PHYS 7895: Quantum Information Theory	Fall 2015

Lecture 18 — October 26, 2015

Prof. Mark M. Wilde *Scribe: Mark M. Wilde*

1 Overview

In the previous lecture, we discussed classical entropy and entropy inequalities.

In this lecture, we discuss several information measures that are important for quantifying the amount of information and correlations that are present in quantum systems. The first fundamental measure that we introduce is the von Neumann entropy. It is the quantum generalization of the Shannon entropy, but it captures both classical and quantum uncertainty in a quantum state. The von Neumann entropy gives meaning to a notion of the *information qubit*. This notion is different from that of the physical qubit, which is the description of a quantum state of an electron or a photon. The information qubit is the fundamental quantum informational unit of measure, determining how much quantum information is present in a quantum system.

The initial definitions here are analogous to the classical definitions of entropy, but we soon discover a radical departure from the intuitive classical notions from the previous chapter: the conditional quantum entropy can be negative for certain quantum states. In the classical world, this negativity simply does not occur, but it takes a special meaning in quantum information theory. Pure quantum states that are entangled have stronger-than-classical correlations and are examples of states that have negative conditional entropy. The negative of the conditional quantum entropy is so important in quantum information theory that we even have a special name for it: the coherent information. We discover that the coherent information obeys a quantum data-processing inequality, placing it on a firm footing as a particular informational measure of quantum correlations.

We then define several other quantum information measures, such as quantum mutual information, that bear similar definitions as in the classical world, but with Shannon entropies replaced with von Neumann entropies. This replacement may seem to make quantum entropy somewhat trivial on the surface, but a simple calculation reveals that a maximally entangled state on two qubits registers *two bits* of quantum mutual information (recall that the largest the mutual information can be in the classical world is *one bit* for the case of two maximally correlated bits).

2 Quantum Entropy

We might expect a measure of the entropy of a quantum system to be vastly different from the classical measure of entropy from the previous chapter because a quantum system possesses not only classical uncertainty but also quantum uncertainty that arises from the uncertainty principle. But recall that the density operator captures both types of uncertainty and allows us to determine probabilities for the outcomes of any measurement on system A. Thus, a quantum measure of uncertainty should be a direct function of the density operator, just as the classical measure of uncertainty is a direct function of a probability density function. It turns out that this function has a strikingly similar form to the classical entropy, as we see below.

Definition 1 (Quantum Entropy)**.** *Suppose that Alice prepares some quantum system A in a state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$. Then the entropy $H(A)_\rho$ of the state is as follows:*

$$H(A)_\rho \equiv -\operatorname{Tr}\{\rho_A \log \rho_A\}. \tag{1}$$

The entropy of a quantum system is also known as the *von Neumann entropy* or the *quantum entropy* but we often simply refer to it as the *entropy.* We can denote it by $H(A)_\rho$ or $H(\rho_A)$ to show the explicit dependence on the density operator ρ_A. The von Neumann entropy has a special relation to the eigenvalues of the density operator, as the following exercise asks you to verify.

Exercise 2. *Consider a density operator ρ_A with the following spectral decomposition:*

$$\rho_A = \sum_x p_X(x)|x\rangle\langle x|_A. \tag{2}$$

Show that the entropy $H(A)_\rho$ is the same as the Shannon entropy $H(X)$ of a random variable X with probability distribution $p_X(x)$.

In our definition of quantum entropy, we use the same notation H as in the classical case to denote the entropy of a quantum system. It should be clear from the context whether we are referring to the entropy of a quantum or classical system.

The quantum entropy admits an intuitive interpretation. Suppose that Alice generates a quantum state $|\psi_y\rangle$ in her lab according to some probability density $p_Y(y)$, corresponding to a random variable Y. Suppose further that Bob has not yet received the state from Alice and does not know which one she sent. The expected density operator from Bob's point of view is then

$$\sigma = \mathbb{E}_Y\{|\psi_Y\rangle\langle\psi_Y|\} = \sum_y p_Y(y)|\psi_y\rangle\langle\psi_y|. \tag{3}$$

The interpretation of the entropy $H(\sigma)$ is that it quantifies Bob's uncertainty about the state Alice sent—his expected information gain is $H(\sigma)$ qubits upon receiving and measuring the state that Alice sends.

2.1 Mathematical Properties of Quantum Entropy

We now discuss several mathematical properties of the quantum entropy: non-negativity, its minimum value, its maximum value, its invariance with respect to isometries, and concavity. The first three of these properties follow from the analogous properties in the classical world because the von Neumann entropy of a density operator is the Shannon entropy of its eigenvalues (see Exercise 2). We state them formally below:

Property 3 (Non-Negativity)**.** *The von Neumann entropy $H(\rho)$ is non-negative for any density operator ρ:*

$$H(\rho) \geq 0. \tag{4}$$

Proof. This follows from non-negativity of Shannon entropy. □

Property 4 (Minimum Value)**.** *The minimum value of the von Neumann entropy is zero, and it occurs when the density operator is a pure state.*

Proof. The minimum value equivalently occurs when the eigenvalues of a density operator are distributed with all the probability mass on one eigenvector and zero on the others, so that the density operator is rank one and corresponds to a pure state. □

Why should the entropy of a pure quantum state vanish? It seems that there is quantum uncertainty inherent in the state itself and that a measure of quantum uncertainty should capture this fact. This last observation only makes sense if we do not know anything about the state that is prepared. But if we know exactly how it was prepared, we can perform a special quantum measurement to verify that the quantum state was prepared, and we do not learn anything from this measurement because the outcome of it is always certain. For example, suppose that Alice prepares the state $|\phi\rangle$ and Bob knows that she does so. He can then perform the following measurement $\{|\phi\rangle\langle\phi|, I - |\phi\rangle\langle\phi|\}$ to verify that she prepared this state. He always receives the first outcome from the measurement and thus never gains any information from it. Thus, in this sense it is reasonable that the entropy of a pure state vanishes.

Property 5 (Maximum Value)**.** *The maximum value of the von Neumann entropy is* $\log d$ *where* d *is the dimension of the system, and it occurs for the maximally mixed state.*

Proof. A proof of the above property is the same as that for the classical case. □

Property 6 (Concavity)**.** *Let* $\rho_x \in \mathcal{D}(\mathcal{H})$ *and let* $p_X(x)$ *be a probability distribution. The entropy is concave in the density operator:*

$$H(\rho) \geq \sum_x p_X(x) H(\rho_x), \tag{5}$$

where $\rho \equiv \sum_x p_X(x)\rho_x$.

The physical interpretation of concavity is as before for classical entropy: entropy can never decrease under a mixing operation. This inequality is a fundamental property of the entropy, and we prove it after developing some important entropic tools.

Property 7 (Isometric Invariance)**.** *Let* $\rho \in \mathcal{D}(\mathcal{H})$ *and* $U : \mathcal{H} \rightarrow \mathcal{H}'$ *be an isometry. The entropy of a density operator is invariant with respect to isometries, in the following sense:*

$$H(\rho) = H(U\rho U^\dagger). \tag{6}$$

Proof. Isometric invariance of entropy follows by observing that the eigenvalues of a density operator are invariant with respect to an isometry. □

3 Joint Quantum Entropy

The joint quantum entropy $H(AB)_\rho$ of the density operator $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ for a bipartite system AB follows naturally from the definition of quantum entropy:

$$H(AB)_\rho \equiv -\operatorname{Tr}\{\rho_{AB} \log \rho_{AB}\}. \tag{7}$$

Now suppose that ρ_{ABC} is a tripartite state, i.e., in $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Then the entropy $H(AB)_\rho$ in this case is defined as above, where $\rho_{AB} = \operatorname{Tr}_C\{\rho_{ABC}\}$. This is a convention that we take throughout. We introduce a few of the properties of joint quantum entropy in the subsections below.

3.1 Marginal Entropies of a Pure Bipartite State

The five properties of quantum entropy in the previous section may give you the impression that the nature of quantum information is not too different from that of classical information. We proved all these properties for the classical case, and their proofs for the quantum case seem similar. The first three even resort to the proofs in the classical case!

Theorem 8 below is where we observe our first radical departure from the classical world. It states that the marginal entropies of a pure bipartite state are equal, while the entropy of the overall state is equal to zero. Recall that the joint entropy $H(X,Y)$ of two random variables X and Y is never less than one of the marginal entropies $H(X)$ or $H(Y)$:

$$H(X,Y) \geq H(X), \qquad H(X,Y) \geq H(Y). \tag{8}$$

The above inequalities follow from the non-negativity of classical conditional entropy. But in the quantum world, these inequalities do not always have to hold, and the following theorem demonstrates that they do not hold for an arbitrary pure bipartite quantum state with Schmidt rank greater than one. The fact that the joint quantum entropy can be less than the marginal quantum entropy is one of the most fundamental differences between classical and quantum information.

Theorem 8. *The marginal entropies $H(A)_\phi$ and $H(B)_\phi$ of a pure bipartite state $|\phi\rangle_{AB}$ are equal:*

$$H(A)_\phi = H(B)_\phi, \tag{9}$$

while the joint entropy $H(AB)_\phi$ vanishes:

$$H(AB)_\phi = 0. \tag{10}$$

Proof. The crucial ingredient for a proof of this theorem is the Schmidt decomposition. Recall that any bipartite state $|\phi\rangle_{AB}$ admits a Schmidt decomposition of the following form:

$$|\phi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\, |i\rangle_A \, |i\rangle_B, \tag{11}$$

where $\{|i\rangle_A\}$ is some orthonormal set of vectors on system A and $\{|i\rangle_B\}$ is some orthonormal set on system B. Recall that the Schmidt rank is equal to the number of non-zero coefficients λ_i. Then the respective marginal states ρ_A and ρ_B on systems A and B are as follows:

$$\rho_A = \sum_i \lambda_i |i\rangle\langle i|_A, \qquad \rho_B = \sum_i \lambda_i |i\rangle\langle i|_B. \tag{12}$$

Thus, the marginal states admit a spectral decomposition with the same eigenvalues. The theorem follows because the von Neumann entropy depends only on the eigenvalues of a given spectral decomposition. □

The theorem applies not only to two systems A and B, but it also applies to any number of systems if we make a bipartite cut of the systems. For example, if the state is $|\phi\rangle_{ABCDE}$, then the following equalities (and others from different combinations) hold by applying Theorem 8:

$$H(A)_\phi = H(BCDE)_\phi, \tag{13}$$
$$H(AB)_\phi = H(CDE)_\phi, \tag{14}$$
$$H(ABC)_\phi = H(DE)_\phi, \tag{15}$$
$$H(ABCD)_\phi = H(E)_\phi. \tag{16}$$

The closest analogy in the classical world to the above property is when we copy a random variable X. That is, suppose that X has a distribution $p_X(x)$ and $\hat{X}$ is some copy of it so that the distribution of the joint random variable $X\hat{X}$ is $p_X(x)\delta_{x,\hat{x}}$. Then the marginal entropies $H(X)$ and $H(\hat{X})$ are both equal. But observe that the joint entropy $H(X\hat{X})$ is also equal to $H(X)$ and this is where the analogy breaks down. That is, there is not a good classical analogy of the notion of purification.

3.2 Additivity

Property 9 (Additivity)**.** *Let $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ and $\sigma_B \in \mathcal{D}(\mathcal{H}_B)$. The quantum entropy is additive for tensor-product states:*

$$H(\rho_A \otimes \sigma_B) = H(\rho_A) + H(\sigma_B). \tag{17}$$

One can verify this property simply by diagonalizing both density operators and resorting to the additivity of the joint Shannon entropies of the eigenvalues.

Additivity is an intuitive property that we would like to hold for any measure of information. For example, suppose that Alice generates a large sequence $|\psi_{x_1}\rangle |\psi_{x_2}\rangle \cdots |\psi_{x_n}\rangle$ of quantum states according to the ensemble $\{p_X(x), |\psi_x\rangle\}$. She may be aware of the classical indices $x_1 x_2 \cdots x_n$, but a third party to whom she sends the quantum sequence may not be aware of these values. The description of the state to this third party is then $\rho \otimes \cdots \otimes \rho$, where $\rho \equiv \mathbb{E}_X \{|\psi_X\rangle\langle\psi_X|\}$, and the quantum entropy of this n-fold tensor product state is $H(\rho \otimes \cdots \otimes \rho) = nH(\rho)$, by applying (17) inductively.

3.3 Joint Quantum Entropy of a Classical–Quantum State

Recall that a classical–quantum state is a bipartite state in which a classical system and a quantum system are classically correlated. An example of such a state is as follows:

$$\rho_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x. \tag{18}$$

The joint quantum entropy of this state takes on a special form that appears similar to entropies in the classical world.

Theorem 10. *The joint entropy $H(XB)_\rho$ of a classical–quantum state, as given in* (18)*, is as follows:*

$$H(XB)_\rho = H(X) + \sum_x p_X(x) H(\rho_B^x), \tag{19}$$

where $H(X)$ is the entropy of a random variable X with distribution $p_X(x)$.

Proof. Consider that

$$H(XB)_\rho = -\operatorname{Tr}\{\rho_{XB} \log \rho_{XB}\}. \tag{20}$$

So we need to evaluate the operator $\log \rho_{XB}$, and we can find a simplified form for it because ρ_{XB} is a classical-quantum state:

$$\begin{aligned}
\log \rho_{XB} &= \log \left[\sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x\right] && (21)\\
&= \log \left[\sum_x |x\rangle\langle x|_X \otimes p_X(x)\rho_B^x\right] && (22)\\
&= \sum_x |x\rangle\langle x|_X \otimes \log\left[p_X(x)\rho_B^x\right]. && (23)
\end{aligned}$$

Then

$$\begin{aligned}
&-\operatorname{Tr}\{\rho_{XB} \log \rho_{XB}\} \\
&= -\operatorname{Tr}\left\{\left[\sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x\right]\left[\sum_{x'} |x'\rangle\langle x'|_X \otimes \log\left[p_X(x')\rho_B^{x'}\right]\right]\right\} && (24)\\
&= -\operatorname{Tr}\left\{\sum_x p_X(x)|x\rangle\langle x|_X \otimes \left(\rho_B^x \log\left[p_X(x)\rho_B^x\right]\right)\right\} && (25)\\
&= -\sum_x p_X(x) \operatorname{Tr}\left\{\rho_B^x \log\left[p_X(x)\rho_B^x\right]\right\}. && (26)
\end{aligned}$$

Consider that

$$\log\left[p_X(x)\rho_B^x\right] = \log\left(p_X(x)\right) I + \log \rho_B^x, \tag{27}$$

which implies that (26) is equal to

$$\begin{aligned}
&-\sum_x p_X(x)\left[\operatorname{Tr}\left\{\rho_B^x \log\left[p_X(x)\right]\right\} + \operatorname{Tr}\left\{\rho_B^x \log \rho_B^x\right\}\right] && (28)\\
&= -\sum_x p_X(x)\left[\log\left[p_X(x)\right] + \operatorname{Tr}\left\{\rho_B^x \log \rho_B^x\right\}\right]. && (29)
\end{aligned}$$

This last line is then equivalent to the statement of the theorem. □

4 Conditional Quantum Entropy

The definition of conditional quantum entropy that has been most useful in quantum information theory is the following simple one, inspired from the relation between joint entropy and marginal entropy in the classical case.

Definition 11 (Conditional Quantum Entropy)**.** *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The conditional quantum entropy $H(A|B)_\rho$ of ρ_{AB} is equal to the difference of the joint quantum entropy $H(AB)_\rho$ and the marginal entropy $H(B)_\rho$:*

$$H(A|B)_\rho \equiv H(AB)_\rho - H(B)_\rho. \tag{30}$$

The above definition is the most natural one, both because it is straightforward to compute for any bipartite state and because it obeys many relations that the classical conditional entropy obeys (such as chaining rules and conditioning reduces entropy). We explore many of these relations in the forthcoming sections. For now, we state "conditioning cannot increase entropy" as the following theorem and tackle its proof later on after developing a few more tools.

Theorem 12 (Conditioning Does Not Increase Entropy)**.** *Consider a bipartite quantum state ρ_{AB}. Then the following inequality applies to the marginal entropy $H(A)_\rho$ and the conditional quantum entropy $H(A|B)_\rho$:*

$$H(A)_\rho \geq H(A|B)_\rho. \tag{31}$$

We can interpret the above inequality as stating that conditioning cannot increase entropy, even if the conditioning system is quantum.

4.1 Conditional Quantum Entropy for Classical–Quantum States

A classical–quantum state is an example of a state for which conditional quantum entropy behaves as in the classical world. Suppose that two parties share a classical–quantum state ρ_{XB} of the form in (18). The system X is classical and the system B is quantum, and the correlations between these systems are entirely classical, determined by the probability distribution $p_X(x)$. Let us calculate the conditional quantum entropy $H(B|X)_\rho$ for this state:

$$H(B|X)_\rho = H(XB)_\rho - H(X)_\rho \tag{32}$$

$$= H(X)_\rho + \sum_x p_X(x) H(\rho_B^x) - H(X)_\rho \tag{33}$$

$$= \sum_x p_X(x) H(\rho_B^x). \tag{34}$$

The first equality follows from Definition 11. The second equality follows from Theorem 10, and the final equality results from algebra.

The above form for conditional entropy is completely analogous with the classical formula and holds whenever the conditioning system is classical.

4.2 Negative Conditional Quantum Entropy

One of the properties of the conditional quantum entropy in Definition 11 that seems counterintuitive at first sight is that it can be negative. This negativity holds for an ebit $|\Phi^+\rangle_{AB}$ shared between Alice and Bob. The marginal state on Bob's system is the maximally mixed state π_B. Thus, the marginal entropy $H(B)$ is equal to one, but the joint entropy vanishes, and so the conditional quantum entropy $H(A|B) = -1$.

What do we make of this result? Well, this is one of the fundamental differences between the classical world and the quantum world, and perhaps is the very essence of the departure from an informational standpoint. The informational statement is that we can sometimes be more certain about the joint state of a quantum system than we can be about any one of its individual parts, and this is the reason that conditional quantum entropy can be negative. This is in fact the same observation that Schrödinger made concerning entangled states:

> "When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction the two representatives [the quantum states] have become entangled. Another way of expressing the peculiar situation is: the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts, even though they may be entirely separate and therefore virtually capable of being 'best possibly known,' i.e., of possessing, each of them, a representative of its own. The lack of knowledge is by no means due to the interaction being insufficiently known — at least not in the way that it could possibly be known more completely — it is due to the interaction itself."

5 Coherent Information

Negativity of the conditional quantum entropy is so important in quantum information theory that we even have an information quantity and a special notation to denote the negative of the conditional quantum entropy:

Definition 13 (Coherent Information)**.** *The coherent information* $I(A\rangle B)_\rho$ *of a bipartite state* $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is as follows:*

$$I(A\rangle B)_\rho \equiv H(B)_\rho - H(AB)_\rho. \tag{35}$$

You should immediately notice that this quantity is the negative of the conditional quantum entropy in Definition 11, but it is perhaps more useful to think of the coherent information not merely as the negative of the conditional quantum entropy, but as an information quantity in its own right. This is why we employ a separate notation for it. The "I" is present because the coherent information is an information quantity that measures quantum correlations, much like the mutual information does in the classical case. For example, we have already seen that the coherent information of an ebit is equal to one. Thus, it is measuring the extent to which we know less about part of a system than we do about its whole. Perhaps surprisingly, the coherent information obeys a quantum data-processing inequality, which gives further support for it having an "I" present in its notation. The Dirac symbol "$\rangle$" is present to indicate that this quantity is a quantum information quantity, having a good meaning really only in the quantum world. The choice of "$\rangle$" over "$\langle$" also indicates a directionality from Alice to Bob, and this notation will make more sense when we begin to discuss the coherent information of a quantum channel.

Exercise 14. *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Consider a purification $|\psi\rangle_{ABE}$ of this state to some environment system E. Show that*

$$I(A\rangle B)_\rho = H(B)_\psi - H(E)_\psi. \tag{36}$$

Thus, there is a sense in which the coherent information measures the difference in the uncertainty of Bob and the uncertainty of the environment.

Exercise 15 (Duality of Conditional Entropy)**.** *Show that $-H(A|B)_\rho = I(A\rangle B)_\rho = H(A|E)_\psi$ for the purification in the above exercise.*

The coherent information can be both negative or positive depending on the bipartite state for which we evaluate it, but it cannot be arbitrarily large or arbitrarily small. The following theorem places a useful bound on its absolute value.

Theorem 16. *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The following bound applies to the absolute value of the conditional entropy $H(A|B)_\rho$:*

$$|H(A|B)_\rho| \leq \log \dim(\mathcal{H}_A). \tag{37}$$

The bounds are saturated for $\rho_{AB} = \pi_A \otimes \sigma_B$, where π_A is the maximally mixed state and $\sigma_B \in \mathcal{D}(\mathcal{H}_B)$, and for $\rho_{AB} = \Phi_{AB}$ (the maximally entangled state).

Proof. We first prove the inequality $H(A|B)_\rho \leq \log \dim(\mathcal{H}_A)$ in two steps:

$$\begin{aligned} H(A|B)_\rho &\leq H(A)_\rho &(38)\\ &\leq \log \dim(\mathcal{H}_A). &(39) \end{aligned}$$

The first inequality follows because conditioning reduces entropy (Theorem 12), and the second inequality follows because the maximum value of the entropy $H(A)_\rho$ is $\log \dim(\mathcal{H}_A)$. We now prove the inequality $H(A|B)_\rho \geq -\log \dim(\mathcal{H}_A)$. Consider a purification $|\psi\rangle_{EAB}$ of the state ρ_{AB}. We then have that

$$\begin{aligned} H(A|B)_\rho &= -H(A|E)_\psi &(40)\\ &\geq -H(A)_\rho &(41)\\ &\geq -\log \dim(\mathcal{H}_A). &(42) \end{aligned}$$

The first equality follows from Exercise 15. The first and second inequalities follow by the same reasons as the inequalities in the previous paragraph. □

Exercise 17 (Conditional Coherent Information)**.** *Consider a tripartite state ρ_{ABC}. Show that*

$$I(A\rangle BC)_\rho = I(A\rangle B|C)_\rho, \tag{43}$$

where $I(A\rangle B|C)_\rho \equiv H(B|C)_\rho - H(AB|C)_\rho$ is the conditional coherent information.

Exercise 18 (Conditional Coherent Information of a Classical–Quantum State)**.** *Suppose we have a classical–quantum state σ_{XAB} where*

$$\sigma_{XAB} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma_{AB}^x, \tag{44}$$

p_X is a probability distribution on a finite alphabet $\mathcal{X}$ and $\sigma_{AB}^x \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ for all $x \in \mathcal{X}$. Show that

$$I(A\rangle BX)_\sigma = \sum_x p_X(x) I(A\rangle B)_{\sigma^x}. \tag{45}$$

6 Quantum Mutual Information

The standard informational measure of correlations in the classical world is the mutual information, and such a quantity plays a prominent role in measuring classical and quantum correlations in the quantum world as well.

Definition 19 (Quantum Mutual Information)**.** *The quantum mutual information of a bipartite state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined as follows:*

$$I(A;B)_\rho \equiv H(A)_\rho + H(B)_\rho - H(AB)_\rho. \tag{46}$$

The following relations hold for quantum mutual information, in analogy with the classical case:

$$\begin{aligned} I(A;B)_\rho &= H(A)_\rho - H(A|B)_\rho & (47)\\ &= H(B)_\rho - H(B|A)_\rho. & (48) \end{aligned}$$

These immediately lead to the following relations between quantum mutual information and the coherent information:

$$\begin{aligned} I(A;B)_\rho &= H(A)_\rho + I(A\rangle B)_\rho & (49)\\ &= H(B)_\rho + I(B\rangle A)_\rho. & (50) \end{aligned}$$

The theorem below gives a fundamental lower bound on the quantum mutual information—we merely state it for now and give a full proof later.

Theorem 20 (Non-Negativity of Quantum Mutual Information)**.** *The quantum mutual information $I(A;B)_\rho$ of any bipartite quantum state ρ_{AB} is non-negative:*

$$I(A;B)_\rho \geq 0. \tag{51}$$

Exercise 21 (Conditioning Does Not Increase Entropy)**.** *Show that non-negativity of quantum mutual information implies that conditioning does not increase entropy (Theorem 12).*

Exercise 22 (Bound on Quantum Mutual Information)**.** *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Prove that the following bound applies to the quantum mutual information:*

$$I(A;B)_\rho \leq 2\log\left[\min\left\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\right\}\right]. \tag{52}$$

What is an example of a state that saturates the bound?

7 Conditional Quantum Mutual Information

We define the conditional quantum mutual information $I(A;B|C)_\rho$ of any tripartite state $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ similarly to how we did in the classical case:

$$I(A;B|C)_\rho \equiv H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho. \tag{53}$$

In what follows, we sometimes abbreviate "conditional quantum mutual information" as CQMI.

One can exploit the above definition and the definition of quantum mutual information to prove a chain rule for quantum mutual information.

Property 23 (Chain Rule for Quantum Mutual Information)**.** *The quantum mutual information obeys a chain rule:*

$$I(A;BC)_\rho = I(A;B)_\rho + I(A;C|B)_\rho. \tag{54}$$

The interpretation of the chain rule is that we can build up the correlations between A and BC in two steps: in a first step, we build up the correlations between A and B, and now that B is available (and thus conditioned on), we build up the correlations between A and C.

Exercise 24. *Use the chain rule for quantum mutual information to prove that*

$$I(A;BC)_\rho = I(AC;B)_\rho + I(A;C)_\rho - I(B;C)_\rho. \tag{55}$$

7.1 Non-negativity of CQMI

In the classical world, non-negativity of conditional mutual information follows trivially from non-negativity of mutual information. The proof of non-negativity of conditional quantum mutual information is far from trivial in the quantum world, unless the conditioning system is classical (see Exercise 26). It is a foundational result that non-negativity of this quantity holds because so much of quantum information theory rests upon this theorem's shoulders (in fact, we could say that this inequality is one of the "bedrocks" of quantum information theory). The list of its corollaries includes the quantum data-processing inequality, the answers to some additivity questions in quantum Shannon theory, the Holevo bound, and others. The proof of Theorem 25 follows directly from monotonicity of quantum relative entropy (Theorem 31), which we prove later. In fact, it is possible to show that monotonicity of quantum relative entropy follows from strong subadditivity as well, so that these two entropy inequalities are essentially equivalent statements.

Theorem 25 (Non-Negativity of CQMI)**.** *Let $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Then the conditional quantum mutual information is non-negative:*

$$I(A;B|C)_\rho \geq 0. \tag{56}$$

This condition is equivalent to the strong subadditivity inequality, so we also refer to this entropy inequality as strong subadditivity.

Exercise 26 (CQMI of Classical–Quantum States)**.** *Consider a classical–quantum state σ_{XAB} of the form in* (44)*. Prove the following relation:*

$$I(A;B|X)_\sigma = \sum_x p_X(x) I(A;B)_{\sigma_x}. \tag{57}$$

Conclude that non-negativity of conditional quantum mutual information is trivial in this special case in which the conditioning system is classical, simply by exploiting non-negativity of quantum mutual information (Theorem 20).

Exercise 27 (Conditioning Does Not Increase Entropy)**.** *Let $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Show that Theorem 25 is equivalent to the following stronger form of Theorem 12:*

$$H(B|C)_\rho \geq H(B|AC)_\rho. \tag{58}$$