1 Distance between states

Now that we have introduced a description of states of a quantum system via density matrices, an important question to ask is the following:

(Q) Given two states $\rho$ and $\sigma$, how well can we distinguish between them?

In order to answer this question we need to introduce a notion of “distance\(^1\) between states, which gives a measure of their distinguishability. The concept of a distance measure between states is essential for determining the efficiency of protocols e.g. data compression or transmission of information over a noisy channel. There are various different notions of distance between states. We will focus on a few of them.

1.1 Trace Distance

Definition (Trace Distance) : The trace distance of two states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is defined as

$$D(\rho, \sigma) := \frac{1}{2} \| \rho - \sigma \|_1,$$

where $\| A \|_1 = \text{Tr}|A|$, with $|A| := \sqrt{A^\dagger A}$.

Consider the difference operator $A = \rho - \sigma$. In its terms, $D(\rho, \sigma) = \frac{1}{2} \text{Tr}|A|$. Let $A$ have the eigenvalue decomposition

$$A = \sum_i a_i |\phi_i\rangle \langle \phi_i|$$

$$= Q - R,$$

where

$$Q = \sum_{i: a_i \geq 0} a_i |\phi_i\rangle \langle \phi_i|$$

$$R = \sum_{i: a_i < 0} (-a_i) |\phi_i\rangle \langle \phi_i|.$$

\(^1\)It need not be a distance in the Euclidean sense.
Note that \( Q \) and \( R \) are positive operators with mutually orthogonal supports. It is easy to see that
\[
D(\rho, \sigma) = \frac{1}{2} (\text{Tr}Q + \text{Tr}R),
\]
(4)
since \( \text{Tr}|A| = \text{Tr}Q + \text{Tr}R \).

**Lemma 1**
\[
D(\rho, \sigma) = \max_{0 \leq P \leq I} \text{Tr}(P(\rho - \sigma)).
\]
(5)

**Proof:** Note that \( \rho - \sigma = Q - R \). Since \( \rho \) and \( \sigma \) are both states, this implies that \( \text{Tr}Q = \text{Tr}R \). From this and (4) it follows that
\[
D(\rho, \sigma) = \text{Tr}Q = \text{Tr}R
\]
(6)

Let \( P \) be a projector onto the support of \( Q \). Then
\[
\text{Tr}(P(\rho - \sigma)) = \text{Tr}(P(Q - R))
\]
\[
= \text{Tr}(PQ)
\]
\[
= \text{Tr}Q
\]
\[
= D(\rho, \sigma).
\]
(7)

In the above we have used the fact that \( R \) and \( Q \) have orthogonal supports.

Conversely, for any \( 0 \leq P \leq I \),
\[
\text{Tr}(P(\rho - \sigma)) = \text{Tr}(P(Q - R))
\]
\[
\leq \text{Tr}(PQ)
\]
\[
\leq \text{Tr}Q
\]
\[
= D(\rho, \sigma),
\]
(8)
since \( P, R \geq 0 \) and \( P \leq I \). Hence, for any \( 0 \leq P \leq I \) we have \( D(\rho, \sigma) \geq \text{Tr}(P(\rho - \sigma)) \), which in turn implies that
\[
D(\rho, \sigma) \geq \max_{0 \leq P \leq I} \text{Tr}(P(\rho - \sigma)),
\]
(9)
and as we saw in the previous part, this bound can be achieved. This yields the statement of the Lemma.
Properties of the trace distance

The trace distance is a metric on the space of density operators:

1. It is symmetric: \( D(\rho, \sigma) = D(\sigma, \rho) \).

2. \( D(\rho, \sigma) = 0 \) if and only if \( \rho = \sigma \).

3. It satisfies the triangle inequality.

Proof is left as an exercise.

Lemma 2 Monotonicity under quantum operations:

\[
D(\Lambda(\rho), \Lambda(\sigma)) \leq D(\rho, \sigma),
\]

for any linear, CPTP map \( \Lambda \).

Proof: Let \( P \) be a projector such that

\[
D(\Lambda(\rho), \Lambda(\sigma)) = \operatorname{Tr}(P(\Lambda(\rho), \Lambda(\sigma)))
\]

Since \( \Lambda \) is trace-preserving,

\[
\operatorname{Tr}\Lambda(Q) = \operatorname{Tr}Q.
\]

Using (6), (11) and (12), and the fact that \( 0 \leq P \leq I, R \geq 0 \), we obtain,

\[
D(\rho, \sigma) = \operatorname{Tr}Q \\
= \operatorname{Tr}\Lambda(Q) \\
\geq \operatorname{Tr}(P\Lambda(Q)) \\
\geq \operatorname{Tr}(P(\Lambda(Q) - \Lambda(R))) \\
= D(\Lambda(\rho), \Lambda(\sigma)).
\]

1.2 Operational significance of the trace distance

One of the fundamental tasks in quantum information theory is (binary) quantum hypothesis testing. It works as follows.

Suppose Alice prepares a quantum system \( A \) in one of two states \( \rho_0 \) and \( \rho_1 \) with equal probability, and sends it to Bob through a noiseless quantum channel. Bob’s aim is to figure out which of the two states the system has been prepared in. For this purpose he does a measurement given by a binary POVM (with POVM elements \( E_0 \) and \( E_1 = I - E_0 \)) on the
system A. Let the POVM element $E_0$ (resp. $E_1$) correspond to his inference being that the state is $\rho_0$ (resp. $\rho_1$). The probability of error in Bob’s inference in this case is given by

$$p_e(\{E_0, E_1\}) = \frac{1}{2} [\text{Tr}(E_0 \rho_1) + \text{Tr}(E_1 \rho_0)]$$

$$= \frac{1}{2} [1 - \text{Tr}(E_0 (\rho_0 - \rho_1))] \quad (14)$$

The minimum probability of error, $p_e^\ast$, is obtained by minimizing the above expression over all possible binary POVMs. Hence the minimum probability of error is given by the following expression:

$$p_e^\ast := \min_{\{E_0, E_1\}_{\text{POVM}}} p_e(\{E_0, E_1\})$$

$$= \frac{1}{2} \left[1 - \max_{0 \leq E_0 \leq I} \text{Tr}(E_0 (\rho_0 - \rho_1))\right]$$

$$= \frac{1}{2} [1 - D(\rho_0, \rho_1)] \quad (15)$$

where in the second line we have made use of the fact that the binary POVM $\{E_0, E_1\}$ is completely specified by the element $E_0$ since $E_1 = I - E_0$, and of course $0 \leq E_0 \leq I$; the last line follows from Lemma 1.

Hence, the maximum success probability $p_{\text{success}}^\ast := 1 - p_e^\ast$ is given by

$$p_{\text{success}}^\ast = \frac{1}{2} [1 + D(\rho_0, \rho_1)] \quad (16)$$

The above equation provides an operational significance to the trace distance in terms of the maximum success probability in distinguishing two quantum states $\rho_0$ and $\rho_1$ in a binary quantum hypothesis testing experiment.

The relation (16) implies that if the two states are such that $D(\rho_0, \rho_1) = 0$ (i.e. the states are indistinguishable), then the maximum success probability is 1/2, and Bob might as well guess randomly which of the two states the system is in. On the other hand, $p_{\text{success}}^\ast$ is maximal in the case in which $D(\rho_0, \rho_1)$ is maximum. In this case the measurement for which $p_{\text{success}}^\ast$ is maximized is a projective measurement consisting of two projection operators – one projecting onto the support of the positive (more precisely, non-negative) part of the difference operator $(\rho_0 - \rho_1)$, and the other projecting onto the support of the negative part of the difference operator $(\rho_0 - \rho_1)$. If $(\rho_0 - \rho_1)$ has the following spectral decomposition:

$$(\rho_0 - \rho_1) = \sum_i a_i |\phi_i\rangle \langle \phi_i|,$$

then the two projection operators are the following

$$P_{\geq 0} := \sum_{i: a_i \geq 0} |\phi_i\rangle \langle \phi_i|$$

$$P_{< 0} = \sum_{i: a_i < 0} |\phi_i\rangle \langle \phi_i| \quad (17)$$
1.3 Fidelity

**Definition (Fidelity)**: The fidelity of two states $\rho, \sigma \in (\mathcal{H})$ is defined as

$$F(\rho, \sigma) := \text{Tr}\left(\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}}\right) = \|\sigma^{1/2}\rho^{1/2}\|_1,$$

where for any operator $\|A\|_1 := \text{Tr}(\sqrt{A^\dagger A})$.

We shall see that $F(\rho, \sigma)$ is symmetric in its arguments (even though this is not obvious from its definition). Further we will prove that $0 \leq F(\rho, \sigma) \leq 1$, with $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$.

The expression for the fidelity takes simple forms in particular cases

(i) If $[\rho, \sigma] = 0$; In this case $\rho$ and $\sigma$ have a simultaneous eigenbasis, which we denote by $\{|e_i\rangle\}_i$. Let their eigenvalue decompositions be given by

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|; \quad \sigma = \sum_i \mu_i |e_i\rangle\langle e_i|$$

Then,

$$F(\rho, \sigma) = \text{Tr}\left[\sum_i \sqrt{\lambda_i \mu_i} |e_i\rangle\langle e_i| \right] = \sum_i \sqrt{\lambda_i \mu_i} = F_{cl}(\lambda, \mu),$$

where $F_{cl}(\lambda, \mu) := \sum_i \sqrt{\lambda_i \mu_i}$ is the classical fidelity of the probability distributions $\lambda = \{\lambda_i\}_i$ and $\mu = \{\mu_i\}_i$, (i.e., the eigenvalue distributions of $\rho$ and $\sigma$).

(ii) If one of the two states is pure:

$$F(|\phi\rangle\langle \phi|, \sigma) = \text{Tr}\left[|\phi\rangle\langle \phi| \sigma \right] = \sqrt{\langle \phi| \sigma| \phi\rangle},$$

since $\langle \phi| \sigma| \phi\rangle$ is a scalar and $|\phi\rangle$ is a normalized vector.

In particular, if both the states are pure, $\rho = |\psi\rangle\langle \psi|$ and $\sigma = |\phi\rangle\langle \phi|$, then the fidelity depends only on the absolute value of the inner product of the states:

$$F(\rho, \sigma) = |\langle \psi| \phi\rangle|.$$

**Exercise**: Prove that the fidelity is invariant under unitary transformations, i.e., for two states $\rho$, $\sigma$ and a unitary operator $U$:

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma).$$
The following theorem (due to Uhlmann) allows us to establish the properties of the fidelity mentioned before:

**Theorem (Uhlmann’s Theorem):** The fidelity of two states \( \rho_A \) and \( \sigma_A \) is given by

\[
F(\rho_A, \sigma_A) = \max_{|\psi_{\rho}^{RA}\rangle, |\psi_{\sigma}^{RA}\rangle} |\langle \psi_{\rho}^{RA} | \psi_{\sigma}^{RA} \rangle|,
\]

where the maximisation ranges over all purifications \( |\psi_{\rho}^{RA}\rangle \), \( |\psi_{\sigma}^{RA}\rangle \) of \( \rho_A \) and \( \sigma_A \) respectively.

We employ the following lemma to prove this theorem:

**Lemma 3** For any operator \( A \in \mathcal{B}(\mathcal{H}) \)

\[
\|A\|_1 = \max_U |\text{Tr}(UA)|,
\]

where \( U \) ranges over all unitaries on \( \mathcal{H} \).

**Proof of Lemma 3:** We need to show that for any unitary operator \( U \in \mathcal{B}(\mathcal{H}) \)

\[
|\text{Tr}(AU)| \leq \text{Tr}|A|,
\]

with equality for some appropriately chosen \( U \).

Let \( A = |A|V \) be the polar decomposition of \( A \). Equality is obviously satisfied in (24) for the choice \( U = V^\dagger \).

To prove the inequality, we use the polar decomposition of \( A \) and the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product:

\[
|\text{Tr}(AU)| = |\text{Tr}(|A|VU)| \\
\leq \sqrt{\text{Tr}|A| \cdot \text{Tr}(U^\dagger V^\dagger |A|VU)} \\
= \text{Tr}|A|,
\]

where we have used the cyclicity of the trace, and the unitarity of \( U \) and \( V \).

We can now proceed to prove Uhlmann’s theorem:

**Proof:** Since all purifications are equivalent up to unitaries on the reference system, the fidelity can be expressed as a maximization over unitaries instead (starting with any two fixed purifications) and the statement of Uhlmann’s theorem can be equivalently expressed as follows:

\[
F(\rho_A, \sigma_A) = \max_U |\langle \psi_{\rho}^{RA} | (U_R^\dagger \otimes I_A)(U_R^\dagger \otimes I_A) |\psi_{\sigma}^{RA} \rangle|,
\]

\[
= \max_U |\langle \psi_{\rho}^{RA} | (U \otimes I_A) |\psi_{\sigma}^{RA} \rangle|.
\]

(27)
To prove (27) choose $|\psi^R_\rho\rangle$, $|\psi^R_\sigma\rangle$ be the canonical purifications of $\rho_A$ and $\sigma_A$ respectively, i.e.,

$$|\psi^R_\rho\rangle = \sqrt{d}(I_R \otimes \sqrt{\rho_A})|\Omega\rangle; \quad |\psi^R_\sigma\rangle = \sqrt{d}(I_R \otimes \sqrt{\sigma_A})|\Omega\rangle,$$

where $d = \dim\mathcal{H}_A = \dim\mathcal{H}_R$ and $|\Omega\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|ii\rangle$ is a MES of Schmidt rank $d$ in $\mathcal{H}_R \otimes \mathcal{H}_A$.

Using these, and identities regarding the MES we obtain

$$|\langle \psi^R_\rho(U \otimes I_A)|\psi^R_\sigma\rangle| = |\text{Tr}(\sqrt{\rho_A}\sqrt{\sigma_A}U^T)| \tag{28}$$

The proof is then completed by using Lemma 3. Details in lecture.

**Conclusions from Uhlmann’s Theorem**

The following relations are a consequence of Uhlmann’s Theorem

1. $0 \leq F(\rho, \sigma) \leq 1$ and $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$.

2. $F(\rho, \sigma) = F(\sigma, \rho)$, (symmetry)

Further the following important lemma provides the monotonicity of the fidelity under partial trace:

**Lemma 4** Let $\rho_{AB}$ and $\sigma_{AB}$ denote states of a bipartite system $AB$, and let $\rho_A$ and $\sigma_A$ denote the corresponding reduced states of the subsystem $A$. Then

$$F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A) \tag{29}$$

**Proof:** By Uhlmann’s Theorem, there exist purifications $|\psi^{ABC}_\rho\rangle$ and $|\psi^{ABC}_\sigma\rangle$ of the states $\rho_{AB}$ and $\sigma_{AB}$ respectively such that

$$F(\rho_{AB}, \sigma_{AB}) = |\langle \psi^{ABC}_\rho|\psi^{ABC}_\sigma\rangle|. \tag{30}$$

Trivially, $|\psi^{ABC}_\rho\rangle$ and $|\psi^{ABC}_\sigma\rangle$ are also purifications of the reduced states $\rho_A$ and $\sigma_A$:

$$\rho_A = \text{Tr}_B\rho_{AB} = \text{Tr}_B\text{Tr}_C|\psi^{ABC}_\rho\rangle\langle \psi^{ABC}_\rho| = \text{Tr}_B|\psi^{ABC}_\rho\rangle\langle \psi^{ABC}_\rho|. \tag{31}$$

Hence, by Uhlmann’s Theorem

$$F(\rho_A, \sigma_A) := \max_{|\psi_\rho\rangle,|\psi_\sigma\rangle} |\langle \psi_\rho|\psi_\sigma\rangle|$$

$$\geq |\langle \psi^{ABC}_\rho|\psi^{ABC}_\sigma\rangle|$$

$$= F(\rho_{AB}, \sigma_{AB}). \tag{32}$$

We will explore relations between trace distance and fidelity in an example sheet.
Appendix: Entanglement Fidelity

Suppose $|\psi_{RA}\rangle \in \mathcal{H}_R \otimes \mathcal{H}_A$ be a purification of the state $\rho \in \mathcal{D}(\mathcal{H}_A)$ of a quantum system $A$. Here $R$ denotes the reference system used for the purification. Let $\Lambda : \mathcal{B}(\mathcal{H}_A) \mapsto \mathcal{B}(\mathcal{H}_A)$ be a linear CPTP map.

(Q) How well is the entanglement between the systems $A$ and $R$ preserved by the quantum operation $\Lambda$?

This can be quantified by the entanglement fidelity $F_e(\rho, \Lambda)$, which is a function of $\rho$ and $\Lambda$ and defined as follows:

$$F_e(\rho, \Lambda) := \langle \psi_{RA}|((\text{id}_R \otimes \Lambda)|\psi_{RA}\rangle\rangle |\psi_{RA}\rangle$$  \hspace{1cm} (33)

By defining $\psi_{RA} = |\psi_{RA}\rangle\langle \psi_{RA}|$, note that

$$F_e(\rho, \Lambda) = \left( F(\psi_{RA}, (\text{id}_R \otimes \Lambda)\psi_{RA}) \right)^2,$$  \hspace{1cm} (34)

that is, the square of the fidelity between the initial and final states of $RA$.

Note that $F_e(\rho, \Lambda)$ depends on $\rho$ and $\Lambda$, and not on the particular purification $|\psi_{RA}\rangle$. To see this, recall that any two purifications $|\psi_{RA}\rangle$ and $|\phi_{RA}\rangle$ of the state $\rho \in \mathcal{D}(\mathcal{H}_A)$ are related by a unitary operation that acts only upon the reference system $R$, i.e., there exists a unitary operator $U \in \mathcal{B}(\mathcal{H}_B)$ such that $|\phi_{RA}\rangle = U \otimes I_A |\psi_{RA}\rangle$. Then, denoting the pure state density matrices corresponding to $|\psi_{RA}\rangle$ and $|\phi_{RA}\rangle$ by $\psi_{RA}$ and $\phi_{RA}$ respectively, and using the invariance of the fidelity under unitary transformations, we obtain that

$$F(\psi_{RA}, (\text{id}_R \otimes \Lambda)\psi_{RA}) = F(\phi_{RA}, (\text{id}_R \otimes \Lambda)\phi_{RA})$$  \hspace{1cm} (35)

and hence the result follows.

The following lemma gives a simple expression for the entanglement fidelity:

**Lemma 5** Let a linear CPTP map $\Lambda : \mathcal{B}(\mathcal{H}) \mapsto \mathcal{B}(\mathcal{H})$ have the following Kraus form

$$\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger,$$  \hspace{1cm} (36)

where $\rho \in \mathcal{D}(\mathcal{H})$. Then

$$F_e(\rho, \Lambda) = \sum_k |\text{Tr}(A_k \rho)|^2$$  \hspace{1cm} (37)

**Proof**

Denoting $\psi_{RA} = |\psi_{RA}\rangle\langle \psi_{RA}|$, we have

$$F_e(\rho, \Lambda) = \langle \psi_{RA}|((\text{id}_R \otimes \Lambda)|\psi_{RA}\rangle\rangle |\psi_{RA}\rangle$$

$$= \sum_k \langle \psi_{RA}|(I \otimes A_k)\psi_{RA}(I \otimes A_k^\dagger)|\psi_{RA}\rangle$$

$$= \sum_k |\langle \psi_{RA}|(I \otimes A_k)|\psi_{RA}\rangle|^2$$  \hspace{1cm} (38)
Now consider the Schmidt decomposition

\[ |\psi_{RA}\rangle = \sum_i \sqrt{\lambda_i} |i_R\rangle |i_A\rangle. \]  

(39)

Obviously,

\[ \rho = \text{Tr}_R |\psi_{RA}\rangle \langle \psi_{RA}| = \sum_i \lambda_i |i_A\rangle \langle i_A| \]. \]

(40)

The identity (37) follows from (38), (39) and (40).

**Lemma 6** If \( \Lambda : \mathcal{B}(\mathcal{H}) \mapsto \mathcal{B}(\mathcal{H}) \) is a linear CPTP map, then for any \( \rho \in \mathcal{D}(\mathcal{H}) \),

\[ F_\varepsilon(\rho, \Lambda) \leq (F(\rho, \Lambda(\rho)))^2 \]

(41)

**Proof**
The proof follows from a direct application of the monotonicity of the fidelity under partial trace.