Lecture 5: Quantum State

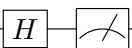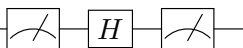Last lecture, we finalized our model of quantum computing, allowing intermediate measurements in addition to local unitary transformations. One complication is that the state of the system is no longer fully captured by a superposition. In order to obtain a full description, we introduced mixed states, which are probability distributions over superpositions. Today, we discuss a better way to represent mixed states, namely density matrices.

1 Solution Exercise #3

You were asked to give an example of two quantum circuits on the same number of qubits that behave the same on all basis states but not on all pure states. There are many solutions. All of them need to use at least one measurement operations. We present two simple ones: one involving one intermediate measurement, and one with only final measurements.

1.1 Solution 1

Consider (a) ─$\boxed{H}$─$\boxed{\measuredangle}$─ vs (b) ─$\boxed{\measuredangle}$─$\boxed{H}$─$\boxed{\measuredangle}$─

In basis states, the first measurement has no effect, so the output for (b) is the same as (a). Starting with basis states $|0\rangle$ and $|1\rangle$, the output distribution is that of a fair coin toss.

Given the pure state $|+\rangle$, the output of (a) is 0, but the output of (b) is a fair coin toss.

1.2 Solution 2

Consider the Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ vs $XH = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

Appending a measurement to the end of each, we get (a) ─$\boxed{H}$─$\boxed{\measuredangle}$─ and (b) ─$\boxed{XH}$─$\boxed{\measuredangle}$─

Given the basis states $|0\rangle$ and $|1\rangle$, both output distributions represent a fair coin toss.

Given the pure state $|+\rangle$ we get:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} |+\rangle = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1-1 \\ 1+-1 \end{bmatrix} = |0\rangle$$

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} |+\rangle = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1+1 \\ 1--1 \end{bmatrix} = |1\rangle$$

Whereas the output of (a) is always 0, the output of (b) is always 1.

1.3 Conclusion

The main takeaway from this exercise is that just describing how a circuit behaves on the basis states is no longer enough. As we argued last lecture, it is enough for unitary circuits, i.e., quantum circuits without measurements.
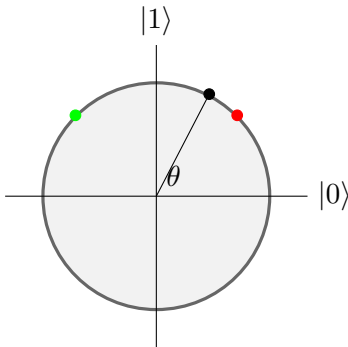
Figure 1: Single qubit system

2 Quantum State

What does "state" mean? State captures all information about a system to predict how a system will behave when future operations are applied in the future. State descriptions must be *sufficient*, meaning that two systems with the same description should be indistinguishable. The descriptions ideally are also *necessary*, meaning there is no redundant information. If two systems have different descriptions, then there is some sequence of operations where the systems behave differently.

2.1 Pure States

Recall that a pure state is given by $|\psi\rangle = \sum_s \alpha_s |s\rangle$ with $\sum_s |\alpha_s|^2 = 1$ representing a superposition of basis states. A measurement in the standard basis is enough to distinguish any two superpositions $|\psi\rangle$ and $|\psi'\rangle$ such that $|\alpha_s| \neq |\alpha'_s|$ for at least one basis states s. This is because the measurement in state $|\psi\rangle$ yields s with probability $|\alpha_s|^2$, and in state $|\psi'\rangle$ with the different probability $|\alpha'_s|^2$. For example, the states corresponding to the red and blue dots in Figure 1 can be distinguished this way.

More generally, we can distinguish $|\psi\rangle$ and $|\psi'\rangle$ as long as they are not parallel, i.e., they differ by more than a global phase. This is because there exists a unitary transformation that brings $|\psi\rangle$ to the basis state $|0^m\rangle$. As the transformation needs to preserve angles, it brings $|\psi'\rangle$ to a pure state that does not have all of its weight on the basis state $|0^m\rangle$. Thus, after the transformation we're in the above situation, in which a measurement can distinguish the two states. An example of this situation are the states corresponding to the red and green dots in Figure 1.

Two superpositions that only differ by a global phase cannot be distinguished physically. This is because measurements of such states yield the same probability distributions, and unitary operations preserve the property of two states only differing by a global phase.

In conclusion, the only redundancy in the representation of pure states as superpositions is the global phase. The phase of one of the basis states with nonzero weight can be set arbitrarily, say to 0 (meaning that the coefficient is real).

2.2 Mixed State

A mixed state can be described as a set of pairs $\{(p_i, |\psi_i\rangle)\}_i$ for a probability distribution p and pure states $|\psi_i\rangle$. Suppose we are in a mixed state $\{(\frac{1}{2}, |\alpha\rangle), (\frac{1}{2}, |\beta\rangle)\}$ consisting of an equal mixture

of $|\alpha\rangle$ and $|\beta\rangle$, where $|\alpha\rangle$ and $|\beta\rangle$ are two orthogonal 1-qubit states with real coefficients. For example, we could have $|\alpha\rangle = |0\rangle$ and $|\beta\rangle = |1\rangle$, or $|\alpha\rangle = |+\rangle$ and $|\beta\rangle = |-\rangle$.

What is the probability that we obtain 0 when we measure the qubit? If we are in the pure state $|\alpha\rangle$ given by the black dot in Figure 1, the probability that we measure 0 is the length square of the projection of $|\alpha\rangle$ onto $|0\rangle$, which is $\cos^2(\theta)$. By the same token, since the state $|\beta\rangle$ is orthogonal to $|\alpha\rangle$, the probability that we measure 0 in that state equals $\cos^2(\theta \pm \pi/2) = \sin^2(\theta)$. Since we are in an equal mixture of $|\alpha\rangle$ and $|\beta\rangle$, the overall probability of measuring 0 is

$$\frac{1}{2}\cos^2(\theta) + \frac{1}{2}\sin^2(\theta) = \frac{1}{2}.$$

Thus, we obtain 0 and 1 with equal probability. Note that this conclusion is independent of θ. In particular, we obtain the same distribution whether we are in the mixed state $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ or in the mixed state $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$, namely a uniform random bit

If we apply some unitary operations before the measurement, the two mixtures we have before the measurement are also both of the same type, i.e., equal mixtures of two orthogonal states with real coefficients, up to possibly changing the phase of the basis vector $|1\rangle$. This is because unitary operations preserve angles, and we can always change the global phase so that the coefficient of $|0\rangle$ is real. As a result, the measurement will again give a uniform random bit. Thus, after the measurement we are in a mixture of 50% $|0\rangle$ and 50% $|1\rangle$, irrespective of the initial angle θ. As soon as we are in the equal mixture of $|0\rangle$ and $|1\rangle$, there is no chance of distinguishing the two states afterwards. Thus, even though the two mixtures $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ and $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$ are different when represented as probability distributions over pure states, they cannot be physically distinguished. Therefore, the way the way of describing mixed states as probability distributions over pure states contains superfluous information. The alternate description using density operators addresses that shortcoming.

3 Density Operators

3.1 Pure States

Consider the pure state $|\psi\rangle$ and measure it in an orthonormal basis $|\phi_j\rangle$. We want to calculate the probability that we see one specific basis state.

$$\Pr[\text{measure } |\phi_j\rangle] = |\langle\psi|\phi_j\rangle|^2 = \langle\phi_j|\psi\rangle \cdot \langle\psi|\phi_j\rangle = \langle\phi_j| \cdot \rho \cdot |\phi_j\rangle \tag{1}$$

where $\rho \doteq |\psi\rangle\langle\psi|$. We can justify the steps in Equation 1 as follows.

1. $\Pr[\text{measure } |\phi_j\rangle] = |\langle\psi|\phi_j\rangle|^2$: The probability of one particular basis state equals the length of the projection of the state on the basis state, squared.

2. $|\langle\psi|\phi_j\rangle|^2 = (\langle\psi|\phi_j\rangle)^* \cdot \langle\psi|\phi_j\rangle = \langle\phi_j|\psi\rangle \cdot \langle\psi|\phi_j\rangle$, where we use the fact that $(AB)^* = B^* A^*$ and $|\psi\rangle^* = \langle\psi|$.

3. $\langle\phi_j|\psi\rangle \cdot \langle\psi|\phi_j\rangle = \langle\phi_j| \cdot (|\psi\rangle\langle\psi|) \cdot |\phi_j\rangle \doteq \langle\phi_j| \cdot \rho \cdot |\phi_j\rangle$ using associativity.

For example, if we are in $|\psi\rangle = |0\rangle$,

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

3

Note that the matrix ρ is the outer product of $|\psi\rangle$ with itself. It has rank one as all of the columns are equal to $|\psi\rangle$ up to some scalar. This will be the **density matrix (operator)** representation of the pure state $|\psi\rangle$. By Equation 1, when we are in the pure state and we know ρ, we know what the probability distribution is of the measurement outcome in any basis $|\phi_i\rangle$.

3.2 Mixed States

For mixed state $\{(p_i, |\psi_i\rangle)\}_i$:

$$\Pr[\text{measure } |\phi_j\rangle] = \sum_i p_i \langle\phi_j| \cdot \rho_i \cdot |\phi_j\rangle = \langle\phi_j| \cdot \left(\sum_i p_i \rho_i\right) \cdot |\phi_j\rangle = \langle\phi_j| \cdot \rho \cdot |\phi_j\rangle$$

where $\rho \doteq \sum_i p_i \cdot \rho_i$. The first step follows because the probability of observing a specific basis state $|\phi_j\rangle$ is the weighted sum of the probabilities of observing $|\phi_j\rangle$ over all components $|\psi_i\rangle$ of the mixture, where the weight of the component $|\psi_i\rangle$ is the one from the mixture; for a given component $|\psi_i\rangle$, the probability of observing $|\phi_j\rangle$ is given by Equation 1 for $|\psi\rangle = |psi_i\rangle$. The second step follows by linearity.

Thus, all information about the probability distribution of measuring a mixed state in any orthonormal basis $|\phi_j\rangle$ is contained in the matrix ρ, called the density operator of the mixed state.

Definition 1. *The density operator of the mixed state $\{(p_i, |\psi_i\rangle)\}_i$ is $\rho \doteq \sum_i p_i |\psi_i\rangle \langle\psi_i|$.*

4 Evolution of Density Operator

If a density operator of the system is given at the beginning of a computation, and the operations are all unitary or measurements, we have enough information to calculate the density operator after the operation. Below, we show how this is true for both unitary operations and partial measurements.

4.1 Unitary Operation

Assume we are in a pure state $|\psi\rangle$ and we apply the unitary operation U: $|\psi'\rangle = U|\psi\rangle$. The density operator after the operation would become:

$$\rho' \doteq |\psi'\rangle \langle\psi'| = (U||\psi\rangle)(U||\psi\rangle)^* = U|\psi\rangle \langle\psi| U^*,$$

where $(U||\psi\rangle)^*$ is the complex conjugate transpose. Note that *ket* turns into a *bra*.

We have that

$$\rho' = U\rho U^* \tag{2}$$

Thus, the new density operator is comprised of the old one with an operation performed upon it.

If we were to start with a mixed state, Equation 2 would remain true by linearity due to a mixed state being a convex combination of the pure states.

4.2 Partial Measurement

A partial measurement can be represented as projections P_s onto subspaces with a subspace for each available outcome s. For example, for a single qubit, there are two possible subspaces due to the possible outcomes $|0\rangle$ and $|1\rangle$. The effect of the measurement on a pure state is that we cancel

all components that are not consistent with the outcome of the measurement, and then renormalize. The probability of a particular outcome s is the squared length of that projection:

$$\Pr[\text{measure } s] = ||P_s \,|\psi\rangle\, ||_2^2.$$

Once we measure s, the new state becomes

$$\frac{P_s \,|\psi\rangle}{||P_s \,|\psi\rangle\, ||_2}.$$

Before we know the outcome of the measurement, the system is in a mixed state with one component for each possible outcome s.

$$\rho' = \sum_s ||P_s \,|\psi\rangle\, ||_2^2 \cdot \frac{P_s \,|\psi\rangle}{||P_s \,|\psi\rangle\, ||_2} \frac{\langle\psi|\, P_s^*}{||P_s \,|\psi\rangle\, ||_2} = \sum_s P_s \,|\psi\rangle \langle\psi|\, P_s^* = \sum_s P_s \,|\psi\rangle \langle\psi|\, P_s,$$

where we used the fact that $P_s^* = P_s$ for an orthogonal projection P_s. We have that

$$\rho' = \sum_s P_s \rho P_s. \tag{3}$$

Once again, the new density operator is comprised of the old one with an operation performed upon it. The formula extends to mixed states by linearity.

Linearity. Note that for both types of operations, unitaries and partial measurements, the new density operator is obtained from the old one through a linear transformation, namely (2) and (3), respectively. In contrast, in the superposition formalism, unitaries act linearly but partial measurements generally do not due to the required renormalization. Apart from the uniqueness of representation for mixed states that we'll argue later in this lecture, the linearity of all operations is another advantage of the density operator formalism.

5   Example Density Operators

5.1   Basis States

$$|0\rangle : \rho_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$|1\rangle : \rho_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

5.2   Pure States

$$|+\rangle \, (= H \,|0\rangle) : \rho_+ = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = H\rho_0 H^*$$

$$|-\rangle \, (= H \,|1\rangle) : \rho_+ = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = H\rho_1 H^*$$

5.3 Mixed States

Consider $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ (e.g., obtained by measuring $|+\rangle$)

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = P_0 \rho_+ P_0 + P_1 \rho_+ P_1$$

Consider $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$. Remember, this mixed state and the previous one are indistinguishable as argued in Section 2.2.

$$\rho = \frac{1}{2}\rho_+ + \frac{1}{2}\rho_- = \frac{1}{2} \cdot \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{1}{2} \cdot \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This is the same matrix we reached in the first mixed state example, which witnesses the fact that the these two mixed states cannot be distinguished.

6 Trace of a Matrix

What follows is a review of the trace of a matrix, a function that calculates the sum of the diagonal elements. Equations 4, 5, and 6 represent properties of the trace

Definition 2. $\mathrm{Tr}(A) = \sum_i A_{ii}$

The trace function has the following elementary properties.

$$\mathrm{Tr}(A^T) = \mathrm{Tr}(A) \tag{4}$$

$$\mathrm{Tr}(AB) = \sum_{i,j} A_{ij} B_{ij} = \mathrm{Tr}(BA) \tag{5}$$

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB) \tag{6}$$

(4) and (5) follow directly from the definition. (6) follows from (5) and associativity of matrix multiplication; it represents a cyclic invariance property of the trace function. The trace function can also be expressed in terms of the eigenvalues of the matrix.

Theorem 1. $\mathrm{Tr}(A) = \sum_i \lambda_i$, *where each eigenvalue λ_i of A is counted according to its algebraic multiplicity.*

The algebraic multiplicity is found by taking the characteristic equation $\det(A - \lambda I) = 0$. The eigenvalues λ are the roots of the characteristic equation, and the algebraic multiplicity of λ is the number of times λ occurs as a root.

Proof (for the case where matrix A has a full basis of eigenvectors). Consider the matrix V whose columns form a basis of eigenvectors of A. We can write $AV = V\Lambda$, where Λ is a diagonal matrix consisting of the eigenvalues of A. We have that

$$\mathrm{Tr}(A) = \mathrm{Tr}(V\Lambda V^{-1}) = \mathrm{Tr}(V^{-1}V\Lambda) = \mathrm{Tr}(\Lambda)$$

The step: $\mathrm{Tr}(V\Lambda V^{-1}) = \mathrm{Tr}(V^{-1}V\Lambda)$ is a result of the cyclicity property (6). $\qquad\square$

7 Properties of Density Operators

The following Properties are true for density operators $\rho \doteq \sum_i p_i |\psi_i\rangle \langle\psi_i|$:

- $\mathrm{Tr}(\rho) = 1$. For pure states this follows from the cyclicity property:

$$\mathrm{Tr}(\rho) = \mathrm{Tr}(|\psi\rangle \langle\psi|) = \mathrm{Tr}(\langle\psi|\psi\rangle) = \mathrm{Tr}(1) = 1.$$

  The result for mixed states follows by linearity and the fact that probabilities add up to 1.

- ρ is Hermitian: $\rho^* = \rho$ This follows right away in case of pure states. For mixed states, the property follows because the p_i's are real numbers.

- ρ is positive semidefinite. That is, for each $|\phi\rangle$: $\langle\phi| \cdot \rho \cdot |\phi\rangle \geq 0$. This can be shown by:

$$\langle\phi| \cdot \rho \cdot |\phi\rangle = \langle\phi| \cdot \sum_i p_i |\psi_i\rangle \langle\psi_i| \cdot |\phi\rangle = \sum_i p_i \langle\phi| |\psi_i\rangle \langle\psi_i| |\phi\rangle = \sum_i p_i |\langle\psi|\phi\rangle|^2 \qquad (7)$$

  $\sum_i p_i |\langle\psi|\phi\rangle|^2 \geq 0$ because all p_i's are nonnegative, and the absolute value squared of $\langle\psi|\phi\rangle$ is also nonnegative.

In fact, the above three properties characterize the set of density operators.

Theorem 2. *A is a density operator iff A is a Hermitian positive semidefinite matrix with* $\mathrm{Tr}(A) = 1$

Proof. We already proved the direction $\Rightarrow$. We now prove $\Leftarrow$.

Since A is Hermitian, it has an orthonormal basis of eigenvectors $|\psi_i\rangle$. We have that $A |\psi_i\rangle = \lambda_i |\psi_i\rangle$ for some $\lambda_i \in \mathbb{C}$. We can compactly write these equations as

$$A = \sum_i \lambda_i \cdot |\psi_i\rangle \langle\psi_i| . \qquad (8)$$

Since A is positive semidefinite, for each i, the quantity $\langle\psi_i| \cdot A \cdot |\psi_i\rangle$ needs to be a nonnegative real. As $\langle\psi_i| \cdot A \cdot |\psi_i\rangle = \lambda_i \langle\psi_i|\psi_i\rangle = \lambda_i$, this means that the eigenvalues λ_i are nonnegative reals. As $\mathrm{Tr}(A) = 1$, they add up to 1. Thus the λ_i's can be thought of as probabilities, and the right-hand side of (8) as the density matrix of the mixed state $\{(\lambda_i, |\psi_i\rangle)\}_i$. $\qquad \square$

Theorem 3. *Mixed states are distinguishable iff density operators differ.*

In other words, there exists some quantum circuit that distinguishes between two states with positive probability iff the states have different density operators. Equivalently, we have the contrapositive: two states always behave the same iff they have the same density operator. Due to the equivalence between distinct states and density operators, we may use the terms interchangeably.

Proof. Assume that two density operators are the same. We showed in the previous sections 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 as a function of the density operator for the current state. Thus, any quantum process operating on two states with the same initial density operators evolves the same for both of the states, results

in the same final density operator for the two final states, and, the same probability for the final measurement. Thus, we cannot distinguish the two states.

Now, consider two distinct density operators ρ_1, ρ_2. Let $\sigma = \rho_1 - \rho_2$. σ is Hermitian because differences 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 measure ρ_1, ρ_2 in the basis $\{|\psi_i\rangle\}_i$.

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

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

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 nonzero difference in probability that we measure state $|\psi_i\rangle$ for the densities ρ_1, ρ_2. $\qquad\square$

8 Exercise #4

In a next lecture we will discuss the use of density operators in systems where some of the qubits are under control of some party (Alice) and the rest of some other party (Bob). In preparation of that lecture, consider an EPR pair (Einstein-Podolsky-Rosen), i.e., a two-component system in that state

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Alice holds the first component, Bob holds the second, and they are separated.

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. Can you describe Alice's view of her qubit by a density operator at each point in time (a), (b), and (c)? Hint for (a): Deferred measurement exercise.