Quantum Algorithms 2/14/2023

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 Solution to Exercise #5

We start with the solution to the homework exercise from last lecture: What rotation represents the action of the Hadamard gate H in the Bloch sphere?

Recall that the density operator ρ of every single-qubit mixed state can be written as

$$\rho = \frac{I + \vec{r}\vec{\sigma}}{2}, \tag{1}$$

where $\vec{r} \in \mathbb{R}^3$ is a vector of length $\|\vec{r}\|_1 \leq 1$. The vector $\vec{r}$ is unique and is called the Bloch vector of ρ.

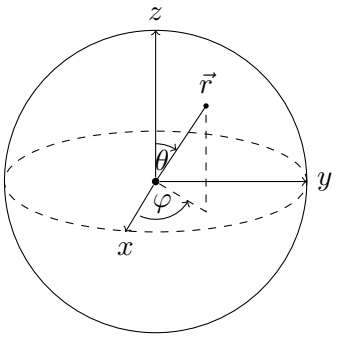

Figure 1: Bloch sphere

Further recall that the action of a unitary operator on ρ in the Bloch sphere is that of a rotation about an axis through the origin, and that

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

First solution. Since $H^1 = I$, the angle of rotation needs to be π, i.e., the rotation is a reflection through an axis. As H maps $|0\rangle$ on the positive z-axis to $|+\rangle$ on the positive z-axis, the axis of reflection needs to go through the bisector of the positive x-axis and the positive z-axis. These two features fully specify the rotation.

Second solution. After applying H, the density operator becomes $\rho' = H\rho H^* = H\rho H$. We plug in (1) and simplify using $HXH = Z$, $HYH = -Y$, and $ZHZ = X$. This shows that the new Bloch vector $\vec{r}'$ equals $(r_z, -r_y, r_x)$, where $\vec{r} \doteq (r_x, r_y, r_z)$. The transformation from $\vec{r}$ to $\vec{r}'$ is that of a reflection through the bisector of the positive x- and z-axis.

Third solution. We saw last lecture that every unitary U on a single qubit can be written in the Pauli basis as

$$U = e^{i\alpha}\left(\cos(\gamma)I + i\sin(\gamma)(\vec{u}\cdot\sigma)\right), \tag{2}$$

where all parameters α, γ, and $\vec{u}$ are real, and $\|\vec{u}\|_2 = 1$. With the restrictions that $\alpha \in [0, 2\pi)$ and $\gamma \in [0, \pi/2]$, the decomposition is unique. The action of U on the Bloch sphere is that of a rotation about $\vec{u}$ over 2γ. We can write H as $U = \frac{1}{\sqrt{2}}(X + Z)$, which shows that that the axis of rotation is $\vec{u} = \frac{1}{\sqrt{2}}(1, 0, 1)$. Moreover, since all Pauli matrices have vanishing traces, in the decomposition (2) we have that $\mathrm{Tr}(U) = e^{i\alpha}\cos(\gamma)\,\mathrm{Tr}(I) = 2e^{i\alpha}\cos(\gamma)$. Since $\mathrm{Tr}(H) = 0$, it follows that $\gamma = \pi/2$, and thus the rotation is over $2\gamma = \pi$.

2 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". Which notion of "closeness" of density operators guarantees 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*

$$d_{\mathrm{stat}}(p_0, p_1) \doteq \max\{|p_0(E) - p_1(E)| \ : \ E \subseteq \{0,1\}^n\}$$

In words, the statistical distance gives the maximum difference in probability that p_0 and p_1 assign to an event. The measure can also be written in terms of the difference in 1-norm when p_0 and p_1 are viewed as vectors of probabilities for all the individual elements of the underlying universe:

$$d_{\mathrm{stat}}(p_0, p_1) = \frac{1}{2}\sum_s |p_0(s) - p_1(s)| = \frac{1}{2}\|p_0 - p_1\|_1$$

The first equation follows because the maximum difference is obtained by taking up in E exactly the elements of the universe for which $p_0 \geq p_1$. In that case

$$\begin{aligned}
\|p_0 - p_1\|_1 \doteq \sum_s |p_0(s) - p_1(s)| &= \sum_{s\in E}(p_0(s) - p_1(s)) + \sum_{s\notin E}(p_1(s) - p_0(s)) \\
&= p_0(E) - p_1(E) + (p_1(\overline{E}) - p_0(\overline{E})) \\
&= p_0(E) - p_1(E) + (1 - p_1(E) - (1 - p_0(E))) \\
&= 2(p_0(E) - p_1(E)).
\end{aligned}$$

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:

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

For a Hermitian matrix like σ, the sum $\sum_i |\lambda_i|$ of the absolute values of the eigenvalues 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 obtained 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.

3 Singular value decomposition

For use in this lecture and much later in the course, we state and derive the singular value decomposition (SVD) of complex matrices. The derivation explains some of the terminology used with SVD, which we review after the proof.

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*

$$A = U\Sigma V^*. \tag{3}$$

Proof. Consider the matrix $B \doteq A^*A$, which has dimension $N \times N$. As B is Hermitian ($B^* = B$), it has an orthonormal basis of eigenvectors with real eigenvalues. Thus, there exist a unitary matrix $V \in \mathbb{C}^{N\times N}$ (consisting of normalized eigenvectors of B) and a diagonal matrix $\Lambda \in \mathbb{R}^{N\times N}$ (containing the eigenvalues of B) such that $BV = V\Lambda$. Moreover, B is positive semidefinite because $\langle\psi|B|\psi\rangle = \|A|\psi\rangle\|^2$ is a nonnegative real for every $|\psi\rangle$. Thus, the eigenvalues λ_i of B are nonnegative and can be written as the square of some other reals: $\lambda_i = \sigma_i^2$ for some $\sigma_i \in \mathbb{R}_{\geq 0}$.

There are r nonzero eigenvalues, where r equals the rank of A^*A and also the rank of A. Without loss of generality these are λ_i for $i \in [r]$.

Consider the matrix $W \doteq AV$, which has dimension $M \times N$. Note that $W^*W = (V^*A^*)(AV) = V^*(BV) = V^*V\Lambda = \Lambda$. This means that the columns of W are orthogonal vectors in $\mathbb{C}^M$. The j-th column, $W_{\cdot,j}$, has 2-norm $\|W_{\cdot,j}\|_2 = \sigma_j$. By normalizing the nonzero columns, setting $U_{\cdot,j} = \frac{1}{\sigma_j} W_{\cdot,j}$ for $jj \in [r]$, and extending with appropriate columns $U_{\cdot,j}$ for the remaining $j \in [M]$ to form a full orthonormal basis for $\mathbb{C}^M$, we obtain a unitary matrix $U \in \mathbb{C}^{M\times M}$ such that $W = U\Sigma$, where $\Sigma \in \mathbb{R}_{\geq 0}^{M\times N}$ is the diagonal matrix with diagonal elements $\sigma_1, \sigma_2 \ldots, \sigma_{\min(M,N)}$. As $W \doteq AV$, we have $AV = U\Sigma$, which is equivalent to (3). □

We make some observations about the uniqueness of the decomposition and introduce the following terminology.

- The values $\sigma_1, \ldots \sigma_r$ are called the *singular values* of A. They are the positive square roots of the nonzero eigenvalues of A^*A, and are uniquely determined. They are usually ordered from largest to smallest: $\sigma_1 \geq \sigma_2 \ldots, \geq \sigma_r > 0$. σ_1 is the largest factor by which the length of a vector gets stretched under the application of A. Similarly, σ_r is the smallest (nonzero) stretch factor.

- The columns V_{*j} of V are eigenvectors of A^*A. The ones corresponding to nonzero eigenvalues are referred to as right singular vectors (because V appears on the right in (3)) and also as row singular vectors (because they appear as rows of V^* in (3)). If all singular values are distinct, then the right/row singular vectors are unique up to a complex unit. Otherwise, there are more degrees of freedom. In any case, the subspace spanned by the right/row singular vectors corresponding to a particular singular value is always the same, irrespective of which right/row singular vectors are chosen.

- The columns U_{*j} of U are eigenvectors of AA^*. The ones corresponding to nonzero eigenvalues are referred to as left singular vectors (because U appears on the left in the decomposition (3)) and also as column singular vectors (because they appear as columns in (3)). Similar uniqueness considerations apply to the right/row singular vectors as to the left/column singular vectors.

- If A is Hermitian, then the singular values of A are the absolute values of the nonzero eigenvalues (which are real but can be negative). The i-th left and right singular vectors of A are also eigenvectors of A belonging to the same eigenvalue, and can be chosen to be the same except for a sign in case the eigenvalue is negative. The same applies for the columns of U and V corresponding to the eigenvalue 0 (if any). In particular, this applies to the setting of Fact 1, as the difference $\rho_1 - \rho_2$ of two density operators is Hermitian.

4 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* $v = 0$.

Note that (1) implies that $\|0\| = 0$, which in combination with (2) implies that $\|v\| = \frac{1}{2}(\|v\|+\|v\|) \geq \frac{1}{2}\|v - v\| = 0$. Another consequence of (1) and (2) is that that the unit ball, i.e., the vectors $v \in V$ with $\|v\| \leq 1$, needs to be convex.

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*

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

Absolute homogeneity and definiteness hold for every positive value of p. The triangle inequality fails for $p < 1$ because the unit ball is not convex. That the triangle inequality holds for $p \geq 1$ is known as Minkowski's inequality (closely related to Hölders inequality).

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 unit ball can be stretched in any direction by applying A to it. Note that any such matrix norm is submultiplicative: $\|AB\| \leq \|A\| \cdot \|B\|$.

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.

An arguably more elementary way to define matrix norms is to view a matrix A as a big vector and apply a vector norm to it. A commonly used matrix norm does this with the 2-norm as the vector norm:

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

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

Note that

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

This shows that the Frobenius norm is a special case of the family of so-called Schatten norms. For any $p \in [1, \infty]$, the *p-Schatten norm* of a matrix A with singular values $\vec{\sigma}$ is given by $\|\vec{\sigma}\|_p$. In fact, the spectral norm is also a Schatten norm, namely corresponding to $p = \infty$. 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. Like all operator norms, the Schatten norms are submultiplicative.

5 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:

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

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 $\|A |\psi\rangle\|_2$, 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, $\|A\rho\|_{\mathrm{Tr}} = \sum_i \sigma_i(A\rho) = \sigma_1(A\rho) = \|A |\psi\rangle\|_2 \leq \|A\|_2$.

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*

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

Moreover, if ρ corresponds to the pure state $|\psi\rangle$, then

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

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

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

and

$$d(p_0, p_1) \leq \|(U_0 - U_1)|\psi\rangle\|_2$$

in case the start state is the pure state $|\psi\rangle$.

6 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 $\tilde{Q}_i$ such that $\left\|\tilde{Q}_i - Q_i\right\|_2 \leq \epsilon$ for some $\epsilon > 0$. The effect of this imprecision at the ith gate on the full system is $U_i = Q_i \otimes I \approx \tilde{U}_i = \tilde{Q}_i \otimes I$

Exercise #6: Show that $\left\|\tilde{U}_i - U_i\right\|_2 = \left\|\tilde{Q}_i - Q_i\right\|_2$.

For the whole circuit $U = U_t U_{t-1} U_{t-2} \cdots U_2 U_1$ we obtain the approximate implementation $\tilde{U} = \tilde{U}_t \tilde{U}_{t-1} \tilde{U}_{t-2} \ldots \tilde{U}_2 \tilde{U}_1$. How do the consecutive errors compound? Defining the partial products $U^{(i)} \doteq U_i U_{i-1} \ldots U_1$ and $\tilde{U}^{(i)} \doteq \tilde{U}_i \ldots \tilde{U}_1$, we have:

$$\begin{aligned}
\|\tilde{U}^{(i)} - U^{(i)}\|_2 &= \|\tilde{U}_i \tilde{U}^{(i-1)} - U_i U^{(i-1)}\|_2 \\
&= \|\tilde{U}_i(\tilde{U}^{(i-1)} - U^{(i-1)}) + (\tilde{U}_i - U_i)U^{(i-1)}\|_2 && \text{(adding zero)} \\
&\leq \|\tilde{U}_i(\tilde{U}^{(i-1)} - U^{(i-1)})\|_2 + \|(\tilde{U}_i - U_i)U^{(i-1)}\|_2 && \text{(triangle inequality)} \\
&= \|\tilde{U}_i - U_i\|_2 + \|\tilde{U}^{(i-1)} - U^{(i-1)}\|_2,
\end{aligned}$$

where the last step can be argued by the fact that a unitary does not change the 2-norm of a vector (and using the operator definition of matrix norm), or that a unitary does not change the singular values (and using the expression for the matrix norm in terms of the singular values), or that operator matrix norms satisfy $\|AB\| \leq \|A\| \cdot \|B\|$ and that a unitary matrix has 2-norm one (for either of the previous reasons). We conclude that consecutive error bounds merely add:

$$\left\|\widetilde{U} - U\right\|_2 \leq \sum_{i=1}^{t} \left\|\widetilde{U}_i - U_i\right\|_2 = \sum_{i=1}^{t} \left\|\widetilde{Q_i} - Q_i\right\|_2 \leq t\epsilon. \tag{4}$$

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

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

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 is the fact that the transition matrices are unitary. For general transition matrices, the errors in individual steps can blow up. For general matrices U_i and $\tilde{U}_i$ satisfying $\|\tilde{U}_i - U_i\| \leq \epsilon_i$, a similar derivation as above (using the additional step of writing $\tilde{U}_i(\tilde{U}^{(i-1)} - U^{(i-1)})$ as $(\tilde{U}_i - U_i)(\tilde{U}^{(i-1)} - U^{(i-1)}) + U_i(\tilde{U}^{(i-1)} - U^{(i-1)})$ and applying the triangle inequality one more time) shows the following bound:

$$\|\tilde{U} - U\|_2 \leq \prod_{i=1}^{t}(\epsilon_i + \|U_i\|_2) - \prod_{i=1}^{t}\|U_i\|_2. \tag{5}$$

Note that for unitary matrices $\|U_i\|_2 = 1$, in which case the general bound (5) yields a somewhat weaker bound that (4), namely $\|\tilde{U} - U\|_2 \leq (\epsilon + 1)^t - 1$.

This relates to how the errors propagate when solving systems of linear equations, which we will discuss later in the course. 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.