| | |
|---|---|
| **PHYS 7895: Quantum Information Theory** | Fall 2015 |

# Lecture 14 — October 14, 2015

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



## 1 Overview

In a previous lecture, we discussed the major noiseless quantum communication protocols such as teleportation, super-dense coding, their coherent versions, and entanglement distribution in detail. Each of these protocols relies on the assumption that noiseless resources are available. For example, the entanglement distribution protocol assumes that a noiseless qubit channel is available to generate a noiseless ebit. This idealization allowed us to develop the main principles of the protocols without having to think about more complicated issues, but in practice, the protocols do not work as expected under the presence of noise.

Given that quantum systems suffer noise in practice, we would like to have a way to determine how well a protocol is performing. The simplest way to do so is to compare the output of an ideal protocol to the output of the actual protocol using a *distance measure* of the two respective output quantum states. That is, suppose that a quantum information-processing protocol should ideally output some quantum state $|\psi\rangle$, but the actual output of the protocol is a quantum state with density operator $\rho$. Then a performance measure $P(\psi, \rho)$ should indicate how close the ideal output is to the actual output. Figure 1 depicts the comparison of an ideal protocol with another protocol that is noisy.

This lecture introduces two distance measures that allow us to determine how close two quantum states are to each other. The first distance measure that we discuss is the *trace distance* and the second is the *fidelity*. (However, note that the fidelity is not a distance measure in the strict mathematical sense—nevertheless, we exploit it as a "closeness" measure of quantum states because it admits an intuitive operational interpretation.) These two measures are mostly interchangeable, but we introduce both because it is often times more convenient in a given situation to use one or the other.

Distance measures are particularly important in quantum Shannon theory because they provide a way for us to determine how well a protocol is performing. Recall that Shannon's method for both the noiseless and noisy coding theorem is to allow for a slight error in a protocol, but to show that this error vanishes in the limit of large block length. Later on when we prove quantum coding theorems, we borrow this technique of demonstrating asymptotically small error, with either the trace distance or the fidelity as the measure of performance.

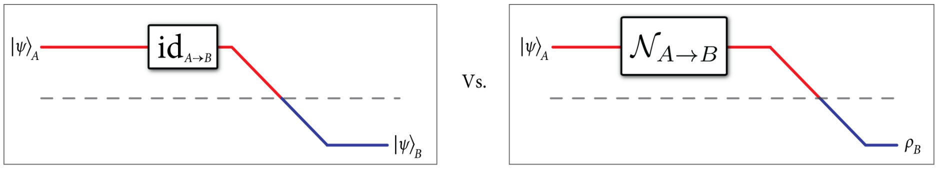


Figure 1: A distance measure quantifies how far the output of a given ideal protocol (depicted on the left) is from an actual protocol that exploits a noisy resource (depicted as the noisy quantum channel $\mathcal{N}_{A\to B}$ on the right).

# 2 Trace Distance

We first introduce the trace distance. Our presentation is somewhat mathematical because we exploit norms on linear operators in order to define it. Despite this mathematical flavor, this section offers an intuitive operational interpretation of the trace distance.

## 2.1 Trace Norm

**Definition 1** (Trace Norm)**.** *The trace norm or Schatten 1-norm $\|M\|_1$ of an operator $M \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ is defined as*

$$\|M\|_1 \equiv \mathrm{Tr}\{|M|\}, \tag{1}$$

*where $|M| \equiv \sqrt{M^\dagger M}$.*

**Proposition 2.** *The trace norm of an operator $M \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ is equal to the sum of its singular values.*

*Proof.* Recall that any function $f$ applied to a Hermitian operator $A$ is as follows:

$$f(A) \equiv \sum_{i:\alpha_i \neq 0} f(\alpha_i)|i\rangle\langle i|, \tag{2}$$

where $\sum_{i:\alpha_i \neq 0} \alpha_i |i\rangle\langle i|$ is a spectral decomposition of $A$. With these two definitions, it is straightforward to show that the trace norm of $M$ is equal to the sum of its singular values. Indeed, let $M = U\Sigma V$ be the singular value decomposition of $M$, where $U$ and $V$ are unitary matrices and $\Sigma$ is a rectangular matrix with the non-negative singular values along the diagonal. Then we can write

$$M = \sum_{i=0}^{d-1} \sigma_i |u_i\rangle\langle v_i|, \tag{3}$$

where $d$ is the rank of $M$, $\{\sigma_i\}$ are the strictly positive singular values of $M$, $\{|u_i\rangle\}$ are the orthonormal columns of $U$ in correspondence with the set $\{\sigma_i\}$, and $\{|v_i\rangle\}$ are the orthonormal rows of $V$ in correspondence with the set $\{\sigma_i\}$. Then

$$M^\dagger M = \left[\sum_{j=0}^{d-1} \sigma_j |v_j\rangle\langle u_j|\right] \left[\sum_{i=0}^{d-1} \sigma_i |u_i\rangle\langle v_i|\right] \tag{4}$$

$$= \sum_{i,j=0}^{d-1} \sigma_j \sigma_i |v_j\rangle\langle u_j||u_i\rangle\langle v_i| \tag{5}$$

$$= \sum_{i=0}^{d-1} \sigma_i^2 |v_i\rangle\langle v_i|, \tag{6}$$

so that

$$\sqrt{M^\dagger M} = \sum_{i=0}^{d-1} \sqrt{\sigma_i^2} |v_i\rangle\langle v_i| = \sum_{i=0}^{d-1} \sigma_i |v_i\rangle\langle v_i|, \tag{7}$$

finally implying that

$$\mathrm{Tr}\left\{|M|\right\} = \sum_{i=0}^{d-1} \sigma_i. \tag{8}$$

This means also that

$$\|M\|_1 \equiv \mathrm{Tr}\{\sqrt{MM^\dagger}\}, \tag{9}$$

because the singular values of $MM^\dagger$ and $M^\dagger M$ are the same (this is the key to one of the homework exercises). One can also easily show that the trace norm of a Hermitian operator is equal to the absolute sum of its eigenvalues. □

The trace norm is indeed a *norm* because it satisfies the following three properties: non-negative definiteness, homogeneity, and the triangle inequality.

**Property 3** (Non-negative Definiteness)**.** *The trace norm of an operator $M$ is non-negative definite:*

$$\|M\|_1 \geq 0. \tag{10}$$

*The trace norm is equal to zero if and only if the operator $M$ is the zero operator:*

$$\|M\|_1 = 0 \quad \Leftrightarrow \quad M = 0. \tag{11}$$

**Property 4** (Homogeneity)**.** *For any constant $c \in \mathbb{C}$,*

$$\|cM\|_1 = |c| \, \|M\|_1 \,. \tag{12}$$

**Property 5** (Triangle Inequality)**.** *For any two operators $M, N \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, the following triangle inequality holds:*

$$\|M + N\|_1 \leq \|M\|_1 + \|N\|_1 \,. \tag{13}$$

Non-negative definiteness follows because the sum of the singular values of an operator is non-negative, and the singular values are all equal to zero (and thus the operator is equal to zero) if and only if the sum of the singular values is equal to zero. Homogeneity follows directly from the fact that $|cM| = |c||M|$. We later give a proof of the triangle inequality (however, for a special case only). Exercise 9 below asks you to prove it for square operators.

Three other important properties of the trace norm are its invariance under isometries, convexity, and a variational characterization. Each of the properties below often arise as useful tools in quantum Shannon theory.

**Property 6** (Isometric Invariance)**.** *The trace norm is invariant under multiplication by isometries $U$ and $V$:*

$$\left\|UMV^\dagger\right\|_1 = \|M\|_1 \,. \tag{14}$$

**Property 7** (Convexity)**.** *For any two operators $M, N \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ and any convex coefficients $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$, the following convexity inequality holds:*

$$\|\lambda_1 M + \lambda_2 N\|_1 \leq \lambda_1 \|M\|_1 + \lambda_2 \|N\|_1 \,. \tag{15}$$

Isometric invariance holds because $M$ and $UMV^\dagger$ have the same singular values. Convexity follows directly from the triangle inequality and homogeneity (thus, any norm is convex in this sense).

**Property 8** (Variational characterization)**.** *For a square operator $M \in \mathcal{L}(\mathcal{H})$, the following variational characterization of the trace norm holds*

$$\|M\|_1 = \max_U |\mathrm{Tr}\{MU\}| \,, \tag{16}$$

*where the optimization is over all unitary operators.*

*Proof.* The above characterization follows by taking a singular value decomposition of $M$ as $M = WDV$, with $W$ and $V$ unitaries and $D$ a diagonal matrix of singular values, and applying the Cauchy–Schwarz inequality to find that

$$\begin{aligned}
|\mathrm{Tr}\{MU\}| &= |\mathrm{Tr}\{WDVU\}| &(17)\\
&= \left|\mathrm{Tr}\left\{\sqrt{D}\sqrt{D}VUW\right\}\right| &(18)\\
&\leq \sqrt{\mathrm{Tr}\left\{\sqrt{D}\sqrt{D}\right\}}\sqrt{\mathrm{Tr}\left\{\left(\sqrt{D}VUW\right)^\dagger \sqrt{D}VUW\right\}} &(19)\\
&= \mathrm{Tr}\{D\} &(20)\\
&= \|M\|_1 \,. &(21)
\end{aligned}$$

The inequality is a consequence of the Cauchy–Schwarz inequality for the Hilbert–Schmidt inner product:

$$\left|\mathrm{Tr}\left\{A^\dagger B\right\}\right| \leq \sqrt{\mathrm{Tr}\{A^\dagger A\}}\sqrt{\mathrm{Tr}\{B^\dagger B\}}. \tag{22}$$

Equality holds by picking $U = V^\dagger W^\dagger$, from which we recover (16). □

**Exercise 9.** *Prove that the triangle inequality (Property 5) holds for square operators $M, N \in \mathcal{L}(\mathcal{H})$. (Hint: Use the characterization in Property 8.)*

## 2.2 Trace Distance from the Trace Norm

The trace norm induces a natural distance measure, called the *trace distance.*

**Definition 10** (Trace Distance)**.** *Given any two operators $M, N \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, the trace distance between them is as follows:*

$$\left\| M - N \right\|_1 . \tag{23}$$

The trace distance is especially useful as a measure of the distinguishability of two quantum states with respective density operators $\rho$ and $\sigma$. The following bounds apply to the trace distance between any two density operators $\rho$ and $\sigma$:

$$0 \leq \left\| \rho - \sigma \right\|_1 \leq 2. \tag{24}$$

Sometimes it is useful to employ the *normalized* trace distance $\frac{1}{2} \left\| \rho - \sigma \right\|_1$, so that $\frac{1}{2} \left\| \rho - \sigma \right\|_1 \in [0, 1]$. The lower bound in (24) applies when two quantum states are equal—quantum states $\rho$ and $\sigma$ are equal to each other if and only if their trace distance is zero. The physical implication of the trace distance being equal to zero is that no measurement can distinguish $\rho$ from $\sigma$. The upper bound in (24) follows from the triangle inequality:

$$\left\| \rho - \sigma \right\|_1 \leq \left\| \rho \right\|_1 + \left\| \sigma \right\|_1 = 2. \tag{25}$$

The trace distance is maximum when $\rho$ and $\sigma$ have support on orthogonal subspaces. Later, we will prove that this is the only case in which this happens, after introducing the fidelity. The physical implication of maximal trace distance is that there exists a measurement that can perfectly distinguish $\rho$ from $\sigma$. We discuss these operational interpretations of the trace distance in more detail in Section 2.4.

**Exercise 11.** *Show that the trace distance is invariant with respect to an isometric quantum channel, in the following sense:*

$$\left\| \rho - \sigma \right\|_1 = \left\| U \rho U^\dagger - U \sigma U^\dagger \right\|_1 , \tag{26}$$

*where $U$ is an isometry. The physical implication of* (28) *is that an isometric quantum channel applied to both states does not increase or decrease the distinguishability of the two states.*

## 2.3 Trace Distance as a Probability Difference

We now state and prove an important lemma that gives an alternative and useful way for characterizing the trace distance. This particular characterization finds application in many proofs of the lemmas that follow concerning trace distance.

**Lemma 12.** *The normalized trace distance $\frac{1}{2} \left\| \rho - \sigma \right\|_1$ between quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is equal to the largest probability difference that two states $\rho$ and $\sigma$ could give to the same measurement outcome $\Lambda$:*

$$\frac{1}{2} \left\| \rho - \sigma \right\|_1 = \max_{0 \leq \Lambda \leq I} \operatorname{Tr} \left\{ \Lambda \left( \rho - \sigma \right) \right\} . \tag{27}$$

*The above maximization is with respect to all positive semi-definite operators $\Lambda \in \mathcal{L}(\mathcal{H})$ that have their eigenvalues bounded from above by one.*

*Proof.* Consider that the difference operator $\rho - \sigma$ is Hermitian and so we can diagonalize it as follows:

$$\rho - \sigma = \sum_i \lambda_i |i\rangle\langle i|,$$

where $\{|i\rangle\}$ is an orthonormal basis of eigenvectors and $\{\lambda_i\}$ is a set of real eigenvalues. Let us define

$$P \equiv \sum_{i:\lambda_i \geq 0} \lambda_i |i\rangle\langle i|, \qquad Q \equiv \sum_{i:\lambda_i < 0} |\lambda_i| \, |i\rangle\langle i|, \tag{28}$$

which implies that $P$ and $Q$ are positive semi-definite and that

$$\rho - \sigma = P - Q. \tag{29}$$

Consider also that $PQ = 0$, and let $\Pi_P$ and $\Pi_Q$ denote the projections onto the supports of $P$ and $Q$, respectively:

$$\Pi_P \equiv \sum_{i:\lambda_i \geq 0} |i\rangle\langle i|, \qquad \Pi_Q \equiv \sum_{i:\lambda_i < 0} |i\rangle\langle i|. \tag{30}$$

Then it follows that

$$\Pi_P P \Pi_P = P, \qquad \Pi_Q Q \Pi_Q = Q, \tag{31}$$

$$\Pi_P Q \Pi_P = 0, \qquad \Pi_Q P \Pi_Q = 0. \tag{32}$$

The following property holds as well:

$$|\rho - \sigma| = |P - Q| = P + Q. \tag{33}$$

because the supports of $P$ and $Q$ are orthogonal and the absolute value of the operator $P - Q$ takes the absolute value of its eigenvalues. Therefore,

$$\begin{aligned} \|\rho - \sigma\|_1 &= \text{Tr}\{|\rho - \sigma|\} && (34)\\ &= \text{Tr}\{P + Q\} && (35)\\ &= \text{Tr}\{P\} + \text{Tr}\{Q\}. && (36) \end{aligned}$$

But

$$\begin{aligned} \text{Tr}\{P\} - \text{Tr}\{Q\} &= \text{Tr}\{P - Q\} && (37)\\ &= \text{Tr}\{\rho - \sigma\} && (38)\\ &= \text{Tr}\{\rho\} - \text{Tr}\{\sigma\} && (39)\\ &= 0. && (40) \end{aligned}$$

where the last equality follows because both quantum states have unit trace. Therefore, $\text{Tr}\{P\} = \text{Tr}\{Q\}$ and

$$\|\rho - \sigma\|_1 = 2 \cdot \text{Tr}\{P\}. \tag{41}$$

Consider then that

$$\begin{aligned} \text{Tr}\{\Pi_P(\rho - \sigma)\} &= \text{Tr}\{\Pi_P(P - Q)\} && (42)\\ &= \text{Tr}\{\Pi_P P\} && (43)\\ &= \text{Tr}\{P\} && (44)\\ &= \frac{1}{2}\|\rho - \sigma\|_1. && (45) \end{aligned}$$

Now we prove that the operator $\Pi_P$ is the maximizing one. Let $\Lambda$ be any positive semi-definite operator with spectrum bounded above by one. Then

$$\begin{aligned}
\operatorname{Tr}\{\Lambda(\rho-\sigma)\} &= \operatorname{Tr}\{\Lambda(P-Q)\} &(46)\\
&\le \operatorname{Tr}\{\Lambda P\} &(47)\\
&\le \operatorname{Tr}\{P\} &(48)\\
&= \frac{1}{2}\|\rho-\sigma\|_1 . &(49)
\end{aligned}$$

The first inequality follows because $\Lambda$ and $Q$ are non-negative and thus $\operatorname{Tr}\{\Lambda Q\}$ is non-negative. The second inequality holds because $\Lambda \le I$. The final equality follows from (43). □

**Exercise 13.** *Let $\rho = |0\rangle\langle 0|$ and $\sigma = |+\rangle\langle +|$. Compute $P$, $Q$, $\Pi_P$, and $\Pi_Q$, as defined in* (30) *and* (32)*, for this choice of $\rho$ and $\sigma$. Compute the trace distance $\|\rho-\sigma\|_1$.*

**Exercise 14.** *Show that the trace norm of any Hermitian operator $\omega$ is given by the following optimization:*

$$\|\omega\|_1 = \max_{-I\le\Lambda\le I} \operatorname{Tr}\{\Lambda\omega\} . \tag{50}$$

## 2.4 Operational Interpretation of the Trace Distance

We now provide an operational interpretation of the trace distance as the distinguishability of two quantum states. The interpretation results from a hypothesis-testing scenario. Suppose that Bob prepares one of two quantum states $\rho_0$ or $\rho_1$ for Alice to distinguish. Suppose further that it is equally likely *a priori* for him to prepare either $\rho_0$ or $\rho_1$. Let $X$ denote the Bernoulli random variable assigned to the prior probabilities so that $p_X(0) = p_X(1) = 1/2$. Alice can perform a binary POVM with elements $\Lambda \equiv \{\Lambda_0, \Lambda_1\}$ to distinguish the two states. That is, Alice guesses the state in question is $\rho_0$ if she receives outcome "0" from the measurement or she guesses the state in question is $\rho_1$ if she receives outcome "1" from the measurement. Let $Y$ denote the Bernoulli random variable assigned to the classical outcomes of her measurement. The success probability $p_{\text{succ}}(\Lambda)$ for this hypothesis testing scenario is the sum of the probability of detecting "0" when the state is $\rho_0$ and the probability of detecting "1" when the state is $\rho_1$:

$$\begin{aligned}
p_{\text{succ}}(\Lambda) &= p_{Y|X}(0|0)p_X(0) + p_{Y|X}(1|1)p_X(1) &(51)\\
&= \operatorname{Tr}\{\Lambda_0\rho_0\}\frac{1}{2} + \operatorname{Tr}\{\Lambda_1\rho_1\}\frac{1}{2}. &(52)
\end{aligned}$$

We can simplify this expression using the completeness relation $\Lambda_0 + \Lambda_1 = I$:

$$\begin{aligned}
p_{\text{succ}}(\Lambda) &= \frac{1}{2}\left(\operatorname{Tr}\{\Lambda_0\rho_0\} + \operatorname{Tr}\{(I-\Lambda_0)\rho_1\}\right) &(53)\\
&= \frac{1}{2}\left(\operatorname{Tr}\{\Lambda_0\rho_0\} + \operatorname{Tr}\{\rho_1\} - \operatorname{Tr}\{\Lambda_0\rho_1\}\right) &(54)\\
&= \frac{1}{2}\left(\operatorname{Tr}\{\Lambda_0\rho_0\} + 1 - \operatorname{Tr}\{\Lambda_0\rho_1\}\right) &(55)\\
&= \frac{1}{2}\left(1 + \operatorname{Tr}\{\Lambda_0(\rho_0-\rho_1)\}\right). &(56)
\end{aligned}$$

Now Alice has freedom in choosing the POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$ to distinguish the states $\rho_0$ and $\rho_1$, and she would like to choose one that maximizes the success probability $p_{\text{succ}}(\Lambda)$. Thus, we can define the success probability with respect to all measurements as follows:

$$p_{\text{succ}} \equiv \max_{\Lambda} p_{\text{succ}}(\Lambda) = \max_{\Lambda} \frac{1}{2} \left(1 + \text{Tr}\left\{\Lambda_0 \left(\rho_0 - \rho_1\right)\right\}\right). \tag{57}$$

We can rewrite the above quantity in terms of the trace distance using its characterization in Lemma 14 because the expression inside of the maximization involves only the operator $\Lambda_0$:

$$p_{\text{succ}} = \frac{1}{2}\left(1 + \frac{1}{2}\left\|\rho_0 - \rho_1\right\|_1\right). \tag{58}$$

Thus, the normalized trace distance has an operational interpretation that it is linearly related to the maximum success probability in distinguishing two quantum states $\rho_0$ and $\rho_1$ in a quantum hypothesis testing experiment. From the above expression for the success probability, it is clear that the states are indistinguishable when $\left\|\rho_0 - \rho_1\right\|_1$ is equal to zero. That is, it is just as good for Alice to guess randomly what the state might be, and in this case, she can do no better than to have 1/2 probability of being correct. On the other hand, the states are perfectly distinguishable when $\left\|\rho_0 - \rho_1\right\|_1$ is maximal and the measurement that distinguishes them consists of two projectors: one projects onto the non-negative eigenspace of $\rho_0 - \rho_1$ and the other projects onto the negative eigenspace of $\rho_0 - \rho_1$. In this sense, we can say that the normalized trace distance is the bias away from random guessing in a hypothesis testing experiment.

**Exercise 15.** *Suppose that the prior probabilities in the above hypothesis-testing scenario are not uniform but are rather equal to $p_0$ and $p_1$. Show that the success probability is instead given by*

$$p_{\text{succ}} = \frac{1}{2}\left(1 + \left\|p_0\rho_0 - p_1\rho_1\right\|_1\right). \tag{59}$$