A'A''B'B''}$ and $V^2_{A'A''B'B''}$ are arbitrary Hermitian operators satisfying the constraints in (18.3.187)–(18.3.189), we conclude (18.3.182). ■

Corollary 18.13 Data Processing under Local Channels

Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Let $\mathcal{K}_{\hat{A}\to A}$, $\mathcal{L}_{\hat{B}\to B}$, $\mathcal{N}_{A'\to A''}$, and $\mathcal{P}_{B'\to B''}$ be local quantum channels, and define the bipartite completely positive map $\mathcal{F}_{\hat{A}\hat{B}\to A''B''}$ as follows:

$$\mathcal{F}_{\hat{A}\hat{B}\to A''B''} := (\mathcal{N}_{A'\to A''} \otimes \mathcal{P}_{B'\to B''})\mathcal{M}_{AB\to A'B'}(\mathcal{K}_{\hat{A}\to A} \otimes \mathcal{L}_{\hat{B}\to B}). \tag{18.3.215}$$

Then

$$C_\beta(\mathcal{F}_{\hat{A}\hat{B}\to A''B''}) \leq C_\beta(\mathcal{M}_{AB\to A'B'}). \tag{18.3.216}$$

PROOF: Apply Propositions 18.11 and 18.12 to find that

$$C_\beta(\mathcal{F}_{\hat{A}\hat{B}\to A''B''}) \\ \leq C_\beta(\mathcal{N}_{A'\to A''} \otimes \mathcal{P}_{B'\to B''}) + C_\beta(\mathcal{M}_{AB\to A'B'}) + C_\beta(\mathcal{K}_{\hat{A}\to A} \otimes \mathcal{L}_{\hat{B}\to B}) \quad (18.3.217)$$
$$= C_\beta(\mathcal{M}_{AB\to A'B'}). \quad (18.3.218)$$

This concludes the proof. ■

Corollary 18.14 Invariance under Local Unitary Channels

Let $\mathcal{M}_{AB\to A'B'}$ be a completely positive bipartite map. Let $\mathcal{U}_A$, $\mathcal{V}_B$, $\mathcal{W}_{A'}$, and $\mathcal{Y}_{B'}$ be local unitary channels, and define the bipartite completely positive map $\mathcal{F}_{\hat{A}\hat{B}\to A''B''}$ as follows:

$$\mathcal{F}_{AB\to A'B'} := (\mathcal{W}_{A'} \otimes \mathcal{Y}_{B'})\mathcal{M}_{AB\to A'B'}(\mathcal{U}_A \otimes \mathcal{V}_B). \quad (18.3.219)$$

Then

$$C_\beta(\mathcal{F}_{AB\to A'B'}) = C_\beta(\mathcal{M}_{AB\to A'B'}). \quad (18.3.220)$$

PROOF: Apply Corollary 18.13 twice to conclude that $C_\beta(\mathcal{M}_{AB\to A'B'}) \geq C_\beta(\mathcal{F}_{AB\to A'B'})$ and $C_\beta(\mathcal{F}_{AB\to A'B'}) \geq C_\beta(\mathcal{M}_{AB\to A'B'})$. ■

Proposition 18.15 Convexity

The measure β is convex, in the following sense:

$$\beta(\mathcal{M}^\lambda_{AB\to A'B'}) \leq \lambda\beta(\mathcal{M}^1_{AB\to A'B'}) + (1-\lambda)\,\beta(\mathcal{M}^0_{AB\to A'B'}), \quad (18.3.221)$$

where $\mathcal{M}^0_{AB\to A'B'}$ and $\mathcal{M}^1_{AB\to A'B'}$ are completely positive bipartite maps, $\lambda \in [0,1]$, and

$$\mathcal{M}^\lambda_{AB\to A'B'} := \lambda\mathcal{M}^1_{AB\to A'B'} + (1-\lambda)\,\mathcal{M}^0_{AB\to A'B'}. \quad (18.3.222)$$

PROOF: Let $S^x_{AA'BB'}$ and $V^x_{AA'BB'}$ satisfy the constraints in (18.3.105) for $\mathcal{M}^x_{AB\to A'B'}$ for $x \in \{0,1\}$. Then

$$S^\lambda_{AA'BB'} := \lambda S^1_{AA'BB'} + (1-\lambda)\,S^0_{AA'BB'}, \quad (18.3.223)$$
$$V^\lambda_{AA'BB'} := \lambda V^1_{AA'BB'} + (1-\lambda)\,V^0_{AA'BB'}, \quad (18.3.224)$$

satisfy the constraints in (18.3.105) for $\mathcal{M}^{\lambda}_{AB\to A'B'}$. Then it follows that

$$\beta(\mathcal{M}^{\lambda}_{AB\to A'B'}) \leq \left\|\operatorname{Tr}_{A'B'}[S^{\lambda}_{AA'BB'}]\right\|_{\infty} \tag{18.3.225}$$

$$\leq \lambda \left\|\operatorname{Tr}_{A'B'}[S^{1}_{AA'BB'}]\right\|_{\infty} + (1-\lambda)\left\|\operatorname{Tr}_{A'B'}[S^{0}_{AA'BB'}]\right\|_{\infty}, \tag{18.3.226}$$

where the second inequality follows from convexity of the ∞-norm. Since the inequality holds for all $S^{x}_{AA'BB'}$ and $V^{x}_{AA'BB'}$ satisfying the constraints in (18.3.105) for $\mathcal{M}^{x}_{AB\to A'B'}$ for $x \in \{0,1\}$, we conclude (18.3.221). ■

18.3.3.2 Related Measures

We now define variations of the bipartite channel measure from (18.3.105). We employ generalized divergences to do so, and in doing so, we arrive at a large number of variations of the basic bipartite channel measure.

Using the generalized channel divergence from Definition 7.81, we define the following:

Definition 18.16 Υ-Measure of Classical Communication for Bipartite Channels

For a bipartite channel $\mathcal{N}_{AB\to A'B'}$, we define the following measure of forward classical communication:

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) := \inf_{\mathcal{M}_{AB\to A'B'}:\beta(\mathcal{M}_{AB\to A'B'})\leq 1} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'}), \tag{18.3.227}$$

where the optimization is with respect to completely positive bipartite maps $\mathcal{M}_{AB\to A'B'}$.

Using the quantum relative entropy, the sandwiched Rényi relative entropy, the Belavkin–Staszewski relative entropy, and the geometric Rényi relative entropy, we then obtain the following respective channel measures: $\Upsilon(\mathcal{N}_{AB\to A'B'})$, $\widetilde{\Upsilon}_{\alpha}(\mathcal{N}_{AB\to A'B'})$, $\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'})$, and $\widehat{\Upsilon}_{\alpha}(\mathcal{N}_{AB\to A'B'})$, defined by substituting $\boldsymbol{D}$ with D, $\widetilde{D}_{\alpha}$, $\widehat{D}$, and $\widehat{D}_{\alpha}$.

We now establish some properties of $\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'})$, analogous to those established earlier for $C_{\beta}(\mathcal{N}_{AB\to A'B'})$ in Section 18.3.3.1. We assume that the underlying generalized divergence satisfies the minimal assumptions in (7.3.34) and (7.3.36); that is, $\boldsymbol{D}(1\|c) \geq 0$ for $c \in (0,1]$ and $\boldsymbol{D}(\rho\|\rho) = 0$ for every state ρ.

Proposition 18.17 Non-Negativity

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \geq 0. \tag{18.3.228}$$

PROOF: We prove the first inequality and the proof of the second inequality is similar. Consider that

$$\begin{aligned}
&\boldsymbol{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \\
&= \inf_{\substack{\mathcal{M}_{AB\to A'B'}: \\ \beta(\mathcal{M}_{AB\to A'B'})\leq 1}} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'}) \\
&\geq \inf_{\substack{\mathcal{M}_{AB\to A'B'}: \\ \beta(\mathcal{M}_{AB\to A'B'})\leq 1}} \boldsymbol{D}(\mathcal{N}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})\|\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})) \\
&\geq \inf_{\substack{\mathcal{M}_{AB\to A'B'}: \\ \beta(\mathcal{M}_{AB\to A'B'})\leq 1}} \boldsymbol{D}(\operatorname{Tr}[\mathcal{N}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]\|\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})]) \\
&= \inf_{\mathcal{M}_{AB\to A'B'}:\beta(\mathcal{M}_{AB\to A'B'})\leq 1} \boldsymbol{D}(1\|\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})])
\end{aligned} \tag{18.3.229}$$

The first inequality follows because $\boldsymbol{D}(\mathcal{N}_{AB\to A'B'}\|\mathcal{M}_{AB\to A'B'})$ involves an optimization over all possible input states, and we have chosen the product of maximally entangled states. The second inequality follows from the data-processing inequality for the generalized divergence. Thus, the inequality follows if we can show that

$$\operatorname{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA}\otimes\Phi_{BS})] \leq 1. \tag{18.3.230}$$

Let λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ be arbitrary Hermitian operators satisfying the constraints in (18.3.114) for $\mathcal{M}_{AB\to A'B'}$. Then, we find that

$$\begin{aligned}
\lambda d_A d_B &= \lambda \operatorname{Tr}_{AB}[I_{AB}] &(18.3.231)\\
&\geq \operatorname{Tr}_{AA'BB'}[S_{AA'BB'}] &(18.3.232)\\
&\geq \operatorname{Tr}_{AA'BB'}[V_{AA'BB'}] &(18.3.233)\\
&= \operatorname{Tr}_{AA'BB'}[T_{BB'}(V_{AA'BB'})] &(18.3.234)\\
&\geq \operatorname{Tr}_{AA'BB'}[T_{BB'}(\Gamma^{\mathcal{M}}_{AA'BB'})] &(18.3.235)\\
&= \operatorname{Tr}_{AA'BB'}[\Gamma^{\mathcal{M}}_{AA'BB'}] &(18.3.236)
\end{aligned}$$

$$= \mathrm{Tr}[\Gamma^{\mathcal{M}}_{AA'BB'}], \tag{18.3.237}$$

which is equivalent to

$$\lambda \geq \mathrm{Tr}[\mathcal{M}_{AB\to A'B'}(\Phi_{RA} \otimes \Phi_{BS})]. \tag{18.3.238}$$

Taking an infimum over λ, $S_{AA'BB'}$, and $V_{AA'BB'}$ satisfying the constraints in (18.3.114) for $\mathcal{M}_{AB\to A'B'}$ and applying the assumption $\beta(\mathcal{M}_{AB\to A'B'}) \leq 1$, we conclude (18.3.230). ■

Proposition 18.18 Stability

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then

$$\Upsilon(\mathcal{N}_{AB\to A'B'}) = \Upsilon(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{N}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}). \tag{18.3.239}$$

PROOF: The definition of the generalized channel divergence in Definition 7.81 implies that it is stable, in the sense that

$$\boldsymbol{D}(\mathcal{N}_{AB\to A'B'} \| \mathcal{M}_{AB\to A'B'}) = \\ \boldsymbol{D}(\mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{N}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}} \,\|\, \mathrm{id}_{\bar{A}\to\tilde{A}} \otimes \mathcal{M}_{AB\to A'B'} \otimes \mathrm{id}_{\bar{B}\to\tilde{B}}), \tag{18.3.240}$$

for every channel $\mathcal{N}_{AB\to A'B'}$ and completely positive map $\mathcal{M}_{AB\to A'B'}$. Combining with Proposition 18.9 and the definition in (18.3.227), we conclude (18.3.239). ■

Proposition 18.19 Zero on Classical Feedback Channels

Let $\overline{\Delta}_{B\to A'}$ be a classical feedback channel:

$$\overline{\Delta}_{B\to A'}(\cdot) := \sum_{i=0}^{d-1} |i\rangle_{A'}\langle i|_B(\cdot)|i\rangle_B\langle i|_{A'}, \tag{18.3.241}$$

where system A' is isomorphic to B and $d = d_{A'} = d_B$. Then

$$\Upsilon(\overline{\Delta}_{B\to A'}) = 0. \tag{18.3.242}$$

PROOF: This follows from Proposition 18.10. Since $\beta(\overline{\Delta}_{B\to A'}) = 1$, we can pick $\mathcal{M}_{B\to A'} = \overline{\Delta}_{B\to A'}$, and then

$$\boldsymbol{D}(\overline{\Delta}_{B\to A'} \| \mathcal{M}_{B\to A'}) = \boldsymbol{D}(\overline{\Delta}_{B\to A'} \| \overline{\Delta}_{B\to A'}) = 0. \tag{18.3.243}$$

So this establishes that $\Upsilon(\overline{\Delta}_{B\to A'}) \leq 0$, and the other inequality $\Upsilon(\overline{\Delta}_{B\to A'}) \geq 0$ follows from Proposition 18.17. ■

Proposition 18.20 Zero on Tensor Products of Local Channels

Let $\mathcal{E}_{A\to A'}$ and $\mathcal{F}_{B\to B'}$ be quantum channels. Then

$$\Upsilon(\mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}) = 0. \tag{18.3.244}$$

Proof: Same argument as given for Proposition 18.19, but use Proposition 18.11 instead. ■

We now establish some properties that are more specific to the Belavkin–Staszewski and geometric Rényi relative entropies (however the first actually holds also for the quantum relative entropy and other quantum Rényi relative entropies).

Proposition 18.21

Let $\mathcal{N}_{AB\to A'B'}$ be a bipartite channel. Then for all $\alpha \in (1,2]$,

$$\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'}) \leq C_\beta(\mathcal{N}_{AB\to A'B'}). \tag{18.3.245}$$

Proof: Pick $\mathcal{M}_{AB\to A'B'} = \frac{1}{\beta(\mathcal{N}_{AB\to A'B'})}\mathcal{N}_{AB\to A'B'}$ in the definition of $\widehat{\Upsilon}(\mathcal{N}_{AB\to A'B'})$ and $\widehat{\Upsilon}_\alpha(\mathcal{N}_{AB\to A'B'})$ and use the fact that, for $c > 0$, $\widehat{D}(\rho\|c\sigma) = \widehat{D}(\rho\|\sigma) - \log_2 c$ and $\widehat{D}_\alpha(\rho\|c\sigma) = \widehat{D}_\alpha(\rho\|\sigma) - \log_2 c$ for all $\alpha \in (1,2]$. We also require the monotonicity in α property from Proposition 7.44. ■

Proposition 18.22 Subadditivity

For bipartite channels $\mathcal{N}^1_{AB\to A'B'}$ and $\mathcal{N}^2_{A'B'\to A''B''}$, the following inequality holds for all $\alpha \in (0,1) \cup (1,2]$:

$$\widehat{\Upsilon}_\alpha(\mathcal{N}^2_{A'B'\to A''B''} \circ \mathcal{N}^1_{AB\to A'B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}^2_{A'B'\to A''B''}) + \widehat{\Upsilon}_\alpha(\mathcal{N}^1_{AB\to A'B'}). \tag{18.3.246}$$

PROOF: This inequality is a direct consequence of the subadditivity inequality in [REF - GEOMETRIC CH RENYI SUBADD], and the fact that if $\mathcal{M}^1$ and $\mathcal{M}^2$ are completely positive bipartite maps satisfying $\beta(\mathcal{M}^1), \beta(\mathcal{M}^2) \leq 1$, then $\beta(\mathcal{M}^2 \circ \mathcal{M}^1) \leq 1$ (see Proposition 18.12). ■

18.3.3.3 Measure of Classical Communication for a Point-to-Point Channel

Let $\mathcal{M}_{A\to B'}$ be a point-to-point completely positive map, which is a special case of a completely positive bipartite map with the Bob input B trivial and the Alice output A' trivial. We first show that β in (18.3.105) reduces to the measure from (12.2.228).

Proposition 18.23

Let $\mathcal{M}_{A\to B'}$ be a point-to-point completely positive map. Then

$$\beta(\mathcal{M}_{A\to B'}) := \inf_{S_{B'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \text{Tr}[S_{B'}] : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ I_A \otimes S_{B'} \pm V_{AB'} \geq 0 \end{array} \right\}. \tag{18.3.247}$$

PROOF: In this case, the systems A' and B are trivial. So then the definition in (18.3.105) reduces to

$$\beta(\mathcal{M}_{A\to B'}) = \inf_{S_{AB'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \|\text{Tr}_{B'}[S_{AB'}]\|_\infty : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ S_{AB'} \pm V_{AB'} \geq 0, \\ S_{AB'} = \pi_A \otimes \text{Tr}_A[S_{AB'}] \end{array} \right\}. \tag{18.3.248}$$

The last constraint implies that the optimization simplifies to

$$\beta(\mathcal{M}_{A\to B'}) = \inf_{S_{AB'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \|\text{Tr}_{B'}[\pi_A \otimes \text{Tr}_A[S_{AB'}]]\|_\infty : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes \text{Tr}_A[S_{AB'}] \pm V_{AB'} \geq 0 \end{array} \right\} \tag{18.3.249}$$

$$= \inf_{S'_{B'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \left\|\text{Tr}_{B'}[\pi_A \otimes S'_{B'}]\right\|_\infty : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} \tag{18.3.250}$$

$$= \inf_{S'_{B'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \operatorname{Tr}[S'_{B'}] \, \|\pi_A\|_\infty : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} \tag{18.3.251}$$

$$= \inf_{S'_{B'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \frac{1}{d_A} \operatorname{Tr}[S'_{B'}] : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S'_{B'} \pm V_{AB'} \geq 0 \end{array} \right\} \tag{18.3.252}$$

$$= \inf_{S_{B'}, V_{AB'} \in \text{Herm}} \left\{ \begin{array}{c} \operatorname{Tr}[S_{B'}] : \\ T_{BB'}(V_{AB'} \pm \Gamma^{\mathcal{M}}_{AB'}) \geq 0, \\ \pi_A \otimes S_{B'} \pm V_{AB'} \geq 0 \end{array} \right\}. \tag{18.3.253}$$

This concludes the proof. ■

More generally, consider that the definition in (18.3.227) becomes as follows for a point-to-point channel $\mathcal{N}_{A \to B'}$:

$$\Upsilon(\mathcal{N}_{A \to B'}) := \inf_{\mathcal{M}_{A \to B'} : \beta(\mathcal{M}_{A \to B'}) \leq 1} \boldsymbol{D}(\mathcal{N}_{A \to B'} \| \mathcal{M}_{A \to B'}), \tag{18.3.254}$$

which leads to the quantities $\widehat{\Upsilon}(\mathcal{N}_{A \to B'})$ and $\widehat{\Upsilon}_\alpha(\mathcal{N}_{A \to B'})$, for which we have the following bounds for $\alpha \in (1, 2]$:

$$\widehat{\Upsilon}(\mathcal{N}_{A \to B'}) \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{A \to B'}) \leq C_\beta(\mathcal{N}_{A \to B'}). \tag{18.3.255}$$

The next proposition is critical for establishing our upper bound proofs in Section 18.3.3.4. It states that if one share of a maximally classically correlated state passes through a completely positive map $\mathcal{M}_{A \to B'}$ for which $\beta(\mathcal{M}_{A \to B'}) \leq 1$, then the resulting operator has a very small chance of passing the comparator test, as defined in (18.3.258). (Recall that we previously used the comparator test in (11.1.37) and (12.1.19).)

Proposition 18.24 Bound for Comparator Test Success Probability

Let

$$\overline{\Phi}_{\hat{A}A} := \frac{1}{d} \sum_{i=0}^{d-1} |i\rangle\langle i|_{\hat{A}} \otimes |i\rangle\langle i|_A \tag{18.3.256}$$

denote the maximally classically correlated state, and let $\mathcal{M}_{A \to B'}$ be a completely positive map $\mathcal{M}_{A\to B'}$ for which $\beta(\mathcal{M}_{A\to B'}) \leq 1$. Then

$$\mathrm{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] \leq \frac{1}{d}, \tag{18.3.257}$$

where $\Pi_{\hat{A}B'}$ is the comparator test:

$$\Pi_{\hat{A}B'} \coloneqq \sum_{i=0}^{d-1} |i\rangle\!\langle i|_{\hat{A}} \otimes |i\rangle\!\langle i|_{B'}, \tag{18.3.258}$$

and the following systems are isomorphic: $\hat{A}$, A, and B'.

PROOF: Recall the expression for $\beta(\mathcal{M}_{A\to B'})$ in (18.3.247). Let $S_{B'}$ and $V_{AB'}$ be arbitrary Hermitian operators satisfying the constraints for $\beta(\mathcal{M}_{A\to B'})$. An application of (4.2.6) implies that

$$\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A}) = \langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}, \tag{18.3.259}$$

where $\tilde{A} \simeq A$. This means that

$$\begin{aligned}
\mathrm{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] &= \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (18.3.260)\\
&= \mathrm{Tr}[T_{B'}(\Pi_{\hat{A}B'})\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes \Gamma^{\mathcal{M}}_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (18.3.261)\\
&= \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes T_{B'}(\Gamma^{\mathcal{M}}_{\tilde{A}B'})|\Gamma\rangle_{A\tilde{A}}] && (18.3.262)\\
&\leq \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes T_{B'}(V_{\tilde{A}B'})|\Gamma\rangle_{A\tilde{A}}] && (18.3.263)\\
&= \mathrm{Tr}[T_{B'}(\Pi_{\hat{A}B'})\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes V_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (18.3.264)\\
&= \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes V_{\tilde{A}B'}|\Gamma\rangle_{A\tilde{A}}] && (18.3.265)\\
&\leq \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes I_{\tilde{A}} \otimes S_{B'}|\Gamma\rangle_{A\tilde{A}}] && (18.3.266)\\
&= \mathrm{Tr}[\Pi_{\hat{A}B'}\langle\Gamma|_{A\tilde{A}}\overline{\Phi}_{\hat{A}A} \otimes I_{\tilde{A}}|\Gamma\rangle_{A\tilde{A}} \otimes S_{B'}] && (18.3.267)\\
&= \mathrm{Tr}[\Pi_{\hat{A}B'}\,\mathrm{Tr}_A[\overline{\Phi}_{\hat{A}A}] \otimes S_{B'}] && (18.3.268)\\
&= \frac{1}{d}\,\mathrm{Tr}[\Pi_{\hat{A}B'}I_{\hat{A}} \otimes S_{B'}] && (18.3.269)\\
&= \frac{1}{d}\,\mathrm{Tr}[S_{B'}]. && (18.3.270)
\end{aligned}$$

Since this holds for all $S_{B'}$ and $V_{AB'}$ satisfying the constraints for $\beta(\mathcal{M}_{A\to B'})$, we conclude that

$$\mathrm{Tr}[\Pi_{\hat{A}B'}\mathcal{M}_{A\to B'}(\overline{\Phi}_{\hat{A}A})] \leq \frac{1}{d}. \tag{18.3.271}$$

This concludes the proof. ■

We finally state another proposition that plays an essential role in our upper bound proofs in Section 18.3.3.4.

Proposition 18.25

Suppose that $\mathcal{N}_{A\to B}$ is a channel with A isomorphic to B that satisfies

$$\frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \leq \varepsilon, \tag{18.3.272}$$

for $\varepsilon \in [0,1)$ and where $\overline{\Phi}_{RB} := \frac{1}{d}\sum_i |i\rangle\langle i|_R \otimes |i\rangle\langle i|_B$ and $d = d_R = d_A = d_B$. Then

$$\log_2 d \leq \inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})), \tag{18.3.273}$$

and for all $\alpha \in (1,2]$,

$$\log_2 d \leq \inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} \widehat{D}_{\alpha}(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{18.3.274}$$

PROOF: We begin by proving (18.3.273). The condition

$$\frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \leq \varepsilon \tag{18.3.275}$$

implies that

$$\mathrm{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon, \tag{18.3.276}$$

where $\Pi_{RB} := \sum_i |i\rangle\langle i|_R \otimes |i\rangle\langle i|_B$ is the comparator test. Indeed, applying a completely dephasing channel $\overline{\Delta}_B(\cdot) := \sum_i |i\rangle\langle i|(\cdot)|i\rangle\langle i|$ to the output of the channel $\mathcal{N}_{A\to B}$ and applying the data-processing inequality for trace distance, we conclude that

$$\varepsilon \geq \frac{1}{2}\left\|\mathcal{N}_{A\to B}(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1 \tag{18.3.277}$$

$$\geq \frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Delta}_B(\overline{\Phi}_{RB})\right\|_1 \tag{18.3.278}$$

$$= \frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1. \tag{18.3.279}$$

Let $\omega_{RB} \coloneqq (\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA})$ and observe that it can be written as

$$\omega_{RB} = \frac{1}{d}\sum_{i,j} p(j|i)|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B \tag{18.3.280}$$

for some conditional probability distribution $p(j|i)$. Then

$$\frac{1}{2}\left\|(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA}) - \overline{\Phi}_{RB}\right\|_1$$

$$= \frac{1}{2}\left\|\frac{1}{d}\sum_{i,j} p(j|i)|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B - \frac{1}{d}\sum_{i,j} \delta_{i,j}|i\rangle\langle i|_R \otimes |j\rangle\langle j|_B\right\|_1 \tag{18.3.281}$$

$$= \frac{1}{2}\frac{1}{d}\sum_i \left\|\sum_j (p(j|i) - \delta_{i,j})|j\rangle\langle j|_B\right\|_1 \tag{18.3.282}$$

$$= \frac{1}{2}\frac{1}{d}\sum_i \left[(1 - p(i|i)) + \sum_{j\neq i} p(j|i)\right] \tag{18.3.283}$$

$$= \frac{1}{d}\sum_i (1 - p(i|i)) \tag{18.3.284}$$

$$= 1 - \sum_i \frac{1}{d}p(i|i). \tag{18.3.285}$$

This implies that

$$\sum_i \frac{1}{d}p(i|i) \geq 1 - \varepsilon. \tag{18.3.286}$$

Now consider that

$$\mathrm{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] = \mathrm{Tr}[\overline{\Delta}_B(\Pi_{RB})\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \tag{18.3.287}$$

$$= \mathrm{Tr}[\Pi_{RB}(\overline{\Delta}_B \circ \mathcal{N}_{A\to B})(\overline{\Phi}_{RA})] \tag{18.3.288}$$

$$= \mathrm{Tr}[\Pi_{RB}\omega_{RB}] \tag{18.3.289}$$

$$= \sum_i \frac{1}{d}p(i|i). \tag{18.3.290}$$

So we conclude that

$$\operatorname{Tr}[\Pi_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon. \tag{18.3.291}$$

Applying the definition of the hypothesis testing relative entropy from Definition 7.65, we conclude that

$$\inf_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} D_H^\varepsilon(\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})\|\mathcal{M}_{A\to B}(\overline{\Phi}_{RA}))$$

$$= \inf_{\substack{\mathcal{M}_{A\to B}:\\ \beta(\mathcal{M}_{A\to B})\leq 1}} \left[-\log_2 \inf_{\Lambda_{RB}\geq 0}\left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] : \\ \operatorname{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon,\ \Lambda_{RB}\leq I_{RB} \end{array}\right\}\right] \tag{18.3.292}$$

$$= -\log_2 \sup_{\substack{\mathcal{M}_{A\to B}:\\ \beta(\mathcal{M}_{A\to B})\leq 1}} \inf_{\Lambda_{RB}\geq 0}\left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] : \\ \operatorname{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon,\ \Lambda_{RB}\leq I_{RB} \end{array}\right\}. \tag{18.3.293}$$

Now consider that

$$\sup_{\substack{\mathcal{M}_{A\to B}:\\ \beta(\mathcal{M}_{A\to B})\leq 1}} \inf_{\Lambda_{RB}\geq 0}\left\{ \begin{array}{c} \operatorname{Tr}[\Lambda_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] : \\ \operatorname{Tr}[\Lambda_{RB}\mathcal{N}_{A\to B}(\overline{\Phi}_{RA})] \geq 1-\varepsilon,\ \Lambda_{RB}\leq I_{RB} \end{array}\right\}$$

$$\leq \sup_{\mathcal{M}_{A\to B}:\beta(\mathcal{M}_{A\to B})\leq 1} \operatorname{Tr}[\Pi_{RB}\mathcal{M}_{A\to B}(\overline{\Phi}_{RA})] \tag{18.3.294}$$

$$\leq \frac{1}{d}, \tag{18.3.295}$$

where the last inequality follows from Proposition 18.24. Then applying a negative logarithm gives (18.3.273).

The inequality in (18.3.274) follows as direct application of the following relationship between hypothesis testing relative entropy and the geometric Rényi relative entropy:

$$D_H^\varepsilon(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{18.3.296}$$

as well as the previous proposition. The proof of (18.3.296) follows the same proof given for Proposition 7.71. ■

18.3.3.4 Proof of Geometric Υ-Information Upper Bound

We now have everything that we need to establish that the geometric Υ-information is an upper bound on the number of bits that can be transmitted by means of a quantum channel assisted by a classical feedback channel. By examining the protocol in Section 18.1, consider that the final state $\omega^p_{M\hat{M}}$ of the protocol can be written as follows:

$$\omega^p_{M\hat{M}} = \mathcal{P}_{M'\to\hat{M}}(\overline{\Phi}^p_{MM'}), \tag{18.3.297}$$

where

$$\mathcal{P}_{M'\to\hat{M}} := \mathcal{D}^n \circ \mathcal{N} \circ \mathcal{E}^{n-1} \circ \overline{\Delta} \circ \mathcal{D}^{n-1} \circ \mathcal{N} \circ \mathcal{E}^{n-2} \circ \overline{\Delta} \circ \mathcal{D}^{n-2} \circ \\ \cdots \circ \mathcal{D}^2 \circ \mathcal{N} \circ \mathcal{E}^1 \circ \overline{\Delta} \circ \mathcal{D}^1 \circ \mathcal{N} \circ \mathcal{E}^0 \circ \mathcal{A}, \tag{18.3.298}$$

and $\mathcal{A}$ is an appending channel that appends the state $\overline{\Delta}_{F_0}(\Psi_{F_0B_0'})$ to the input state $\overline{\Phi}^p_{MM'}$. In (18.3.298), we have omitted all system labels for simplicity.

We now state the main result of this section:

Theorem 18.26

Fix $n \in \mathbb{N}$, $\varepsilon \in [0,1)$, and $\alpha \in (1,2]$, and let $\mathcal{N}_{A\to B}$ be a quantum channel. For all $(n, |\mathcal{M}|, \varepsilon)$ classical-feedback-assisted classical communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\frac{\log_2 |\mathcal{M}|}{n} \leq \widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B}) + \frac{\alpha}{n\,(\alpha-1)} \log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{18.3.299}$$

where $\widehat{\Upsilon}_\alpha(\mathcal{N}_{A\to B})$ is the geometric Υ-information of $\mathcal{N}_{A\to B}$, as defined in (18.3.254), with $\boldsymbol{D}$ set to $\widehat{D}_\alpha$.

Proof: Consider an arbitrary $(n, |\mathcal{M}|, \varepsilon)$ protocol of the form described in Section 18.1, with final state as given in (18.3.297). Let the distribution p over the messages be the uniform distribution. Since the condition $p^*_{\text{err}}(\mathcal{C}) \leq \varepsilon$ holds, with p^*_{err} defined in (18.1.12), we can apply (18.3.274) of Proposition 18.25 to conclude that

$$\log_2 |\mathcal{M}| \leq \inf_{\mathcal{M}_{M'\to\hat{M}}:\beta(\mathcal{M}_{M'\to\hat{M}})\leq 1} \widehat{D}_\alpha(\mathcal{P}_{M'\to\hat{M}}(\overline{\Phi}_{MM'}) \| \mathcal{M}_{M'\to\hat{M}}(\overline{\Phi}_{MM'}))$$

$$+ \frac{\alpha}{\alpha - 1} \log_2\!\left(\frac{1}{1-\varepsilon}\right) \tag{18.3.300}$$

$$\leq \widehat{\Upsilon}_\alpha(\mathcal{P}_{M' \to \hat{M}}) + \frac{\alpha}{\alpha - 1} \log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{18.3.301}$$

where the second inequality follows from the definition in with $\boldsymbol{D}$ set to $\widehat{D}_\alpha$. Eq. (18.3.298) indicates that the whole protocol is a serial composition of bipartite channels. Then we find that

$$\widehat{\Upsilon}_\alpha(\mathcal{P}_{M' \to \hat{M}})$$

$$= \widehat{\Upsilon}_\alpha(\mathcal{D}^n \circ \mathcal{N} \circ \mathcal{E}^{n-1} \circ \overline{\Delta} \circ \mathcal{D}^{n-1} \circ \mathcal{N} \circ \mathcal{E}^{n-2} \circ \overline{\Delta} \circ \mathcal{D}^{n-2} \circ \tag{18.3.302}$$

$$\cdots \circ \mathcal{D}^2 \circ \mathcal{N} \circ \mathcal{E}^1 \circ \overline{\Delta} \circ \mathcal{D}^1 \circ \mathcal{N} \circ \mathcal{E}^0 \circ \mathcal{A}) \tag{18.3.303}$$

$$\leq n\widehat{\Upsilon}_\alpha(\mathcal{N}) + n\widehat{\Upsilon}_\alpha(\overline{\Delta}) + \sum_{i=1}^{n} \widehat{\Upsilon}_\alpha(\mathcal{D}^i) + \sum_{i=0}^{n-1} \widehat{\Upsilon}_\alpha(\mathcal{E}^i) + \widehat{\Upsilon}_\alpha(\mathcal{A}) \tag{18.3.304}$$

$$= n\widehat{\Upsilon}_\alpha(\mathcal{N}). \tag{18.3.305}$$

The inequality follows from Proposition 18.22. The last equality follows from Propositions 18.18, 18.20, and 18.19 because each encoding channel $\mathcal{E}^i$ and decoding channel $\mathcal{D}^i$ is a local channel and $\overline{\Delta}$ is a classical feedback channel. We also implicitly used the stability property in Proposition 18.18. Putting everything together, we conclude that

$$\log_2 |\mathcal{M}| \leq n\widehat{\Upsilon}_\alpha(\mathcal{N}) + \frac{\alpha}{\alpha - 1} \log_2\!\left(\frac{1}{1-\varepsilon}\right), \tag{18.3.306}$$

which is equivalent to the desired bound in (18.3.299). ■

18.4 Classical Capacity of a Quantum Channel Assisted by Classical Feedback

In this section, we analyze the asymptotic case of feedback-assisted communication, in which we allow for an arbitrary large number of rounds of feedback. The definitions in this section are similar to those in previous chapters, and so we keep this section brief.

Definition 18.27 Achievable Rate for Classical-Feedback-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an achievable rate for classical-feedback-assisted classical communication over $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, all $\delta > 0$, and all sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ classical-feedback-assisted classical communication protocol.

Definition 18.28 Classical-Feedback-Assisted Classical Capacity of a Quantum Channel

The classical-feedback-assisted classical capacity of a quantum channel $\mathcal{N}$, denoted by $C_{\text{CFB}}(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$C_{\text{CFB}}(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{18.4.1}$$

Definition 18.29 Strong Converse Rate for Classical-Feedback-Assisted Classical Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a strong converse rate for classical-feedback-assisted classical communication over $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, all $\delta > 0$, and all sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ classical-feedback-assisted classical communication protocol.

Definition 18.30 Strong Converse Classical-Feedback-Assisted Classical Capacity of a Quantum Channel

The strong converse classical-feedback-assisted classical capacity of a quantum channel $\mathcal{N}$, denoted by $\widetilde{C}_{\text{CFB}}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{C}_{\text{CFB}}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{18.4.2}$$

We conclude several theorems, based on the bounds given in Section 16.32:

Theorem 18.31 Classical-Feedback-Assisted Classical Capacity of Entanglement-Breaking Channels

For an entanglement-breaking channel $\mathcal{N}$, its classical-feedback-assisted classical capacity $C_{\text{CFB}}(\mathcal{N})$ and its strong converse quantum-feedback-assisted classical capacity are both equal to its Holevo information $\chi(\mathcal{N})$, i.e.,

$$C_{\text{CFB}}(\mathcal{N}) = \widetilde{C}_{\text{CFB}}(\mathcal{N}) = \chi(\mathcal{N}), \tag{18.4.3}$$

where $\chi(\mathcal{N})$ is defined in (7.11.106).

Proof: The lower bound $\chi(\mathcal{N}) \leq C_{\text{CFB}}(\mathcal{N})$ follows from Theorem 12.13 (i.e., not making use of the classical feedback channel at all). The upper bound $\widetilde{C}_{\text{CFB}}(\mathcal{N}) \leq \chi(\mathcal{N})$ follows from (18.3.6) of Theorem 18.2, and by reasoning similar to that given in the proof of Theorem 12.19. ■

Theorem 18.32 Average Entropy Weak Converse Bound for Classical Capacity Assisted by Classical Feedback

Let $\mathcal{N}_{A\to B}$ be a quantum channel that can be written as the following probabilistic mixture of quantum channels:

$$\mathcal{N}_{A\to B} = \sum_x p_X(x)\mathcal{N}^x_{A\to B}, \tag{18.4.4}$$

where p_X is a probability distribution and $\{\mathcal{N}^x_{A\to B}\}_x$ is a set of quantum channels. The following upper bound holds for the classical capacity of a quantum channel assisted by classical feedback:

$$C_{\text{CFB}}(\mathcal{N}_{A\to B}) \leq \sup_{\rho_A} \sum_x p_X(x) H(\mathcal{N}^x_{A\to B}(\rho_A)). \tag{18.4.5}$$

Proof: This is a direct consequence of Theorem 18.6 and reasoning similar to that given for the proof around (12.2.43). ■

Theorem 18.33 **Geometric Υ-Information Strong Converse Bound for Classical Capacity Assisted by Classical Feedback**

The following upper bound holds for the strong converse classical capacity of a quantum channel $\mathcal{N}_{A\to B}$ assisted by classical feedback:

$$\widetilde{C}_{\text{CFB}}(\mathcal{N}_{A\to B}) \le \widehat{\Upsilon}(\mathcal{N}_{A\to B}), \tag{18.4.6}$$

where $\widehat{\Upsilon}(\mathcal{N}_{A\to B})$ is the Υ-information defined from the Belavkin–Staszewski relative entropy (see (18.3.254) and Definition 7.51).

Proof: This is a direct consequence of the upper bound in Theorem 18.26 and reasoning similar to that given in the proof of Theorem . We also require the fact that the geometric Rényi relative entropy converges to the Belavkin–Staszewski relative entropy in the limit as $\alpha \to 1$ (see Proposition 7.52). ■

18.5 Examples

In this section, we briefly provide some examples of channels for which we evaluate the capacity upper bounds in Section 18.4. We begin with the quantum erasure channel (see Section 4.5.2). Recall that a quantum erasure channel acts as follows on an input density operator ρ:

$$\mathcal{E}_p(\rho) \coloneqq (1-p)\,\rho + p|e\rangle\langle e|, \tag{18.5.1}$$

where $p \in [0,1]$ is the erasure probability and $|e\rangle\langle e|$ is an erasure state orthogonal to every possible input. Let d be the dimension of the input to the channel. By inspection, we see that the erasure channel is a probabilistic mixture of an identity channel and a channel that traces out the input and replaces with the erasure state. Thus, we apply Theorem 18.32 to conclude that

$$\begin{aligned} C_{\text{CFB}}(\mathcal{E}_p) &\le \sup_{\rho_A} \left[(1-p)\,H(\text{id}(\rho_A)) + pH(|e\rangle\langle e|)\right] &(18.5.2)\\ &= (1-p)\sup_{\rho_A} H(\text{id}(\rho_A)) &(18.5.3)\\ &= (1-p)\log_2 d. &(18.5.4)\end{aligned}$$

Since this upper bound is an achievable for classical communication over the erasure channel without feedback (see Theorem 12.33), we then conclude that

$$C(\mathcal{E}_p) = C_{\text{CFB}}(\mathcal{E}_p) = (1-p)\log_2 d. \tag{18.5.5}$$

That is, classical feedback does not increase the classical capacity of the erasure channel.

Finally, we evaluate the bound in Theorem 18.33 for the qubit depolarizing channel. Recall from Section 16.32 that it is defined as

$$\mathcal{D}^p(X) \coloneqq (1-p)\,X + p\,\mathrm{Tr}[X]\pi, \tag{18.5.6}$$

$$\pi \coloneqq I/d. \tag{18.5.7}$$

It was already established in Section 16.32 that $\Upsilon(\mathcal{D}^p)$ is an upper bound on its (unassisted) classical capacity, and we discussed in Section 16.32 how the Holevo information is equal to its classical capacity. What we find now is that $\Upsilon(\mathcal{D}^p)$ is an upper bound on its classical capacity assisted by a classical feedback channel. Figure 18.1 plots this upper bound and also plots the Holevo information lower bound when $d = 2$. The latter is given by $1 - h_2(p/2)$, where h_2 is the binary entropy function. Note that the depolarizing channel is entanglement breaking for $p \geq \frac{d}{d+1}$. As such, the bounds from Theorem 18.31 apply, so that, for $p \geq \frac{d}{d+1}$, the Holevo information $1 - h_2(p/2)$ is equal to the classical capacity assisted by classical feedback.

18.6 Summary

In this chapter, we developed the general theory of classical communication over a quantum channel assisted by classical feedback from receiver to sender. Our main focus was on establishing upper bounds on this capacity. The main findings of this chapter are as follows:

1. We first proved that classical feedback does not enhance the classical capacity of an entanglement-breaking channel.

2. Next, we established that the average output entropy of a channel is a weak converse upper bound on the feedback-assisted capacity. The method for establishing this average entropy bound involves identifying an information

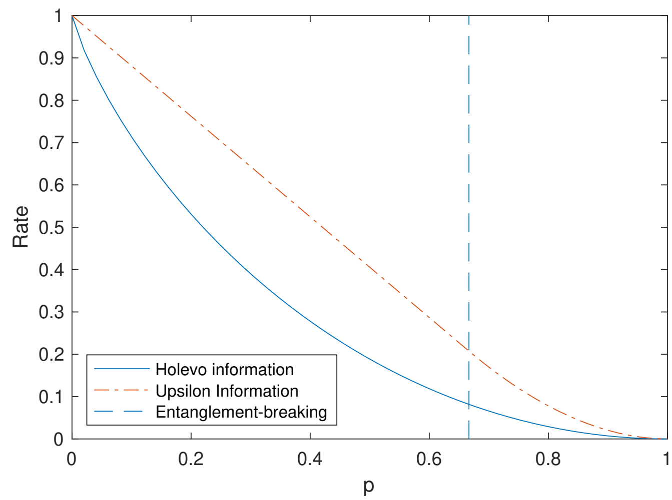


FIGURE 18.1: Lower and upper bounds on the classical-feedback-assisted classical capacity of the qubit depolarizing channel in (18.5.6), with $d = 2$. The dashed vertical line indicates that the qubit depolarizing channel is entanglement breaking for $p \geq 2/3$, so that the Holevo information is equal to the feedback-assisted capacity for these values, according to Theorem 18.31.

measure that has two key properties: 1) it does not increase under a one-way local operations and classical communication channel from the receiver to the sender and 2) a quantum channel from sender to receiver cannot increase the information measure by more than the maximum average output entropy of the channel. This information measure can be understood as the sum of two terms, with one corresponding to classical correlation and the other to entanglement.

3. We finally established a general strong converse upper bound on the feedback-assisted capacity, in terms of the geometric Υ-information of a quantum channel. The main method for doing was to devise an information measure for bipartite channels that is equal to zero for classical feedback channels and products of local channels.

18.7 Bibliographic Notes

The classical capacity of a quantum channel assisted by a classical feedback channel was first studied by Bowen and Nagarajan (2005), who proved that classical feedback does not increase the capacity of entanglement-breaking channels. This result was strengthened to a strong converse statement by Ding and Wilde (2018). Smith and Smolin (2009) provided an example of a channel for which classical feedback can signficantly enhance the classical capacity. Bennett et al. (2006) related the feedback-assisted capacity to other capacities in quantum Shannon theory, and García-Patrón et al. (2018) related it to other notions of feedback-assisted capacity. Ding et al. (2019) established the entropy upper bound on the feedback-assisted capacity, and Ding et al. (2023) established the geometric Υ-information upper bound on the strong converse feedback-assisted capacity.

Chapter 19

LOCC-Assisted Quantum Communication

This chapter develops an important variation of quantum communication, in which we allow the sender and receiver the free use of classical communication. That is, in between every use of a given quantum communication channel $\mathcal{N}_{A\to B}$, the sender and receiver are allowed to perform local operations and classical communication (LOCC). For this reason, the capacity considered in this chapter is called the LOCC-assisted quantum capacity.

The practical motivation for this kind of feedback-assisted quantum capacity comes from the fact that, these days, classical communication is rather cheap and plentiful. Thus, from a resource-theoretic perspective, it can be sensible to simply allow classical communication as a free resource (similar to how we did for entanglement-assisted communication in Chapter 11). Then our goal is to place informative bounds on the rate at which quantum information can be communicated from the sender to the receiver in this setting. Furthermore, these bounds are relevant for understanding and placing limitations on the speed at which distributed quantum computation can be carried out.

In order to establish upper bounds on LOCC-assisted quantum capacity, we revisit the concept of amortization introduced in Section 17.1.3. However, in this context, we proceed somewhat differently, instead employing entanglement measures to quantify how much entanglement can be generated by multiple uses of a quantum channel. Then we define the amortized entanglement of a quantum channel as the largest difference between the output and input entanglement of the channel.

In particular, two entanglement measures on which we focus are the squashed entanglement and the Rains relative entropy, as well as a variant of the latter called max-Rains relative entropy. One key property of the squashed entanglement and the max-Rains relative entropy is that these entanglement measures do not increase under amortization, similar to how we previously observed that the mutual information of a channel does not increase under amortization. This key property leads to the conclusion that these entanglement measures can serve as upper bounds on the LOCC-assisted quantum capacity of a quantum channel, and arriving at this conclusion is one of the main goals of this chapter. At the end of the chapter, we demonstrate the utility of the squashed entanglement and Rains family of bounds by evaluating them for several example quantum channels of interest.

Combining Entanglement Distillation and Teleportation to Obtain a Quantum Communication Protocol

We can use entanglement distillation along with teleportation in the asymptotic setting to obtain a lower bound on the number of transmitted qubits in a quantum communication protocol. The strategy is as follows; see Figure 19.1.

1. Alice prepares several copies, say n, of a pure state $\psi_{\tilde{A}A}$, with the dimension of $\tilde{A}$ equal to the dimension of A, and sends each of the A systems through the channel $\mathcal{N}_{A\to B}$ to Bob.

2. Alice and Bob now share n copies of the state $\omega_{\tilde{A}B} = \mathcal{N}_{A\to B}(\psi_{\tilde{A}A})$. They perform a one-way entanglement distillation protocol to convert these mixed entangled states to a perfect, maximally entangled state $\Phi_{\hat{A}\hat{B}}$ of Schmidt rank $d \geq 2$. Roughly speaking, as shown in Chapter 13, a Schmidt rank of $2^{nI(\tilde{A}\rangle B)_\omega}$ is achievable for n copies of $\omega_{\tilde{A}B}$, as $n \to \infty$.

3. Using the distilled maximally entangled state, along with $2\log_2 d$ bits of classical communication, Alice and Bob perform the quantum teleportation protocol to transmit the A' part of an arbitrary pure state $\Psi_{RA'}$ from Alice to Bob, with $d_{A'} = d$.

Since there are n uses of the channel in this strategy, we see that as $n \to \infty$, the *rate* of this strategy (the number of qubits per channel use) is $\frac{\log_2 d}{n} = I(A\rangle B)_\omega$. By optimizing over all initial pure states $\psi_{\tilde{A}A}$ prepared by Alice, we find that, in the asymptotic setting, the rate $\sup_{\psi_{\tilde{A}A}} I(\tilde{A}\rangle B)_\omega = I^c(\mathcal{N})$ is achievable, where we

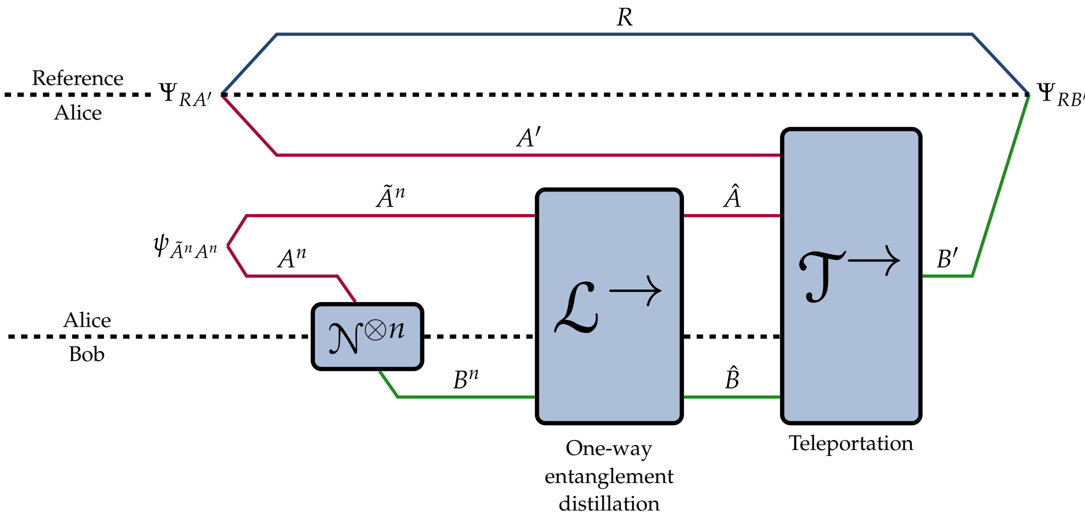


FIGURE 19.1: Sketch of a quantum communication protocol over n uses of the channel $\mathcal{N}$, which makes use of entanglement distillation and teleportation. The arrow "$\rightarrow$" indicates that the channel contains classical communication from Alice to Bob only. We show in [] that, in the asymptotic limit, this strategy achieves the quantum capacity of $\mathcal{N}$.

recognize the coherent information $I^c(\mathcal{N})$ of the channel $\mathcal{N}$, which we define in (7.11.107). Note that the strategy we have outlined here also involves classical communication between Alice and Bob during the entanglement distillation step and during the teleportation step. The results of Section 14.1.3 and Section 14.2.1 show that this strategy is nonetheless a valid quantum communication protocol, in the sense that the rate $I^c(\mathcal{N})$ can be achieved by a strategy that does not make use of forward classical communication.

Now, Alice can do better than prepare the state $\psi_{\tilde{A}A}^{\otimes n}$ and send each A system through the channel: she can prepare a state $\psi_{\tilde{A}_1\cdots\tilde{A}_n A_1\cdots A_n} \equiv \psi_{\tilde{A}^n A^n}$ such that the systems $A_1, \dots, A_n$ being sent through $\mathcal{N}$ are *entangled*. One can then achieve a rate of $\frac{1}{n}\sup_{\psi_{\tilde{A}^n A^n}} I(\tilde{A}^n \rangle B^n)_\omega = \frac{1}{n} I^c(\mathcal{N}^{\otimes n})$. Since we are free to use the channel $\mathcal{N}$ as many times as we want, we can optimize over n to obtain the communication rate

$$\sup_{n\in\mathbb{N}} \frac{1}{n} I^c(\mathcal{N}^{\otimes n}) =: I^c_{\text{reg}}(\mathcal{N}), \tag{19.0.1}$$

and it is this *regularized* coherent information of $\mathcal{N}$ that is optimal for quantum communication over the channel $\mathcal{N}$. In other words, $Q(\mathcal{N}) = I^c_{\text{reg}}(\mathcal{N})$, and we prove this in Section 14.2.2.

19.1 n-Shot LOCC-Assisted Quantum Communication Protocol

This section discusses the most general form for an n-shot LOCC-assisted quantum communication protocol.

Before starting, we should clarify that the goal of such a protocol is to produce an approximate maximally entangled state, with the Schmidt rank of the ideal target state as large as possible. Due to the assumption of free classical communication, as well as the quantum teleportation protocol discussed in Section 5.1, generating an approximate maximally entangled state is equivalent to generating an approximate identity quantum channel, such that the dimension of the ideal target identity channel is equal to the Schmidt rank of the target maximally entangled state. To make this statement more quantitative, suppose that $\omega_{A'B'}$ is an approximate maximally entangled state; i.e., suppose that it is ε-close in normalized trace distance to a maximally entangled state $\Phi_{A'B'}$ of Schmidt rank d:

$$\frac{1}{2}\|\omega_{A'B'} - \Phi_{A'B'}\|_1 \leq \varepsilon. \tag{19.1.1}$$

Let $\mathcal{T}_{AA'B'\to B}$ denote the one-way LOCC channel corresponding to quantum teleportation (as discussed around (5.1.24)). Then by applying (5.1.25), it follows that teleportation over the ideal resource state $\Phi_{A'B'}$ realizes an identity channel $\mathrm{id}_{A\to B}$ of dimension d:

$$\mathcal{T}_{AA'B'\to B}((\cdot)\otimes\Phi_{A'B'}) = \mathrm{id}_{A\to B}(\cdot). \tag{19.1.2}$$

Let $\mathcal{T}^{\omega}_{A\to B}$ denote the channel realized by teleportation over the unideal state $\omega_{A'B'}$:

$$\mathcal{T}^{\omega}_{A\to B}(\cdot) := \mathcal{T}_{AA'B'\to B}((\cdot)\otimes\omega_{A'B'}). \tag{19.1.3}$$

We would then like to determine the deviation of the ideal channel from $\mathcal{T}^{\omega}_{A\to B}$, and to do so, we can employ the normalized diamond distance. Then consider that, from the data-processing inequality for trace distance,

$$\begin{aligned}
&\left\|\mathcal{T}^{\omega}_{A\to B} - \mathrm{id}_{A\to B}\right\|_{\diamond}\\
&= \sup_{\psi_{RA}}\left\|\mathcal{T}^{\omega}_{A\to B}(\psi_{RA}) - \mathrm{id}_{A\to B}(\psi_{RA})\right\|_1 &(19.1.4)\\
&= \sup_{\psi_{RA}}\|\mathcal{T}_{AA'B'\to B}(\psi_{RA}\otimes\omega_{A'B'}) - \mathcal{T}_{AA'B'\to B}(\psi_{RA}\otimes\Phi_{A'B'})\|_1 &(19.1.5)
\end{aligned}$$

$$\leq \sup_{\psi_{RA}} \|\psi_{RA} \otimes \omega_{A'B'} - \psi_{RA} \otimes \Phi_{A'B'}\|_1 \tag{19.1.6}$$

$$= \|\omega_{A'B'} - \Phi_{A'B'}\|_1 \leq 2\varepsilon, \tag{19.1.7}$$

so that we arrive at the desired statement mentioned above:

$$\frac{1}{2}\|\omega_{A'B'} - \Phi_{A'B'}\|_1 \leq \varepsilon \qquad \Rightarrow \qquad \frac{1}{2}\left\|\mathcal{T}^{\omega}_{A\to B} - \mathrm{id}_{A\to B}\right\|_\diamond \leq \varepsilon. \tag{19.1.8}$$

Thus, for the above reason, we focus exclusively on LOCC-assisted protocols whose aim is to generate an approximate maximally entangled state. In what follows, all bipartite cuts for separable states or LOCC channels should be understood as being between Alice's and Bob's systems.

A protocol for LOCC-assisted quantum communication is depicted in Figure [REF], and it is defined by the following elements:

$$(\rho^{(1)}_{A_1'A_1B_1'}, \{\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'}\}_{i=2}^n, \mathcal{L}^{(n+1)}_{A_n'B_nB_n'\to M_AM_B}), \tag{19.1.9}$$

where $\rho^{(1)}_{A_1'A_1B_1'}$ is a separable state, $\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'}$ is an LOCC channel for $i \in \{2, \ldots, n\}$, and $\mathcal{L}^{(n+1)}_{A_n'B_nB_n'\to M_AM_B}$ is a final LOCC channel that generates the approximate maximally entangled state in systems M_A and M_B. Let $\mathcal{C}$ denote all of these elements, which together constitute the LOCC-assisted quantum communication code. All systems with primed labels should be understood as local quantum memory or scratch registers that Alice or Bob can employ in this information-processing task. They are also assumed to be finite-dimensional, but could be arbitrarily large. The unprimed systems are the ones that are either input to or output from the quantum communication channel $\mathcal{N}_{A\to B}$.

The LOCC-assisted quantum communication protocol begins with Alice and Bob performing an LOCC channel $\mathcal{L}^{(1)}_{\emptyset\to A_1'A_1B_1'}$, which leads to the separable state $\rho^{(1)}_{A_1'A_1B_1'}$ mentioned above, where A_1' and B_1' are systems that are finite-dimensional but arbitrarily large. The system A_1 is such that it can be fed into the first channel use. Alice sends system A_1 through the first channel use, leading to a state

$$\omega^{(1)}_{A_1'B_1B_1'} := \mathcal{N}_{A_1\to B_1}(\rho^{(1)}_{A_1'A_1B_1'}). \tag{19.1.10}$$

Alice and Bob then perform the LOCC channel $\mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'}$, which leads to the state

$$\rho^{(2)}_{A_2'A_2B_2'} := \mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'}(\omega^{(1)}_{A_1'B_1B_1'}). \tag{19.1.11}$$

Alice sends system A_2 through the second channel use $\mathcal{N}_{A_2\to B_2}$, leading to the state

$$\omega^{(2)}_{A'_2B_2B'_2} \coloneqq \mathcal{N}_{A_2\to B_2}(\rho^{(2)}_{A'_2A_2B'_2}). \tag{19.1.12}$$

This process iterates: the protocol uses the channel n times. In general, we have the following states for all $i \in \{2, \ldots, n\}$:

$$\rho^{(i)}_{A'_iA_iB'_i} \coloneqq \mathcal{L}^{(i)}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iB'_i}(\omega^{(i-1)}_{A'_{i-1}B_{i-1}B'_{i-1}}), \tag{19.1.13}$$

$$\omega^{(i)}_{A'_iB_iB'_i} \coloneqq \mathcal{N}_{A_i\to B_i}(\rho^{(i)}_{A'_iA_iB'_i}), \tag{19.1.14}$$

where $\mathcal{L}^{(i)}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iB'_i}$ is an LOCC channel. The final step of the protocol consists of an LOCC channel $\mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to M_AM_B}$, which generates the systems M_A and M_B for Alice and Bob, respectively. The protocol's final state is as follows:

$$\omega_{M_AM_B} \coloneqq \mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to M_AM_B}(\omega^{(n)}_{A'_nB_nB'_n}). \tag{19.1.15}$$

The goal of the protocol is for the final state $\omega_{M_AM_B}$ to be close to a maximally entangled state, and we define the quantum error probability of the code as follows:

$$q_{\text{err}}(\mathcal{C}) \coloneqq 1 - F(\omega_{M_AM_B}, \Phi_{M_AM_B}) \tag{19.1.16}$$

$$= 1 - \langle\Phi|_{M_AM_B}\omega_{M_AM_B}|\Phi\rangle_{M_AM_B}, \tag{19.1.17}$$

where F denotes the quantum fidelity (Definition 6.5) and the maximally entangled state $\Phi_{M_AM_B} = |\Phi\rangle\langle\Phi|_{M_AM_B}$ is defined from

$$|\Phi\rangle_{M_AM_B} \coloneqq \frac{1}{\sqrt{M}}\sum_{m=1}^{M} |m\rangle_{M_A} \otimes |m\rangle_{M_B}, \tag{19.1.18}$$

such that it has Schmidt rank M. Intuitively, the quantum error probability $q_{\text{err}}(\mathcal{C})$ is equal to the probability that one obtains the outcome "not maximally entangled state $\Phi_{M_AM_B}$" when performing the test or measurement $\{\Phi_{M_AM_B}, I_{M_AM_B} - \Phi_{M_AM_B}\}$ on the final state $\omega_{M_AM_B}$ of the protocol.

Definition 19.1 **(n, M, ε) LOCC-Assisted Quantum Communication Protocol**

Let $\mathcal{C} := (\rho^{(1)}_{A_1' A_1 B_1'}, \{\mathcal{L}^{(i)}_{A_{i-1}' B_{i-1} B_{i-1}' \to A_i' A_i B_i'}\}_{i=2}^{n}, \mathcal{L}^{(n+1)}_{A_n' B_n B_n' \to M_A M_B})$ be the elements of an n-round LOCC-assisted quantum communication protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an (n, M, ε) protocol, with $\varepsilon \in [0,1]$, if $q_{\text{err}}(\mathcal{C}) \leq \varepsilon$.

19.1.1 Lower Bound on the Number of Transmitted Qubits

[IN PROGRESS]

one-shot lower bound in terms of coherent information of a state. This achieves coherent information of a channel, as well as reverse coherent information of a channel.

19.1.2 Amortized Entanglement as a General Upper Bound for LOCC-Assisted Quantum Communication Protocols

It is an interesting question to determine whether the inequality opposite to that in Lemma 10.4 holds, and as the following proposition demonstrates, this question is intimately related to finding useful upper bounds on the rate at which maximal entanglement can be distilled by employing an LOCC-assisted quantum communication protocol. The following proposition represents our fundamental starting point when analyzing limitations on LOCC-assisted quantum communication protocols.

Proposition 19.2

Let $\mathcal{N}_{A\to B}$ be a quantum channel, let $\varepsilon \in [0,1]$, and let E be an entanglement measure that is equal to zero for all separable states. For an (n, M, ε) LOCC-assisted quantum communication protocol with final state $\omega_{M_A M_B}$, the following bound holds

$$E(M_A; M_B)_\omega \leq n \cdot E^{\mathcal{A}}(\mathcal{N}). \tag{19.1.19}$$

PROOF: Consider an LOCC-assisted quantum communication protocol as presented in Section 19.1. From the monotonicity of the entanglement measure E with respect to LOCC channels, we find that

$$
\begin{aligned}
&E(M_A;M_B)_\omega \\
&\quad \leq E(A'_n;B_nB'_n)_{\omega^{(n)}} && (19.1.20)\\
&\quad = E(A'_n;B_nB'_n)_{\omega^{(n)}} - E(A'_1A_1;B'_1)_{\rho^{(1)}} && (19.1.21)\\
&\quad = E(A'_n;B_nB'_n)_{\omega^{(n)}} + \left[\sum_{i=2}^{n} E(A'_iA_i;B'_i)_{\rho^{(i)}} - E(A'_iA_i;B'_i)_{\rho^{(i)}}\right]\\
&\qquad\quad - E(A'_1A_1;B'_1)_{\rho^{(1)}} && (19.1.22)\\
&\quad \leq \sum_{i=1}^{n}\left[E(A'_i;B_iB'_i)_{\omega^{(i)}} - E(A'_iA_i;B'_i)_{\rho^{(i)}}\right] && (19.1.23)\\
&\quad \leq n\cdot E^{\mathcal{A}}(\mathcal{N}). && (19.1.24)
\end{aligned}
$$

The first equality follows because the state $\rho^{(1)}_{A'_1A_1B'_1}$ is a separable state, and by assumption, the entanglement measure E vanishes for all such states. The second equality follows by adding and subtracting equal terms. The second inequality follows because $E(A'_iA_i;B'_i)_{\rho^{(i)}} \leq E(A'_{i-1};B_{i-1}B'_{i-1})_{\omega^{(i-1)}}$ for all $i \in \{2,\ldots,n\}$, due to monotonicity of the entanglement measure E with respect to LOCC channels. The final inequality follows from the definition of amortized entanglement and the fact that the states $\omega^{(i)}_{A'_iB_iB'_i}$ and $\rho^{(i)}_{A'_iA_iB'_i}$ are particular states to consider in its optimization. ■

The inequality in (19.1.19) states that the entanglement of the final output state $\omega_{M_AM_B}$, as quantified by E, cannot exceed n times the amortized entanglement of the channel $\mathcal{N}_{A\to B}$. Intuitively, the only resource allowed in the protocol, which has the potential to generate entanglement, is the quantum communication channel $\mathcal{N}_{A\to B}$. All of the LOCC channels allowed for free have no ability to generate entanglement on their own. Thus, the entanglement of the final state should not exceed the largest possible amount of entanglement that could ever be generated with n calls to the channel, and this largest entanglement is exactly the amortized entanglement of the channel.

Observe that the bound in Proposition 19.2 depends on the final state $\omega_{M_AM_B}$, and thus it is not a universal bound, depending only on the parameters n, M, and ε, because this state in turn depends on the entire protocol. Similar to the upper bounds established in previous chapters, it is desirable to refine this bound such that it depends only on n, M, and ε, which are the parameters characterizing any generic LOCC-assisted quantum communication protocol. In the forthcoming sections, we consider particular entanglement measures, such as squashed entanglement and Rains relative entropy, which allow us to relate the parameters M and ε to the final state $\omega_{M_A M_B}$.

To end this section, we note here that the bound in Proposition 19.2 simplifies for teleportation-simulable channels and for entanglement measures that are subadditive with respect to states and equal to zero for all separable states. This conclusion is a consequence of Propositions 10.6 and 19.2:

Corollary 19.3 Reduction by Teleportation

Let E_S be an entanglement measure that is subadditive with respect to states and equal to zero for all separable states. Let $\mathcal{N}_{A \to B}$ be a channel that is teleportation-simulable with associated resource state $\theta_{RB'}$. Let $\varepsilon \in [0, 1]$. For an (n, M, ε) LOCC-assisted quantum communication protocol with final state $\omega_{M_A M_B}$, the following bound holds

$$E_S(M_A; M_B)_\omega \leq n \cdot E_S(R; B')_\theta. \tag{19.1.25}$$

19.1.3 Squashed Entanglement Upper Bound on the Number of Transmitted Qubits

We now establish the squashed entanglement upper bound on the number of qubits that a sender can transmit to a receiver by employing an LOCC-assisted quantum communication protocol:

Theorem 19.4 *n*-Shot Squashed Entanglement Upper Bound

Let $\mathcal{N}_{A \to B}$ be a quantum channel, and let $\varepsilon \in [0, 1)$. For all (n, M, ε) LOCC-assisted quantum communication protocols over the channel $\mathcal{N}_{A \to B}$, the following bound holds

$$\log_2 M \leq \frac{1}{1 - \sqrt{\varepsilon}} \left[n \cdot E_{\text{sq}}(\mathcal{N}) + g_2(\sqrt{\varepsilon}) \right]. \tag{19.1.26}$$

PROOF: Consider an arbitrary (n, M, ε) LOCC-assisted quantum communication protocol over the channel $\mathcal{N}_{A\to B}$, as defined in Section 19.1. The squashed entanglement is an entanglement measure (monotone under LOCC as shown in Theorem 9.33) and it is equal to zero for separable states (Proposition 9.4.5). Thus, Proposition 19.2 applies, and we find that

$$E_{\text{sq}}(M_A; M_B)_\omega \leq n \cdot E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) = n \cdot E_{\text{sq}}(\mathcal{N}), \tag{19.1.27}$$

where the equality follows from Theorem 10.20. Applying Definition 19.1 leads to

$$F(\Phi_{M_A M_B}, \omega_{M_A M_B}) \geq 1 - \varepsilon. \tag{19.1.28}$$

As a consequence of Proposition 9.38, we find that

$$\begin{aligned} &E_{\text{sq}}(M_A; M_B)_\omega \\ &\geq E_{\text{sq}}(M_A; M_B)_\Phi - \left[\sqrt{\varepsilon} \log_2 \min\{|M_A|, |M_B|\} + g_2(\sqrt{\varepsilon})\right] && (19.1.29) \\ &= \log_2 M - \left[\sqrt{\varepsilon} \log_2 M + g_2(\sqrt{\varepsilon})\right] && (19.1.30) \\ &= (1 - \sqrt{\varepsilon}) \log_2 M - g_2(\sqrt{\varepsilon}). && (19.1.31) \end{aligned}$$

The first equality follows from Proposition 9.36. We can finally rearrange the established inequality $n \cdot E_{\text{sq}}(\mathcal{N}) \geq (1 - \sqrt{\varepsilon}) \log_2 M - g_2(\sqrt{\varepsilon})$ to be in the form stated in the theorem. ■

19.2 *n*-Shot PPT-Assisted Quantum Communication Protocol

Recalling the completely PPT-preserving channels of Definition 4.27 as forming a superset of LOCC channels, we can also consider quantum communication protocols assisted by completely PPT-preserving channels (abbreviated as C-PPT-P channels). Such a PPT-assisted protocol has exactly the same form as given in Section 19.1, but it is instead defined by the following elements:

$$(\rho^{(1)}_{A'_1 A_1 B'_1}, \{\mathcal{P}^{(i)}_{A'_{i-1} B_{i-1} B'_{i-1} \to A'_i A_i B'_i}\}_{i=2}^n, \mathcal{P}^{(n+1)}_{A'_n B_n B'_n \to M_A M_B}), \tag{19.2.1}$$

where $\rho^{(1)}_{A'_1 A_1 B'_1}$ is a PPT state, $\mathcal{P}^{(i)}_{A'_{i-1} B_{i-1} B'_{i-1} \to A'_i A_i B'_i}$ is a C-PPT-P channel for $i \in \{2, \ldots, n\}$, and $\mathcal{P}^{(n+1)}_{A'_n B_n B'_n \to M_A M_B}$ is a final C-PPT-P channel that generates an approximate maximally entangled state in systems M_A and M_B. Denoting the final state of the protocol again by $\omega_{M_A M_B}$, the criterion for such a protocol is again given by (19.1.16). We then arrive at the following definition:

Definition 19.5 (n, M, ε) PPT-Assisted Quantum Communication Protocol

Let $\mathcal{C} := (\rho^{(1)}_{A'_1 A_1 B'_1}, \{\mathcal{P}^{(i)}_{A'_{i-1} B_{i-1} B'_{i-1} \to A'_i A_i B'_i}\}_{i=2}^n, \mathcal{P}^{(n+1)}_{A'_n B_n B'_n \to M_A M_B})$ be the elements of an n-round PPT-assisted quantum communication protocol over the channel $\mathcal{N}_{A \to B}$. The protocol is called an (n, M, ε) protocol, with $\varepsilon \in [0, 1]$, if $q_{\text{err}}(\mathcal{C}) \le \varepsilon$.

Since every LOCC channel is a C-PPT-P channel, we can make the following observation immediately:

REMARK: Every (n, M, ε) LOCC-assisted quantum communication protocol is also an (n, M, ε) PPT-assisted quantum communication protocol.

As a consequence of the above observation, any converse bound or limitation that we establish for an arbitrary (n, M, ε) PPT-assisted protocol is also a converse bound for an (n, M, ε) LOCC-assisted protocol.

By the exact same reasoning as in the proof of Proposition 19.2, but replacing LOCC channels with C-PPT-P channels, we arrive at the following proposition:

Proposition 19.6

Let $\mathcal{N}_{A \to B}$ be a quantum channel, let $\varepsilon \in [0, 1]$, and let E be an entanglement measure that is monotone under completely PPT-preserving channels and is equal to zero for all PPT states. For an (n, M, ε) PPT-assisted quantum communication protocol with final state $\omega_{M_A M_B}$, the following bound holds

$$E(M_A; M_B)_\omega \le n \cdot E^{\mathcal{A}}(\mathcal{N}). \tag{19.2.2}$$

Recalling Definition 4.32, a channel $\mathcal{N}_{A \to B}$ with input system A and output system B is defined to be PPT-simulable with associated resource state $\omega_{RB'}$ if the following equality holds for all input states ρ_A:

$$\mathcal{N}_{A \to B}(\rho_A) = \mathcal{P}_{ARB' \to B}(\rho_A \otimes \omega_{RB'}), \tag{19.2.3}$$

where $\mathcal{P}_{ARB'\to B}$ is a completely PPT-preserving channel between the sender, who has systems A and R, and the receiver, who has system B'.

By the same reasoning used to arrive at Corollary 19.3, but replacing LOCC channels with completely PPT-preserving ones, we find the following:

Corollary 19.7

Let E_S denote an entanglement measure that is monotone non-increasing with respect to completely PPT-preserving channels, subadditive with respect to states, and equal to zero for all PPT states. Let $\mathcal{N}_{A\to B}$ be a channel that is PPT-simulable with associated resource state $\theta_{RB'}$. Let $\varepsilon \in [0,1]$. For an (n, M, ε) PPT-assisted quantum communication protocol with final state $\omega_{M_A M_B}$, the following bound holds

$$E_S(M_A; M_B)_\omega \leq n \cdot E_S(R; B')_\theta. \tag{19.2.4}$$

19.2.1 Rényi–Rains Information Upper Bounds on the Number of Transmitted Qubits

We now establish the max-Rains information upper bound on the number of qubits that a sender can transmit to a receiver by employing a PPT-assisted quantum communication protocol:

Theorem 19.8 ***n*-Shot Max-Rains Upper Bound**

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all (n, M, ε) PPT-assisted quantum communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\log_2 M \leq n \cdot R_{\max}(\mathcal{N}) + \log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{19.2.5}$$

Proof: Consider an arbitrary (n, M, ε) LOCC-assisted quantum communication protocol over the channel $\mathcal{N}_{A\to B}$, as defined in Section 19.1. The max-Rains relative entropy is an entanglement measure (monotone under completely PPT-preserving channels as shown in Proposition 9.28), and it is equal to zero for PPT states. Thus, Proposition 19.6 applies, and we find that

$$R_{\max}(M_A; M_B)_\omega \leq n \cdot R^{\mathcal{A}}_{\max}(\mathcal{N}) = n \cdot R_{\max}(\mathcal{N}), \tag{19.2.6}$$

where the equality follows from Theorem 10.18. Applying Definition 19.5 leads to

$$F(\Phi_{M_A M_B}, \omega_{M_A M_B}) \geq 1 - \varepsilon. \tag{19.2.7}$$

As a consequence of Propositions 13.6 and 7.71, we find that

$$\log_2 M \leq R^{\varepsilon}(M_A; M_B)_\omega \tag{19.2.8}$$

$$\leq R_{\max}(M_A; M_B)_\omega + \log_2\left(\frac{1}{1-\varepsilon}\right). \tag{19.2.9}$$

Combining (19.2.6) and (19.2.9), we conclude the proof. ■

At this point, it is worthwhile to compare the squashed-entanglement bound in Theorem 19.4 with the max-Rains information bound in Theorem 19.8. First, both bounds hold for all quantum channels, and so this is an advantage that they both possess. The squashed entanglement and max-Rains information are rather different quantities, and so the quantities on their own can vary based on the channel for which they are evaluated. The squashed-entanglement bound in Theorem 19.4 is a weak-converse bound, whereas the max-Rains information bound in Theorem 19.8 is a strong-converse bound. The max-Rains information bound has the advantage that it is efficiently computable by a semi-definite program, whereas it is not known how to compute the squashed entanglement. However, one can apply Proposition 9.37 to see that the squashed entanglement bound gives a whole host of upper bounds related to the choice of a squashing channel, and one can potentially obtain tight bounds by making a clever choice of a squashing channel.

For channels that are PPT-simulable with associated resource states, as recalled in (19.2.3), we obtain upper bounds that can be even stronger:

Theorem 19.9 ***n*-Shot Rényi–Rains Upper Bounds for PPT-Simulable Channels**

Let $\mathcal{N}_{A\to B}$ be a quantum channel that is PPT-simulable with associated resource state $\theta_{SB'}$, and let $\varepsilon \in [0, 1)$. For all (n, M, ε) PPT-assisted quantum communication protocols over the channel $\mathcal{N}_{A\to B}$, the following bounds hold for all

$\alpha > 1$:

$$\log_2 M \leq n \cdot \widetilde{R}_\alpha(S; B')_\theta + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{19.2.10}$$

$$\log_2 M \leq \frac{1}{1-\varepsilon}\left[n \cdot R(S; B')_\theta + h_2(\varepsilon)\right]. \tag{19.2.11}$$

Proof: Consider an arbitrary (n, M, ε) LOCC-assisted quantum communication protocol over the channel $\mathcal{N}_{A\to B}$, as defined in Section 19.1. The Rényi–Rains relative entropy and Rains relative entropy are monotone non-increasing under completely PPT-preserving channels (Proposition 9.25), equal to zero for PPT states, and subadditive with respect to states (Proposition 9.25). As such, Corollary 19.7 applies, and we find for $\alpha > 1$ that

$$\widetilde{R}_\alpha(M_A; M_B)_\omega \leq n \cdot \widetilde{R}_\alpha(S; B')_\theta, \tag{19.2.12}$$

$$R(M_A; M_B)_\omega \leq n \cdot R(S; B')_\theta. \tag{19.2.13}$$

Applying Definition 19.5 leads to

$$F(\Phi_{M_A M_B}, \omega_{M_A M_B}) \geq 1 - \varepsilon. \tag{19.2.14}$$

As a consequence of Proposition 13.6, we have that

$$\log_2 M \leq R^\varepsilon(M_A; M_B)_\omega. \tag{19.2.15}$$

Applying Propositions 7.70 and 7.71, we find that

$$\log_2 M \leq \widetilde{R}_\alpha(M_A; M_B)_\omega + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{19.2.16}$$

$$\log_2 M \leq \frac{1}{1-\varepsilon}\left[R(M_A; M_B)_\omega + h_2(\varepsilon)\right]. \tag{19.2.17}$$

Putting together (19.2.12), (19.2.13), (19.2.16), and (19.2.17) concludes the proof. ■

19.3 LOCC- and PPT-Assisted Quantum Capacities of Quantum Channels

In this section, we analyze the asymptotic capacities, and as before, the upper bounds for the asymptotic capacities are straightforward consequences of the non-asymptotic bounds given in Sections 19.1.3 and 19.2.1. The definitions of these capacities are similar to what we have given previously, and so we only state them here briefly.

Definition 19.10 Achievable Rate for LOCC-Assisted Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an achievable rate for LOCC-assisted quantum communication over $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, all $\delta > 0$, and all sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ LOCC-assisted quantum communication protocol.

Definition 19.11 LOCC-Assisted Quantum Capacity of a Quantum Channel

The LOCC-assisted quantum capacity of a quantum channel $\mathcal{N}$, denoted by $Q^{\leftrightarrow}(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$Q^{\leftrightarrow}(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{19.3.1}$$

Definition 19.12 Weak Converse Rate for LOCC-Assisted Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a weak converse rate for LOCC-assisted quantum communication over $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

Definition 19.13 Strong Converse Rate for LOCC-Assisted Quantum Communication

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a strong converse rate for LOCC-assisted quantum communication over $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, all $\delta > 0$, and all sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ LOCC-assisted quantum communication protocol.

Definition 19.14 Strong Converse LOCC-Assisted Quantum Capacity of a Quantum Channel

The strong converse LOCC-assisted quantum capacity of a quantum channel $\mathcal{N}$, denoted by $\widetilde{Q}^{\leftrightarrow}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{Q}^{\leftrightarrow}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{19.3.2}$$

We have the exact same definitions for PPT-assisted quantum communication, and we use the notation $Q^{\leftrightarrow}_{\text{PPT}}$ to refer to the PPT-assisted quantum capacity and $\widetilde{Q}^{\leftrightarrow}_{\text{PPT}}$ for the strong converse PPT-assisted quantum capacity.

Recall that, by definition, the following bounds hold

$$Q^{\leftrightarrow}(\mathcal{N}) \leq \widetilde{Q}^{\leftrightarrow}(\mathcal{N}) \leq \widetilde{Q}^{\leftrightarrow}_{\text{PPT}}(\mathcal{N}), \tag{19.3.3}$$

$$Q^{\leftrightarrow}(\mathcal{N}) \leq Q^{\leftrightarrow}_{\text{PPT}}(\mathcal{N}) \leq \widetilde{Q}^{\leftrightarrow}_{\text{PPT}}(\mathcal{N}). \tag{19.3.4}$$

As a direct consequence of the bound in Theorem 19.4 and methods similar to those given in the proof of Theorem 11.23, we find the following:

Theorem 19.15 Squashed-Entanglement Weak-Converse Bound

The squashed entanglement of a channel $\mathcal{N}$ is a weak converse rate for LOCC-assisted quantum communication:

$$Q^{\leftrightarrow}(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N}). \tag{19.3.5}$$

As a direct consequence of the bound in Theorem 19.8 and methods similar to those given in Section 11.2.3, we find that

Theorem 19.16 Max-Rains Strong-Converse Bound

The max-Rains information of a channel $\mathcal{N}$ is a strong converse rate for PPT-assisted quantum communication:

$$\widetilde{Q}^{\leftrightarrow}_{\text{PPT}}(\mathcal{N}) \leq R_{\max}(\mathcal{N}). \tag{19.3.6}$$

As a direct consequence of the bound in Theorem 19.9 and methods similar to those given in Section 11.2.3, we find that

Theorem 19.17 Rains Strong-Converse Bound for PPT-Simulable Channels

Let $\mathcal{N}$ be a quantum channel that is PPT-simulable with associated resource state $\theta_{SB'}$. Then the Rains information of $\mathcal{N}$ is a strong converse rate for PPT-assisted quantum communication:

$$\widetilde{Q}^{\leftrightarrow}_{\text{PPT}}(\mathcal{N}) \leq R(\mathcal{N}). \tag{19.3.7}$$

19.4 Examples

[IN PROGRESS]

erasure channel - get Rains information as a strong converse rate - will match lower bound in terms of reverse coherent information (Hayashi called this pseudo-coherent information)

covariant dephasing channels - get Rains information as a strong converse rate and then coherent information matches this (will evaluate Rains information bound in unassisted quantum capacity chapter)

depolarizing channel - evaluate Rains information

use squashed entanglement to give upper bound for amplitude damping channel

19.5 Bibliographic Notes

The concept of LOCC-assisted quantum communication over a quantum channel was presented in (Bennett et al., 1996c, Section V). The same Section V of (Bennett et al., 1996c) also showed how to use the notion of teleportation simulation of a quantum channel and entanglement measures in order to bound the LOCC-assisted quantum capacity from above by a resource state that can realize the channel by teleportation simulation. Müller-Hermes (2012) presented a more detailed analysis of this bounding technique. Other papers that make use of the teleportation-simulation technique in this and other contexts include those by Horodecki et al.

(1999); Gottesman and Chuang (1999); Zhou et al. (2000); Bowen and Bose (2001); Takeoka et al. (2002); Giedke and Ignacio Cirac (2002); Wolf et al. (2007); Niset et al. (2009); Chiribella et al. (2009); Soeda et al. (2011); Leung and Matthews (2015); Pirandola et al. (2017); Takeoka et al. (2016); Wilde et al. (2017); Takeoka et al. (2017); Kaur and Wilde (2017).

A precise mathematical definition of an LOCC-assisted quantum communication protocol conducted over a quantum channel was presented in (Müller-Hermes, 2012, Definition 12) and (Takeoka et al., 2014, Section IV).

That the entanglement of the final state of an n-round LOCC-assisted quantum communication is bounded from above by n times the channel's amortized entanglement (Proposition 19.2) was anticipated by Bennett et al. (2003) and proven by Kaur and Wilde (2017). Corollary 19.3 was anticipated by Bennett et al. (1996c) and presented in more detail in (Müller-Hermes, 2012, Chapter 4), while the form in which we have presented it here is closely related to the presentation by Kaur and Wilde (2017).

The n-round PPT-assisted quantum communication protocols presented in Section 19.2 were considered by Kaur and Wilde (2017), with PPT-assisted quantum communication over a single or parallel use of a quantum channel considered by Leung and Matthews (2015); Wang and Duan (2016b); Wang et al. (2019b). The bound in Proposition 19.6 was established by Kaur and Wilde (2017).

Theorem 19.4 is due to Takeoka et al. (2014).

Wang and Duan (2016a) defined a semi-definite programming upper bound on distillable entanglement of a bipartite state, and Wang and Duan (2016b) defined a semi-definite programming upper bound on the quantum capacity of a quantum channel. Wang et al. (2019b) observed that the quantity defined by Wang and Duan (2016a) is equal to the max-Rains relative entropy, while also observing that the quantity defined by Wang and Duan (2016b) is equal to the max-Rains information of a quantum channel. Berta and Wilde (2018) established the max-Rains information as an upper bound on the n-round non-asymptotic PPT-assisted quantum capacity (Theorem 19.8). The upper bounds in Theorem 19.9 are due to Kaur and Wilde (2017).

Chapter 20

Secret Key Agreement

This chapter continues with the theme of feedback-assisted communication. Here, we consider secret-key-agreement protocols, where the goal is for the sender and receiver of a quantum channel $\mathcal{N}_{A\to B}$ to establish secret key, by using the quantum channel $\mathcal{N}_{A\to B}$ along with the free use of public classical communication. That is, between every channel use in a secret-key-agreement protocol, the sender and receiver are allowed to perform local operations and public classical communication (LOPC). The notion of capacity developed in this chapter is known as secret-key-agreement capacity.

The secret key distilled in such a secret-key-agreement protocol should be protected from an eavesdropper. The model we assume here is that the eavesdropper is quite powerful, having access to the full environment of every use of the quantum channel $\mathcal{N}_{A\to B}$, as well as a copy of all of the classical data exchanged by the sender and receiver when they conduct a round of LOPC. To understand this model in a physical context, suppose that the quantum channel connecting the sender and receiver is a fiber-optic cable, which we can model as a bosonic loss channel, and suppose that the sender employs quantum states of light to distill a secret key with the receiver. Then, in this eavesdropper model, we are assuming that all of the light that does not make it to the receiver is collected by the eavesdropper in a quantum memory. In this way, the secret key distilled by such a protocol is guaranteed to be secure against a quantum-enabled eavesdropper.

The practical motivation for secret-key-agreement protocols is related to the motivation that we considered in the previous chapter on LOCC-assisted quantum communication. Classical communication is cheap and plentiful these days, and so from a resource-theoretic perspective, it seems sensible to allow it for free in a theoretical model of communication. Once a secret key has been established, it can be used in conjuction with the well known one-time pad protocol as a scheme for private communication of an arbitrary message that has the same size as the key. Thus, as a consequence of the one-time pad protocol, it follows that secret key agreement and private communication are equivalent information-processing tasks when public classical communication is available for free. Furthermore, the model of LOPC-assisted secret key agreement is essentially the same model considered in quantum key distribution, which is one of the most famous applications in quantum information science. One of the main goals of this chapter is to place bounds on the rate at which secret-key-agreement is possible. Due to the strong connection between the model of secret-key-agreement and quantum key distribution, these bounds then place limitations on the rates at which it is possible to generate secret key in a quantum key distribution protocol.

The main method for placing limitations on the rates of secret-key-agreement protocols is similar to the approach that we took in the last chapter. In fact, there are many parallels. We again use the concept of amortization and entanglement measures, such as squashed entanglement and relative entropy of entanglement (and several variants of the latter).

However, the main difference between this chapter and the previous one is that the communication model is different. As discussed above, a secret-key-agreement protocol is a three-party protocol, consisting of the legitimate sender and receiver, as well as the eavesdropper. Thus, *a priori*, it is not obvious how to connect entanglement measures, which are used in two-party protocols, to secret-key-agreement protocols. To overcome this problem, we exploit the purification principle to establish a powerful equivalence between tripartite secret-key-agreement protocols and bipartite private-state distillation protocols. After doing so, we can apply entanglement measures to bound the rate at which it is possible to distill bipartite private states in a bipartite private-state distillation protocol, and then by appealing to the aforementioned equivalence, we can bound the rate at which it is possible to distill secret key in a tripartite secret-key-agreement protocol.

The main conclusion of this chapter is that entanglement measures such as squashed entanglement and relative entropy of entanglement (and the latter's variations) are upper bounds on the secret-key-agreement capacity of quantum channels. At the end of the chapter, we evaluate these bounds for various channels of interest in order to determine the fundamental limitations on secret key agreement for these channels.

20.1 *n*-Shot Secret-Key-Agreement Protocol

We begin by discussing the most general form for a secret-key-agreement protocol conducted over a quantum channel. The most important point to clarify before starting is the communication model, in particular, to address the question of who has access to what. First, we suppose that there is a quantum channel $\mathcal{N}_{A\to B}$ connecting the legitimate sender Alice to the legitimate receiver Bob. Alice has exclusive access to the input system A, and Bob has exclusive access to the output system B. As we know from Chapter 4, every quantum channel has an isometric channel $\mathcal{U}_{A\to BE}$ extending it, such that the original channel $\mathcal{N}_{A\to B}$ is recovered by tracing over the purifying or environment system E:

$$\mathcal{N}_{A\to B} = \mathrm{Tr}_E \circ \mathcal{U}_{A\to BE}. \tag{20.1.1}$$

Taking the same perspective as that in Chapter 16, with the idea that a powerful, fully quantum eavesdropper could have access to every system to which the legitimate parties do not have access, we suppose that the quantum eavesdropper has access to the environment system E. Furthermore, in a secret-key-agreement protocol, the legitimate parties are allowed to use a public, classical communication channel, in addition to the quantum channel $\mathcal{N}_{A\to B}$, in order to generate a secret key. Since this channel is public, we suppose that the eavesdropper has access to all of the classical data exchanged between the legitimate parties.

In more detail, an n-shot protocol for secret key agreement consists of n calls to the quantum channel $\mathcal{N}_{A\to B}$, interleaved by LOPC channels. Since all of the classical data exchanged between Alice and Bob is assumed to be public and available to the eavesdropper as well, we call these channels "LOPC" channels, which is an abbreviation of "local operations and public communication." In fact, a protocol for secret key agreement has essentially the same structure as a protocol for LOCC-assisted quantum communication, as discussed in Section 19.1, with the exception that the systems at the end should hold a secret key instead of a maximally entangled state.

A protocol for secret key agreement is depicted in Figure [REF],and it consists of the following elements:

$$(\rho_{A_1'A_1B_1'Y_1}, \{\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}' \to A_i'A_iB_i'Y_i}\}_{i=2}^{n}, \mathcal{L}^{(n+1)}_{A_n'B_nB_n' \to K_AK_BY_{n+1}}). \tag{20.1.2}$$

All systems labeled by A belong to Alice, those labeled by B belong to Bob, and those labeled by Y are classical systems belonging to Eve, representing a copy of the classical data exchanged by Alice and Bob in a round of LOPC. In the above, $\rho_{A_1'A_1B_1'Y_1}$ is a separable state, $\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}' \to A_i'A_iB_i'Y_i}$ is an LOPC channel for $i \in \{2, \dots, n\}$, and $\mathcal{L}^{(n+1)}_{A_n'B_nB_n' \to K_AK_BY_{n+1}}$ is a final LOPC channel that generates the approximate secret key in systems K_A and K_B. Let $\mathcal{C}$ denote all of these elements, which together constitute the secret-key-agreement protocol. As with LOCC-assisted quantum communication, all systems with primed labels should be understood as local quantum memory or scratch registers that Alice and Bob can employ in this information-processing task. We also assume that they are finite-dimensional, yet arbitrarily large. The unprimed systems are the ones that are either input to or output from the quantum communication channel $\mathcal{N}_{A\to B}$.

The secret-key-agreement protocol begins with Alice and Bob performing an LOPC channel $\mathcal{L}^{(1)}_{\emptyset \to A_1'A_1B_1'Y_1}$, which leads to the separable state $\rho^{(1)}_{A_1'A_1B_1'Y_1}$ mentioned above, where A_1' and B_1' are systems that are finite-dimensional yet arbitrarily large. In particular, the state $\rho^{(1)}_{A_1'A_1B_1'Y_1}$ has the following form:

$$\rho^{(1)}_{A_1'A_1B_1'Y_1} \coloneqq \sum_{y_1} p_{Y_1}(y_1)\tau^{y_1}_{A_1'A_1} \otimes \zeta^{y_1}_{B_1'} \otimes |y_1\rangle\langle y_1|_{Y_1}, \tag{20.1.3}$$

where Y_1 is a classical random variable corresponding to the message exchanged between Alice and Bob, which is needed to establish this state. The classical system Y_1 belongs to the eavesdropper. Also, $\{\tau^{y_1}_{A_1'A_1}\}_{y_1}$ and $\{\zeta^{y_1}_{B_1'}\}_{y_1}$ are sets of quantum states and p_{Y_1} is a probability distribution. Note that the reduced state for Alice and Bob is a generic separable state of the following form:

$$\rho^{(1)}_{A_1'A_1B_1'} = \sum_{y_1} p_{Y_1}(y_1)\tau^{y_1}_{A_1'A_1} \otimes \zeta^{y_1}_{B_1'}. \tag{20.1.4}$$

The system A_1 of $\rho^{(1)}_{A_1'A_1B_1'Y_1}$ is such that it can be fed into the first channel use. Alice then sends system A_1 through the first channel use, leading to a state

$$\omega^{(1)}_{A_1'B_1B_1'E_1Y_1} \coloneqq \mathcal{U}^{\mathcal{N}}_{A_1\to B_1E_1}(\rho^{(1)}_{A_1'A_1B_1'Y_1}). \tag{20.1.5}$$

Note that we write the channel use as the isometric channel $\mathcal{U}^{\mathcal{N}}_{A_1\to B_1E_1}$ that extends $\mathcal{N}_{A_1\to B_1}$, since we would like to incorporate the eavesdropper's system E_1 explicitly into the description of the protocol. Alice and Bob then perform the LOPC channel $\mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'Y_2}$, which leads to the state

$$\rho^{(2)}_{A_2'A_2B_2'E_1Y_1Y_2} \coloneqq \mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'Y_2}(\omega^{(1)}_{A_1'B_1B_1'E_1Y_1}). \tag{20.1.6}$$

The LOPC channel $\mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'Y_2}$ can be written as

$$\mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'Y_2} \coloneqq \sum_{y_2} \mathcal{E}^{y_2}_{A_1'\to A_2'A_2} \otimes \mathcal{F}^{y_2}_{B_1B_1'\to B_2'} \otimes |y_2\rangle\langle y_2|_{Y_2}. \tag{20.1.7}$$

In the above, $\{\mathcal{E}^{y_2}_{A_1'\to A_2'A_2}\}_{y_2}$ and $\{\mathcal{F}^{y_2}_{B_1B_1'\to B_2'}\}_{y_2}$ are sets of completely positive maps such that the sum map $\sum_{y_2} \mathcal{E}^{y_2}_{A_1'\to A_2'A_2} \otimes \mathcal{F}^{y_2}_{B_1B_1'\to B_2'}$ is trace preserving. The classical system Y_2 represents the eavesdropper's copy of the classical data exchanged by Alice and Bob in this round of LOPC. Note that the reduced channel acting on Alice and Bob's systems is as follows:

$$\mathcal{L}^{(2)}_{A_1'B_1B_1'\to A_2'A_2B_2'} = \sum_{y_2} \mathcal{E}^{y_2}_{A_1'\to A_2'A_2} \otimes \mathcal{F}^{y_2}_{B_1B_1'\to B_2'}. \tag{20.1.8}$$

Alice sends system A_2 through the second channel use $\mathcal{U}^{\mathcal{N}}_{A_2\to B_2E_2}$, leading to the state

$$\omega^{(2)}_{A_2'B_2B_2'E_1E_2Y_1Y_2} \coloneqq \mathcal{U}^{\mathcal{N}}_{A_2\to B_2E_2}(\rho^{(2)}_{A_2'A_2B_2'E_1Y_1Y_2}). \tag{20.1.9}$$

This process iterates: the protocol uses the channel n times. In general, we have the following states for all $i \in \{2, \dots, n\}$:

$$\rho^{(i)}_{A_i'A_iB_i'E_1^{i-1}Y_1^i} \coloneqq \mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'Y_i}(\omega^{(i-1)}_{A_{i-1}'B_{i-1}B_{i-1}'E_1^{i-1}Y_1^{i-1}}), \tag{20.1.10}$$

$$\omega^{(i)}_{A_i'B_iB_i'E_1^iY_1^i} \coloneqq \mathcal{U}^{\mathcal{N}}_{A_i\to B_iE_i}(\rho^{(i)}_{A_i'A_iB_i'E_1^{i-1}Y_1^i}), \tag{20.1.11}$$

where $\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'Y_i}$ is an LOPC channel that can be written as

$$\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'Y_i} \coloneqq \sum_{y_i} \mathcal{E}^{y_i}_{A_{i-1}'\to A_i'A_i} \otimes \mathcal{F}^{y_i}_{B_{i-1}B_{i-1}'\to B_i'} \otimes |y_i\rangle\langle y_i|_{Y_i}. \tag{20.1.12}$$

In the above, $\{\mathcal{E}^{y_i}_{A_{i-1}'\to A_i'A_i}\}_{y_i}$ and $\{\mathcal{F}^{y_i}_{B_{i-1}B_{i-1}'\to B_i'}\}_{y_i}$ are sets of completely positive maps such that the sum map $\sum_{y_i} \mathcal{E}^{y_i}_{A_{i-1}'\to A_i'A_i} \otimes \mathcal{F}^{y_i}_{B_{i-1}B_{i-1}'\to B_i'}$ is trace preserving.

The classical system Y_i represents the eavesdropper's copy of the classical data exchanged by Alice and Bob in this round of LOPC. Note that the reduced channel acting on Alice and Bob's systems is as follows:

$$\mathcal{L}^{(i)}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iB'_i} = \sum_{y_i} \mathcal{E}^{y_i}_{A'_{i-1}\to A'_iA_i} \otimes \mathcal{F}^{y_i}_{B_{i-1}B'_{i-1}\to B'_i}. \tag{20.1.13}$$

In (20.1.10)–(20.1.11), we have employed the following shorthands: $E_1^i \equiv E_1 \cdots E_i$ and $Y_1^i \equiv Y_1 \cdots Y_i$. The final step of the protocol consists of an LOPC channel $\mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to K_AK_BY_{n+1}}$, which generates the key systems K_A and K_B for Alice and Bob, respectively. The protocol's final state is as follows:

$$\omega_{K_AK_BE_1^nY_1^{n+1}} := \mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to K_AK_BY_{n+1}}(\omega^{(n)}_{A'_nB_nB'_nE_1^nY_1^n}). \tag{20.1.14}$$

Note that the final LOPC channel can be written as

$$\mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to K_AK_BY_{n+1}} := \sum_{y_{n+1}} \mathcal{E}^{y_{n+1}}_{A'_n\to K_A} \otimes \mathcal{F}^{y_{n+1}}_{B_nB'_n\to K_B} \otimes |y_{n+1}\rangle\langle y_{n+1}|_{Y_{n+1}}, \tag{20.1.15}$$

and the reduced final channel acting on Alice and Bob's systems is as follows:

$$\mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to K_AK_B} = \sum_{y_{n+1}} \mathcal{E}^{y_{n+1}}_{A'_n\to K_A} \otimes \mathcal{F}^{y_{n+1}}_{B_nB'_n\to K_B}. \tag{20.1.16}$$

The goal of the protocol is for the final state $\omega_{K_AK_BE_1^nY_1^{n+1}}$ to be nearly indistinguishable from a tripartite secret-key state, and we define the privacy error of the code to be as follows:

$$p_{\mathrm{err}}(\mathcal{C}) := 1 - F(\omega_{K_AK_BE_1^nY_1^{n+1}}, \overline{\Phi}_{K_AK_B} \otimes \sigma_{E_1^nY_1^{n+1}}), \tag{20.1.17}$$

where $\sigma_{E_1^nY_1^{n+1}}$ is some state of the eavesdropper's systems, F denotes the quantum fidelity (Definition 6.5) and the maximally classically correlated state $\overline{\Phi}_{K_AK_B}$ is defined as

$$\overline{\Phi}_{K_AK_B} := \frac{1}{K}\sum_{k=1}^{K} |k\rangle\langle k|_{K_A} \otimes |k\rangle\langle k|_{K_B}. \tag{20.1.18}$$

Intuitively, the privacy error $p_{\mathrm{err}}(\mathcal{C})$ quantifies how distinguishable the final state $\omega_{K_AK_BE_1^nY_1^{n+1}}$ is from an ideal tripartite secret-key state $\overline{\Phi}_{K_AK_B} \otimes \sigma_{E_1^nY_1^{n+1}}$, in which the key values in K_A and K_B are perfectly correlated and uniformly random and in tensor product with the eavesdropper's systems $E_1^n Y_1^{n+1}$. For an ideal tripartite secret-key state, it is difficult for an eavesdropper to guess the value of the key by observing the content of her quantum systems $E_1^n Y_1^{n+1}$. In fact, the chance for an eavesdropper to guess the key value of an ideal secret-key state is equal to $1/K$, which is no better than random guessing.

Due to the isometric invariance of the fidelity and the fact that all isometric extensions of a channel are related by an isometry acting on the environment system, the privacy error in (20.1.17) is invariant under any choice of an isometric channel $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ that extends the original channel $\mathcal{N}_{A\to B}$. Thus, the relevant performance parameters for a secret-key agreement protocol do not change with the particular isometric extension chosen. This is to be expected since the actual information that the eavesdropper gains in the protocol does not depend on the particular isometric extension chosen.

Definition 20.1 (n, K, ε) Secret-Key-Agreement Protocol

Let $(\rho^{(1)}_{A_1'A_1B_1'Y_1}, \{\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'Y_i}\}_{i=2}^n, \mathcal{L}^{(n+1)}_{A_n'B_nB_n'\to K_AK_BY_{n+1}})$ be the elements of an n-round LOPC-assisted secret-key-agreement protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an (n, K, ε) protocol, with $\varepsilon \in [0, 1]$, if the privacy error $p_{\text{err}}(\mathcal{C}) \le \varepsilon$.

20.1.1 Equivalence between Secret Key Agreement and LOPC-Assisted Private Communication

The goal of a secret-key-agreement protocol is for Alice and Bob to establish an approximation of an ideal secret key, the latter begin uniformly distributed, perfectly correlated, and independent of Eve's quantum systems. What is the use of this secret key? As it turns out, it can be used for private communication by means of the one-time pad protocol. This in turn means that secret key agreement and private classical communication are equivalent information processing tasks when public classical communication is available for free, and the goal of this section is to clarify this point.

An LOPC-assisted private communication protocol uses a quantum channel n times along with public classical communication to transmit an arbitrary message of size K privately from Alice to Bob in such a way that the fidelity of the actual state at the end of the protocol and the ideal state is no smaller than $1 - \varepsilon$. In more detail, let $\overline{\Phi}^p_{M_A M_B}$ denote the following state in which there is an arbitrary distribution p over the message:

$$\overline{\Phi}^p_{M_A M_B} \coloneqq \sum_{m=1}^{K} p(m)|m\rangle\langle m|_{M_A} \otimes |m\rangle\langle m|_{M_B}. \tag{20.1.19}$$

Let $\omega^p_{M_A M_B E_1^n Y_1^{n+1}}$ denote the final state of the protocol, which is defined in the same way as (20.1.14), with the exception that the message distribution p is no longer uniform. Then an (n, K, ε) LOPC-assisted private communication protocol is defined similarly to an (n, K, ε) secret-key-agreement protocol as given above, except that the following inequality holds

$$\max_{p:\mathcal{M}\to[0,1]} 1 - F(\omega^p_{M_A M_B E_1^n Y_1^{n+1}}, \overline{\Phi}^p_{M_A M_B} \otimes \sigma_{E_1^n Y_1^{n+1}}) \leq \varepsilon. \tag{20.1.20}$$

where the maximization is over all message distributions p and $\sigma_{E_1^n Y_1^{n+1}}$ is some fixed state of the eavesdropper's systems that is independent of the message transmitted.

By the use of the one-time pad protocol, it follows that an (n, K, ε) secret-key-agreement protocol leads to an (n, K, ε) LOPC-assisted private communication protocol. To see how the one-time pad protocol works in conjunction with secret key agreement, suppose that Alice and Bob have completed an (n, K, ε) secret-key-agreement protocol as described in the previous section, with the final state $\omega_{K_A K_B E_1^n Y_1^{n+1}}$ satisfying $p_{\text{err}}(\mathcal{C}) \leq \varepsilon$. Alice then brings in her local message registers M_A and $M_{A'}$, so that the overall quantum state is

$$\overline{\Phi}^p_{M_A M_{A'}} \otimes \omega_{K_A K_B E_1^n Y_1^{n+1}} \tag{20.1.21}$$

The one-time pad protocol is an LOPC protocol in which Alice then performs the following classical computation, represented as a quantum channel, on her message register $M_{A'}$ and her key register K_A:

$$\sum_{k,m} |m \oplus k\rangle_{C_A}\langle m|_{M_{A'}}\langle k|_{K_A}(\cdot)|m\rangle_{M_{A'}}|k\rangle_{K_A}\langle m \oplus k|_{C_A}, \tag{20.1.22}$$

where the addition $\oplus$ is modulo K. She then transmits the classical register C_A over a public classical channel to Bob. Eve can make a copy $C_{A'}$ of this classical register containing the value $m \oplus k$, but since Bob's key register K_B is not available to her, the register $C_{A'}$ is nearly independent of Alice's message register M_A (depending on how small ε is). Bob then performs the following classical computation, represented as a quantum channel, on his received register C_A and his key register K_B:

$$\sum_{c,k} |c \ominus k\rangle_{M_B} \langle c|_{C_A} \langle k|_{K_B} (\cdot) |c\rangle_{C_A} |k\rangle_{K_B} \langle c \ominus k|_{M_B}, \tag{20.1.23}$$

where the subtraction $\ominus$ is modulo K. Let $\omega^p_{M_A M_B E_1^n Y_1^{n+1} C_{A'}}$ denote the final state of the protocol. By applying the data-processing inequality to (20.1.20), as well as the fact mentioned above that $C_{A'}$ is independent of M_A and M_B in the ideal case, the following inequality holds

$$\max_{p:\mathcal{M}\to[0,1]} 1 - F(\omega^p_{M_A M_B E_1^n Y_1^{n+1} C_{A'}}, \overline{\Phi}^p_{M_A M_B} \otimes \sigma_{E_1^n Y_1^{n+1} C_{A'}}) \leq \varepsilon, \tag{20.1.24}$$

where $\sigma_{E_1^n Y_1^{n+1} C_{A'}}$ is a fixed state of the eavesdropper's systems. Thus, an (n, K, ε) secret-key-agreement protocol leads to an (n, K, ε) LOPC-assisted private communication protocol, as claimed.

The other implication is trivial: while employing an (n, K, ε) LOPC-assisted private communication protocol, Alice can choose the distribution of the message to be uniform, and then an (n, K, ε) LOPC-assisted private communication protocol leads to an (n, K, ε) secret-key-agreement protocol. Thus, it follows that LOPC-assisted private communication and secret key agreement are equivalent whenever public classical communication is available for free.

20.2 Equivalence between Secret Key Agreement and LOCC-Assisted Private-State Distillation

There is a deep and powerful equivalence between a secret-key-agreement protocol as described above and a protocol that uses LOCC assistance to distill a bipartite private state (recall Definition 15.4). This equivalence is helpful in the analysis of secret-key-agreement protocols, in the sense that one can use tools from entanglement theory in order to establish bounds on the rate at which secret key agreement is possible.

The main idea behind this equivalence is to apply the purification principle to a secret-key-agreement protocol and then examine the consequences. That is, we can purify each step of the secret-key-agreement protocol discussed in the previous section, and then we can examine various reduced states at each step. In what follows, we detail such a purified protocol.

20.2.1 The Purified Protocol

To begin with, recall that the initial state $\rho^{(1)}_{A'_1A_1B'_1}$ of a secret-key-agreement protocol is a separable state of the form in (20.1.3). The state $\rho^{(1)}_{A'_1A_1B'_1}$ can be purified as follows

$$|\rho^{(1)}\rangle_{A'_1A_1S_{A_1}B'_1S_{B_1}Y_1} \coloneqq \sum_{y_1} \sqrt{p_{Y_1}(y_1)}|\tau^{y_1}\rangle_{A'_1A_1S_{A_1}} \otimes |\zeta^{y_1}\rangle_{B'_1S_{B_1}} \otimes |y_1\rangle_{Y_1}, \quad (20.2.1)$$

where the systems S_{A_1} and S_{B_1} are known as local "shield" systems. In principle, the shield systems S_{A_1} and S_{B_1} could be held by Alice and Bob, respectively, and the states $|\tau^{y_1}\rangle_{A'_1A_1S_{A_1}}$ and $|\zeta^{y_1}\rangle_{B'_1S_{B_1}}$ purify $\tau^{y_1}_{A'_1A_1}$ and $\zeta^{y_1}_{B'_1}$ in (20.1.3), respectively. We assume without loss of generality that the shield systems contain a coherent classical copy of the classical random variable Y_1, such that tracing over systems S_{A_1} and S_{B_1} recovers the original state in (20.1.3). As before, Eve possesses system Y_1, which contains a coherent classical copy of the classical data exchanged.

Each LOPC channel $\mathcal{L}^{(i)}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iB'_i}$ for $i \in \{2, \dots, n\}$ is of the form in (20.1.12) and can be purified to an isometry in the following way:

$$U^{\mathcal{L}^{(i)}}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iS_{A_i}B'_iS_{B_i}Y_i} \coloneqq \sum_{y_i} U^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}} \otimes U^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}} \otimes |y_i\rangle_{Y_i}, \quad (20.2.2)$$

where $\{U^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}}\}_{y_i}$ and $\{U^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}\}_{y_i}$ are collections of linear operators, each of which is a contraction, that is,

$$\|U^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}}\|_\infty, \ \|U^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}\|_\infty \leq 1, \quad (20.2.3)$$

such that the linear operator in (20.2.2) is an isometry.

It is important to note here that the isometry $U^{\mathcal{L}^{(i)}}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iS_{A_i}B'_iS_{B_i}Y_i}$ results from purifying each step of an LOPC channel. That is, an LOPC channel is implemented as a sequence of one-way LOPC channels, which each consist of a generalized measurement by one party, classical communication of the measurement outcome to the other, and a channel by the other party, conditioned on the outcome of the measurement. So when purifying the LOPC channel, we purify each of these steps, and the resulting purified channel is what is represented in (20.2.2).

The systems S_{A_i} and S_{B_i} in (20.2.2) are shield systems belonging to Alice and Bob, respectively, and we assume without loss of generality that they contain a coherent classical copy of the classical random variable Y_i, such that tracing over the systems S_{A_i} and S_{B_i} recovers the original LOPC channel in (20.1.12). As before, Y_i is a system held by Eve, containing a coherent classical copy of the classical data exchanged in this round.

Thus, a purification of the state $\rho^{(i)}_{A'_iA_iB'_i}$ after each LOPC channel is as follows:

$$|\rho^{(i)}\rangle_{A'_iA_iS_{A_1^i}B'_iS_{B_1^i}E_1^{i-1}Y_1^i} \coloneqq U^{\mathcal{L}^{(i)}}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iS_{A_i}B'_iS_{B_i}Y_i}|\omega^{(i-1)}\rangle_{A'_{i-1}B_{i-1}B'_{i-1}S_{A_1^{i-1}}S_{B_1^{i-1}}E_1^{i-1}Y_1^{i-1}}, \quad (20.2.4)$$

where we have employed the shorthands $S_{A_1^i} \equiv S_{A_1}\cdots S_{A_i}$ and $S_{B_1^i} \equiv S_{B_1}\cdots S_{B_i}$, with a similar shorthand for E_1^{i-1} and Y_1^i as before. A purification of the state $\omega^{(i)}_{A'_iB_iB'_i}$ after each use of the channel $\mathcal{N}_{A\to B}$ is

$$|\omega^{(i)}\rangle_{A'_iB_iS_{A_1^i}B'_iS_{B_1^i}E_1^iY_1^i} \coloneqq U^{\mathcal{N}}_{A_i\to B_iE_i}|\rho^{(i)}\rangle_{A'_iA_iS_{A_1^i}B'_iS_{B_1^i}E_1^{i-1}Y_1^i}, \quad (20.2.5)$$

where $U^{\mathcal{N}}_{A_i\to B_iE_i}$ is an isometric extension of the ith channel use $\mathcal{N}_{A_i\to B_i}$.

The final LOPC channel takes the form in (20.1.15), and it can be purified to an isometry similarly as

$$U^{\mathcal{L}^{(n+1)}}_{A'_nB_nB'_n\to K_AS_{A_{n+1}}K_BS_{B_{n+1}}Y_{n+1}} \coloneqq \sum_{y_{n+1}} U^{\mathcal{E}^{y_{n+1}}}_{A'_n\to K_AS_{A_{n+1}}} \otimes U^{\mathcal{F}^{y_{n+1}}}_{B_nB'_n\to K_BS_{B_{n+1}}} \otimes |y_{n+1}\rangle_{Y_{n+1}}. \quad (20.2.6)$$

The systems $S_{A_{n+1}}$ and $S_{B_{n+1}}$ are again shield systems belonging to Alice and Bob, respectively, and we assume again that they contain a coherent classical copy of the classical random variable Y_{n+1}, such that tracing over $S_{A_{n+1}}$ and $S_{B_{n+1}}$ recovers the original LOPC channel in (20.1.15). As before, Y_{n+1} is a system held by Eve, containing a coherent classical copy of the classical data exchanged in this round.

The final state at the end of the purified protocol is a pure state $|\omega\rangle_{K_AS_AK_BS_BE^nY^{n+1}}$, given by

$$|\omega\rangle_{K_AS_AK_BS_BE^nY^{n+1}} \coloneqq$$

$$U^{\mathcal{L}^{(n+1)}}_{A'_nB_nB'_n\to K_AS_{A_{n+1}}K_BS_{B_{n+1}}Y_{n+1}}|\omega^{(n)}\rangle_{A'_nB_nS_{A_1^n}B'_nS_{B_1^n}E_1^nY_1^n}. \quad (20.2.7)$$

Alice is in possession of the key system K_A and the shield systems $S_A \equiv S_{A_1}\cdots S_{A_{n+1}}$, Bob possesses the key system K_B and the shield systems $S_B \equiv S_{B_1}\cdots S_{B_{n+1}}$, and Eve holds the environment systems $E^n \equiv E_1\cdots E_n$. Additionally, Eve has coherent copies $Y^{n+1} \equiv Y_1\cdots Y_{n+1}$ of all the classical data exchanged.

20.2.2 LOCC-Assisted Bipartite Private-State Distillation

As a consequence of the purification principle, on the one hand, if we trace over the shield systems at every step, then we simply recover the original tripartite secret-key-agreement protocol detailed in Section 20.1. On the other hand, suppose that we instead trace over all of Eve's systems at each step. Due to the fact that each state of the Y systems is a coherent classical copy, the resulting reduced states consist of a classical mixture of various states of Alice and Bob's systems, as would arise in an LOCC-assisted quantum communication protocol.

It is worthwhile to examine how each step changes after tracing over Eve's systems of the purified protocol. For the first step, the reduced state of Alice and Bob's systems is a separable state of the following form:

$$\rho^{(1)}_{A'_1A_1S_{A_1}B'_1S_{B_1}} = \sum_{y_1} p_{Y_1}(y_1)\tau^{y_1}_{A'_1A_1S_{A_1}} \otimes \zeta^{y_1}_{B'_1S_{B_1}}, \quad (20.2.8)$$

where $\tau^{y_1}_{A'_1A_1S_{A_1}} = |\tau^{y_1}\rangle\!\langle\tau^{y_1}|_{A'_1A_1S_{A_1}}$ and $\zeta^{y_1}_{B'_1S_{B_1}} = |\zeta^{y_1}\rangle\!\langle\zeta^{y_1}|_{B'_1S_{B_1}}$. Tracing over Eve's system Y_i of each isometry $U^{\mathcal{L}^{(i)}}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iS_{A_i}B'_iS_{B_i}Y_i}$ leads to the following LOCC channel:

$$\mathcal{L}^{(i)}_{A'_{i-1}B_{i-1}B'_{i-1}\to A'_iA_iS_{A_i}B'_iS_{B_i}} = \sum_{y_i} \mathcal{U}^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}} \otimes \mathcal{U}^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}, \quad (20.2.9)$$

where

$$\mathcal{U}^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}}(\cdot) \coloneqq U^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}}(\cdot)[U^{\mathcal{E}^{y_i}}_{A'_{i-1}\to A'_iA_iS_{A_i}}]^\dagger, \quad (20.2.10)$$

$$\mathcal{U}^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}(\cdot) \coloneqq U^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}(\cdot)[U^{\mathcal{F}^{y_i}}_{B_{i-1}B'_{i-1}\to B'_iS_{B_i}}]^\dagger. \quad (20.2.11)$$

Tracing over Eve's system Y_{n+1} of the final isometry leads to the following LOCC channel:

$$\mathcal{L}^{(n+1)}_{A'_nB_nB'_n\to K_AS_{A_{n+1}}K_BS_{B_{n+1}}} = \sum_{y_{n+1}} \mathcal{U}^{\mathcal{E}^{y_{n+1}}}_{A'_n\to K_AS_{A_{n+1}}} \otimes \mathcal{U}^{\mathcal{F}^{y_{n+1}}}_{B_nB'_n\to K_BS_{B_{n+1}}}, \quad (20.2.12)$$

with a similar convention as in (20.2.10)–(20.2.11) for the maps $\mathcal{U}^{\mathcal{E}^{y_{n+1}}}_{A'_n \to K_A S_{A_{n+1}}}$ and $\mathcal{U}^{\mathcal{F}^{y_{n+1}}}_{B_n B'_n \to K_B S_{B_{n+1}}}$.

The states at every step of the protocol are then given by the following for all $i \in \{2, \ldots, n\}$:

$$\rho^{(i)}_{A'_i A_i S_{A_1^i} B'_i S_{B_1^i}} := \mathcal{L}^{(i)}_{A'_{i-1} B_{i-1} B'_{i-1} \to A'_i A_i S_{A_i} B'_i S_{B_i}} (\omega^{(i-1)}_{A'_{i-1} S_{A_1^{i-1}} B_{i-1} B'_{i-1} S_{B_1^{i-1}}}), \tag{20.2.13}$$

$$\omega^{(i)}_{A'_i S_{A_1^i} B_i B'_i S_{B_1^i}} := \mathcal{N}_{A_i \to B_i} (\rho^{(i)}_{A'_i A_i S_{A_1^i} B'_i S_{B_1^i}}), \tag{20.2.14}$$

and the final state of the protocol is given by

$$\omega_{K_A S_A K_B S_B} := \mathcal{L}^{(n+1)}_{A'_n B_n B'_n \to K_A S_{A_{n+1}} K_B S_{B_{n+1}}} (\omega^{(n)}_{A'_n B_n S_{A_1^n} B'_n S_{B_1^n}}). \tag{20.2.15}$$

Finally, by employing (20.1.17) and Proposition 15.7, the following condition holds

$$p_{\text{err}}(\mathcal{C}) = 1 - F(\omega_{K_A S_A K_B S_B}, \gamma_{K_A S_A K_B S_B}), \tag{20.2.16}$$

where $\gamma_{K_A S_A K_B S_B}$ is a bipartite private state of the form in Theorem 15.5. Thus, applying Definition 20.1, it follows that

$$F(\omega_{K_A S_A K_B S_B}, \gamma_{K_A S_A K_B S_B}) \geq 1 - \varepsilon. \tag{20.2.17}$$

We can now make a critical observation. By tracing over Eve's systems E^n and Y^{n+1} at every step of the purified protocol as we did above, it is clear that the resulting protocol is an LOCC-assisted protocol that distills an approximate bipartite private state on the systems $K_A S_A K_B S_B$, with performance parameter given by (20.2.17). Indeed, the initial state in (20.2.8) is a separable state, and the channels in (20.2.9) and (20.2.12) are LOCC channels, and so the protocol has the same form as an LOCC-assisted quantum communication protocol, as we studied in the previous chapter. However, the goal of this LOCC-assisted bipartite private-state distillation is not as stringent as it was in the previous chapter, for LOCC-assisted quantum communication. Namely, it is only necessary to distill an approximate bipartite private state and not necessarily an approximate maximally entangled state; but keep in mind that a maximally entangled state is a particular kind of bipartite private state. Thus, starting with a tripartite secret-key-agreement protocol, we can apply the purification principle, then trace over the systems of the eavesdropper, and the result is a bipartite private-state distillation protocol assisted by LOCC.

Alternatively, this reasoning can go in the opposite direction. Suppose instead that we had started with an LOCC-assisted bipartite private-state distillation protocol of the above form. Then we could purify it as we did in the previous subsection, and after doing so, we could trace over Alice and Bob's shield systems. If the protocol satisfies the condition in (20.2.17), then the resulting protocol would be an (n, K, ε) tripartite secret-key-agreement protocol, which follows as a consequence of the equivalence between approximate bipartite private states and tripartite secret-key states, as given in Proposition 15.7.

As a consequence of this equivalence between tripartite secret-key-agreement protocols and bipartite private-state distillation protocols, we can employ the tools of entanglement theory in order to analyze bipartite private-state distillation protocols, in a way similar to how we did in the last chapter. For example, if our goal is to determine upper bounds on the rate at which it is possible to generate secret key in a secret-key-agreement protocol, then we can employ an entanglement measure to analyze the equivalent bipartite private-state distillation protocol in order to accomplish the goal. In fact, this is exactly what we accomplish in this chapter, demonstrating that the squashed entanglement of a channel serves as an upper bound on secret-key rates, and that variations of the relative entropy of entanglement, similar in spirit to the Rains relative entropy, serve as upper bounds on secret-key rates as well.

20.2.2.1 Unboundedness of Shield Systems in a Bipartite Private-State Distillation Protocol

One observation that we make here is that the shield systems in a bipartite private-state distillation protocol are finite-dimensional, yet arbitrarily large. That is, there is no bound that we can establish on their dimension for a generic private-state distillation protocol, and this unboundedness is a consequence of the fact that the shield systems result from purifying the local memory or scratch registers of Alice and Bob, which in turn have no bound on their dimension. This unboundedness poses a challenge when trying to establish upper bounds on the rate at which secret key agreement, or equivalently, bipartite private-state distillation is possible. However, there are methods for handling this unboundedness that we detail later.

20.2.3 Relation between Secret Key Agreement and LOCC-Assisted Quantum Communication

Due to the fact that a maximally entangled state is a particular kind of bipartite private state and due to the equivalence between secret key agreement and LOCC-assisted bipartite private-state distillation, we arrive at the following conclusion, which relates LOCC-assisted quantum communication to secret key agreement:

Proposition 20.2

Let $\mathcal{N}_{A\to B}$ be a quantum channel, let $n, K \in \mathbb{N}$, and let $\varepsilon \in [0,1]$. Then an (n, K, ε) LOCC-assisted quantum communication protocol is also an (n, K, ε) protocol for secret key agreement.

This statement is a rather simple observation, but it has consequences for capacities related to these tasks. That is, from this statement, we can conclude that the LOCC-assisted quantum capacity of a given quantum channel is bounded from above by its secret-key-agreement capacity. Thus, any upper bound established on the secret-key-agreement capacity of a quantum channel is also an upper bound on its LOCC-assisted quantum capacity. Furthermore, if a given quantum quantity is a lower bound on the LOCC-assisted quantum capacity of a quantum channel, then it is also a lower bound on its secret-key-agreement capacity.

20.2.4 *n*-Shot Secret-Key-Agreement Protocol Assisted by Public Separable Channels

In Section 19.2, we generalized the notion of an LOCC-assisted quantum communication protocol to one that is assisted by PPT-preserving channels. We note here that we can consider a similar kind of generalization for secret-key-agreement protocols.

Recall from Section 4.6.2 that any LOCC channel $\mathcal{L}_{AB}$ can be written as a separable channel of the following form:

$$\mathcal{L}_{AB\to A'B'} = \sum_y \mathcal{E}^y_{A\to A'} \otimes \mathcal{F}^y_{B\to B'}, \tag{20.2.18}$$

where $\{\mathcal{E}^y_{A\to A'}\}_y$ and $\{\mathcal{F}^y_{B\to B'}\}_y$ are sets of completely positive maps such that the sum map $\sum_y \mathcal{E}_A^y \otimes \mathcal{F}_B^y$ is trace preserving. However, the converse statement is not true. That is, it is not possible in general to implement an arbitrary separable channel of the form above as an LOCC channel.

Thus, we can allow for a slight generalization of a secret-key-agreement protocol to one that is assisted by public separable channels. Indeed, we define a public separable channel to be the following generalization of an LOPC channel:

$$\mathcal{L}_{AB\to A'B'Y} = \sum_y \mathcal{E}_{A\to A'}^y \otimes \mathcal{F}_{B\to B'}^y \otimes |y\rangle\langle y|_Y, \tag{20.2.19}$$

where Alice and Bob have access to the A and B systems, respectively, and the eavesdropper has access to the system Y. The only requirement for a public separable channel is that $\{\mathcal{E}_{A\to A'}^y\}_y$ and $\{\mathcal{F}_{B\to B'}^y\}_y$ are sets of completely positive maps such that the sum map $\sum_y \mathcal{E}_A^y \otimes \mathcal{F}_B^y$ is trace preserving. Similar to the distinction between LOCC and separable channels, it is not possible in general to implement a public separable channel via local operations and public classical communication.

The main point that we make in this section is that we can generalize a secret-key-agreement protocol to be assisted by public separable channels rather than just LOPC channels. For fixed privacy error, the resulting protocol achieves a rate of communication that is either the same or higher than that achieved by an LOPC-assisted protocol, due to the fact that every LOPC channel is a public separable channel. Such a protocol is defined in the same way as we did in Section 20.1, and then we arrive at the following definition:

Definition 20.3 (n, K, ε) Secret-Key-Agreement Protocol Assisted by Public Separable Channels

Let $\mathcal{C} := (\rho^{(1)}_{A_1'A_1B_1'Y_1}, \{\mathcal{L}^{(i)}_{A_{i-1}'B_{i-1}B_{i-1}'\to A_i'A_iB_i'Y_i}\}_{i=2}^n, \mathcal{L}^{(n+1)}_{A_n'B_nB_n'\to K_AK_BY_{n+1}})$ be the elements of an n-round public-separable-assisted secret-key-agreement protocol over the channel $\mathcal{N}_{A\to B}$. The protocol is called an (n, K, ε) protocol, with $\varepsilon \in [0, 1]$, if the privacy error $p_{\text{err}}(\mathcal{C}) \le \varepsilon$.

Furthermore, the equivalence between secret key agreement and bipartite private-state distillation, as outlined in Sections 20.2.1 and 20.2.2, still holds under this generalization (one can check that all of the steps given in Sections 20.2.1 and 20.2.2 still hold). However, the correspondence changes as follows: to any tripartite

(n, K, ε) secret-key-agreement protocol assisted by public separable channels, there exists an (n, K, ε) bipartite private-state distillation protocol assisted by separable channels and vice versa. Thus, we can again employ the tools of entanglement theory to analyze secret-key-agreement protocols assisted by public separable channels.

20.2.5 Amortized Entanglement Bound for Secret-Key-Agreement Protocols

Due to the equivalence between tripartite secret-key-agreement protocols and bipartite private-state distillation protocols, we can use the tools of entanglement theory to establish upper bounds on the rate at which secret-key-agreement is possible. Namely, we can apply the idea of amortized entanglement from the previous chapter in order to establish a generic upper bound in terms of an amortized entanglement measure. In fact, by the same steps used to arrive at Proposition 19.2, we find the following bound:

Proposition 20.4

Let $\mathcal{N}_{A\to B}$ be a quantum channel, let $\varepsilon \in [0,1]$, and let E be an entanglement measure that is equal to zero for all separable states. For an (n, K, ε) secret-key-agreement protocol, the following bound holds

$$E(K_A S_A; K_B S_B)_\omega \le n \cdot E^{\mathcal{A}}(\mathcal{N}), \tag{20.2.20}$$

where $\omega_{K_A S_A K_B S_B}$ is the final state resulting from the equivalent (n, K, ε) bipartite private-state distillation protocol (see (20.2.15)) and $E^{\mathcal{A}}(\mathcal{N})$ is the amortized entanglement of the channel $\mathcal{N}_{A\to B}$, as given in Definition 10.3.

Just as the bound from Proposition 19.2 depends on the final state of the LOCC-assisted quantum communication protocol, the same is true for the bound in (20.2.20). The bound is thus not a universal bound (a universal bound would depend only on the protocol parameters n, K, and ε). Thus, one of the main goals of the forthcoming sections is to employ particular entanglement measures in order to arrive at universal bounds for secret-key-agreement protocols.

We should also observe that the quantity $E(K_A S_A; K_B S_B)_\omega$ in the bound in

(20.2.20) can be understood as quantifying amount of entanglement between the systems K_AS_A and K_BS_B. As such, the shield systems S_A and S_B are involved, and they can have arbitrarily large dimension. Thus, one must account for this in the analysis of the entanglement $E(K_AS_A;K_BS_B)_\omega$. For example, in the previous chapter, we analyzed the analogous quantity $E(M_A;M_B)_\omega$ by employing squashed entanglement. In particular, since the state $\omega_{M_AM_B}$ there was an approximate maximally entangled state, we applied the uniform continuity of squashed entanglement from Proposition 9.38 in order to evaluate $E(M_A;M_B)_\omega$. When we did so, the dimension of the maximally entangled state appeared in the continuity bound, and this was acceptable there because the dimension of the maximally entangled state is directly related to the rate of entanglement distillation. However, it is not clear whether we can take such an approach, via uniform continuity of squashed entanglement, when analyzing bipartite private-state distillation, due to the fact that the shield systems do not necessarily have a bounded dimension. As such, we employ another method to analyze such protocols.

Just as the bound in Proposition 19.2 simplifies for teleportation-simulable channels and particular entanglement measures, the same is true for the bound given in Proposition 20.4, by employing the same reasoning:

Corollary 20.5 Reduction by Teleportation

Let E_S denote an entanglement measure that is subadditive with respect to states (Definition 9.1.9) and equal to zero for all separable states. Let $\mathcal{N}_{A\to B}$ be a channel that is LOCC-simulable with associated resource state $\theta_{RB'}$ (Definition 4.25). Let $\varepsilon\in[0,1]$. For an (n,K,ε) secret-key-agreement protocol, the following bound holds

$$E_S(K_AS_A;K_BS_B)_\omega \le n\cdot E_S(R;B')_\theta, \tag{20.2.21}$$

where $\omega_{K_AS_AK_BS_B}$ is the final state resulting from the equivalent (n,K,ε) bipartite private-state distillation protocol.

Just as we found bounds that apply to PPT-assisted quantum communication in terms of entanglement measures that are monotone with respect to PPT-preserving channels, we can also find bounds that apply to secret-key-agreement protocols that are assisted by public separable channels:

Proposition 20.6

Let $\mathcal{N}_{A\to B}$ be a quantum channel, let $\varepsilon \in [0,1]$, and let E be an entanglement measure that is monotone non-increasing with respect to separable channels and equal to zero for all separable states. For an (n, K, ε) secret-key-agreement protocol assisted by public separable channels, the following bound holds

$$E(K_A S_A; K_B S_B)_\omega \leq n \cdot E^{\mathcal{A}}(\mathcal{N}), \tag{20.2.22}$$

where $\omega_{K_A S_A K_B S_B}$ is the final state resulting from the equivalent (n, K, ε) bipartite private-state distillation protocol assisted by separable channels and $E^{\mathcal{A}}(\mathcal{N})$ is the amortized entanglement of the channel $\mathcal{N}_{A\to B}$, as given in Definition 10.3.

Finally, this bound again simplifies for channels that are simulable by the action of a separable channel on a resource state $\theta_{RB'}$ (separable-simulable channels):

Corollary 20.7

Let E_S denote an entanglement measure that is that is monotone non-increasing with respect to separable channels, subadditive with respect to states (Definition 9.1.9), and equal to zero for separable states. Let $\mathcal{N}_{A\to B}$ be a channel that is separable-simulable with associated resource state $\theta_{RB'}$ (Definition 4.26). Let $\varepsilon \in [0,1]$. For an (n, K, ε) secret-key-agreement protocol assisted by public separable channels, the following bound holds

$$E_S(K_A S_A; K_B S_B)_\omega \leq n \cdot E_S(R; B')_\theta, \tag{20.2.23}$$

where $\omega_{K_A S_A K_B S_B}$ is the final state resulting from the equivalent (n, K, ε) bipartite private-state distillation protocol assisted by separable channels.

20.3 Squashed Entanglement Upper Bound on the Number of Transmitted Private Bits

We now employ the squashed entanglement in order to bound the number of private bits that an n-shot secret-key-agreement protocol can generate. We have already shown in Section 9.4 that the squashed entanglement satisfies all of the requirements needed to apply it in Proposition 20.4. Namely, it is equal to zero for separable states, it is an entanglement measure (non-increasing under the action of an LOCC channel), and the squashed entanglement of a channel does not increase under amortization (Theorem 10.20). Putting all of these items together, we can already conclude the following bound for an (n, K, ε) secret-key-agreement protocol:

$$E_{\mathrm{sq}}(K_A S_A; K_B S_B)_\omega \leq n \cdot E_{\mathrm{sq}}(\mathcal{N}), \tag{20.3.1}$$

where $E_{\mathrm{sq}}(\mathcal{N})$ is the squashed entanglement of a channel (Definition 10.1) and $\omega_{K_A S_A K_B S_B}$ is the final state resulting from the equivalent (n, K, ε) bipartite private-state distillation protocol. Thus, what remains is to evaluate the squashed entanglement of an approximate bipartite private state, and this is the main technical problem that we consider in this section before concluding that squashed entanglement is an upper bound on secret-key rates.

20.3.1 Squashed Entanglement and Approximate Private States

This subsection establishes Proposition 20.9, which is an upper bound on the logarithm of the dimension K of a key system of an ε-approximate private state, as given in Definition 15.6, in terms of its squashed entanglement, plus another term depending only on ε and $\log_2 K$.

In what follows, we suppose that $\gamma_{AA'BB'}$ is a private state with key systems AB and shield systems $A'B'$. Recall from Theorem 15.5 that a private state of $\log_2 K$ private bits can be written in the following form:

$$\gamma_{ABA'B'} = U_{ABA'B'} \left(\Phi_{AB} \otimes \sigma_{A'B'}\right) U^\dagger_{ABA'B'}, \tag{20.3.2}$$

where Φ_{AB} is a maximally entangled state of Schmidt rank K

$$\Phi_{AB} := \frac{1}{K} \sum_{i,j} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B, \tag{20.3.3}$$

and

$$U_{ABA'B'} = \sum_{i,j} |i\rangle\langle i|_A \otimes |j\rangle\langle j|_B \otimes U^{ij}_{A'B'} \tag{20.3.4}$$

is a controlled unitary known as a "twisting unitary," with each $U^{ij}_{A'B'}$ a unitary operator. Due to the fact that the maximally entangled state Φ_{AB} is unextendible, any extension $\gamma_{AA'BB'E}$ of a private state $\gamma_{AA'BB'}$ necessarily has the following form:

$$\gamma_{AA'BB'E} = U_{AA'BB'} \left(\Phi_{AB} \otimes \sigma_{A'B'E}\right) U^{\dagger}_{AA'BB'}, \tag{20.3.5}$$

where $\sigma_{A'B'E}$ is an extension of $\sigma_{A'B'}$.

We start with the following lemma, which applies to any extension of a bipartite private state:

Lemma 20.8

Let $\gamma_{AA'BB'}$ be a bipartite private state, and let $\gamma_{AA'BB'E}$ be an extension of it, as given above. Then the following identity holds for any such extension:

$$2\log_2 K = I(A;BB'|E)_{\gamma} + I(A';B|AB'E)_{\gamma}. \tag{20.3.6}$$

Proof: First consider that the following identity holds as a consequence of two applications of the chain rule for conditional quantum mutual information:

$$\begin{aligned} I(AA';BB'|E)_{\gamma} &= I(A;BB'|E)_{\gamma} + I(A';BB'|AE)_{\gamma} \\ &= I(A;BB'|E)_{\gamma} + I(A';B'|AE)_{\gamma} + I(A';B|B'AE)_{\gamma}. \end{aligned} \tag{20.3.7}$$

Combined with the following identity, which holds for an extension $\gamma_{AA'BB'E}$ of a private state $\gamma_{AA'BB'}$,

$$I(AA';BB'|E)_{\gamma} = 2\log_2 K + I(A';B'|AE)_{\gamma}, \tag{20.3.8}$$

we recover the statement in (20.3.6). So it remains to prove (20.3.8). By definition, we have that

$$I(AA';BB'|E)_{\gamma} = H(AA'E)_{\gamma} + H(BB'E)_{\gamma} - H(E)_{\gamma} - H(AA'BB'E)_{\gamma}. \tag{20.3.9}$$

By applying (20.3.3)–(20.3.5), we can write $\gamma_{AA'BB'E}$ as follows:

$$\gamma_{AA'BB'E} = \frac{1}{K}\sum_{i,j} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B \otimes U^{ii}_{A'B'}\sigma_{A'B'E}(U^{jj}_{A'B'})^{\dagger}. \tag{20.3.10}$$

Tracing over system B leads to the following state:

$$\gamma_{AA'B'E} = \frac{1}{K}\sum_{i} |i\rangle\langle i|_A \otimes \gamma^{i}_{A'B'E}, \tag{20.3.11}$$

where

$$\gamma^i_{A'B'E} := U^{ii}_{A'B'}\sigma_{A'B'E}(U^{ii}_{A'B'})^\dagger. \tag{20.3.12}$$

Similarly, tracing over system A of $\gamma_{AA'BB'E}$ leads to

$$\gamma_{BA'B'E} = \frac{1}{K}\sum_i |i\rangle\langle i|_B \otimes \gamma^i_{A'B'E}. \tag{20.3.13}$$

So these and the chain rule for conditional entropy imply that

$$H(AA'E)_\gamma = H(A)_\gamma + H(A'E|A)_\gamma = \log_2 K + H(A'E|A)_\gamma. \tag{20.3.14}$$

Similarly, we have that

$$H(BB'E)_\gamma = \log_2 K + H(B'E|B)_\gamma = \log_2 K + H(B'E|A)_\gamma, \tag{20.3.15}$$

where we have used the symmetries in (20.3.11)–(20.3.13). Since $\gamma_E = \gamma^i_E$ for all i (this is a consequence of $\gamma_{ABA'B'}$ being an ideal private state), we find that

$$H(E)_\gamma = \frac{1}{K}\sum_i H(E)_{\gamma^i} = H(E|A)_\gamma. \tag{20.3.16}$$

Finally, we have that

$$H(AA'BB'E)_\gamma = H(ABA'B'E)_{\Phi\otimes\sigma} \tag{20.3.17}$$

$$= H(AB)_\Phi + H(A'B'E)_\sigma \tag{20.3.18}$$

$$= \frac{1}{K}\sum_i H(A'B'E)_{\gamma^i} \tag{20.3.19}$$

$$= H(A'B'E|A)_\gamma. \tag{20.3.20}$$

The first equality follows from unitary invariance of quantum entropy. The second equality follows because the entropy is additive for tensor-product states. The third equality follows because $H(AB)_\Phi = 0$ since Φ_{AB} is a pure state, and $\sigma_{A'B'E}$ is related to $\gamma^i_{A'B'E}$ by the unitary $U^{ii}_{A'B'}$. The final equality follows by applying (20.3.11), and the fact that conditional entropy is a convex combination of entropies for a classical-quantum state where the conditioning system is classical.

Combining (20.3.9), (20.3.14), (20.3.15), (20.3.16), (20.3.20), and the fact that

$$I(A';B'|AE)_\gamma = H(A'E|A)_\gamma + H(B'E|A)_\gamma - H(E|A)_\gamma - H(A'B'E|A)_\gamma, \tag{20.3.21}$$

we recover (20.3.8). ■

We can now establish the squashed entanglement bound for an approximate bipartite private state:

Proposition 20.9

Let $\gamma_{AA'BB'}$ be a private state, with key systems AB and shield systems $A'B'$, and let $\omega_{AA'BB'}$ be an ε-approximate private state, in the sense that

$$F(\gamma_{AA'BB'}, \omega_{AA'BB'}) \geq 1 - \varepsilon \tag{20.3.22}$$

for $\varepsilon \in [0, 1]$. Suppose that $|A| = |B| = K$. Then

$$(1 - 2\sqrt{\varepsilon}) \log_2 K \leq E_{\text{sq}}(AA'; BB')_{\omega} + 2g_2(\sqrt{\varepsilon}), \tag{20.3.23}$$

where

$$g_2(\delta) := (\delta + 1) \log_2(\delta + 1) - \delta \log_2 \delta. \tag{20.3.24}$$

PROOF: By applying Uhlmann's theorem for fidelity (Theorem 6.8) and the inequalities relating trace distance and fidelity from Theorem 6.14, for a given extension $\omega_{AA'BB'E}$ of $\omega_{AA'BB'}$, there exists an extension $\gamma_{AA'BB'E}$ of $\gamma_{AA'BB'}$ such that

$$\frac{1}{2} \|\gamma_{AA'BB'E} - \omega_{AA'BB'E}\|_1 \leq \sqrt{\varepsilon}. \tag{20.3.25}$$

Defining $f_1(\delta, K) := 2\delta \log_2 K + 2g_2(\delta)$, we then find that

$$\begin{aligned}
2 \log_2 K &= I(A; BB'|E)_{\gamma} + I(A'; B|AB'E)_{\gamma} && (20.3.26)\\
&\leq I(A; BB'|E)_{\omega} + I(A'; B|AB'E)_{\omega} + 2f_1(\sqrt{\varepsilon}, K) && (20.3.27)\\
&\leq I(A; BB'|E)_{\omega} + I(A'; B|AB'E)_{\omega} \\
&\qquad + I(A'; B'|AE)_{\omega} + 2f_1(\sqrt{\varepsilon}, K) && (20.3.28)\\
&= I(AA'; BB'|E)_{\omega} + 2f_1(\sqrt{\varepsilon}, K). && (20.3.29)
\end{aligned}$$

The first equality follows from Lemma 20.8. The first inequality follows from two applications of Proposition 7.10 (uniform continuity of conditional mutual information). The second inequality follows because $I(A'; B'|AE)_{\omega} \geq 0$ (this is strong subadditivity from Theorem 7.6). The last equality is a consequence of the chain rule for conditional mutual information, as used in (20.3.7). Since the inequality

$$2 \log_2 K \leq I(AA'; BB'|E)_{\omega} + 2f_1(\sqrt{\varepsilon}, K) \tag{20.3.30}$$

holds for any extension of $\omega_{AA'BB'}$, the statement of the proposition follows. ■

20.3.2 Squashed Entanglement Upper Bound

We now establish the squashed entanglement upper bound on the number of private bits that a sender can transmit to a receiver by employing a secret-key-agreement protocol. The proof is similar to that of Theorem 19.4, but it instead invokes Proposition 20.9.

Theorem 20.10 ***n*-Shot Squashed Entanglement Upper Bound**

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0, 1/4)$. For all (n, K, ε) secret-key-agreement protocols over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\log_2 M \leq \frac{1}{1-2\sqrt{\varepsilon}}\left[n \cdot E_{\text{sq}}(\mathcal{N}) + 2g_2(\sqrt{\varepsilon})\right]. \tag{20.3.31}$$

PROOF: Given an arbitrary (n, K, ε) secret-key-agreement protocol as outlined in Section 20.1, we consider its equivalent (n, K, ε) LOCC-assisted bipartite private-state distillation protocol, as outlined in Section 20.2.2. The squashed entanglement is an entanglement measure (monotone under LOCC as shown in Theorem 9.33) and it is equal to zero for separable states (Proposition 9.4.5). Thus, Proposition 20.4 applies, and we find that

$$E_{\text{sq}}(K_A S_A; K_B S_B)_\omega \leq n \cdot E_{\text{sq}}^{\mathcal{A}}(\mathcal{N}) = n \cdot E_{\text{sq}}(\mathcal{N}), \tag{20.3.32}$$

where the equality follows from Theorem 10.20. Applying Definition 20.1 and (20.2.17) leads to

$$F(\gamma_{K_A S_A K_B S_B}, \omega_{K_A S_A K_B S_B}) \geq 1 - \varepsilon, \tag{20.3.33}$$

where $\gamma_{K_A S_A K_B S_B}$ is an exact private state of $\log_2 K$ private bits. As a consequence of Proposition 20.9, we find that

$$E_{\text{sq}}(K_A S_A; K_B S_B)_\omega \geq (1 - 2\sqrt{\varepsilon}) \log_2 K - 2g_2(\sqrt{\varepsilon}). \tag{20.3.34}$$

Putting together (20.3.32) and (20.3.34), we arrive at the statement of the theorem. ■

20.4 Relative Entropy of Entanglement Upper Bounds on the Number of Transmitted Private Bits

We now establish the max-relative entropy of entanglement bound on the number of private bits that a sender can transmit to a receiver by employing a secret-key-agreement protocol assisted by public separable channels:

Theorem 20.11 ***n*-Shot Max-Relative Entropy of Entanglement Upper Bound**

Let $\mathcal{N}_{A\to B}$ be a quantum channel, and let $\varepsilon \in [0,1)$. For all (n, K, ε) secret-key-agreement protocols assisted by public separable channels, over the channel $\mathcal{N}_{A\to B}$, the following bound holds

$$\log_2 K \le n \cdot E_{\max}(\mathcal{N}) + \log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{20.4.1}$$

PROOF: Given an arbitrary (n, K, ε) secret-key-agreement protocol assisted by public separable channels as outlined in Section 20.2.4, we consider its equivalent (n, K, ε) separable-assisted bipartite private-state distillation protocol. The max-relative entropy of entanglement is an entanglement measure (monotone under separable channels channels as shown in Proposition 9.16) and it is equal to zero for separable states. Thus, Proposition 20.4 applies, and we find that

$$E_{\max}(K_A S_A; K_B S_B)_\omega \le n \cdot E^{\mathcal{A}}_{\max}(\mathcal{N}) = n \cdot E_{\max}(\mathcal{N}), \tag{20.4.2}$$

where the equality follows from Theorem 10.16. Applying Definition 20.1 and (20.2.17) leads to

$$F(\gamma_{K_A S_A K_B S_B}, \omega_{K_A S_A K_B S_B}) \ge 1 - \varepsilon, \tag{20.4.3}$$

where $\gamma_{K_A S_A K_B S_B}$ is an exact private state of $\log_2 K$ private bits. As a consequence of Propositions 15.15 and 7.71, we find that

$$\log_2 K \le E_R^{\varepsilon}(S_A K_A; S_B K_B)_\omega \tag{20.4.4}$$

$$\le E_{\max}(S_A K_A; S_B K_B)_\omega + \log_2\!\left(\frac{1}{1-\varepsilon}\right). \tag{20.4.5}$$

Combining (20.4.2) and (20.4.5), we conclude the proof. ■

For channels that are separable-simulable with associated resource states, as given in Definition 4.26, we obtain upper bounds that can be even stronger:

Theorem 20.12 ***n*-Shot Rényi–REE Upper Bounds for Separable-Simulable Channels**

Let $\mathcal{N}_{A\to B}$ be a quantum channel that is separable-simulable with associated resource state $\theta_{SB'}$, and let $\varepsilon \in [0,1)$. For all (n, K, ε) secret-key-agreement protocols assisted by public separable channels, over the channel $\mathcal{N}_{A\to B}$, the following bounds hold for all $\alpha > 1$:

$$\log_2 K \leq n \cdot \widetilde{E}_\alpha(S; B')_\theta + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{20.4.6}$$

$$\log_2 K \leq \frac{1}{1-\varepsilon} \left[n \cdot E_R(S; B')_\theta + h_2(\varepsilon)\right]. \tag{20.4.7}$$

Proof: Given an arbitrary (n, K, ε) secret-key-agreement protocol assisted by public separable channels as outlined in Section 20.2.4, we consider its equivalent (n, K, ε) separable-assisted bipartite private-state distillation protocol. The Rényi relative entropy of entanglement and relative entropy of entanglement are monotone non-increasing under separable channels (Proposition 9.16), equal to zero for separable states, and subadditive with respect to states (Proposition 9.16). As such, Corollary 20.7 applies, and we find for $\alpha > 1$ that

$$\widetilde{E}_\alpha(S_A K_A; S_B K_B)_\omega \leq n \cdot \widetilde{E}_\alpha(S; B')_\theta, \tag{20.4.8}$$

$$E_R(S_A K_A; S_B K_B)_\omega \leq n \cdot E_R(S; B')_\theta. \tag{20.4.9}$$

Applying Definition 20.1 and (20.2.17) leads to

$$F(\gamma_{K_A S_A K_B S_B}, \omega_{K_A S_A K_B S_B}) \geq 1 - \varepsilon, \tag{20.4.10}$$

where $\gamma_{K_A S_A K_B S_B}$ is an exact private state of $\log_2 K$ private bits. As a consequence of Proposition 15.15, we have that

$$\log_2 K \leq E_R^\varepsilon(S_A K_A; S_B K_B)_\omega. \tag{20.4.11}$$

Applying Propositions 7.70 and 7.71, we find that

$$\log_2 K \leq \widetilde{E}_\alpha(S_A K_A; S_B K_B)_\omega + \frac{\alpha}{\alpha - 1} \log_2\left(\frac{1}{1-\varepsilon}\right), \tag{20.4.12}$$

$$\log_2 K \leq \frac{1}{1-\varepsilon}\left[E_R(S_A K_A; S_B K_B)_\omega + h_2(\varepsilon)\right]. \tag{20.4.13}$$

Putting together (20.4.8), (20.4.9), (20.4.12), and (20.4.13) concludes the proof. ■

20.5 Secret-Key-Agreement Capacities of Quantum Channels

In this section, we analyze the asymptotic capacities, and as before, the upper bounds for the asymptotic capacities are straightforward consequences of the non-asymptotic bounds given in Sections 20.3.2 and 20.4. The definitions of these capacities are similar to what we have given previously, and so we only state them here briefly.

Definition 20.13 Achievable Rate for Secret Key Agreement

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called an achievable rate for secret key agreement over $\mathcal{N}$ if for all $\varepsilon \in (0, 1]$, all $\delta > 0$, and all sufficiently large n, there exists an $(n, 2^{n(R-\delta)}, \varepsilon)$ secret-key-agreement protocol.

Definition 20.14 Secret-Key-Agreement Capacity of a Quantum Channel

The secret-key-agreement capacity of a quantum channel $\mathcal{N}$, denoted by $P^{\leftrightarrow}(\mathcal{N})$, is defined as the supremum of all achievable rates, i.e.,

$$P^{\leftrightarrow}(\mathcal{N}) := \sup\{R : R \text{ is an achievable rate for } \mathcal{N}\}. \tag{20.5.1}$$

Definition 20.15 Weak Converse Rate for Secret Key Agreement

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a weak converse rate for secret key agreement over $\mathcal{N}$ if every $R' > R$ is not an achievable rate for $\mathcal{N}$.

Definition 20.16 Strong Converse Rate for Secret Key Agreement

Given a quantum channel $\mathcal{N}$, a rate $R \in \mathbb{R}^+$ is called a strong converse rate for secret key agreement over $\mathcal{N}$ if for all $\varepsilon \in [0, 1)$, all $\delta > 0$, and all sufficiently large n, there does not exist an $(n, 2^{n(R+\delta)}, \varepsilon)$ secret-key-agreement protocol.

Definition 20.17 Strong Converse Secret-Key-Agreement Capacity of a Quantum Channel

The strong converse secret-key-agreement capacity of a quantum channel $\mathcal{N}$, denoted by $\widetilde{P}^{\leftrightarrow}(\mathcal{N})$, is defined as the infimum of all strong converse rates, i.e.,

$$\widetilde{P}^{\leftrightarrow}(\mathcal{N}) := \inf\{R : R \text{ is a strong converse rate for } \mathcal{N}\}. \tag{20.5.2}$$

We have the exact same definitions for secret key agreement assisted by public separable channels, and we use the notation $P_{\text{SEP}}^{\leftrightarrow}$ to refer to the public-separable-assisted secret-key-agreement capacity and $\widetilde{P}_{\text{SEP}}^{\leftrightarrow}$ for the strong converse secret-key-agreement capacity assisted by public separable channels.

Recall that, by definition, the following bounds hold

$$P^{\leftrightarrow}(\mathcal{N}) \leq \widetilde{P}^{\leftrightarrow}(\mathcal{N}) \leq \widetilde{P}_{\text{SEP}}^{\leftrightarrow}(\mathcal{N}), \tag{20.5.3}$$

$$P^{\leftrightarrow}(\mathcal{N}) \leq P_{\text{SEP}}^{\leftrightarrow}(\mathcal{N}) \leq \widetilde{P}_{\text{SEP}}^{\leftrightarrow}(\mathcal{N}). \tag{20.5.4}$$

As a direct consequence of the bound in Theorem 20.10 and methods similar to those given in the proof of Theorem 11.23, we find the following:

Theorem 20.18 Squashed-Entanglement Weak-Converse Bound

The squashed entanglement of a channel $\mathcal{N}$ is a weak converse rate for secret key agreement:

$$P^{\leftrightarrow}(\mathcal{N}) \leq E_{\text{sq}}(\mathcal{N}). \tag{20.5.5}$$

As a direct consequence of the bound in Theorem 20.11 and methods similar to those given in Section 11.2.3, we find that

Theorem 20.19 Max-Relative Entropy of Entanglement Strong-Converse Bound

The max-relative entropy of entanglement of a channel $\mathcal{N}$ is a strong converse rate for secret key agreement assisted by public separable channels:

$$\widetilde{P}^{\leftrightarrow}_{\text{SEP}}(\mathcal{N}) \leq E_{\max}(\mathcal{N}). \tag{20.5.6}$$

As a direct consequence of the bound in Theorem 20.12 and methods similar to those given in Section 11.2.3, we find that

Theorem 20.20 Relative Entropy of Entanglement Strong-Converse Bound for Separable-Simulable Channels

Let $\mathcal{N}$ be a quantum channel that is separable-simulable with associated resource state $\theta_{SB'}$. Then the relative entropy of entanglement of $\theta_{SB'}$ is a strong converse rate for secret key agreement assisted by public separable channels:

$$\widetilde{P}^{\leftrightarrow}_{\text{SEP}}(\mathcal{N}) \leq E_R(S;B')_{\theta}. \tag{20.5.7}$$

20.6 Examples

[IN PROGRESS]

20.7 Bibliographic Notes

Quantum key distribution is one of the first examples of a secret-key-agreement protocol conducted over a quantum channel (Bennett and Brassard, 1984; Ekert, 1991). Secret key agreement was considered in classical information theory by Maurer (1993); Ahlswede and Csiszár (1993). Secret key distillation from a bipartite quantum state was studied by a number of researchers, including Devetak and Winter (2005); Horodecki et al. (2005a); Christandl (2006); Horodecki et al. (2008a, 2009a); Christandl et al. (2007, 2012), well before secret key agreement over quantum channels was considered. One of the seminal insights in this domain, that a tripartite secret-key-distillation protocol is equivalent to a bipartite private-state distillation protocol, was made by Horodecki et al. (2005a, 2009a). These authors also established the relative entropy of entanglement (Vedral and Plenio, 1998) as an upper bound on the rate of a secret-key-distillation protocol. In earlier work, Curty et al. (2004) observed that a separable state has no distillable secret key.

Secret key agreement over a quantum channel was formally defined by Takeoka et al. (2014), who also observed that the aforementioned insight extends to this setting, with more details being given by Kaur and Wilde (2017). The issue of unbounded shield systems resulting from secret-key-distillation or secret-key-agreement protocols was somewhat implicit in (Horodecki et al., 2005a, 2009a) and discussed in more detail by Christandl et al. (2012); Wilde (2016); Wilde et al. (2017). The relation between LOCC-assisted quantum communication and secret key agreement was discussed by Wilde et al. (2017). The notion of secret key agreement assisted by public separable channels is original to this book (including the observation that various previously known bounds apply to these more general protocols).

The use of teleportation simulation and relative entropy of entanglement as a method for bounding rates of secret-key-agreement protocols was presented by Pirandola et al. (2017).

The amortized entanglement bound for secret-key-agreement protocols (Proposition 20.4) was contributed by Kaur and Wilde (2017). Corollary 20.5, as presented here, is due to Kaur and Wilde (2017).

The squashed entanglement upper bound on the rate of a secret-key-agreement protocol was established by Takeoka et al. (2014); Wilde (2016). Lemma 20.8 and Proposition 20.9 are due to Wilde (2016).

The use of sandwiched relative entropy of entanglement for bounding key rates was contributed by Wilde et al. (2017). The max-relative entropy of entanglement of a state was introduced by Datta (2009b,a), and the generalization to channels by Christandl and Müller-Hermes (2017). Lemmas 9.21 and 10.10 were established by Berta and Wilde (2018). Proposition 10.16 was proven by Christandl and Müller-Hermes (2017). The proof that we follow here was given by Berta and Wilde (2018). The interpretation of Proposition 10.16 in terms of "amortization collapse" was given by Berta and Wilde (2018). Theorem 20.11 was proven by Christandl and Müller-Hermes (2017). The precise statement of Theorem 20.12 is due to Wilde et al. (2017); Kaur and Wilde (2017) (although it was stated for

LOCC-simulable channels in these papers).

Summary

[IN PROGRESS]

Appendix A

Analyzing General Communication Scenarios

[IN PROGRESS]

Bibliography

A. Acín, E. Bagan, M. Baig, Ll. Masanes, and R. Muñoz Tapia. Multiple-copy two-state discrimination with individual measurements. *Physical Review A*, 71:032338, March 2005. doi: 10.1103/PhysRevA.71.032338. URL https://link.aps.org/doi/10.1103/PhysRevA.71.032338.

C. Adami and N. J. Cerf. von Neumann capacity of noisy quantum channels. *Physical Review A*, 56:3470–3483, November 1997. doi: 10.1103/PhysRevA.56.3470. URL https://link.aps.org/doi/10.1103/PhysRevA.56.3470.

Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum Circuits with Mixed States. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, page 20–30, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 0897919629. doi: 10.1145/276698.276708. URL https://doi.org/10.1145/276698.276708.

Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. I. Secret sharing. *IEEE Transactions on Information Theory*, 39:1121–1132, July 1993. URL https://ieeexplore.ieee.org/document/243431.

Robert Alicki and Mark Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37:L55–L57, January 2004. URL https://doi.org/10.1088/0305-4470/37/5/l01.

Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119:120506, September 2017. doi: 10.1103/PhysRevLett.119.120506. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.120506.

Anurag Anshu, Mario Berta, Rahul Jain, and Marco Tomamichel. A minimax approach to one-shot entropy inequalities. *Journal of Mathematical Physics*, 60:122201, December 2019. doi: 10.1063/1.5126723. URL https://doi.org/10.1063/1.5126723.

Anurag Anshu, Rahul Jain, and Naqueeb A. Warsi. Building blocks for communication over noisy quantum networks. *IEEE Transactions on Information Theory*, 65:1287–1306, February 2019. doi: 10.1109/TIT.2018.2851297. URL https://ieeexplore.ieee.org/document/8399830.

Anurag Anshu, Rahul Jain, and Naqueeb A. Warsi. On the near-optimality of one-shot classical communication over quantum channels. *Journal of Mathematical Physics*, 60:012204, January 2019. doi: 10.1063/1.5039796. URL https://doi.org/10.1063/1.5039796.

Huzihiro Araki. On an inequality of Lieb and Thirring. *Letters in Mathematical Physics*, 19:167–170, February 1990. ISSN 1573-0530. doi: 10.1007/BF01045887. URL https://doi.org/10.1007/BF01045887.

S. Arora, E. Hazan, and S. Kale. Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 339–348, 2005. doi: 10.1109/SFCS.2005.35. URL https://ieeexplore.ieee.org/document/1530726.

Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 227–236, New York, NY, USA, June 2007. Association for Computing Machinery. ISBN 9781595936318. doi: 10.1145/1250790.1250823. URL https://doi.org/10.1145/1250790.1250823.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8:121–164, 2012. doi: 10.4086/toc.2012.v008a006. URL http://www.theoryofcomputing.org/articles/v008a006.

Koenraad Audenaert, Bart De Moor, Karl Gerd H. Vollbrecht, and Reinhard F. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. *Physical Review A*, 66:032310, September 2002. doi: 10.1103/PhysRevA.66.032310. URL http://link.aps.org/doi/10.1103/PhysRevA.66.032310.

Koenraad M. R. Audenaert. Comparisons between quantum state distinguishability measures. *Quantum Information and Computation*, 14:31–38, January 2014. ISSN 1533-7146. URL http://dl.acm.org/citation.cfm?id=2600498.2600500.

Koenraad M. R. Audenaert and Jens Eisert. Continuity bounds on the quantum relative entropy. *Journal of Mathematical Physics*, 46:102104, October 2005. doi: 10.1063/1.2044667. URL https://doi.org/10.1063/1.2044667.

Koenraad M. R. Audenaert, John Calsamiglia, Ramon Muñoz Tapia, Emilio Bagan, Lluis Masanes, Antonio Acin, and Frank Verstraete. Discriminating states: The quantum Chernoff bound. *Physical Review Letters*, 98:160501, April 2007. doi: 10.1103/PhysRevLett.98.160501. URL http://link.aps.org/doi/10.1103/PhysRevLett.98.160501.

Koenraad M. R. Audenaert, Milàn Mosonyi, and Frank Verstraete. Quantum state discrimination bounds for finite sample size. *Journal of Mathematical Physics*, 53:122205, December 2012. doi: 10.1063/1.4768252. URL https://doi.org/10.1063/1.4768252.

Masashi Ban, Kouichi Yamazaki, and Osamu Hirota. Accessible information in combined and sequential quantum measurementson a binary-state signal. *Physical Review A*, 55:22–26, January 1997. doi: 10.1103/PhysRevA.55.22. URL https://link.aps.org/doi/10.1103/PhysRevA.55.22.

Howard Barnum, Michael A. Nielsen, and Benjamin Schumacher. Information transmission through a noisy quantum channel. *Physical Review A*, 57:4153–4175, June 1998. doi: 10.1103/PhysRevA.57.4153. URL https://link.aps.org/doi/10.1103/PhysRevA.57.4153.

Howard Barnum, Emanuel Knill, and Michael A. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46:1317–1329, July 2000. URL https://ieeexplore.ieee.org/document/850671.

David Beckman, Daniel Gottesman, Michael A. Nielsen, and John Preskill. Causal and localizable quantum operations. *Physical Review A*, 64:052309, October 2001. doi: 10.1103/PhysRevA.64.052309. URL https://link.aps.org/doi/10.1103/PhysRevA.64.052309.

Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54:122202, December 2013. URL https://aip.scitation.org/doi/10.1063/1.4838855.

Salman Beigi, Nilanjana Datta, and Felix Leditzky. Decoding quantum information via the Petz recovery map. *Journal of Mathematical Physics*, 57:082203, August 2016. URL https://doi.org/10.1063/1.4961515.

V. P. Belavkin and P. Staszewski. C*-algebraic generalization of relative entropy and entropy. *Annales de l'I.H.P. Physique théorique*, 37:51–58, 1982. URL http://eudml.org/doc/76163.

Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, second edition, 2017. doi: 10.1017/9781139207010.

Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *International Conference on Computer System and Signal Processing, IEEE, 1984*, pages 175–179, 1984.

Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881–2884, November 1992. doi: 10.1103/PhysRevLett.69.2881. URL https://link.aps.org/doi/10.1103/PhysRevLett.69.2881.

Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, March 1993. URL https://link.aps.org/doi/10.1103/PhysRevLett.70.1895.

Charles H. Bennett, Gilles Brassard, Claude Crepeau, and Ueli M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41:1915–1923, November 1995. doi: 10.1109/18.476316. URL https://ieeexplore.ieee.org/document/476316.

Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53:2046–2052, April 1996a. doi: 10.1103/PhysRevA.53.2046. URL https://link.aps.org/doi/10.1103/PhysRevA.53.2046.

Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76:722–725, January 1996b. doi: 10.1103/PhysRevLett.76.722. URL https://link.aps.org/doi/10.1103/PhysRevLett.76.722.

Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54:3824–3851, November 1996c. doi: 10.1103/PhysRevA.54.3824. URL https://link.aps.org/doi/10.1103/PhysRevA.54.3824.

Charles H. Bennett, David P. DiVincenzo, and John A. Smolin. Capacities of quantum erasure channels. *Physical Review Letters*, 78:3217–3220, April 1997. doi: 10.1103/PhysRevLett.78.3217. URL https://link.aps.org/doi/10.1103/PhysRevLett.78.3217.

Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K. Wootters. Quantum nonlocality without entanglement. *Physical Review A*, 59:1070–1091, February 1999a. doi: 10.1103/PhysRevA.59.1070. URL http://link.aps.org/doi/10.1103/PhysRevA.59.1070.

Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83:3081–3084, October 1999b. doi: 10.1103/PhysRevLett.83.3081. URL https://link.aps.org/doi/10.1103/PhysRevLett.83.3081.

Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48:2637–2655, October 2002. URL https://ieeexplore.ieee.org/document/1035117.

Charles H. Bennett, Aram W. Harrow, Debbie W. Leung, and John A. Smolin. On the capacities of bipartite Hamiltonians and unitary gates. *IEEE Transactions on Information Theory*, 49:1895–1911, August 2003. ISSN 0018-9448. doi: 10.1109/TIT.2003.814935. URL https://ieeexplore.ieee.org/document/1214070.

Charles H. Bennett, Igor Devetak, Peter W. Shor, and John A. Smolin. Inequalities and separations among assisted capacities of quantum channels. *Physical Review Letters*, 96:150502, April 2006. URL https://link.aps.org/doi/10.1103/PhysRevLett.96.150502.

Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60:2926–2959, May 2014. URL https://ieeexplore.ieee.org/document/6757002.

Dominic W. Berry. Qubit channels that achieve capacity with two states. *Physical Review A*, 71:032334, March 2005. URL https://link.aps.org/doi/10.1103/PhysRevA.71.032334.

Mario Berta. Single-shot quantum state merging. Diploma thesis, ETH Zurich, February 2008.

Mario Berta and Mark M. Wilde. Amortization does not enhance the max-Rains information of a quantum channel. *New Journal of Physics*, 20:053044, May 2018. doi: 10.1088/1367-2630/aac153. URL https://doi.org/10.1088/1367-2630/aac153.

Mario Berta, Omar Fawzi, and Marco Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107:2239–2265, December 2017. URL https://doi.org/10.1007/s11005-017-0990-7.

Reinhold A. Bertlmann and Philipp Krammer. Bloch vectors for qudits. *Journal of Physics A: Mathematical and Theoretical*, 41:235303, May 2008. doi: 10.1088/1751-8113/41/23/235303. URL https://doi.org/10.1088/1751-8113/41/23/235303.

Rajendra Bhatia. *Matrix Analysis*. Springer New York, 1997. doi: 10.1007/978-1-4612-0653-8.

Igor Bjelakovic and Rainer Siegmund-Schultze. Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb, July 2003.

Garry Bowen. Quantum feedback channels. *IEEE Transactions on Information Theory*, 50: 2429–2434, October 2004. URL https://ieeexplore.ieee.org/document/1337116.

Garry Bowen and Sougato Bose. Teleportation as a depolarizing quantum channel, relative entropy, and classical capacity. *Physical Review Letters*, 87:267901, December 2001. doi: 10.1103/PhysRevLett.87.267901. URL https://link.aps.org/doi/10.1103/PhysRevLett.87.267901.

Garry Bowen and Rajagopal Nagarajan. On feedback and the classical capacity of a noisy quantum channel. *IEEE Transactions on Information Theory*, 51:320–324, January 2005. URL https://ieeexplore.ieee.org/document/1365361.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Fernando G. S. L. Brandao and Nilanjana Datta. One-shot rates for entanglement manipulation under non-entangling maps. *IEEE Transactions on Information Theory*, 57:1754–1760, March 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2104531. URL https://ieeexplore.ieee.org/document/5714245.

Fernando G.S.L. Brandao, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 306:805–830, September 2011. ISSN 0010-3616. doi: 10.1007/s00220-011-1302-1. URL http://dx.doi.org/10.1007/s00220-011-1302-1.

Sarah Brandsen, Mengke Lian, Kevin D. Stubbs, Narayanan Rengaswamy, and Henry D. Pfister. Adaptive Procedures for Discriminating Between Arbitrary Tensor-Product Quantum States. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1933–1938, 2020. doi: 10.1109/ISIT44484.2020.9174234. URL https://ieeexplore.ieee.org/abstract/document/9174234.

Samuel L. Braunstein and H. J. Kimble. Teleportation of continuous quantum variables. *Physical Review Letters*, 80:869–872, January 1998. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.869.

Samuel L. Braunstein, Giacomo M. D'Ariano, Gerard J. Milburn, and Massimiliano F. Sacchi. Universal teleportation with a twist. *Physical Review Letters*, 84:3486–3489, April 2000. doi: 10.1103/PhysRevLett.84.3486. URL https://link.aps.org/doi/10.1103/PhysRevLett.84.3486.

Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, 2002.

Dorje Brody and Bernhard Meister. Minimum decision cost for quantum ensembles. *Physical Review Letters*, 76:1–5, January 1996. doi: 10.1103/PhysRevLett.76.1. URL https://link.aps.org/doi/10.1103/PhysRevLett.76.1.

Francesco Buscemi and Nilanjana Datta. The quantum capacity of channels with arbitrarily correlated noise. *IEEE Transactions on Information Theory*, 56:1447–1460, March 2010a. ISSN 0018-9448. doi: 10.1109/TIT.2009.2039166. URL https://ieeexplore.ieee.org/document/5429118.

Francesco Buscemi and Nilanjana Datta. Distilling entanglement from arbitrary resources. *Journal of Mathematical Physics*, 51:102201, October 2010b. doi: http://dx.doi.org/10.1063/1.3483717. URL http://scitation.aip.org/content/aip/journal/jmp/51/10/10.1063/1.3483717.

Mark S. Byrd and Navin Khaneja. Characterization of the positivity of the density matrix in terms of the coherence vector representation. *Physical Review A*, 68:062322, December 2003. doi: 10.1103/PhysRevA.68.062322. URL https://link.aps.org/doi/10.1103/PhysRevA.68.062322.

Ning Cai, Andreas Winter, and Raymond W. Yeung. Quantum privacy and quantum wiretap channels. *Problems of Information Transmission*, 40:318–336, October 2004. ISSN 0032-9460. URL http://dx.doi.org/10.1007/s11122-005-0002-x.

Gianfranco Cariolaro and Tomaso Erseghe. *Pulse Position Modulation*. John Wiley & Sons, Inc., 2003. ISBN 9780471219286. doi: 10.1002/0471219282.eot394. URL http://dx.doi.org/10.1002/0471219282.eot394.

Eric A. Carlen. Trace inequalities and quantum entropy: An introductory course. *Contemporary Mathematics*, 529:73–140, 2010. URL http://www.ueltschi.org/AZschool/notes/EricCarlen.pdf.

Filippo Caruso and Vittorio Giovannetti. Degradability of bosonic Gaussian channels. *Physical Review A*, 74:062307, December 2006. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.74.062307.

Nicholas J. Cerf and Christoph Adami. Negative entropy and information in quantum mechanics. *Physical Review Letters*, 79:5194–5197, December 1997. doi: 10.1103/PhysRevLett.79.5194. URL https://link.aps.org/doi/10.1103/PhysRevLett.79.5194.

Nicolas J. Cerf and Chris Adami. Information theory of quantum entanglement and measurement. *Physica D: Nonlinear Phenomena*, 120:62–81, September 1998. ISSN 0167-2789. doi: https://doi.org/10.1016/S0167-2789(98)00045-1. URL http://www.sciencedirect.com/science/article/pii/S0167278998000451.

Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 12 1952. doi: 10.1214/aoms/1177729330. URL https://doi.org/10.1214/aoms/1177729330.

Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Realization schemes for quantum instruments in finite dimensions. *Journal of Mathematical Physics*, 50:042101, April 2009. doi: 10.1063/1.3105923. URL http://dx.doi.org/10.1063/1.3105923.

Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything you always wanted to know about LOCC (but were afraid to ask). *Communications in Mathematical Physics*, 328:303–326, May 2014. ISSN 1432-0916. doi: 10.1007/s00220-014-1953-9. URL http://dx.doi.org/10.1007/s00220-014-1953-9.

Eric Chitambar, Julio I. de Vicente, Mark W. Girard, and Gilad Gour. Entanglement manipulation and distillability beyond LOCC. *Journal of Mathematical Physics*, 61:042201, April 2020. URL https://doi.org/10.1063/1.5124109.

Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and Its Applications*, 10:285–290, 1975. URL https://www.sciencedirect.com/science/article/pii/0024379575900750.

Matthias Christandl. *The Structure of Bipartite Quantum States: Insights from Group Theory and Cryptography*. PhD thesis, University of Cambridge, April 2006.

Matthias Christandl and Alexander Müller-Hermes. Relative entropy bounds on quantum, private and repeater capacities. *Communications in Mathematical Physics*, 353:821–852, July 2017. doi: 10.1007/s00220-017-2885-y. URL https://doi.org/10.1007/s00220-017-2885-y.

Matthias Christandl and Andreas Winter. 'Squashed entanglement': An additive entanglement measure. *Journal of Mathematical Physics*, 45:829–840, March 2004. URL https://doi.org/10.1063/1.1643788.

Matthias Christandl, Artur Ekert, Michal Horodecki, Pawel Horodecki, Jonathan Oppenheim, and Renato Renner. Unifying classical and quantum key distillation. *Proceedings of the 4th Theory of Cryptography Conference, Lecture Notes in Computer Science*, 4392:456–478, February 2007. URL https://doi.org/10.1007/978-3-540-70936-7_25.

Matthias Christandl, Norbert Schuch, and Andreas Winter. Entanglement of the antisymmetric state. *Communications in Mathematical Physics*, 311:397–422, March 2012. URL https://doi.org/10.1007/s00220-012-1446-7.

Benoît Collins. Moments and cumulants of polynomial random variables on unitarygroups, the Itzykson-Zuber integral, and free probability. *International Mathematics Research Notices*, 2003:953–982, 01 2003. ISSN 1073-7928. doi: 10.1155/S107379280320917X. URL https://doi.org/10.1155/S107379280320917X.

Benoît Collins and Piotr Śniady. Integration with Respect to the Haar Measure on Unitary, Orthogonal and Symplectic Group. *Communications in Mathematical Physics*, 264:773–795, June 2006. ISSN 1432-0916. doi: 10.1007/s00220-006-1554-3. URL https://doi.org/10.1007/s00220-006-1554-3.

Tom Cooney, Milan Mosonyi, and Mark M. Wilde. Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *Communications in Mathematical Physics*, 344:797–829, June 2016. URL https://doi.org/10.1007/s00220-016-2645-4.

John Cortese. Relative entropy and single qubit Holevo-Schumacher-Westmoreland channel capacity, July 2002.

Imre Csiszár and Janos Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24:339–348, May 1978. URL https://ieeexplore.ieee.org/document/1055892.

Toby Cubitt, David Elkouss, William Matthews, Maris Ozols, David Pérez-García, and Sergii Strelchuk. Unbounded number of channel uses may be required to detect quantum capacity. *Nature Communications*, 6:6739, 2015. URL https://doi.org/10.1038/ncomms7739.

Toby S. Cubitt, Mary Beth Ruskai, and Graeme Smith. The structure of degradable quantum channels. *Journal of Mathematical Physics*, 49:102104, 2008. URL https://doi.org/10.1063/1.2953685.

Marcos Curty, Maciej Lewenstein, and Norbert Lütkenhaus. Entanglement as a precondition for secure quantum key distribution. *Physical Review Letters*, 92:217903, May 2004. doi: 10.1103/PhysRevLett.92.217903. URL https://link.aps.org/doi/10.1103/PhysRevLett.92.217903.

Nilanjana Datta. Max-relative entropy of entanglement, alias log robustness. *International Journal of Quantum Information*, 7:475–491, January 2009a. URL https://www.worldscientific.com/doi/abs/10.1142/S0219749909005298.

Nilanjana Datta. Min- and max-relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55:2816–2826, June 2009b. URL https://ieeexplore.ieee.org/document/4957651.

Nilanjana Datta and Min-Hsiu Hsieh. One-shot entanglement-assisted quantum and classical communication. *IEEE Transactions on Information Theory*, 59:1929–1939, March 2013. URL https://ieeexplore.ieee.org/document/6359930.

Nilanjana Datta and Felix Leditzky. A limit of the quantum Rényi divergence. *Journal of Physics A: Mathematical and Theoretical*, 47:045304, January 2014. URL http://stacks.iop.org/1751-8121/47/i=4/a=045304.

Nilanjana Datta, Milan Mosonyi, Min-Hsiu Hsieh, and Fernando G. S. L. Brandão. A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels. *IEEE Transactions on Information Theory*, 59:8014–8026, December 2013. ISSN 0018-9448. doi: 10.1109/TIT.2013.2282160. URL https://ieeexplore.ieee.org/document/6670246.

Nilanjana Datta, Marco Tomamichel, and Mark M. Wilde. On the second-order asymptotics for entanglement-assisted communication. *Quantum Information Processing*, 15:2569–2591, June 2016. URL https://doi.org/10.1007/s11128-016-1272-5.

Edward B. Davies and J. T. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics*, 17:239–260, 1970. URL https://doi.org/10.1007/BF01647093.

John de Pillis. Linear transformations which preserve Hermitian and positive semidefinite operators. *Pacific Journal of Mathematics*, 23:129–137, 1967. doi: pjm/1102991990. URL https://projecteuclid.org/journals/pacific-journal-of-mathematics/volume-23/issue-1/Linear-transformations-which-preserve-hermitian-and-positive-semidefinite-operators/pjm/1102991990.full.

Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51:44–55, January 2005. doi: 10.1109/TIT.2004.839515. URL https://ieeexplore.ieee.org/document/1377491.

Igor Devetak and Peter W. Shor. The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information. *Communications in Mathematical Physics*, 256:287–303, June 2005. URL https://doi.org/10.1007/s00220-005-1317-6.

Igor Devetak and Andreas Winter. Distillation of secret key and entanglement from quantum states. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 461:207–235, January 2005. doi: 10.1098/rspa.2004.1372. URL https://doi.org/10.1098/rspa.2004.1372.

Igor Devetak, Marius Junge, Christopher King, and Mary Beth Ruskai. Multiplicativity of completely bounded p-norms implies a new additivity result. *Communications in Mathematical Physics*, 266:37–63, August 2006. URL https://doi.org/10.1007/s00220-006-0034-0.

Dawei Ding and Mark M. Wilde. Strong converse for the feedback-assisted classical capacity of entanglement-breaking channels. *Problems of Information Transmission*, 54:1–19, 2018. URL https://doi.org/10.1134/S0032946018010015.

Dawei Ding, Yihui Quek, Peter W. Shor, and Mark M. Wilde. Entropy bound for the classical capacity of a quantum channel assisted by classical feedback. In *Proceedings of the 2019 IEEE International Symposium on Information Theory*, pages 250–254, Paris, France, July 2019. URL https://ieeexplore.ieee.org/document/8849604.

Dawei Ding, Sumeet Khatri, Yihui Quek, Peter W. Shor, Xin Wang, and Mark M. Wilde. Bounding the forward classical capacity of bipartite quantum channels. *IEEE Transactions on Information Theory*, 69(5):3034–3061, May 2023. doi: 10.1109/TIT.2022.3233924. URL https://ieeexplore.ieee.org/document/10005080.

David P. DiVincenzo, Peter W. Shor, and John A. Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57:830–839, February 1998. doi: 10.1103/PhysRevA.57.830. URL https://link.aps.org/doi/10.1103/PhysRevA.57.830.

David P. DiVincenzo, Peter W. Shor, John A. Smolin, Barbara M. Terhal, and Ashish V. Thapliyal. Evidence for bound entangled states with negative partial transpose. *Physical Review A*, 61, May 2000. doi: 10.1103/physreva.61.062312. URL https://link.aps.org/doi/10.1103/PhysRevA.61.062312.

Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69:022308, February 2004. doi: 10.1103/PhysRevA.69.022308. URL https://link.aps.org/doi/10.1103/PhysRevA.69.022308.

Frederic Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, University of Montreal, April 2010.

Frédéric Dupuis and Mark M. Wilde. Swiveled Rényi entropies. *Quantum Information Processing*, 15:1309–1345, March 2016. ISSN 1573-1332. doi: 10.1007/s11128-015-1211-x. URL http://dx.doi.org/10.1007/s11128-015-1211-x.

Frederic Dupuis, Lea Kraemer, Philippe Faist, Joseph M. Renes, and Renato Renner. Generalized entropies. *XVIIth International Congress on Mathematical Physics*, pages 134–153, 2013. URL https://doi.org/10.1142/9789814449243_0008.

Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. One-shot decoupling. *Communications in Mathematical Physics*, 328:251–284, May 2014. ISSN 1432-0916. doi: 10.1007/s00220-014-1990-4. URL http://dx.doi.org/10.1007/s00220-014-1990-4.

Wolfgang Dür, J. Ignacio Cirac, Maciej Lewenstein, and Dagmar Bruß. Distillability and partial transposition in bipartite systems. *Physical Review A*, 61:062313, May 2000. doi: 10.1103/PhysRevA.61.062313. URL https://link.aps.org/doi/10.1103/PhysRevA.61.062313.

T. Eggeling, D. Schlingemann, and Reinhard F. Werner. Semicausal operations are semilocalizable. *Europhysics Letters*, 57:782–788, March 2002. doi: 10.1209/epl/i2002-00579-4. URL https://doi.org/10.1209%2Fepl%2Fi2002-00579-4.

Harold G. Eggleston. *Convexity*. Cambridge University Press, 1958.

Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67:661–663, August 1991. URL https://link.aps.org/doi/10.1103/PhysRevLett.67.661.

David Elkouss and Sergii Strelchuk. Superadditivity of private information for any number of uses of the channel. *Physical Review Letters*, 115:040501, July 2015. doi: 10.1103/PhysRevLett.115.040501. URL http://link.aps.org/doi/10.1103/PhysRevLett.115.040501.

Kun Fang and Hamza Fawzi. Geometric Rényi divergence and its applications in quantum channel capacities. *Communications in Mathematical Physics*, 384:1615–1677, June 2021. doi: 10.1007/s00220-021-04064-4. URL https://doi.org/10.1007/s00220-021-04064-4.

William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, third edition, 1968.

Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54:122201, December 2013. URL https://doi.org/10.1063/1.4838835.

Bert Fristedt and Lawrence Gray. *A Modern Approach to Probability Theory*. Birkhäuser, Boston, 1997. ISBN 978-1-4899-2839-9.

Christopher A. Fuchs and Carlton M. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3:345–356, 1995. doi: 10.1007/BF02228997. URL https://doi.org/10.1007/BF02228997.

Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum mechanical states. *IEEE Transactions on Information Theory*, 45:1216–1227, May 1998. URL https://ieeexplore.ieee.org/document/761271.

Jun-Ichi Fujii, Masatoshi Fujii, and Ritsuo Nakamoto. Jensen's operator inequality and its application. *Sûrikaisekikenkyûsho Kôkyûroku*, 1396:85–93, March 2004. URL http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1396-10.pdf.

Jingliang Gao. Quantum union bounds for sequential projective measurements. *Physical Review A*, 92:052331, November 2015. doi: 10.1103/PhysRevA.92.052331. URL https://link.aps.org/doi/10.1103/PhysRevA.92.052331.

Raúl García-Patrón, Stefano Pirandola, Seth Lloyd, and Jeffrey H. Shapiro. Reverse coherent information. *Physical Review Letters*, 102:210501, May 2009. doi: 10.1103/PhysRevLett.102.210501. URL https://link.aps.org/doi/10.1103/PhysRevLett.102.210501.

Raul García-Patrón, William Matthews, and Andreas Winter. Quantum enhancement of randomness distribution. *IEEE Transactions on Information Theory*, 64:4664–4673, June 2018. URL https://ieeexplore.ieee.org/document/8328871.

I. Gelfand and Mark Aronovich Naimark. On the imbedding of normed rings into the ring of operators in Hilbert space. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):197–217, 1943. URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=sm&paperid=6155&option_lang=eng.

Sevag Gharibian. Strong NP-hardness of the quantum separability problem. *Quantum Information and Computation*, 10:343–360, March 2010. ISSN 1533-7146. URL https://doi.org/10.26421/QIC10.3-4-11.

Géza Giedke and J. Ignacio Cirac. Characterization of Gaussian operations and distillation of Gaussian states. *Physical Review A*, 66:032316, September 2002. doi: 10.1103/PhysRevA.66.032316. URL https://link.aps.org/doi/10.1103/PhysRevA.66.032316.

Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71:062310, June 2005. URL https://link.aps.org/doi/10.1103/PhysRevA.71.062310.

Vittorio Giovannetti and Rosario Fazio. Information-capacity description of spin-chain correlations. *Physical Review A*, 71:032314, March 2005. doi: 10.1103/PhysRevA.71.032314. URL https://link.aps.org/doi/10.1103/PhysRevA.71.032314.

Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Achieving the Holevo bound via sequential measurements. *Physical Review A*, 85:012302, January 2012. URL https://link.aps.org/doi/10.1103/PhysRevA.85.012302.

Daniel Gottesman and Isaac L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390–393, November 1999. doi: 10.1038/46503. URL https://doi.org/10.1038/46503.

Markus Grassl, Thomas Beth, and Thomas Pellizzari. Codes for the quantum erasure channel. *Physical Review A*, 56:33–38, July 1997. doi: 10.1103/PhysRevA.56.33. URL https://link.aps.org/doi/10.1103/PhysRevA.56.33.

Manish Gupta and Mark M. Wilde. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. *Communications in Mathematical Physics*, 334:867–887, March 2015. URL https://doi.org/10.1007/s00220-014-2212-9.

Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69:448–484, 2004. ISSN 0022-0000. doi: https://doi.org/10.1016/j.jcss.2004.06.003. URL http://www.sciencedirect.com/science/article/pii/S0022000004000893.

Brian C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Springer New York, 2013. ISBN 9781461471165.

Frank Hansen and Gert K. Pedersen. Jensen's operator inequality. *Bulletin of the London Mathematical Society*, 35:553–564, July 2003. ISSN 1469-2120. doi: 10.1112/S0024609303002200. URL http://dx.doi.org/10.1112/S0024609303002200.

Aram W. Harrow. The church of the symmetric subspace, 2013.

Matthew B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255–257, April 2009. URL https://doi.org/10.1038/nphys1224.

Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, September 1996. doi: 10.1103/PhysRevA.54.1869. URL https://link.aps.org/doi/10.1103/PhysRevA.54.1869.

Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76:062301, December 2007. doi: 10.1103/PhysRevA.76.062301. URL https://link.aps.org/doi/10.1103/PhysRevA.76.062301.

Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation*. Springer, second edition, 2017.

Masahito Hayashi and Hiroshi Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49:1753–1768, July 2003. URL https://ieeexplore.ieee.org/document/1023343.

Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246:359–374, April 2004. URL https://doi.org/10.1007/s00220-004-1049-z.

Patrick Hayden, Michał Horodecki, Andreas Winter, and Jon Yard. A decoupling approach to the quantum capacity. *Open Systems & Information Dynamics*, 15:7–19, 2008a. doi: 10.1142/S1230161208000043. URL https://doi.org/10.1142/S1230161208000043.

Patrick Hayden, Peter W. Shor, and Andreas Winter. Random quantum codes from Gaussian ensembles and an uncertainty relation. *Open Systems & Information Dynamics*, 15:71–89, March 2008b. URL https://doi.org/10.1142/S1230161208000079.

Teiko Heinosaari and Mário Ziman. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press, 2012.

Carl W. Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10:254–291, 1967. ISSN 0019-9958. URL https://doi.org/10.1016/S0019-9958(67)90302-6.

Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969. ISSN 0022-4715. doi: 10.1007/BF01007479. URL https://doi.org/10.1007/BF01007479.

Carl W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.

Fumio Hiai and Milán Mosonyi. Different quantum f-divergences and the reversibility of quantum operations. *Reviews in Mathematical Physics*, 29:1750023, August 2017. URL https://doi.org/10.1142/S0129055X17500234.

Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143:99–114, December 1991. URL https://doi.org/10.1007/BF02100287.

B. L. Higgins, A. C. Doherty, S. D. Bartlett, G. J. Pryde, and H. M. Wiseman. Multiple-copy state discrimination: Thinking globally, acting locally. *Physical Review A*, 83:052314, May 2011. doi: 10.1103/PhysRevA.83.052314. URL https://link.aps.org/doi/10.1103/PhysRevA.83.052314.

Foek T. Hioe and Joseph H. Eberly. n-level coherence vector and higher conservation laws in quantum optics and quantum mechanics. *Physical Review Letters*, 47:838–841, September 1981. doi: 10.1103/PhysRevLett.47.838. URL https://link.aps.org/doi/10.1103/PhysRevLett.47.838.

Alexander S. Holevo. An analogue of statistical decision theory and noncommutative probability theory. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 26:133–149, 1972a. URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=mmo&paperid=260&option_lang=eng.

Alexander S. Holevo. On quasiequivalence of locally normal states. *Theoretical and Mathematical Physics*, 13:1071–1082, November 1972b. ISSN 1573-9333. URL https://doi.org/10.1007/BF01035528.

Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973. URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=ppi&paperid=903&option_lang=eng.

Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44:269–273, January 1998. URL https://ieeexplore.ieee.org/document/651037.

Alexander S. Holevo. On entanglement assisted classical capacity. *Journal of Mathematical Physics*, 43:4326–4333, September 2002a. URL https://doi.org/10.1063/1.1495877.

Alexander S. Holevo. Remarks on the classical capacity of quantum channel, December 2002b.

Alexander S. Holevo. Multiplicativity of p-norms of completely positive maps and the additivity problem in quantum information theory. *Russian Mathematical Surveys*, 61:301–339, 2006. URL https://doi.org/10.1070/rm2006v061n02abeh004313.

Alexander S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*, volume 16. Walter de Gruyter, 2013.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Matrix Analysis. Cambridge University Press, 2013. ISBN 9780521839402.

Karol Horodecki, Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. Secure key from bound entanglement. *Physical Review Letters*, 94:160502, April 2005a. doi: 10.1103/PhysRevLett.94.160502. URL http://link.aps.org/doi/10.1103/PhysRevLett.94.160502.

Karol Horodecki, Michał Horodecki, Paweł Horodecki, Debbie Leung, and Jonathan Oppenheim. Quantum key distribution based on private states: Unconditional security over untrusted channels with zero quantum capacity. *IEEE Transactions on Information Theory*, 54:2604–2620, June 2008a. ISSN 0018-9448. doi: 10.1109/TIT.2008.921870. URL https://ieeexplore.ieee.org/document/4529275.

Karol Horodecki, Michał Horodecki, Paweł Horodecki, Debbie Leung, and Jonathan Oppenheim. Unconditional privacy over channels which cannot convey quantum information. *Physical Review Letters*, 100:110502, March 2008b. doi: 10.1103/PhysRevLett.100.110502. URL http://link.aps.org/doi/10.1103/PhysRevLett.100.110502.

Karol Horodecki, Michal Horodecki, Pawel Horodecki, and Jonathan Oppenheim. General paradigm for distilling classical key from quantum states. *IEEE Transactions on Information Theory*, 55: 1898–1929, April 2009a. URL https://ieeexplore.ieee.org/document/4802308.

Michał Horodecki. Simplifying monotonicity conditions for entanglement measures. *Open Systems & Information Dynamics*, 12:231–237, September 2005. URL https://doi.org/10.1007/s11080-005-0920-5.

Michał Horodecki and Paweł Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Physical Review A*, 59:4206–4216, June 1999. doi: 10.1103/PhysRevA.59.4206. URL http://link.aps.org/doi/10.1103/PhysRevA.59.4206.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223:1–8, November 1996. doi: 10.1016/s0375-9601(96)00706-2. URL https://www.sciencedirect.com/science/article/pii/S0375960196007062.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a 'bound' entanglement in nature? *Physical Review Letters*, 80:5239–5242, June 1998. doi: 10.1103/physrevlett.80.5239. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.5239.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A*, 60:1888–1898, September 1999. doi: 10.1103/PhysRevA.60.1888. URL https://link.aps.org/doi/10.1103/PhysRevA.60.1888.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Unified approach to quantum capacities: Towards quantum noisy coding theorem. *Physical Review Letters*, 85:433–436, July 2000. doi: 10.1103/PhysRevLett.85.433. URL https://link.aps.org/doi/10.1103/PhysRevLett.85.433.

Michal Horodecki, Peter W. Shor, and Mary Beth Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15:629–641, 2003. URL https://doi.org/10.1142/S0129055X03001709.

Michal Horodecki, Jonathan Oppenheim, and Andreas Winter. Partial quantum information. *Nature*, 436:673–676, August 2005b. URL https://doi.org/10.1038/nature03909.

Michal Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, January 2007. URL https://doi.org/10.1007/s00220-006-0118-x.

Pawel Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Physics Letters A*, 232:333–339, August 1997. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(97)00416-7. URL http://www.sciencedirect.com/science/article/pii/S0375960197004167.

Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, June 2009b. doi: 10.1103/RevModPhys.81.865. URL https://link.aps.org/doi/10.1103/RevModPhys.81.865.

Lawrence M. Ioannou. Computational complexity of the quantum separability problem. *Quantum Information and Computation*, 7:335–370, May 2007. ISSN 1533-7146. URL https://doi.org/10.26421/QIC7.4-5.

Raban Iten, Joseph M. Renes, and David Sutter. Pretty good measures in quantum information theory. *IEEE Transactions on Information Theory*, 63:1270–1279, February 2017. doi: 10.1109/TIT.2016.2639521. URL https://ieeexplore.ieee.org/document/7782776.

Vojkan Jaksic, Yoshiko Ogata, Claude-Alain Pillet, and Robert Seiringer. Quantum hypothesis testing and non-equilibrium statistical mechanics. *Reviews in Mathematical Physics*, 24:1230002, 2012. URL https://doi.org/10.1142/S0129055X12300026.

A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3:275–278, 1972. ISSN 0034-4877. doi: https://doi.org/10.1016/0034-4877(72)90011-0. URL https://www.sciencedirect.com/science/article/pii/0034487772900110.

Anna Jencova. A relation between completely bounded norms and conjugate channels. *Communications in Mathematical Physics*, 266:65–70, August 2006. URL https://doi.org/10.1007/s00220-006-0035-z.

Anna Jencova. Quantum hypothesis testing and sufficient subalgebras. *Letters in Mathematical Physics*, 93:15–27, 2010. URL https://doi.org/10.1007/s11005-010-0398-0.

Vishal Katariya and Mark M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. *Quantum Information Processing*, 20:78, April 2021. URL https://doi.org/10.1007/s11128-021-02992-7.

Eneet Kaur and Mark M. Wilde. Amortized entanglement of a quantum channel and approximately teleportation-simulable channels. *Journal of Physics A: Mathematical and Theoretical*, July 2017. URL http://iopscience.iop.org/10.1088/1751-8121/aa9da7.

Johannes Henricus Bernardus Kemperman. Strong converses for a general memoryless channel with feedback. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes*, 1971.

Leonid G Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20:53–72, 1980. ISSN 0041-5553. URL https://www.sciencedirect.com/science/article/pii/0041555380900610.

Sumeet Khatri, Eneet Kaur, Saikat Guha, and Mark M. Wilde. Second-order coding rates for key distillation in quantum key distribution, October 2019.

Sumeet Khatri, Kunal Sharma, and Mark M. Wilde. Information-theoretic aspects of the generalized amplitude-damping channel. *Physical Review A*, 102:012401, July 2020. URL https://link.aps.org/doi/10.1103/PhysRevA.102.012401.

Gen Kimura. The Bloch vector for N-level systems. *Physics Letters A*, 314:339–349, 2003. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(03)00941-1. URL https://www.sciencedirect.com/science/article/pii/S0375960103009411.

Christopher King. Maximal p-norms of entanglement breaking channels. *Quantum Information and Computation*, 3:186–190, 2003a. URL https://doi.org/10.26421/QIC3.2-9.

Christopher King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49:221–229, January 2003b. ISSN 0018-9448. URL https://ieeexplore.ieee.org/document/1159773.

Christopher King, Keiji Matsumoto, Michael Nathanson, and Mary Beth Ruskai. Properties of conjugate channels with applications to additivity and multiplicativity. *Markov Processes and Related Fields*, 13:391–423, 2007. URL http://math-mprf.org/journal/articles/id1123/. J. T. Lewis memorial issue.

Alexei Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52:1191–1249, 1997. URL https://doi.org/10.1070/rm1997v052n06abeh002155.

Oskar Klein. Zur Quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre. *Z. Physik*, 72:767–775, 1931.

Rochus Klesse. Approximate quantum error correction, random codes, and quantum channel capacity. *Physical Review A*, 75:062315, June 2007. doi: 10.1103/PhysRevA.75.062315. URL https://link.aps.org/doi/10.1103/PhysRevA.75.062315.

Rochus Klesse. A random coding based proof for the quantum coding theorem. *Open Systems & Information Dynamics*, 15:21–45, March 2008. URL https://doi.org/10.1142/S1230161208000055.

Robert Koenig and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103:070504, August 2009. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.070504.

Robert Koenig, Renato Renner, and Christian Schaffner. The Operational Meaning of Min- and Max-Entropy. *IEEE Transactions on Information Theory*, 55:4337–4347, September 2009. URL https://ieeexplore.ieee.org/document/5208530.

Pieter Kok, W. J. Munro, Kae Nemoto, T. C. Ralph, Jonathan P. Dowling, and G. J. Milburn. Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics*, 79:135–174, January 2007. doi: 10.1103/RevModPhys.79.135. URL https://link.aps.org/doi/10.1103/RevModPhys.79.135.

Hidetoshi Komiya. Elementary proof for Sion's minimax theorem. *Kodai Mathematical Journal*, 11:5–7, 1988. URL https://doi.org/10.2996/kmj/1138038812.

Karl Kraus. *States, Effects and Operations: Fundamental Notions of Quantum Theory,*. Springer Verlag, 1983.

Dennis Kretschmann and Reinhard F. Werner. Tema con variazioni: quantum channel capacity. *New Journal of Physics*, 6:26, 2004. URL http://stacks.iop.org/1367-2630/6/i=1/a=026.

Erwin Kreyszig. *Introductory Functional Analysis with Applications*. Wiley Classics Library. Wiley, 1989. ISBN 9780471504597.

Lev Landau. Das dämpfungsproblem in der wellenmechanik. *Zeitschrift für Physik*, 45:430–441, May 1927. ISSN 0044-3328. URL https://doi.org/10.1007/BF01343064.

Oscar E. Lanford, III and Derek W. Robinson. Mean entropy of states in quantum-statistical mechanics. *Journal of Mathematical Physics*, 9:1120–1125, July 1968. doi: 10.1063/1.1664685. URL https://doi.org/10.1063/1.1664685.

Jimmie D. Lawson and Yongdo Lim. The geometric mean, matrices, metrics, and more. *The American Mathematical Monthly*, 108:797–812, November 2001. doi: 10.1080/00029890.2001.11919815. URL https://doi.org/10.1080/00029890.2001.11919815.

Felix Leditzky. *Relative entropies and their use in quantum information theory*. PhD thesis, University of Cambridge, November 2016.

Felix Leditzky. Distillable key of degradable states. unpublished, August 2019. private email communication.

Felix Leditzky, Nilanjana Datta, and Graeme Smith. Useful states and entanglement distillation. *IEEE Transactions on Information Theory*, 64:4689–4708, July 2018. doi: 10.1109/TIT.2017.2776907. URL https://ieeexplore.ieee.org/document/8119865.

Felix Leditzky, Eneet Kaur, Nilanjana Datta, and Mark M. Wilde. Approaches for approximate additivity of the Holevo information of quantum channels. *Physical Review A*, 97:012332, January 2018. doi: 10.1103/PhysRevA.97.012332. URL https://link.aps.org/doi/10.1103/PhysRevA.97.012332.

Yin Tat Lee, Aaron Sidford, and Sam Chiu Wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *IEEE 56th Annual Symposium on the Foundations of Computer Science*, pages 1049–1065, October 2015. URL https://ieeexplore.ieee.org/document/7354442.

Matthew S. Leifer. Conditional density operators and the subjectivity of quantum operations. *AIP Conference Proceedings*, 889:172–186, February 2007. doi: 10.1063/1.2713456. URL https://aip.scitation.org/doi/abs/10.1063/1.2713456.

Matthew S. Leifer and Robert W. Spekkens. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Physical Review A*, 88:052130, November 2013. doi: 10.1103/PhysRevA.88.052130. URL http://link.aps.org/doi/10.1103/PhysRevA.88.052130.

Matthew S. Leifer, Leah Henderson, and Noah Linden. Optimal entanglement generation from quantum operations. *Physical Review A*, 67:012306, January 2003. doi: 10.1103/physreva.67.012306. URL https://link.aps.org/doi/10.1103/PhysRevA.67.012306.

Debbie Leung and William Matthews. On the power of PPT-preserving and non-signalling codes. *IEEE Transactions on Information Theory*, 61:4486–4499, August 2015. ISSN 0018-9448. doi: 10.1109/TIT.2015.2439953. URL https://ieeexplore.ieee.org/document/7115934.

Ke Li and Andreas Winter. Squashed entanglement, k-extendibility, quantum Markov chains, and recovery maps. *Foundations of Physics*, 48:910–924, February 2018. URL https://doi.org/10.1007/s10701-018-0143-6.

Hou Li-Zhen and Fang Mao-Fa. Entanglement-assisted classical capacity of a generalized amplitude damping channel. *Chinese Physics Letters*, 24:2482, 2007a. URL http://stacks.iop.org/0256-307X/24/i=9/a=006.

Hou Li-Zhen and Fang Mao-Fa. The Holevo capacity of a generalized amplitude-damping channel. *Chinese Physics*, 16:1843, 2007b. URL http://stacks.iop.org/1009-1963/16/i=7/a=006.

Elliot H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11:267–288, December 1973. URL https://doi.org/10.1016/0001-8708(73)90011-X.

Elliott H. Lieb and Mary Beth Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14:1938–1941, 1973a. URL https://doi.org/10.1063/1.1666274.

Elliott H. Lieb and Mary Beth Ruskai. A fundamental property of quantum-mechanical entropy. *Physical Review Letters*, 30:434–436, March 1973b. doi: 10.1103/PhysRevLett.30.434. URL https://link.aps.org/doi/10.1103/PhysRevLett.30.434.

Elliott H. Lieb and Walter E. Thirring. *Inequalities for the Moments of the Eigenvalues of the Schrodinger Hamiltonian and Their Relation to Sobolev Inequalities*, pages 269–304. Princeton University Press, 1976. doi: doi:10.1515/9781400868940-014. URL https://doi.org/10.1515/9781400868940-014.

Göran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40:147–151, June 1975. ISSN 0010-3616. doi: 10.1007/BF01609396. URL http://dx.doi.org/10.1007/BF01609396.

Zi-Wen Liu and Andreas Winter. Resource theories of quantum channels and the universal role of resource erasure, April 2019.

Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55:1613–1622, March 1997. doi: 10.1103/PhysRevA.55.1613. URL https://link.aps.org/doi/10.1103/PhysRevA.55.1613.

Per-Olov Löwdin. On the nonorthogonality problem. volume 5 of *Advances in Quantum Chemistry*, pages 185–199. Academic Press, 1970. doi: https://doi.org/10.1016/S0065-3276(08)60339-1. URL https://www.sciencedirect.com/science/article/pii/S0065327608603391.

Per-Olov Löwdin. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *The Journal of Chemical Physics*, 18:365–375, 1950. doi: 10.1063/1.1747632. URL https://doi.org/10.1063/1.1747632.

Keiji Matsumoto. A new quantum version of f-divergence, 2013.

Keiji Matsumoto. Quantum fidelities, their duals, and convex analysis, August 2014a.

Keiji Matsumoto. On the condition of conversion of classical probability distribution families into quantum families, December 2014b.

Keiji Matsumoto. A new quantum version of f-divergence. In Masanao Ozawa, Jeremy Butterfield, Hans Halvorson, Miklós Rédei, Yuichiro Kitajima, and Francesco Buscemi, editors, *Reality and Measurement in Algebraic Quantum Theory*, volume 261, pages 229–273, Singapore, 2018. Springer Singapore. ISBN 9789811324864 9789811324871. doi: 10.1007/978-981-13-2487-1_10. URL http://link.springer.com/10.1007/978-981-13-2487-1_10. Series Title: Springer Proceedings in Mathematics & Statistics.

William Matthews and Stephanie Wehner. Finite blocklength converse bounds for quantum channels. *IEEE Transactions on Information Theory*, 60:7317–7329, November 2014. URL https://ieeexplore.ieee.org/document/6891222.

Ueli M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39:733–742, May 1993. URL https://ieeexplore.ieee.org/document/256484.

I. Mayer. On Löwdin's method of symmetric orthogonalization. *International Journal of Quantum Chemistry*, 90:63–65, 2002. doi: https://doi.org/10.1002/qua.981. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.981.

Simon Milz and Kavan Modi. Quantum Stochastic Processes and Quantum non-Markovian Phenomena. *PRX Quantum*, 2:030201, July 2021. doi: 10.1103/PRXQuantum.2.030201. URL https://link.aps.org/doi/10.1103/PRXQuantum.2.030201.

Adam Miranowicz and Satoshi Ishizaka. Closed formula for the relative entropy of entanglement. *Physical Review A*, 78:032310, September 2008. doi: 10.1103/PhysRevA.78.032310. URL https://link.aps.org/doi/10.1103/PhysRevA.78.032310.

Gert Molière and Max Delbrück. *Statistische Quantenmechanik und Thermodynamik*. Berlin: Verlag der Akademie der Wissenschaften, 1935.

Ciara Morgan and Andreas Winter. 'Pretty strong' converse for the quantum capacity of degradable channels. *IEEE Transactions on Information Theory*, 60:317–333, January 2014. URL https://ieeexplore.ieee.org/document/6663606.

Milan Mosonyi and Nilanjana Datta. Generalized relative entropies and the capacity of classical-quantum channels. *Journal of Mathematical Physics*, 50:072104, July 2009. doi: 10.1063/1.3167288. URL http://dx.doi.org/10.1063/1.3167288.

Milán Mosonyi and Fumio Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57:2474–2487, April 2011. URL https://ieeexplore.ieee.org/document/5730573.

Milán Mosonyi and Tomohiro Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Communications in Mathematical Physics*, 334:1617–1648, March 2015. URL https://doi.org/10.1007/s00220-014-2248-x.

Milán Mosonyi and Dénes Petz. Structure of sufficient quantum coarse-grainings. *Letters in Mathematical Physics*, 68:19–30, April 2004. ISSN 1573-0530. URL https://doi.org/10.1007/s11005-004-4072-2.

Alexander Müller-Hermes. Transposition in quantum information theory. Master's thesis, Technical University of Munich, September 2012.

Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new generalization and some properties. *Journal of Mathematical Physics*, 54:122203, December 2013. URL https://doi.org/10.1063/1.4838856.

Hiroshi Nagaoka. The converse part of the theorem for quantum Hoeffding bound, November 2006.

Mark Aronovich Naimark. Spectral functions of a symmetric operator. *Izv. Akad. Nauk SSSR Ser. Mat.*, 4:277–318, 1940. URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=im&paperid=3745&option_lang=eng.

Michael A. Nielsen. Continuity bounds for entanglement. *Physical Review A*, 61:064301, April 2000. URL https://link.aps.org/doi/10.1103/PhysRevA.61.064301.

Michael A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303:249 – 252, 2002. ISSN 0375-9601. doi: DOI:10.1016/S0375-9601(02)01272-0. URL https://www.sciencedirect.com/science/article/pii/S0375960102012720.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

Julien Niset, Jaromír Fiurasek, and Nicolas J. Cerf. No-go theorem for Gaussian quantum error correction. *Physical Review Letters*, 102:120501, March 2009. doi: 10.1103/PhysRevLett.102.120501. URL http://link.aps.org/doi/10.1103/PhysRevLett.102.120501.

Michael Nussbaum and Arleta Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37:1040–1057, 2009. doi: 10.1214/08-AOS593. URL https://doi.org/10.1214/08-AOS593.

Tomohiro Ogawa and Hiroshi Nagaoka. *Strong Converse and Stein's Lemma in Quantum Hypothesis Testing*, pages 28–42. World Scientific, 2005. doi: 10.1142/9789812563071_0003. URL https://www.worldscientific.com/doi/abs/10.1142/9789812563071_0003.

Samad Khabbazi Oskouei, Stefano Mancini, and Mark M. Wilde. Union bound for quantum information processing. *Proceedings of the Royal Society A*, 475:20180612, January 2019. doi: 10.1098/rspa.2018.0612. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0612.

Masanao Ozawa. Quantum measuring processes of continuous observables. *Journal of Mathematical Physics*, 25:79–87, 1984. URL https://doi.org/10.1063/1.526000.

Vern Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2003. doi: 10.1017/CBO9780511546631.

Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413–1415, August 1996. doi: 10.1103/PhysRevLett.77.1413. URL http://link.aps.org/doi/10.1103/PhysRevLett.77.1413.

Dénes Petz. Quasi-entropies for States of a von Neumann Algebra. *Publications of the Research Institute for Mathematical Sciences*, 21:787–800, 1985. doi: 10.2977/prims/1195178929. URL https://doi.org/10.2977/prims/1195178929.

Dénes Petz. Quasi-entropies for finite quantum systems. *Reports in Mathematical Physics*, 23:57–65, 1986a. URL https://doi.org/10.1016/0034-4877(86)90067-4.

Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105:123–131, March 1986b. ISSN 1432-0916. URL https://doi.org/10.1007/BF01212345.

Dénes Petz. Sufficiency of channels over von Neumann algebras. *Quarterly Journal of Mathematics*, 39:97–108, 1988. ISSN 1464-3847. URL https://doi.org/10.1093/qmath/39.1.97.

Dénes Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15:79, March 2003. URL https://doi.org/10.1142/S0129055X03001576.

Dénes Petz and Mary Beth Ruskai. Contraction of generalized relative entropy under stochastic mappings on matrices. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1:83–89, January 1998. URL https://doi.org/10.1142/S0219025798000077.

Marco Piani, Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Properties of quantum nonsignaling boxes. *Physical Review A*, 74:012305, July 2006. doi: 10.1103/PhysRevA.74.012305. URL https://link.aps.org/doi/10.1103/PhysRevA.74.012305.

Stefano Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi. Fundamental limits of repeaterless quantum communications. *Nature Communications*, 8:15043, 2017. URL https://doi.org/10.1038/ncomms15043.

Martin B. Plenio. Logarithmic negativity: A full entanglement monotone that is not convex. *Physical Review Letters*, 95:090503, August 2005. doi: 10.1103/PhysRevLett.95.090503. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.090503.

Martin B. Plenio and Shashank Virmani. An introduction to entanglement measures. *Quantum Information & Computation*, 7:1–51, January 2007. ISSN 1533-7146. URL https://doi.org/10.26421/QIC7.1-2-1.

Martin B. Plenio, Shashank Virmani, and P. Papadopoulos. Operator monotones, the reduction criterion and the relative entropy. *Journal of Physics A: Mathematical and General*, 33:L193, June 2000. URL http://stacks.iop.org/0305-4470/33/i=22/a=101.

Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation*, pages 1327–1333, September 2010. URL https://ieeexplore.ieee.org/abstract/document/5707067.

Haoyu Qi, Kunal Sharma, and Mark M. Wilde. Entanglement-assisted private communication over quantum broadcast channels. *Journal of Physics A: Mathematical and Theoretical*, 51:374001, August 2018a. doi: 10.1088/1751-8121/aad5f3. URL https://doi.org/10.1088/1751-8121/aad5f3.

Haoyu Qi, Qing-Le Wang, and Mark M. Wilde. Applications of position-based coding to classical communication over quantum channels. *Journal of Physics A*, 51:444002, November 2018b. URL https://doi.org/10.1088/1751-8121/aae290.

Lu-Feng Qiao, Alexander Streltsov, Jun Gao, Swapan Rana, Ruo-Jing Ren, Zhi-Qiang Jiao, Cheng-Qiu Hu, Xiao-Yun Xu, Ci-Yu Wang, Hao Tang, Ai-Lin Yang, Zhi-Hao Ma, Maciej Lewenstein, and Xian-Min Jin. Entanglement activation from quantum coherence and superposition. *Physical Review A*, 98:052351, November 2018. doi: 10.1103/PhysRevA.98.052351. URL https://link.aps.org/doi/10.1103/PhysRevA.98.052351.

Jaikumar Radhakrishnan, Pranab Sen, and Naqueeb Ahmad Warsi. One-shot private classical capacity of quantum wiretap channel: Based on one-shot quantum covering lemma, March 2017.

Eric M. Rains. Entanglement purification via separable superoperators, 1998.

Eric M. Rains. Bound on distillable entanglement. *Physical Review A*, 60:179–184, July 1999a. doi: 10.1103/PhysRevA.60.179. URL http://link.aps.org/doi/10.1103/PhysRevA.60.179.

Eric M. Rains. Rigorous treatment of distillable entanglement. *Physical Review A*, 60:173–178, July 1999b. doi: 10.1103/PhysRevA.60.173. URL https://link.aps.org/doi/10.1103/PhysRevA.60.173.

Eric M. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47:2921–2933, November 2001. URL https://ieeexplore.ieee.org/document/959270.

Alexey E. Rastegin. Relative error of state-dependent cloning. *Physical Review A*, 66:042304, October 2002. doi: 10.1103/PhysRevA.66.042304. URL http://link.aps.org/doi/10.1103/PhysRevA.66.042304.

Alexey E. Rastegin. A lower bound on the relative error of mixed-state cloning and related operations. *Journal of Optics B: Quantum and Semiclassical Optics*, 5:S647, December 2003. URL http://stacks.iop.org/1464-4266/5/i=6/a=017.

Alexey E. Rastegin. Sine distance for quantum states, February 2006.

Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics*, volume I: Functional Analysis. Academic Press, 1981. ISBN 9780080570488.

Joseph M. Renes and Renato Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57:7377–7385, November 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2162226. URL https://ieeexplore.ieee.org/document/5967913.

Joseph M. Renes and Renato Renner. One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys. *IEEE Transactions on Information Theory*, 58:1985–1991, March 2012. doi: 10.1109/TIT.2011.2177589. URL https://ieeexplore.ieee.org/document/6157080.

Renato Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zürich, December 2005.

Luca Rigovacca, Go Kato, Stefan Baeuml, Myungshik S. Kim, William J. Munro, and Koji Azuma. Versatile relative entropy bounds for quantum networks. *New Journal of Physics*, 20:013033, January 2018. URL https://doi.org/10.1088/1367-2630/aa9fcf.

Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970. ISBN 9780691015866.

B. Rosgen and J. Watrous. On the hardness of distinguishing mixed-state quantum computations. In *20th Annual IEEE Conference on Computational Complexity (CCC'05)*, pages 344–354, 2005. doi: 10.1109/CCC.2005.21. URL https://ieeexplore.ieee.org/document/1443098.

Sheldon Ross. *Introduction to Probability Models*. Academic Press, 12 edition, 2019. ISBN 978-0-12-814346-9.

Aidan Roy and A. J. Scott. Unitary designs and codes. *Designs, Codes and Cryptography*, 53: 13–31, October 2009. doi: 10.1007/s10623-009-9290-2. URL https://doi.org/10.1007/s10623-009-9290-2.

Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1976. ISBN 9780070856134.

Mary Beth Ruskai. Inequalities for quantum entropy: A review with conditions for equality. *Journal of Mathematical Physics*, 43:4358–4375, September 2002. doi: 10.1063/1.1497701. URL https://doi.org/10.1063/1.1497701.

Massimiliano F. Sacchi. Optimal discrimination of quantum operations. *Physical Review A*, 71: 062340, June 2005. doi: 10.1103/PhysRevA.71.062340. URL https://link.aps.org/doi/10.1103/PhysRevA.71.062340.

J. J. Sakurai. *Modern Quantum Mechanics*. Addison–Wesley Publishing Company, Inc., revised edition, 1994.

Benjamin Schumacher. Sending entanglement through noisy quantum channels. *Physical Review A*, 54:2614–2628, October 1996. doi: 10.1103/PhysRevA.54.2614. URL https://link.aps.org/doi/10.1103/PhysRevA.54.2614.

Benjamin Schumacher and Michael A. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54:2629–2635, October 1996. doi: 10.1103/PhysRevA.54.2629. URL https://link.aps.org/doi/10.1103/PhysRevA.54.2629.

Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, July 1997. URL https://link.aps.org/doi/10.1103/PhysRevA.56.131.

Benjamin Schumacher and Michael D. Westmoreland. Approximate quantum error correction. *Quantum Information Processing*, 1:5–12, April 2002. ISSN 1573-1332. doi: 10.1023/A:1019653202562. URL https://doi.org/10.1023/A:1019653202562.

Pranab Sen. Achieving the Han–Kobayashi inner bound for the quantum interference channel by sequential decoding. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 736–740, September 2012. doi: 10.1109/ISIT.2012.6284656. URL https://ieeexplore.ieee.org/document/6284656.

Claude Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2:S8–S19, September 1956. URL https://ieeexplore.ieee.org/document/1056798.

Claude E. Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28:656–715, October 1949. doi: 10.1002/j.1538-7305.1949.tb00928.x. URL https://ieeexplore.ieee.org/document/6769090.

Naresh Sharma. Equality conditions for the quantum f-relative entropy and generalized data processing inequalities. *Quantum Information Processing*, 11:137–160, 2012. ISSN 2157-8095. URL https://doi.org/10.1007/s11128-011-0238-x.

Naresh Sharma and Naqueeb Ahmad Warsi. Fundamental bound on the reliability of quantum information transmission. *Physical Review Letters*, 110:080501, February 2013. doi: 10.1103/PhysRevLett.110.080501. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.080501.

Yaoyun Shi and Xiaodi Wu. Epsilon-net method for optimizations over separable states. In Artur Czumaj, Kurt Mehlhorn, Andrew Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming*, pages 798–809, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31594-7. URL https://doi.org/10.1007/978-3-642-31594-7_67.

Maksim E. Shirokov. Tight uniform continuity bounds for the quantum conditional mutual information, for the Holevo quantity, and for capacities of quantum channels. *Journal of Mathematical Physics*, 58:102202, October 2017. doi: 10.1063/1.4987135. URL https://doi.org/10.1063/1.4987135.

Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52:R2493–R2496, October 1995. doi: 10.1103/PhysRevA.52.R2493. URL https://link.aps.org/doi/10.1103/PhysRevA.52.R2493.

Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43:4334–4340, 2002a. doi: 10.1063/1.1498000. URL https://doi.org/10.1063/1.1498000.

Peter W. Shor. The quantum channel capacity and coherent information. In *Lecture Notes, MSRI Workshop on Quantum Computation*, 2002b.

Peter W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246:453–472, 2004. URL https://doi.org/10.1007/s00220-004-1071-1.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8:171–176, March 1958. URL https://msp.org/pjm/1958/8-1/p14.xhtml.

Graeme Smith. Private classical capacity with a symmetric side channel and its application to quantum cryptography. *Physical Review A*, 78:022306, August 2008. doi: 10.1103/PhysRevA.78.022306. URL https://link.aps.org/doi/10.1103/PhysRevA.78.022306.

Graeme Smith and John A. Smolin. Degenerate quantum codes for Pauli channels. *Physical Review Letters*, 98:030501, January 2007. doi: 10.1103/PhysRevLett.98.030501. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.030501.

Graeme Smith and John A. Smolin. Extensive nonadditivity of privacy. *Physical Review Letters*, 103:120503, September 2009. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.120503.

Graeme Smith and Jon Yard. Quantum communication with zero-capacity channels. *Science*, 321:1812–1815, September 2008. URL https://science.sciencemag.org/content/321/5897/1812.

Graeme Smith, Joseph M. Renes, and John A. Smolin. Structured codes improve the Bennett-Brassard-84 quantum key rate. *Physical Review Letters*, 100:170502, April 2008. doi: 10.1103/PhysRevLett.100.170502. URL https://link.aps.org/doi/10.1103/PhysRevLett.100.170502.

Graeme Smith, John A. Smolin, and Jon Yard. Quantum communication with Gaussian channels of zero quantum capacity. *Nature Photonics*, 5:624–627, August 2011. URL https://doi.org/10.1038/nphoton.2011.203.

R. R. Smith. Completely Bounded Maps between C*-Algebras. *Journal of the London Mathematical Society*, s2-27:157–166, 02 1983. ISSN 0024-6107. doi: 10.1112/jlms/s2-27.1.157. URL https://doi.org/10.1112/jlms/s2-27.1.157.

Akihito Soeda, Peter S. Turner, and Mio Murao. Entanglement cost of implementing controlled-unitary operations. *Physical Review Letters*, 107:180501, October 2011. doi: 10.1103/physrevlett.107.180501. URL https://link.aps.org/doi/10.1103/PhysRevLett.107.180501.

Benjamin Steinberg. *Representation Theory of Finite Groups: An Introductory Approach*. Springer New York, 2011. ISBN 9781461407768.

W. Forrest Stinespring. Positive Functions on C*-Algebras. *Proceedings of the American Mathematical Society*, 6:211–216, 1955. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/2032342.

Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press and SIAM, fifth edition, May 2016.

Ruslan L. Stratonovich. Information capacity of a quantum communications channel. i. *Soviet Radiophysics*, 8:82–91, January 1965. ISSN 1573-9120. doi: 10.1007/BF01038470. URL https://doi.org/10.1007/BF01038470.

David Sutter, Volkher B. Scholz, Andreas Winter, and Renato Renner. Approximate degradable quantum channels. *IEEE Transactions on Information Theory*, 63:7832–7844, December 2017. URL https://ieeexplore.ieee.org/document/8046086.

Masahiro Takeoka, Masashi Ban, and Masahide Sasaki. Quantum channel of continuous variable teleportation and nonclassicality of quantum states. *Journal of Optics B: Quantum and Semiclassical Optics*, 4:114, April 2002. URL http://stacks.iop.org/1464-4266/4/i=2/a=306.

Masahiro Takeoka, Saikat Guha, and Mark M. Wilde. The squashed entanglement of a quantum channel. *IEEE Transactions on Information Theory*, 60:4987–4998, August 2014. ISSN 0018-9448. URL https://ieeexplore.ieee.org/document/6832533.

Masahiro Takeoka, Kaushik P. Seshadreesan, and Mark M. Wilde. Unconstrained distillation capacities of a pure-loss bosonic broadcast channel. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2484–2488, July 2016. doi: 10.1109/ISIT.2016.7541746. URL https://ieeexplore.ieee.org/document/7541746.

Masahiro Takeoka, Kaushik P. Seshadreesan, and Mark M. Wilde. Unconstrained capacities of quantum key distribution and entanglement distillation for pure-loss bosonic broadcast channels. *Physical Review Letters*, 119:150501, October 2017. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.150501.

Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*. Springer, 2015.

Marco Tomamichel and Masahito Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59:7693–7710, November 2013. ISSN 0018-9448. doi: 10.1109/TIT.2013.2276628. URL https://ieeexplore.ieee.org/document/6574274.

Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55:5840–5847, December 2009. URL https://ieeexplore.ieee.org/document/5319753.

Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality Between Smooth Min- and Max-Entropies. *IEEE Transactions on Information Theory*, 56:4674–4681, September 2010. URL https://ieeexplore.ieee.org/document/5550419.

Marco Tomamichel, Mario Berta, and Joseph M. Renes. Quantum coding with finite resources. *Nature Communications*, 7:11419, May 2016. URL https://doi.org/10.1038/ncomms11419.

Marco Tomamichel, Mark M. Wilde, and Andreas Winter. Strong converse rates for quantum communication. *IEEE Transactions on Information Theory*, 63:715–727, January 2017. doi: 10.1109/tit.2016.2615847. URL https://ieeexplore.ieee.org/document/7586115.

Robert R. Tucci. Quantum entanglement and conditional information transmission, September 1999.

Robert R. Tucci. Entanglement of distillation and conditional mutual information, Februrary 2002.

Armin Uhlmann. The 'Transition Probability' in the State Space of a *-Algebra. *Reports on Mathematical Physics*, 9:273–279, April 1976. URL https://www.sciencedirect.com/science/article/pii/0034487776900604.

Michael L. Ulrey. Sequential coding for channels with feedback. *Information and Control*, 32: 93–100, October 1976. URL https://doi.org/10.1016/S0019-9958(76)90129-7.

Hisaharu Umegaki. Conditional expectations in an operator algebra IV (entropy and information). *Kodai Mathematical Seminar Reports*, 14:59–85, 1962. URL https://doi.org/10.2996/kmj/1138844604.

Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996. doi: 10.1137/1038003. URL https://doi.org/10.1137/1038003.

Gonzalo Vazquez-Vilar. Multiple quantum hypothesis testing expressions and classical-quantum channel converse bounds. In *2016 IEEE International Symposium on Information Theory*, pages 2854–2857, Barcelona, Spain, 2016. URL https://ieeexplore.ieee.org/document/7541820.

Vlatko Vedral and Martin B. Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57:1619–1633, March 1998. doi: 10.1103/PhysRevA.57.1619. URL http://link.aps.org/doi/10.1103/PhysRevA.57.1619.

Vlatko Vedral, Martin B. Plenio, M. A. Rippin, and Peter L. Knight. Quantifying entanglement. *Physical Review Letters*, 78:2275–2279, March 1997. doi: 10.1103/PhysRevLett.78.2275. URL https://link.aps.org/doi/10.1103/PhysRevLett.78.2275.

Sergio Verdu. On channel capacity per unit cost. *IEEE Transactions on Information Theory*, 36:1019–1030, 1990. doi: 10.1109/18.57201. URL https://ieeexplore.ieee.org/document/57201.

Gilbert S. Vernam. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *Transactions of the American Institute of Electrical Engineers*, 45:295–301, 1926. URL https://ieeexplore.ieee.org/document/5061224.

Guifré Vidal. Entanglement monotones. *Journal of Modern Optics*, 47:355–376, 2000. doi: 10.1080/09500340008244048. URL https://www.tandfonline.com/doi/abs/10.1080/09500340008244048.

Guifré Vidal and Reinhard F. Werner. Computable measure of entanglement. *Physical Review A*, 65:032314, February 2002. doi: 10.1103/PhysRevA.65.032314. URL https://link.aps.org/doi/10.1103/PhysRevA.65.032314.

Johann von Neumann. Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1: 245–272, 1927a. URL http://eudml.org/doc/59230.

Johann von Neumann. Thermodynamik quantenmechanischer gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 102: 273–291, 1927b.

Johann von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, December 1928. ISSN 1432-1807. URL https://doi.org/10.1007/BF01448847.

Johann von Neumann. *Mathematische grundlagen der quantenmechanik*. Verlag von Julius Springer Berlin, 1932.

Michael Walter, David Gross, and Jens Eisert. Multi-partite entanglement, 2016.

Kun Wang, Xin Wang, and Mark M. Wilde. Quantifying the unextendibility of entanglement, November 2019a.

Ligong Wang and Renato Renner. One-shot classical-quantum capacity and hypothesis testing. *Physical Review Letters*, 108:200501, 2012. doi: 10.1103/PhysRevLett.108.200501. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.200501.

Xin Wang and Runyao Duan. Improved semidefinite programming upper bound on distillable entanglement. *Physical Review A*, 94:050301, November 2016a. doi: 10.1103/physreva.94.050301. URL https://link.aps.org/doi/10.1103/PhysRevA.94.050301.

Xin Wang and Runyao Duan. A semidefinite programming upper bound of quantum capacity. In *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, July 2016b. doi: 10.1109/isit.2016.7541587. URL https://ieeexplore.ieee.org/document/7541587.

Xin Wang and Mark M. Wilde. Resource theory of asymmetric distinguishability. *Physical Review Research*, 1:033170, December 2019. doi: 10.1103/PhysRevResearch.1.033170. URL http://arxiv.org/abs/1905.11629.

Xin Wang and Mark M. Wilde. α-logarithmic negativity. *Physical Review A*, 102:032416, September 2020. doi: 10.1103/PhysRevA.102.032416. URL https://link.aps.org/doi/10.1103/PhysRevA.102.032416.

Xin Wang, Wei Xie, and Runyao Duan. Semidefinite programming strong converse bounds for classical capacity. *IEEE Transactions on Information Theory*, 64:640–653, January 2018. ISSN 0018-9448. doi: 10.1109/TIT.2017.2741101. URL https://ieeexplore.ieee.org/document/8012535.

Xin Wang, Kun Fang, and Runyao Duan. Semidefinite programming converse bounds for quantum communication. *IEEE Transactions on Information Theory*, 65:2583–2592, April 2019b. URL https://ieeexplore.ieee.org/document/8482492.

Xin Wang, Kun Fang, and Marco Tomamichel. On converse bounds for classical communication over quantum channels. *IEEE Transactions on Information Theory*, 65:4609–4619, July 2019c. doi: 10.1109/TIT.2019.2898656. URL https://ieeexplore.ieee.org/document/8638816.

John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5:217–238, November 2009. doi: 10.4086/toc.2009.v005a011. URL http://www.theoryofcomputing.org/articles/v005a011.

John Watrous. Simpler semidefinite programs for completely bounded norms. *Chicago Journal of Theoretical Computer Science*, July 2013. URL http://cjtcs.cs.uchicago.edu/articles/2013/8/contents.html.

John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142.

R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43:4353–4357, 2002. doi: 10.1063/1.1498491. URL https://doi.org/10.1063/1.1498491.

Reinhard F. Werner. An application of Bell's inequalities to a quantum state extension problem. *Letters in Mathematical Physics*, 17:359–363, May 1989a. doi: 10.1007/BF00399761. URL https://doi.org/10.1007/BF00399761.

Reinhard F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40:4277–4281, October 1989b. doi: 10.1103/PhysRevA.40.4277. URL http://link.aps.org/doi/10.1103/PhysRevA.40.4277.

Reinhard F. Werner. All teleportation and dense coding schemes. *Journal of Physics A: Mathematical and General*, 34:7081, September 2001. URL http://stacks.iop.org/0305-4470/34/i=35/a=332.

Mark M. Wilde. Sequential decoding of a general classical-quantum channel. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 469, September 2013. ISSN 1364-5021. doi: 10.1098/rspa.2013.0259. URL https://doi.org/10.1098/rspa.2013.0259.

Mark M. Wilde. Squashed entanglement and approximate private states. *Quantum Information Processing*, 15:4563–4580, November 2016. ISSN 1573-1332. doi: 10.1007/s11128-016-1432-7. URL http://dx.doi.org/10.1007/s11128-016-1432-7.

Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, second edition, 2017a. URL https://doi.org/10.1017/CBO9781139525343.

Mark M. Wilde. Position-based coding and convex splitting for private communication over quantum channels. *Quantum Information Processing*, 16:264, October 2017b. URL https://doi.org/10.1007/s11128-017-1718-4.

Mark M. Wilde. Strong and uniform convergence in the teleportation simulation of bosonic Gaussian channels. *Physical Review A*, 97:062305, June 2018a. doi: 10.1103/PhysRevA.97.062305. URL https://link.aps.org/doi/10.1103/PhysRevA.97.062305.

Mark M. Wilde. Optimized quantum f-divergences and data processing. *Journal of Physics A*, 51:374002, September 2018b. URL https://doi.org/10.1088/1751-8121/aad5a1.

Mark M. Wilde and Haoyu Qi. Energy-constrained private and quantum capacities of quantum channels. *IEEE Transactions on Information Theory*, 64:7802–7827, December 2018. URL https://ieeexplore.ieee.org/document/8541091.

Mark M. Wilde and Andreas Winter. Strong converse for the quantum capacity of the erasure channel for almost all codes. In Steven T. Flammia and Aram W. Harrow, editors, *9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2014)*, volume 27 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 52–66, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-73-6. doi: 10.4230/LIPIcs.TQC.2014.52. URL http://drops.dagstuhl.de/opus/volltexte/2014/4806.

Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics*, 331:593–622, October 2014. URL https://doi.org/10.1007/s00220-014-2122-x.

Mark M. Wilde, Marco Tomamichel, and Mario Berta. Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory*, 63:1792–1817, March 2017. URL https://ieeexplore.ieee.org/document/7807212.

Andreas Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347: 291–313, October 2016. URL https://doi.org/10.1007/s00220-016-2609-8.

Michael M. Wolf, David Pérez-García, and Geza Giedke. Quantum capacities of bosonic channels. *Physical Review Letters*, 98:130501, March 2007. doi: 10.1103/PhysRevLett.98.130501. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.130501.

Jacob Wolfowitz. *Coding Theorems of Information Theory*, volume 31 of *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. Springer, 1964.

William K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Physical Review Letters*, 80:2245–2248, March 1998. doi: 10.1103/PhysRevLett.80.2245. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.2245.

Aaron D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54:1355–1387, October 1975. URL https://ieeexplore.ieee.org/document/6772207.

Dong Yang. A simple proof of monogamy of entanglement. *Physics Letters A*, 360:249–250, 2006. ISSN 0375-9601. doi: https://doi.org/10.1016/j.physleta.2006.08.027. URL http://www.sciencedirect.com/science/article/pii/S0375960106012801.

Jon Yard, Patrick Hayden, and Igor Devetak. Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions. *IEEE Transactions on Information Theory*, 54:3091–3113, July 2008. URL https://ieeexplore.ieee.org/document/4545000.

Haidong Yuan and Chi-Hang Fred Fung. Fidelity and Fisher Information on Quantum Channels. *New Journal of Physics*, 19:113039, November 2017. doi: 10.1088/1367-2630/aa874c. URL http://stacks.iop.org/1367-2630/19/i=11/a=113039?key=crossref.c8abb94f653e6d572133885d9e0b86b0.

Horace Yuen, Robert Kennedy, and Melvin Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21:125–134, March 1975. URL https://ieeexplore.ieee.org/document/1055351.

Sisi Zhou and Liang Jiang. An exact correspondence between the quantum Fisher information and the Bures metric, October 2019. URL http://arxiv.org/abs/1910.08473.

Xinlan Zhou, Debbie W. Leung, and Isaac L. Chuang. Methodology for quantum logic gate construction. *Physical Review A*, 62:052316, oct 2000. doi: 10.1103/PhysRevA.62.052316. URL https://link.aps.org/doi/10.1103/PhysRevA.62.052316.

Karol Zyczkowski, Paweł Horodecki, Anna Sanpera, and Maciej Lewenstein. Volume of the set of separable states. *Physical Review A*, 58:883–892, August 1998. doi: 10.1103/PhysRevA.58.883. URL https://link.aps.org/doi/10.1103/PhysRevA.58.883.

Index