8.372 Quantum Information Science III **Fall 2024**

Lecture 2: September 10, 2024

Scribe: David D. Dai and Yeongwoo Hwang *Trace distance, fidelity, metrics*

2.1 Norms

2.1.1 General Properties

Norms measure "how big" an object is. They have three properties:

1. $\|cx\| = |c| \cdot \|x\| \quad \forall c \in \mathcal{C}$ (homogeneous)
2. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
3. $\|x\| = 0 \leftrightarrow x = 0$, where 0 is the additive identity (separating)

If an operation satisfies the first two properties but not the third separating property, it is called a seminorm. A space equipped with a valid metric is a metric space.

2.1.2 Vector Norms

An important class of norms on vectors $\mathcal{C}^d$ are the L_p norms:

$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}}. \tag{2.1}$$

There are a few important special cases. The L^1 norm is the sum of the absolute values of the entries (Manhattan distance), the L^2 norm is the Euclidean norm, and the L^∞ norm is the maximum of the absolute values of the entries. Intuitively, p tells us how much the larger entries are weighed relative to the smaller entries. The L^∞ norm ignores all but the largest entry[1], while the L^1 norm treats all entries equally. Additionally, we require $p \geq 1$; $p < 1$ violates the triangle inequality, as can be seen easily for $x = (1, 0)$ and $y = (0, 1)$.

We call L_p and L_q *dual* if $1/p + 1/q = 1$. Then Hölder's inequality states:

$$\|x\|_p = \max_{\|y\|_q = 1} |\langle x, y \rangle| . \tag{2.2}$$

We will not prove Hölder's inequality but can inspect a few cases. $p = 2$ and $q = 2$ are dual, yielding:

$$\|x\|_2 = \max_{\|y\|_2 = 1} |\langle x, y \rangle| . \tag{2.3}$$

[1]Note that duplicated entries are not a problem. Even if there are m copies of the largest entry, the factor of m is suppressed by the $1/p$ power.

This is consistent with the Cauchy-Schwarz inequality: $\langle x, y\rangle \leq \sqrt{\langle x, x\rangle \langle y, y\rangle}$ with saturation if and only if $x \propto y$. $p = 1$ and $q = \infty$ are also dual, yielding:

$$\begin{aligned} \|x\|_1 &= \max_{\|y\|_\infty = 1} |\langle x, y\rangle| , \\ \sum_i |x_i| &= \max_{\max(|y_i|)=1} |\langle x, y\rangle| . \end{aligned} \tag{2.4}$$

This makes sense; if $x = (r_1 e^{i\theta_1}, r_2 e^{i\theta_2} \dots r_d e^{i\theta_d})$ for positive r and θ, then the maximum is achieved by $y = (e^{-i\theta_1}, e^{-i\theta_2} \dots e^{-i\theta_d})$. $p = \infty$ and $q = 1$ are also dual, yielding:

$$\begin{aligned} \|x\|_\infty &= \max_{\|y\|_1 = 1} |\langle x, y\rangle| , \\ \max(x_i) &= \max_{\sum_i |y_i| = 1} |\langle x, y\rangle| . \end{aligned} \tag{2.5}$$

This also makes sense; the maximum is achieved by $y_i = \delta_{i,\mathrm{argmax}(|x_i|)}$.

2.1.3 Matrix Norms

For some operator X, the Schatten p-norm S_p is:

$$\|X\|_{S_p} \equiv \|X\|_p = \|\Sigma(X)\|_p, \tag{2.6}$$

where $\Sigma(X)$ are the singular values of X. The Schatten p-norm of X is the L_p norm of X's singular values. All of the S_p norms have the nice property that they are invariant under left or right matrix multiplication by a unitary, because this does not change the singular values. If X is Hermitian, the singular values may be replaced with eigenvalues.

$S_\infty(X)$ corresponds to the maximum singular value of X, which is also the maximum factor by which X can stretch a vector by:

$$\|X\|_\infty = \max \Sigma(X) = \max_{\|v\|_2 = 1} \|Xv\|. \tag{2.7}$$

S_1 and S_2 can be expressed without using the SVD:

$$\|X\|_1 = \mathrm{Tr}\sqrt{X^\dagger X}, \|X\|_2 = \sqrt{\mathrm{Tr}\, X^\dagger X}, \tag{2.8}$$

where the square root is well-defined because $X^\dagger X$ is positive semi-definite. It is easy to show that Eq. 2.8 is consistent with Eq. 2.6 by plugging in $X = U\Sigma V^\dagger$.

We note that proving the triangle inequality for S_p is nontrivial.

2.1.4 Some Useful Sets

The unit sphere S and ball B with respect to some norm $\|\cdot\|$ are:

$$S = \{x : \|x\| = 1\}, \tag{2.9}$$

$$B = \{x : \|x\| \leq 1\}. \tag{2.10}$$

Below are a few sets commonly encountered in quantum information science:

- Pure quantum states: $S(L_2)$,
- Classical probability distributions: $S(L_1) \cap \{\text{nonnegative entries}\}$,
- Density matrices: $S(S_1) \cap \{\text{positive semidefinite}\}$,
- Measurement operators: $B(S_\infty) \cap \{\text{positive semidefinite}\}$.

2.2 Comparing Probability Distributions

2.2.1 Total Variation Distance

The total variation distance (TVD) between two probability distributions p and q is:

$$T(p,q) = \frac{1}{2}\|p-q\|_1 = \frac{1}{2}\sum_i |p_i - q_i|\,. \tag{2.11}$$

Note that

$$T(p,q) = \frac{1}{2}\sum_i |p_i - q_i| \le \frac{1}{2}\sum_i (p_i + q_i) = 1, \tag{2.12}$$

so $T(p,q) \in [0,1]$.

$T(p,q)$ has a nice operation definition. Consider a guessing game where we are given a random variable X, drawn either from distribution p or q with equal prior probability $1/2$. Given X's value, how often can we correctly guess which distribution it was drawn from? Bayes' rule gives the probability that X was drawn from p given $X = i$:

$$P(X \text{ from } p | X = i) = \frac{p_i}{p_i + q_i}. \tag{2.13}$$

The best strategy is to guess p if $P(X \text{ from } p|X=i) > 1/2$ and q otherwise, so the probability that we guess correctly given $X = i$ is

$$P(\text{correct}|X=i) = \max\left(\frac{p_i}{p_i+q_i}, \frac{q_i}{p_i+q_i}\right) = \frac{1}{2} + \frac{|p_i - q_i|}{2(p_i+q_i)}. \tag{2.14}$$

Then the probability that we guess correctly in general is:

$$\begin{aligned} P(\text{correct}) &= \sum_i P(\text{correct}|X=i)P(X=i) \\ &= \sum_i \left[\frac{1}{2} + \frac{|p_i - q_i|}{2(p_i+q_i)}\right]\left[\frac{p_i+q_i}{2}\right] \\ &\frac{1}{2} + \frac{T(p,q)}{2}. \end{aligned} \tag{2.15}$$

For example, if $T(p,q) = 1/2$, then we can correctly guess whether a random variable was drawn from p or q three-quarters of the time.

2.2.2 Fidelity (Bhattacharyya Coefficient)

An alternative way of comparing probability distributions is the fidelity:

$$F(p,q) = \langle \sqrt{p}, \sqrt{q} \rangle \,, \tag{2.16}$$

where $\sqrt{p}$ is the element-wise square root of the vector of probabilities. The square root is necessary to guarantee that $F(p,p) = 1$ for all p, something which would not be true for $\langle p, p \rangle$.

If our random variable comes from concatenating two independent random variables, i.e. $i = (a,b), p_i = p_a p_b$, then the fidelity factorizes:

$$\begin{aligned} F(p_i, q_i) &= \sum_i \sqrt{p_i q_i} \\ &= \sum_{a,b} \sqrt{p_a p_b q_a q_b} \\ &= F(p_a, q_a) F(p_b, q_b). \end{aligned} \tag{2.17}$$

The TVD notably lacks this property. The total variation distance and fidelity also satisfy the inequalities

$$1 - F \leq T \leq \sqrt{2(1-F)}. \tag{2.18}$$

Even through T doesn't factorize, Eq. 2.18 allows us to bound $T(p^{\otimes n}, q^{\otimes n})$, where $p^{\otimes n}$ means the probability distribution corresponding to drawing n times from p. In particular, T approaches 1 exponentially in n.

2.3 Quantum Distinguishability

What is the appropriate metric via which we should compare quantum states? A first guess could be the ℓ_2 vector norm, i.e. $\| \, |p\rangle - |q\rangle \, \|_2$. This is equal to $\sqrt{2(1 - \mathrm{Re}(\langle p|q\rangle)}$ and has an undesirable sensitivity to relative phase. By maximizing over the global phase, we obtain the closeness measure,

$$| \, \langle p|q\rangle \, |$$

which we'll define as the *fidelity* between $|p\rangle$ and $|q\rangle$. However, this definition also has a drawback, which is that there is no nice "operational" interpretation of fidelity. For that, we introduce the *trace distance*

Definition 2.3.1 (Trace Distance)**.** *Let ρ, σ be two mixed states. The trace distance between ρ, σ is denoted $T(\rho, \sigma)$ and is defined equivalently as,*

$$T(\rho,\sigma) = \frac{1}{2} \|\rho - \sigma\|_1 \tag{2.19}$$

$$= \max_{0 \preceq M \preceq \mathbb{I}} \mathrm{tr}[M(\rho - \sigma)] \tag{2.20}$$

This metric has some nice properties,

- (Unitary Invariance) $T(V\rho V^\dagger, V\sigma V^\dagger) = T(\rho,\sigma)$
- (Data Processing Inequality or Monotonicity) $T(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq T(\rho,\sigma)$

In fact, by defining our channel via the measurement obtaining the maximum in (2.20) we can saturate the monotonicity bound:

$$\mathcal{E}(\rho) := \text{tr}[M\rho]\,|0\rangle\langle 0| + \text{tr}[(\mathbb{I} - M)\rho]\,|1\rangle\langle 1|$$

Note that we've define trace distance over mixed states, whereas our definition of fidelity was limited to pure states. We can generalize to mixed states as follows,

Definition 2.3.2 (Fidelity)**.** *Let ρ, σ be two mixed states. The fidelity between ρ, σ is denoted $F(\rho, \sigma)$ and is defined as*

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \text{tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]$$

Some nice properties of fidelity are,

- (Fuchs-van de Graaf Inequalities) $1 - F \leq T \leq \sqrt{1 - F^2}$
- (DPI or Monotonicity) $F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq F(\rho, \sigma)$

We'll prove the Fuchs-van de Graaf inequalities in the problem set. Interestingly, fidelity *does not* satisfy the triangle inequality (and thus is not a metric); however, $\arccos(F(\cdot, \cdot))$ does.

2.3.1 Uhlmann's Theorem

The definition we gave for fidelity is quite cumbersome; in practice it can be very annoying to compute the square roots of matrices. Uhlmann's theorem gives a nice alternative characterization of the fidelity between mixed states.

Theorem 2.3.1 (Uhlmann's Theorem)**.** *Let ρ, σ be two mixed states defined over a quantum register A. Then,*

$$F(\rho, \sigma) = \max_{\substack{|\rho\rangle_{AB} \text{ s.t. } \text{tr}_B[|\rho\rangle\langle\rho|]=\rho \\ |\sigma\rangle_{AB} \text{ s.t. } \text{tr}_B[|\sigma\rangle\langle\sigma|]=\sigma}} |\,\langle\rho|\sigma\rangle\,|$$

i.e. the mixed state fidelity between ρ and σ is the maximum pure state fidelity between purifications of ρ and σ.

Before giving the proof of this theorem, we define the "canonical" purification of a mixed state. To do so, we define the (unnormalized) maximally entangled state between two registers A and B of equal dimension d as,

$$|\Gamma\rangle = \sum_{i=1}^{d} |i\rangle_A\,|i\rangle_B$$

Definition 2.3.3 (Canonical Purification)**.** *For a mixed state ρ, its canonical purification is denoted by $|\phi^\rho\rangle$ and defined as,*

$$|\phi^\rho\rangle := (\sqrt{\rho}_A \otimes \mathbb{I}_B)\,|\Gamma\rangle_{AB}$$

The fact that $\text{tr}_B[\phi^\rho] = \rho$ can be verified by a simple computation. We'll also need the following lemma, which, intuitively, we should think of as the matrix analogue of "tuning" the phases of a probability distribution to obtain the ℓ_1 norm.

Lemma 2.3.1.

$$\max_U |\,\text{tr}[AU]| = \|A\|_1$$

Proof. Take the singular-value decomposition of A to obtain $A = UDW^\dagger$. Then, $\mathrm{tr}[AU] = \mathrm{tr}\left[DW^\dagger UV\right]$, where we've used the cyclic property of the trace. Rather than maximizing over U, consider maximizing over $U = W\tilde{U}V^\dagger$. This is equivalent as for an original $U^\star$, we can set $\tilde{U} = W^\dagger U^\star V$ and then $U = U^\star$. Thus,

$$\max_U \mathrm{tr}\left[DW^\dagger UV\right] = \max_{U=W\tilde{U}V^\dagger} \mathrm{tr}\left[DW^\dagger UV\right] = \max_{\tilde{U}} \mathrm{tr}\left[D\tilde{U}\right]$$

But since D is a diagonal matrix, the RHS is just $\max_U \sum_i D_{i,i}U_{i,i} \leq \mathrm{tr}[|D|] = \|A\|_1$. □

We now give the proof of Uhlmann's theorem.

Proof of Theorem 2.3.1. Recall that all purifications of a mixed state ρ_A as $|\rho\rangle_{AB}$ are equivalent under a unitary on just the B register. As a result, we can replace $\max_{|\rho\rangle,|\sigma\rangle} |\langle\rho|\sigma\rangle|$ with

$$\max_{U,V} |\langle\phi^\rho| (\mathbb{I}\otimes U_B)(\mathbb{I}\otimes V_B) |\phi^\sigma\rangle| \tag{2.21}$$

But U_BV_B is just another unitary and can think of this as fixing $|\phi^\rho\rangle$ and only maximizing over a single unitary (which is equivalent to maximizing over purifications of σ). Then,

$$\begin{aligned}
(2.21) &= \max_U |\langle\phi^\rho| (\mathbb{I}\otimes U) |\phi^\sigma\rangle| && (2.22)\\
&= \max_U \langle\Gamma| (\sqrt{\rho}\otimes\mathbb{I})(\mathbb{I}\otimes U)(\sqrt{\sigma}\otimes\mathbb{I}) |\Gamma\rangle && (2.23)\\
&= \max_U \langle\Gamma| (\sqrt{\rho}\sqrt{\sigma})\otimes U |\Gamma\rangle && (2.24)\\
&= \max_U \mathrm{tr}\left[\sqrt{\rho}\sqrt{\sigma}U^\top\right] && (2.25)\\
&= \|\sqrt{\rho}\sqrt{\sigma}\|_1 && (2.26)
\end{aligned}$$

where in (2.25) we've used that the maximally mixed state over registers A, B satisfies $(\mathbb{I}\otimes U)|\Gamma\rangle = (U^\top\otimes\mathbb{I})|\Gamma\rangle$. The last equality uses Lemma 2.3.1. □

2.4 No-go Theorem for Bit Commitment

To conclude the lecture, we revisit the no-go theorem from the first lecture and relax the hiding condition so that Bob is allowed some small probability of recover Alice's commitment. Formally, let's say that an honest Alice commits to the states $|\psi_0\rangle_{AB}$ or $|\psi_1\rangle_{AB}$. Then, a limit on Bob's ability to distinguish these two states corresponds to requiring,

$$T(\mathrm{tr}_A[\psi_0], \mathrm{tr}_A[\psi_1]) \leq \varepsilon \implies F(\mathrm{tr}_A[\psi_0], \mathrm{tr}_A[\psi_1]) \geq 1-\varepsilon$$

Since $|\psi_0\rangle_{AB}$ and $|\psi_1\rangle_{AB}$ are purifications of $\mathrm{tr}_A[\psi_0]$ and $\mathrm{tr}_A[\psi_1]$, we know that there exists a unitary U such that,

$$F((U\otimes\mathbb{I})|\psi_0\rangle, |\psi_1\rangle) \geq 1-\varepsilon$$

Define $|\psi_{\text{fake}}\rangle := (U\otimes\mathbb{I})|\psi_0\rangle$. Converting back to trace distance, we have that

$$T(\psi_{\text{fake}}, \psi_1) \leq \sqrt{2\varepsilon} \overset{\text{monotonicitiy}}{\implies} T(\mathcal{E}_{\text{reveal}}(\psi_{\text{fake}}), \mathcal{E}_{\text{reveal}}(\psi_1)) \leq \sqrt{2\varepsilon}$$

We conclude that Bob cannot distinguish between $|\psi_{\text{fake}}\rangle$, which corresponds to $|\psi_0\rangle$ with a unitary applied to only Alice's side, and $|\psi_1\rangle$. Thus, this protocol is not binding.