CS 880: Quantum Algorithms	2/16/2021

Lecture 6: Quantum distance

Instructor: Dieter van Melkebeek — Scribe: Owen Levin

In today's lecture, we make use of *density operators* discussed in the previous lecture in two key areas. First we study the structure of quantum systems involving multiple parties each holding different parts of the system's information. Density operators yield a natural way to express the evolution of the system for either party. The exercise posed in the previous lecture is used as a segue into this topic. Then we go on to characterize various matrix and vector norms, which we use in order to evaluate how errors in implementation of a quantum circuit affect the results we measure. Here, density operators play a key role in allowing us to bound the difference between outputs of similar systems.

1 Density operators recap

Definition 1 (Density operator). *Let $\{|\psi_i\rangle\}_i$ be a collection of pure states. Then for any convex combination of these states, $\{(p_i, |\psi_i\rangle)\}$, i.e. for any mixed or pure state, we define the corresponding* density operator *ρ as the following matrix:*

$$\rho \doteq \sum_i p_i |\psi_i\rangle \langle\psi_i| .$$

An equivalent characterization of the set of such matrices is given in the following theorem.

Theorem 1 (Characterization of density operators). *A matrix ρ is a* density operator *if and only if*

- *ρ is Hermitian,*
- *ρ is positive semi-definite, and*
- $\operatorname{Tr} \rho = 1$

1.1 Evolution of a quantum system's density operator

Two key aspects of describing a quantum system is how the states change when we (1) apply unitary operators, and (2) perform measurements. We show that density operators behave nicely with both of these actions.

For that, we need to describe the density operator ρ' corresponding to the state $|\psi'\rangle$ obtained from state $|\psi\rangle$ either by applying a unitary operation to $|\psi\rangle$ or by making a measurement of $|\psi\rangle$.

Let's start with applying a unitary operation U to the state $|\psi\rangle$. The new state is $|\psi'\rangle = U|\psi\rangle$, so the corresponding density operator is

$$\begin{aligned}\rho' &= U|\psi\rangle (U|\psi\rangle)^* \\ &= U|\psi\rangle\langle\psi| U^* \\ &= U\rho U^*.\end{aligned}$$

Because U is a linear operator, this extends to any linear combination of states $\{|\psi_i\rangle\}_i$. As a corollary, any convex combination of states is also covered, so we get the result for all mixed states for free. Thus, for any density operator ρ and unitary U we have $\rho \overset{U}{\mapsto} \rho' = U\rho U^*$

Similarly, for partial measurements P_s of ρ via projections onto a subset of the basis, we have $\rho \mapsto \sum_s P_s \rho P_s$, where P_s is the projection onto state $|\psi_s\rangle$.

To see how the probability distributions on states changes, suppose we make a measurement of a state $|\psi\rangle$ whose density operator is ρ. We measure with respect to some orthogonal basis $\{|\phi_1\rangle, \ldots, |\phi_k\rangle\}$. The state is a linear combination of the basis vectors, say $|\psi\rangle = \sum_i \alpha_i |\phi_i\rangle$. We observe the state $|\phi_i\rangle$ with probability $p_i = |\alpha_i|^2$, so the new state is a mixed state $\left\{\left(|\alpha_i|^2, |\phi_i\rangle\right)\right\}_i$, and its corresponding density operator is

$$\rho' = \sum_i |\alpha_i|^2 |\phi_i\rangle \langle\phi_i| = \sum_i |\phi_i\rangle \alpha_i \overline{\alpha_i} \langle\phi_i| \tag{1}$$

Note that if we multiply ρ on the right with $|\phi_j\rangle$ and on the left with $\langle\phi_j|$, we get the probability that we observe $|\phi_j\rangle$. This follows from the second summation in (1) because $\langle\phi_j|\phi_i\rangle = 1$ if $i = j$, and is zero otherwise. Thus, another way of writing (1) is

$$\rho' = \sum_i \langle\phi_i| \rho |\phi_i\rangle |\phi_i\rangle \langle\phi_i| .$$

Once again, we can apply linearity to get the resulting density operator when we observe a mixed state.

Theorem 2. *Two states behave identically if and only if they have the same density operator. Equivalently, we have the contrapositive: for any pair of distinct density operators ρ_1, ρ_2, there exists some quantum circuit that distinguishes between the two with positive probability.*

Remark 3. Due to the equivalence between distinct states and density operators, we may use the terms interchangably.

Proof (Theorem 2). Assume that two density operators are the same. We just showed in the previous paragraphs that we only need the density operator in order to describe the outcome of some quantum process. We gave an expression for the density operator corresponding to the next state of the system. Thus, any quantum process operating on two states with the same density operators evolves the same for both of the states, results in the same final density operator for the two final states, and, most importantly, the probability of observing a string x is the same for both states. Thus, since we rely on observations to decide on the output of quantum algorithms, we cannot tell from the distribution of the observations which state we were in at the beginning.

Now, consider two distinct density operators ρ_1, ρ_2. Let $\sigma = \rho_1 - \rho_2$. σ is Hermitian because sums of Hermitian operators are Hermitian. Therefore, σ can be diagonalized over a basis of orthonormal eigenvectors $\{|\psi_i\rangle\}_i$:

$$\sigma = \sum_i \lambda_i |\psi_i\rangle \langle\psi_i| \quad \text{for } \vec{\lambda} \neq \vec{0}.$$

Now, we observe ρ_1, ρ_2 in the basis $\{|\psi_i\rangle\}_i$.

$$\Pr[\text{observe } |\psi_i\rangle \text{ in state } \rho_1] = \langle\psi_i| \rho_1 |\psi_i\rangle$$

$$\Pr\left[\text{observe } |\psi_i\rangle \text{ in state } \rho_2\right] = \langle\psi_i|\, \rho_2\, |\psi_i\rangle$$

These must differ for at least one i because the expansion of their difference σ on the basis element $|\psi_i\rangle$ is λ_i. Since $\vec{\lambda} \neq \vec{0}$, this difference will be nonzero in some component i. Thus, there is a positive difference in probability that we observe state $|\psi_i\rangle$ for the densities ρ_1, ρ_2. □

2 Solution to exercise

Exercise 1. *Consider the EPR pair (Einstein-Podolsky-Rosen): a two-component system in state*

$$\tfrac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right),$$

where Alice holds first component, and separately Bob holds the second.

1. *Describe the density operator for the entire system*
   (a) *at the start,*
   (b) *after Bob's component is measured,and*
   (c) *after the outcome is announced.*
2. *Describe Alice's view at each of the above points.*

2.1 Solution to part 1.

We describe the system using density operators. In part (a) we have

$$\rho_a = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

After measuring the second qubit, we have

$$\rho_b = \tfrac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} + \tfrac{1}{2} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that this is just the partial measurement projection of ρ_a onto just the first qubit: $\rho_b = P_{|*0\rangle}\rho_a P_{|*0\rangle} + P_{|*1\rangle}\rho_a P_{|*1\rangle}$. Thus, for outcome $b \in \{0,1\}$, the density operator after b is announced is given by $P_{|*b\rangle}\rho_a P_{|*b\rangle}$ (up to normalization). Thus, ρ_c can be either of two options:

$$\rho_c \in \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}$$

2.2 Solution to part 2.

We describe the system using density operators. In part (a) we have

$$\rho_a = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

From Alice's perspective, this is $\rho_a^{(\text{Alice})} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ After measuring the second qubit, we have

$$\rho_b = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

From Alice's perspective, this is again $\rho_b^{(\text{Alice})} \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \rho_a^{(\text{Alice})}$ She doesn't perceive any change when Bob makes his measurement. ρ_c can be either of two options:

$$\rho_c \in \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}.$$

From Alice's perspective, these are $\rho_c^{(\text{Alice})} \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$.

2.3 Discussion of exercise

Remark 4 (Spooky action at a distance). The collapse of an entire system (two parties') after a local observation by only one of two parties for the EPR pair and similar systems has been called "spooky action at a distance" by some. Einstein, Podolsky, and Rosen described a similar experiments with change occurring faster than light speed. That is, if Alice and Bob are separated by some distance, then the instantaneous change in density operator from Bob's measurement changed the whole system. However, because Alice cannot perceive any change in her density operator until she hears what state Bob measured, there's no actual information traveling faster than light here.

Remark 5 (Probabilistic spooky action). The same phenomenon occurs in classical probabilistic computation. If we set up the same system with probabilistic measurement and separated the two bits far apart, then we achieve the same result.

Remark 6 (Hidden variables). It should be noted that in the quantum setting, there is still *more* going on than in the classical setting. In the classical setting, such things can be described by hidden variables such as secret coin flips which neither party can see, but determine all the information. This is *not* the case for quantum systems. A well known collection of result codifying the fact that quantum systems are not described by probabilistic hidden variables can be found by reading about Bell's theorem or Bell's inequality.

3 Two-party systems

In a two-party system, two parties, Alice and Bob, have access to two different parts A and B of a quantum register. Alice applies unitary transformations and observations to her part of the register without affecting Bob's part, and vice versa. The general form of the state is $\sum_{s,t} \alpha_{s,t} |s\rangle |t\rangle$ where the first component (the state $|s\rangle$) belongs to Alice and the second component belongs to Bob. To Alice, the state of the system looks like a mixed state over all possible states that Bob's part of the quantum register could be in. Thus, Alice's state is

$$\left\{ \left(\Pr[t], \frac{\sum_s \alpha_{s,t} |s\rangle}{\sqrt{\Pr[t]}} \right) \right\}_t \quad \text{where} \quad \Pr[t] = \sum_s |\alpha_{s,t}|^2,$$

and there is a symmetric expression for Bob's state.

Suppose ρ is the density operator for the whole system. Then we wish to reduce ρ to $\rho^{(\text{Alice})}$, the density operator for Alice. We call this the *reduced density operator.*

$$\begin{aligned} \rho^{(\text{Alice})} &= \sum_t \Pr[t] \cdot \frac{\sum_s \alpha_{s,t} |s\rangle}{\sqrt{\Pr[t]}} \cdot \frac{\sum_{s'} \overline{\alpha_{s',t}} \langle s'|}{\sqrt{\Pr[t]}} \\ &= \sum_{s,s'} \left(\sum_t \alpha_{s,t} \overline{\alpha_{s',t}} \right) |s\rangle \langle s'| \\ &= \sum_{s,s'} \left(\sum_t \rho_{\underbrace{(s,t)}_{\substack{\text{row}\\ \text{index}}}, \underbrace{(s',t)}_{\substack{\text{column}\\ \text{index}}}} \right) |s\rangle \langle s'| \, . \doteq \mathrm{Tr}_B(\rho) \end{aligned} \tag{2}$$

This expression can be extended by linearity to any mixed combined state. The final expression in (2) $\mathrm{Tr}_B(\rho)$ is read as *trace with respect to B of ρ* and is known as a partial trace. This may look a little confusing, so let's look at an example.

Example 7. *Suppose Alice and Bob operate on a two-qubit system, where the first qubit A belongs to Alice and the second qubit B belongs to Bob. The density operator is*

$$\rho = \begin{pmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho_{01,00} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho_{10,00} & \rho_{10,01} & \rho_{10,10} & \rho_{10,11} \\ \rho_{11,00} & \rho_{11,01} & \rho_{11,10} & \rho_{11,11} \end{pmatrix}.$$

Then the trace with respect to B is the matrix

$$\mathrm{Tr}_B(\rho) = \begin{pmatrix} \rho_{00,00} + \rho_{01,01} & \rho_{00,10} + \rho_{01,11} \\ \rho_{10,00} + \rho_{11,01} & \rho_{10,10} + \rho_{11,11} \end{pmatrix}.$$

We see that the top left entry of $\mathrm{Tr}_B(\rho)$ is the trace of a submatrix of ρ where Alice's part of the first index (i.e., the first bit of the first index in our case) is fixed to s and Alice's part of the second index is fixed to s'. Using this observation, we see that the trace with respect to A is the matrix

$$\mathrm{Tr}_A(\rho) = \begin{pmatrix} \rho_{00,00} + \rho_{10,10} & \rho_{00,01} + \rho_{10,11} \\ \rho_{01,00} + \rho_{11,10} & \rho_{01,01} + \rho_{11,11} \end{pmatrix}.$$

Example 8. *Consider the density operator for the EPR pair* $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$*:*

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Then we have $\mathrm{Tr}_A(\rho) = \mathrm{Tr}_B(\rho) = \frac{1}{2}I$.

4 Distance between density operators

Suppose we have implemented some quantum system in the real world. In real world situations due to noise or errors, theoretically identical operators may be different though "close". For which sense of "close"do outputs of density operators produce the same or similar output?

4.1 Statistical distance

As a toy model for the situation, consider two distinct densities ρ_b for $b \in \{0,1\}$. Let p_b be the probability distribution on $\{0,1\}^n$ resulting from full observation of ρ_b. One notion of distance we could use is the statistical distance between their probability distributions.

Definition 2 (Statistical distance). *The* statistical distance *between two distributions* p_0, p_1 *is given by*

$$d_{\text{stat}}(p_0, p_1) \doteq \max\{|p_0(E) - p_1(E)| \;:\; E \subseteq \{0,1\}^n\} = \frac{1}{2}\sum_s |p_0(s) - p_1(s)| = \frac{1}{2}\|p_0 - p_1\|_1$$

Now, we investigate the statistical distance between distributions arising from density operators, and we connect it back to a norm on the density operators' difference. First, recall that $p_b(s) = \langle s|\rho_b|s\rangle$. Then consider the deviation $\sigma \doteq \rho_0 - \rho_1$ expressed in some orthonormal basis of pure states $\{|\psi_i\rangle\}_i$: $\sigma = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$. For any state s, this allows us to write $p_0(s) - p_1(s) = \langle s|\sigma|s\rangle = \sum_i \lambda_i |\langle\psi_i|s\rangle|^2$. Now, we can sum over s to compute the 1-norm:

$$\begin{aligned}
\|p_0 - p_1\|_1 &= \sum_s \left|\sum_i |\langle\psi_i|s\rangle|^2\right| \\
&\le \sum_s \sum_i |\lambda_i| \, |\langle\psi_i|s\rangle|^2 && \text{(by the triangle inequality)} \\
&= \sum_i |\lambda_i| \underbrace{\sum_s |\langle\psi_i|s\rangle|^2}_{\substack{=1 \text{ since this is just} \\ \text{the probability of} \\ \text{being in } any \text{ state } s}} \\
&= \sum_i |\lambda_i| \doteq \|\sigma\|_{\mathrm{Tr}}
\end{aligned}$$

We can conclude for probability distributions p_0, p_1 arising from density operators ρ_0, ρ_1 that $d_{\text{stat}}(p_0, p_1) \le \frac{1}{2}\|\rho_0 - \rho_1\|_{\mathrm{Tr}}$. We aim now to connect $\|\cdot\|_{\mathrm{Tr}}$ to more familiar norms, such as the 2-norm. In particular, we will see that if ρ is a density operator and U_0, U_1 are unitary, then for $\rho_b \doteq U_b \rho U_b^*$ the trace norm $\|\rho_0 - \rho_1\|_{\mathrm{Tr}}$ will be bounded above by $2\|U_0 - U_1\|_2$.

4.2 Singular value decomposition

Here we describe Singular value decomposition (SVD) of complex matrices, as well as some of the terminology used with SVD.

Theorem 9 (SVD). *For any matrix $A \in \mathbb{C}^{M\times N}$ there exists unitary matrices $U \in \mathbb{C}^{M\times M}, V \in \mathbb{C}^{N\times N}$ as well as diagonal matrix $\Sigma \in \mathbb{R}^{M\times N}_{\geq 0}$ such that*

$$A = U\Sigma V^*.$$

4.2.1 SVD Terminology

Definition 3 (Singular values). *Σ is given by* $\operatorname{diag}(\vec{\sigma}) = \operatorname{diag}(\sigma_1, \sigma_2, \ldots)$ *where $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ are called the* singular values *of A. the singular values $\sigma_1, \sigma_2, \ldots, \sigma_k$ are sometimes called the* top k singular values.

Proposition 10. *Singular values are square roots of the eigenvalues of A^*A.*

Definition 4 (column/left singular vectors). *$U_{*,j}$ are the* column singular vectors*, and are sometimes also called* left singular vectors.

Definition 5 (row/right singular vectors). *$V_{*,j}$ are the* row singular vectors*, and are sometimes also called* right singular vectors.

Corollary 11. *If A is Hermitian, then (1) The singular values are absolute values of the eigenvalues of A, and (2) The left and right singular vectors coincide up to scalar multiplication.*

4.3 Vector and matrix norms

Definition 6 (norm). *A* norm *is a map $\|\cdot\|$ from a vectorspace V to $\mathbb{R}$ satisfying **(1), (2),** and**(3)** for any $u, v \in V$ and $\alpha \in \mathbb{R}$.*

(1) absolute homogeneity: $\|\alpha v\| = |\alpha| \, \|v\|$,

(2) triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$,

(3) definiteness: *If $\|v\| = 0$, then $v = 0$.*

Example 12 (Vector p-norms). *For $p \in [1, \infty)$ the following is a p-norm on $x \in \mathbb{C}^n$:*

$$\|x\|_p \doteq \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}.$$

taking the limit as $p \to \infty$, we can extend this to $p \in [1, \infty]$ with

$$\|x\|_\infty \doteq \max\{|x_i|\}_i.$$

Example 13 (Operator norms). *Fix your favorite $p \in [1, \infty]$ and use it to define the operator p-norm for matrices*

$$\|A\|_p \doteq \sup \left\{ \|Ax\|_p \ : \ \|x\|_p = 1 \right\}$$

Think of this as the most a ball of p-norm 1 can be stretched in any direction by applying A to it. In the special case where $p = 2$, this maximum stretch is just the furthest stretch in Euclidean space. In this case the maximum stretch is just the largest singular value of A. Therefore $\|A\|_2 = \sigma_1$.

Example 14 (Frobenius norm). *The Frobenius norm is norm norm we get when we consider A as a big vector and apply the usual 2-norm to it:*

$$\|A\|_F \doteq \sqrt{\sum_{i,j} |A_{ij}|^2} = \sqrt{\operatorname{Tr}(A^* A)} = \sqrt{\sum_i \sigma_i^2} = \|\vec{\sigma}\|_2$$

Example 15 (Schatten norms). *The Frobenius norm is actually just a special case of a family of norms. For any $p \in [1, \infty]$, the p-*Schatten norm *of a matrix A with singular values $\vec{\sigma}$ is given by $\|\vec{\sigma}\|_p$.*

Remark 16 (Unitary invariance of Schatten norms). A useful property of Schatten norms is that they are invariant under unitary transformations as the singular values are unaffected. One way to see this is through the SVD.

Example 17 (Trace/Nuclear norm). *The trace norm of a matrix A is given by*

$$\|A\|_{\mathrm{Tr}} \doteq \operatorname{Tr}\left(\sqrt{A^* A}\right) = \sum_i \sigma_i = \|\vec{\sigma}\|_1 .$$

4.4 Connecting trace norms and 2 norms

We know that given two density operators $\rho_0\, \rho_1$, we could compute the statistical distance between probability distributions of their observations: $d_{\mathrm{stat}}(p_0,\, p_1)$. We saw previously this is bounded above by the half trace norm of the density operators' difference:

$$d_{\mathrm{stat}}(p_0,\, p_1) \leq \frac{1}{2} \|\rho_0 - \rho_1\|_{\mathrm{Tr}} .$$

Let's dive a little deeper now. Suppose we have initial state ρ, and then we apply either of two unitary operators U_b for $b \in \{0, 1\}$. Then if $\rho_b = U_b \rho U_b^*$, the result of a unitary operator acting on ρ, how different can the two results be?

Lemma 18. *Given a density operator ρ, let $\rho_b = U_b \rho U_b^*$ for $b \in \{0, 1\}$. Then*

$$\|\rho_0 - \rho_1\|_{\mathrm{Tr}} \leq 2 \|(U_0 - U_1)\, \rho\|_{\mathrm{Tr}} .$$

Proof (Lemma 18). This is a straightforward computation.

$$\begin{aligned}
\|\rho_0 - \rho_1\|_{\mathrm{Tr}} &= \|U_0 \rho U_0^* - U_2 \rho U_1^*\|_{\mathrm{Tr}} && \text{(expanding } \rho_b\text{)} \\
&= \|U_0 \rho\,(U_0^* - U_1^*) + (U_0 - U_1)\, \rho U_1^*\|_{\mathrm{Tr}} && \text{(adding zero)} \\
&\leq \|U_0 \rho\,(U_0^* - U_1^*)\|_{\mathrm{Tr}} + \|(U_0 - U_1)\, \rho U_1^*\|_{\mathrm{Tr}} && \text{(triangle inequality)} \\
&\leq \|\rho\,(U_0^* - U_1^*)\|_{\mathrm{Tr}} + \|(U_0 - U_1)\, \rho\|_{\mathrm{Tr}} && \text{(singular values unchanged by unitary)} \\
&= 2 \|(U_0 - U_1)\, \rho\|_{\mathrm{Tr}} && \text{(conjugation preserves norms),}
\end{aligned}$$

which proves the result. $\square$

We can take this result further though.

Theorem 19. *Given a density operator ρ, let $\rho_b = U_b \rho U_b^*$ for $b \in \{0,1\}$ and unitary U_b,*

$$\|\rho_0 - \rho_1\|_{\mathrm{Tr}} \leq 2 \|U_0 - U_1\|_2 .$$

Corollary 20. *Given a density operator ρ, let $\rho_b = U_b \rho U_b^*$ for $b \in \{0,1\}$, and let p_b be the induced probability distributions on observed states for ρ_b.*

$$d_{\mathrm{stat}}(p_0, p_1) \leq \frac{1}{2} \|\rho_0 - \rho_1\|_{\mathrm{Tr}} \leq \|(U_0 - U_1)\|_2 .$$

Proof (Theorem 19). First we prove the result for a matrix A acting on a pure state $\rho \doteq |\psi\rangle \langle\psi|$, which is a rank 1 matrix. Then the action would be given by $A|\psi\rangle$.

Note that when we apply ρ to $|\psi\rangle$, we get $(|\psi\rangle\langle\psi|)|\psi\rangle = |\psi\rangle(\langle\psi||\psi\rangle) = |\psi\rangle$. When we apply ρ to any $|\phi\rangle$ that is orthogonal to $|\psi\rangle$, we get $(|\psi\rangle\langle\psi|)|\phi\rangle = |\psi\rangle(\langle\psi||\phi\rangle) = |\psi\rangle \cdot 0$, which is the zero vector.

This implies there is an orthonormal basis containing $|\psi\rangle$ in which one basis vector, namely $|\psi\rangle$, is stretched by $A\rho$ by a factor of $\|A|\psi\rangle\|_2 = \|A\rho\|_2$, and the other vectors are shrunk by $A\rho$ to the zero vector. This means that $A\rho$ has one singular vector of value $\sigma_1 = \|A|\psi\rangle\|_2$, and the other ones are all zero. Thus, $\|A\rho\|_{\mathrm{Tr}} = \sum_i \sigma_i \geq \sigma_1 = \sigma_1 == \|A|\psi\rangle\|_2 \leq \|A\|_2$.

Now we extend by linearity to mixed states. Consider the mixed state $\rho = \sum_j p_j \rho_j$ where ρ_j are pure states. $\|A\rho\|_{\mathrm{Tr}} = \left\|\sum_j p_j A \rho_j\right\|_{\mathrm{Tr}} \leq \sum_j p_j \|A\rho_j\|_{\mathrm{Tr}} \leq \|A\|_{\mathrm{Tr}}$, Where the first inequality is the triangle inequality, and the second comes from the fact that $\sum_j p_j = 1$ combined with our result on individual pure states. □

5 Quantum gate precision

Suppose we have a unitary circuit with quantum gates Q_i for $i \in [t]$. Then any implementation of Q_i may have some imprecision and instead realize $\widetilde{Q_i}$ such that $\left\|\widetilde{Q_i} - Q_i\right\|_2 \leq \epsilon$ for some $\epsilon > 0$.

The overall effect of this imprecision at the ith gate is $U_i = Q_i \otimes I \approx \widetilde{U}_i = \widetilde{Q_i} \otimes I$. Then $\left\|\widetilde{U}_i - U_i\right\|_2 = \left\|\widetilde{Q_i} - Q_i\right\|_2$.

For the whole circuit $U = U_t U_{t-1} U_{t-2} \cdots U_2 U_1$ we have approximate implementation $\widetilde{U} = \widetilde{U_t}\widetilde{U_{t-1}}\widetilde{U_{t-2}} \cdots \widetilde{U_2}\widetilde{U_1}$. Given this, we should investigate how consecutive errors compound.

$$\begin{aligned}
\left\|\widetilde{U_{i+1}}\widetilde{U}_i - U_{i+1}U_i\right\|_2 &= \left\|\widetilde{U_{i+1}}\left(\widetilde{U}_i - U_i\right) + \left(\widetilde{U_{i+1}} - U_{i+1}\right)U_i\right\|_2 \\
&\leq \left\|\widetilde{U_{i+1}}\left(\widetilde{U}_i - U_i\right)\right\|_2 + \left\|\left(\widetilde{U_{i+1}} - U_{i+1}\right)U_i\right\|_2 \qquad \text{(triangle inequality)} \\
&\leq \left\|\left(\widetilde{U}_i - U_i\right)\right\|_2 + \left\|\left(\widetilde{U_{i+1}} - U_{i+1}\right)\right\|_2, \qquad (3)
\end{aligned}$$

where (3) follows from the fact unitary matrix multiplication doesn't affect singular values. We can conclude that consecutive errors purely add. This allows the bound

$$\left\|\widetilde{U} - U\right\|_2 \leq \sum_{i=1}^{t} \left\|\widetilde{U}_i - U_i\right\| = \sum_{i=1}^{t} \left\|\widetilde{Q_i} - Q_i\right\| \leq t\epsilon,$$

which proves the following theorem.

Theorem 21. *If each of t gates is implemented to within ϵ precision in 2-norm, Then*

$$d_{\text{stat}}(\widetilde{p}, p) \leq \frac{1}{2}\|\widetilde{\rho}-\rho\|_{\text{Tr}} \leq \left\|\widetilde{U}-U\right\|_2 \leq t\epsilon.$$

Remark 22. Contrast this with general series of matrix multiplications where instead of $t\epsilon$, the error bound would be on the order of the product of 2-norms of the matrices times $t\epsilon$.

Note that when solving systems of linear equations the error iscontrolled by the condition numbers of each matrix, with the condition number being the ratio of top singular value to smallest singular value. Again, unitary matrices save us there since they have the smallest possible condition number of 1, because all of their singular values are equal to 1 in absolute value.