Quantum Computing 2/15/2022

Lecture 7: Quantum distance

Instructor: Dieter van Melkebeek

In this lecture we investigate how errors propagate through a quantum circuit. We measure distances between probability distributions using that statistical distance, and distances between density operators and unitaries using various matrix norms. We characterize common norms after a review of the singular value decomposition of a matrix, and then establish bounds on the error propagation.

1 From states to output distributions

Suppose we have implemented some quantum system in the real world. Due to noise or errors, theoretically identical operators may be different though "close". For which sense of "closeness" of density operators guarantee closeness of the resulting output distributions?

Let us first recall the notion of closeness of probability distributions based on the statistical distance.

Definition 1 (Statistical distance). *The* statistical distance *between two distributions p_0, p_1 is given by*

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

We investigate the statistical distance between distributions arising from density operators, and connect it back to a norm on the density operators' difference.

Consider two density operators ρ_b for $b \in \{0,1\}$, and the probability distributions p_b they induce on the possible outcomes after a full measurement. First, recall that $p_b(s) = \langle s|\,\rho_b\,|s\rangle$. Then consider the deviation $\sigma \doteq \rho_0 - \rho_1$, which is a Hermitian matrix and therefore has an orthonormal basis of eigenstates $\{|\psi_i\rangle\}_i$. We can express σ in terms of its eigenstates as follows: $\sigma = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$. 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:

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

The sum $\sum_i |\lambda_i|$ of the absolute values of the eigenvalues of σ is known as the trace norm of σ, denoted $\|\sigma\|_{\mathrm{Tr}}$. We will explain the terminology and connect with other norms after a review of the singular value decomposition, but let us first state the result we obtained.

Fact 1. *Let p_0, p_1 be the probability distributions on the outcomes obtaining by full measurements of the quantum states ρ_0, ρ_1, respectively. Then*

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

where ρ_1 and ρ_2 are viewed as density matrices.

2 Singular value decomposition

We describe the singular value decomposition (SVD) of complex matrices, as well as some of the terminology used with SVD.

Theorem 2 (Singular Value Decomposition (SVD)). *For any matrix $A \in \mathbb{C}^{M\times N}$ there exist unitary matrices $U \in \mathbb{C}^{M\times M}, V \in \mathbb{C}^{N\times N}$ as well as a diagonal matrix $\Sigma \in \mathbb{R}^{M\times N}_{\geq 0}$ such that*

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

We use the following terminology:

- Σ is given by $\mathrm{diag}(\vec{\sigma}) = \mathrm{diag}(\sigma_1, \sigma_2, \ldots)$ 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 the *top* k *singular values*. Note that the singular values are the square roots of the eigenvalues of A^*A.
- $U_{*,j}$ are the *column singular vectors*, and are also called *left singular vectors*.
- $V_{*,j}$ are the *row singular vectors*, and are also called *right singular vectors*.

Note that if A is Hermitian, then:

- The singular values are absolute values of the eigenvalues of A.
- The left and right singular vectors coincide up to scalar multiplication.

3 Vector and matrix norms

Recall the requirements for a norm:

Definition 2 (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 $v = 0$.

We make use of the following norms for vectors:

Definition 3 (Vector p-norms). *For $p \in [1, \infty)$ the following is the p-norm of $x \in \mathbb{C}^n$:*

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

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

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

Given a vector norm $\|\cdot\|$, we can generically define an induced matrix norm $\|\cdot\|$, known as the operator norm:

$$\|A\| \doteq \max \{\|Ax\| \ : \ \|x\| = 1\} .$$

Think of $\|A\|$ as the most a ball of norm 1 can be stretched in any direction by applying A to it. In particular, we use $\|A\|_p$ for $p \in [1, \infty]$ to denote the operator norm induced by $\|\cdot\|_p$ as the vector norm.

Definition 4 (Operator norm). *For a matrix A and any $p \in [1, \infty]$,*

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

In the special where $p = 2$, the maximum stretch is just the furthest stretch in Euclidean space, which is given by the largest singular value of A, so $\|A\|_2 = \sigma_1$. This norm is often referred to a the spectral norm.

Another commonly used matrix norm is what we get when we view the matrix A as a big vector and apply the 2-norm to this vector:

Definition 5 (Frobenius norm). *The Frobenius norm of a matrix A is:*

$$\|A\|_F \doteq \sqrt{\sum_{i,j} |A_{ij}|^2}.$$

Note that

$$\sqrt{\sum_{i,j} |A_{ij}|^2} = \sqrt{\mathrm{Tr}\left(A^* A\right)} = \sqrt{\sum_i \sigma_i^2} = \|\vec{\sigma}\|_2 .$$

The Frobenius norm is actually just a special case of a family of norms known as Schatten norms. For any $p \in [1, \infty]$, the *p-Schatten norm* of a matrix A with singular values $\vec{\sigma}$ is given by $\|\vec{\sigma}\|_p$. Another special case of the Schatten norms that is often used and has its own name, corresponds to the case $p = 1$.

Definition 6 (Trace/Nuclear norm). *The trace norm (also known as nuclear norm) of a matrix A is given by*

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

A useful property of Schatten norms is that they are invariant under unitary transformations as the singular values are unaffected by such transformations. One way to see this is through the SVD.

4 From unary operators to states

Suppose that we want to apply a unitary operator to a given state, but we only manage to realize a close approximation and thus in reality apply a different unitary. How different can the resulting states be? We derive a good upper bound in this section.

Suppose we have initial state ρ, and then we apply either of two unitary operators U_0 or U_1, resulting in the states ρ_0 or ρ_1, respectively. We know from previous lectures that $\rho_b = U_b \rho U_b^*$ for $b \in \{0, 1\}$. We have the following:

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

In order to upper bound $\|\left(U_0 - U_1\right) \rho\|_{\mathrm{Tr}}$ as a function of the distance between U_0 and U_1, we analyze $\|A\rho\|_{\mathrm{Tr}}$ for a generic matrix A, and apply the result with $A = U_0 - U_1$.

First consider the case of a pure state $\rho \doteq |\psi\rangle \langle\psi|$, which is a rank 1 matrix, and analyze the effect of $A\rho$. Note that when we apply ρ to $|\psi\rangle$, we get $\left(|\psi\rangle \langle\psi|\right) |\psi\rangle = |\psi\rangle \left(\langle\psi| |\psi\rangle\right) = |\psi\rangle$. When we apply ρ to any $|\phi\rangle$ that is orthogonal to $|\psi\rangle$, we get $\left(|\psi\rangle \langle\psi|\right) |\phi\rangle = |\psi\rangle \left(\langle\psi| |\phi\rangle\right) = |\psi\rangle \cdot 0$, 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 $\sigma_1(A\rho) = \|A |\psi\rangle\|_2$, and the other ones are all zero. Thus, $\|A\rho\|_{\mathrm{Tr}} = \sum_i \sigma_i(A\rho) \geq \sigma_1(A\rho) = \|A |\psi\rangle\|_2 \leq \|A\|_2$.

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 one comes from the fact that $\sum_j p_j = 1$ combined with our result on individual pure states.

We conclude:

Fact 3. *Let ρ_0 and ρ_1 be the states obtained by applying the unitary matrices U_0 and U_1 to a common start state ρ, respectively. Then*

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

Combined with Fact 1, we obtain the following upper found on the statistical distance between the output distributions p_0 and p_1 obtained by measuring states ρ_0 and ρ_1, respectively:

$$d(p_0, p_1) \leq \|\left(U_0 - U_1\right)\|_2 .$$

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 $U_i = Q_i \otimes I \approx \widetilde{U}_i = \widetilde{Q_i} \otimes I$, and the operator definition of the matrix norm implies that $\left\|\widetilde{U}_i - U_i\right\|_2 = \left\|\widetilde{Q_i} - Q_i\right\|_2$.

For the whole circuit $U = U_t U_{t-1} U_{t-2} \cdots U_2 U_1$ we obtain the approximate implementation $\widetilde{U} = \widetilde{U_t}\widetilde{U_{t-1}}\widetilde{U_{t-2}} \cdots \widetilde{U_2}\widetilde{U_1}$. How do the consecutive errors compound?

$$\begin{aligned}
\left\|\widetilde{U_{i+1}}\widetilde{U}_i - U_{i+1}U_i\right\|_2 &= \left\|\widetilde{U_{i+1}}\left(\widetilde{U}_i - U_i\right) + \left(\widetilde{U_{i+1}} - U_{i+1}\right)U_i\right\|_2 && \text{(adding zero)} \quad (1)\\
&\leq \left\|\widetilde{U_{i+1}}\left(\widetilde{U}_i - U_i\right)\right\|_2 + \left\|\left(\widetilde{U_{i+1}} - U_{i+1}\right)U_i\right\|_2 && \text{(triangle inequality)}\\
&\leq \left\|\left(\widetilde{U}_i - U_i\right)\right\|_2 + \left\|\left(\widetilde{U_{i+1}} - U_{i+1}\right)\right\|_2, && (2)
\end{aligned}$$

where the last step follows from the fact unitary matrix multiplication doesn't affect singular values. We can conclude that consecutive error bounds purely add:

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

In combination with Fact 1 and Fact 3 we have shown that

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

In words:

Theorem 4. *If each of t gates is implemented to within ϵ precision in 2-norm, then the output distribution differs from the correct one by at most $t\epsilon$ in statistical distance.*

The reason why the errors only add up (as opposed to being blown up) is the fact that the transition matrices are unitary. For general transition matrices, the error bound would be on the order of the product of the differences in 2-norm times $t\epsilon$.

This relates to how the errors propagate when solving systems of linear equations. There the error is controlled by the condition number of the matrix, with the condition number being the ratio of top singular value to smallest singular value. Unitary matrices similarly save us there since they have the smallest possible condition number of 1; as they maintain inner products, all of their singular values are equal to 1 in absolute value.