

Quantum States and Partial Trace

Contents

6.1 Quantum States	50
6.1.1 Pure States	50
6.1.2 Mixed States	51
6.1.3 Density Operators (Matrices)	51
6.1.4 Properties of Pure and Mixed States	53
6.2 Trace and Partial Trace	55
6.2.1 Expectation Values as Traces of Density Operators	57
6.3 In Closing	58

The notion of a quantum state is critical to both quantum information processing (quantum computing) and quantum information theory. In previous seminars, we’ve only been exposed to one particular kind of quantum state—a *pure state*. This is where we have “complete knowledge” of what the state is. (Even though we may be “surprised” by measurement outcomes, we know for sure what the underlying state is.) In contrast, a *mixed state* is a statistical mixture of pure states.

We’ll formally define these concepts throughout the seminar, which will lead us to the important notion of a *density operator* (matrix). Finally, we’ll mention some things these definitions enable us to talk about, including entanglement, distance between quantum states, fidelity, and (partial) trace. These tools are indispensable in quantum computing, in particular the analysis and design of quantum algorithms and study of noise in quantum systems.

Some linear algebra is assumed in this seminar. We will formally cover partial trace (and for completeness trace as well), but knowing the trace of an operator is assumed in our coverage of density operators. Familiarity with hermitian operators and eigenvalues/eigenvectors is assumed.

6.1 Quantum States

We first formally define the type of quantum states we have seen, for example, in the quantum teleportation algorithm, phase estimation algorithm, and Grover’s algorithm. This definition should be familiar.

6.1.1 Pure States

Pure quantum states can be thought of as definite quantum states. That is, states in which we have complete knowledge (statistically speaking) of the underlying wavefunction.

Definition 6.1 (Pure state). A **pure state** is a quantum state whose wavefunction is known with certainty.

Note that complete knowledge of the underlying wavefunction does not give complete knowledge of measurement statistics. Quantum measurements are always statistical (sometimes called *noisy*) due to the underlying postulates of quantum mechanics which we have previously seen. While this measurement noise or uncertainty is always present, we can introduce a different type of uncertainty by allowing for a statistical distribution of wavefunctions. This is the concept of mixed states, formally defined below.

6.1.2 Mixed States

We begin with the formal definition of a mixed state.

Definition 6.2 (Mixed state). A **mixed state** is a statistical distribution of pure states. We can write a mixed state as

$$\{(p_i, |\psi_i\rangle)\}_{i=1}^n \quad (6.1)$$

where the p_i 's form a valid probability distribution ($\sum_{i \in [n]} p_i = 1$) and each $|\psi_i\rangle$ is a valid (pure) state.

Example 1: Example of a Mixed State.

Suppose you are in the lab, and your lab mate tells you she'll prepare the $|0\rangle$ state for you with probability $2/3$ and the $|1\rangle$ state with probability $1/3$. This then defines a *mixed state* which we may write as

$$\{(2/3, |0\rangle), (1/3, |1\rangle)\}. \quad (6.2)$$

Note that this is completely different from the state

$$(2/3)|0\rangle + (1/3)|1\rangle \quad (6.3)$$

(which is not a valid quantum state because it is not normalized). Even if it was normalized (what is the normalization factor?), the mixed state (6.2) is still completely different from (6.3). Why is this?

Equation 6.2 refers to the situation in which we have the state $|0\rangle$ with probability $2/3$, and we have the state $|1\rangle$ with probability $1/3$. There is no situation in which we ever have a superposition of $|0\rangle$ and $|1\rangle$. In contrast, equation (6.3) refers to the situation in which we *definitely have* the superposition state $(2/3N)|0\rangle + (1/3N)|1\rangle$ where N is the normalization factor.

This may seem like a subtle point, but it is crucial to understand. Shortly, we'll introduce a mathematical tool that makes it easier to see the difference between pure and mixed states. However, before moving on, make sure you understand the difference at the conceptual level.

6.1.3 Density Operators (Matrices)

Density operators provide a convenient mathematical representation of mixed states. In the above example, we saw that we vehemently do *not* want to write a mixed state as $\sum_i p_i |\psi_i\rangle$, as this defines an (unnormalized) pure state. The definition of a mixed state looks similar to this but with one important difference—we take the outer product of the states $|\psi_i\rangle$ in the weighted sum.

Before giving a definition, we make a note on the terminology *density operator* vs. *density matrix*. The distinction is the same as elsewhere in linear algebra: a matrix is a representation of a (linear) operator *in a particular basis*. This difference between matrices and operators is subtle but important. In some sense, an operator is “more general” than a matrix, for an operator can be expressed in any basis, while a matrix is an operator expressed in some particular basis. Important notes aside, we'll stick to the standard convention that the terms *density operator* and *density matrix* are interchangeable throughout this seminar and future ones.

Definition 6.3 (Density Operator). A **density operator** is a hermitian operator with no negative eigenvalues^a and unit trace. That is, ρ is a density operator if and only if

$$\rho^\dagger = \rho \quad \text{and} \quad \rho v = \lambda v \iff \lambda \geq 0 \quad \text{and} \quad \text{Tr}[\rho] = 1. \quad (6.4)$$

Note that the middle condition is sometimes written in the shorthand notation $\rho \geq 0$. The most general form of a density operator (in a finite dimensional Hilbert space) can be written

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (6.5)$$

where $\sum_i p_i = 1$ and $\langle\psi_i|\psi_j\rangle = \delta_{ij}$.

^aNote that a hermitian operator with no negative eigenvalues is often called a *positive semidefinite operator*. A *positive operator* is hermitian with all eigenvalues greater than zero. The shorthand notations $\rho \geq 0$ and $\rho > 0$, respectively, are sometimes used for these cases.

Theorem 6.1 (Properties of Density Operators.). Let ρ_1 and ρ_2 be density operators, and let U be a unitary operator. Then, the following hold:

- (1) $\rho_1 \otimes \rho_2$ is a density operator.
- (2) $U\rho U^\dagger$ is a density operator.

Proof. We first prove (1) by showing each of the three conditions in (6.4) hold. First, we have that

$$(\rho_1 \otimes \rho_2)^\dagger = \rho_1^\dagger \otimes \rho_2^\dagger = \rho_1 \otimes \rho_2$$

where in the second step we used the fact that ρ_1 and ρ_2 are density operators. Now, suppose that $v_1 \otimes v_2$ is an eigenvector of $\rho_1 \otimes \rho_2$ with eigenvalue λ . That is, $(\rho_1 \otimes \rho_2)(v_1 \otimes v_2) = \lambda(v_1 \otimes v_2)$. Then,

$$(\rho_1 \otimes \rho_2)(v_1 \otimes v_2) = \rho_1 v_1 \otimes \rho_2 v_2 = \lambda_1 v_1 \otimes \lambda_2 v_2 = \lambda_1 \lambda_2 (v_1 \otimes v_2).$$

Since $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, we must have $\lambda = \lambda_1 \lambda_2 \geq 0$. Finally, we can see that

$$\text{Tr}[\rho_1 \otimes \rho_2] = \text{Tr}[\rho_1] \otimes \text{Tr}[\rho_2] = (1)(1) = 1.$$

Thus, $\rho_1 \otimes \rho_2$ is a density operator.

We now prove (2). Letting $\rho' := U\rho U^\dagger$, we first have

$$\rho'^\dagger = (U\rho U^\dagger)^\dagger = U\rho^\dagger U^\dagger = U\rho U^\dagger = \rho'.$$

Now suppose that $\rho'v = \lambda v$. Then,

$$U\rho U^\dagger v = \lambda v \implies \rho(U^\dagger v) = \lambda(U^\dagger v)$$

where we have multiplied by U^\dagger on the left. This says that $U^\dagger v$ is an eigenvector of ρ with eigenvalue λ . Since ρ is a density operator, we have $\lambda \geq 0$. Last, using cyclicity of the trace, we see that

$$\text{Tr}[\rho'] = \text{Tr}[U\rho U^\dagger] = \text{Tr}[\rho U^\dagger U] = \text{Tr}[\rho] = 1].$$

□

Exercise 63: Let $A \in \mathbb{C}^{n \times n}$ be nonzero so that $\text{Tr}[AA^\dagger] > 0$. Prove that

$$\rho := \frac{AA^\dagger}{\text{Tr}[AA^\dagger]} \quad (6.6)$$

is a density operator.

We can now describe pure states and mixed states in terms of density operators. For a pure state, we have one wavefunction $|\psi\rangle$ with probability $p = 1$. Thus, we have the following:

Theorem 6.2 (Density matrix of a pure state). The density matrix of a pure state $\{(1, |\psi\rangle)\}$ is given by

$$\rho = |\psi\rangle\langle\psi| \quad (6.7)$$

where $|\psi\rangle$ is a normalized wavefunction.

For a mixed state, we can weight each outer product by the probability to get a density operator. Thus, we have the following:

Theorem 6.3 (Density matrix of a mixed state). The density matrix of a mixed state $\{(p_i, |\psi_i\rangle)\}$ is given by

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (6.8)$$

These mathematical descriptions of pure and mixed states will be useful for proving properties about them. Note one particularly nice feature of density operators is that they allow us to treat pure states and mixed states with the same mathematical formalism. Indeed, we can think of pure states as a special case of mixed states where we're summing over one state with unit probability.

6.1.4 Properties of Pure and Mixed States

Before we state and prove properties of pure and mixed states, we first give a definition.

Definition 6.4 (Purity). The **purity** of a quantum state described by the density operator ρ is defined to be $\text{Tr}[\rho^2]$.

As the name implies, the *purity* gives some measure of “how pure” a quantum state is—that is, how close the state is to a pure state. We will soon see that pure states have purity one and mixed states have purity less than one.

Theorem 6.4 (Properties of pure states). Let ρ be a pure state. Then,

- (1) $\rho^2 = \rho$.
- (2) $\text{Tr}[\rho] = 1$.
- (3) $\text{Tr}[\rho^2] = 1$.
- (4) $\rho \geq 0$.

Proof. For a pure state, $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle$ is a correctly normalized wavefunction—that is, $\langle\psi|\psi\rangle = 1$. (Note that ρ is a projector.) Thus, we have that

$$\rho^2 = |\psi\rangle\langle\psi||\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|. \quad (6.9)$$

Since ρ is a density operator, we know that (2) is true. Combining (1) and (2) proves (3), and (4) is true again because ρ is a density operator. \square

We now state the analogue of this theorem for mixed states.

Theorem 6.5 (Properties of mixed states). Let ρ be a mixed state. Then,

- (1) $\rho^2 \neq \rho$ in general.
- (2) $\text{Tr}[\rho] = 1$.
- (3) $\text{Tr}[\rho^2] < 1$.
- (4) $\rho \geq 0$.

Proof. This proof is left as an exercise. \square

Exercise 64: Prove Theorem 6.5.

Example 2: Calculating the purity of a mixed state.

Let $\langle a|b\rangle = 0$ and consider the mixed state described by the density operator

$$\rho = \frac{1}{2}|a\rangle\langle a| + \frac{1}{2}|b\rangle\langle b|. \quad (6.10)$$

We can calculate

$$\rho^2 = \frac{1}{4}|a\rangle\langle a| + \frac{1}{4}|b\rangle\langle b|. \quad (6.11)$$

Thus, the purity of this state is $\text{Tr}[\rho^2] = 1/2$. Since the purity is not equal to one, this must be a mixed state (which we of course already knew). In general, calculating the purity is a good way of finding out whether or not the state is pure or mixed.

Now that we have some mathematical formalism, we can see the difference between the two states described in Example 1.

Example 3: Differences between pure and mixed states.

In a previous example, we considered the pure state described by the wavefunction

$$|\psi\rangle = (2/3N)|0\rangle + (1/3N)|1\rangle \quad (6.12)$$

(where N is a normalization factor) and the mixed state described by the statistical distribution

$$\{(2/3, |0\rangle), (1/3, |1\rangle)\}. \quad (6.13)$$

We can now calculate the density operators for both of the states.

For the pure state, we have

$$\rho_1 = [(2/3N)|0\rangle + (1/3N)|1\rangle][(2/3N)\langle 0| + (1/3N)\langle 1|] = \frac{1}{9N^2} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \quad (6.14)$$

and for the mixed state we have

$$\rho_2 = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1| \begin{bmatrix} 2/3 & 0 \\ 0 & 1/3 \end{bmatrix} \quad (6.15)$$

These density operators are clearly distinct—and so are the states!

While this example illustrates some of the differences between pure and mixed states, one should not believe that mixed states are diagonal in the computational basis.

Exercise 65: Consider the mixed state described by

$$\{(1 - 2\epsilon)|+\rangle, (\epsilon, |0\rangle), (\epsilon, |1\rangle)\} \quad (6.16)$$

where $0 < \epsilon \ll 1$ is a real number. (Recall that $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$.) The interpretation of this state is thus that we prepare the coherent $|+\rangle$ state with high probability or either of the incoherent computational basis states with small probability.

(1) Prove that the density operator for this state is

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 1 - 2\epsilon \\ 1 - 2\epsilon & 1 \end{bmatrix} \quad (6.17)$$

(2) Prove that this density operator satisfies the four properties of mixed states listed in Theorem 6.5.

6.2 Trace and Partial Trace

We've been using the trace above in our treatment of density operators. Here we'll define it formally to motivate the *partial trace*, an extremely useful operation in quantum information theory. The partial trace is in some respect the “inverse” of the tensor product—while the tensor product is used to form larger composite systems, the partial trace is used to form smaller (composite) systems by “ignoring” one or more constituent systems.

If you're like me, the first time someone defined a trace for you, it was the sum of diagonal elements of a matrix.

Definition 6.5 (Trace of a matrix). The **trace of a matrix** is the sum of its diagonal elements.

One can show that the trace is *basis independent*. Indeed it must be if we're thinking about the trace of an operator, else it would be ill-defined.

Exercise 66: Show that the trace of a matrix is independent of which basis the matrix is expressed in. A special case of this is when the matrix is expressed in its *eigenbasis*, from which we see that the trace of a matrix is equal to the sum of the eigenvalues of said matrix.

Thus, we can define the trace of an operator:

Definition 6.6 (Trace of an operator). The **trace of an operator** T is

$$\text{Tr}[T] := \sum_i \langle i|T|i\rangle \quad (6.18)$$

where $\{|i\rangle\}$ is *any* orthonormal basis.

The partial trace is going to be defined very similarly. First, the partial trace only makes sense when we have some notion of “partial.” For us, as quantum information theorists, we know that we build up multiparticle systems via a tensor product. For example, if ρ_A represents the state of particle A and ρ_B represents the state of particle B , then the state of both particles is given by

$$\rho_{AB} = \rho_A \otimes \rho_B \quad (6.19)$$

We could take the trace of the *entire* system ρ_{AB} in the same way we defined the trace of an operator above. But, what if we wanted to take the trace of only one *component* of the system? For example, what if we just wanted to *trace out* (standard terminology) the B particle? We would do this if we were only interested in the state of A or knowing some property about the state of A , *regardless* of what the state of B was.

Example 4: Conceptual example of the partial trace.

If you're familiar with marginal probability distributions, this example may help elucidate the conceptual understanding behind the partial trace. Suppose we have a joint (discrete) probability distribution $p(X, Y)$ enumerated by the table below.

X/Y	y_1	y_2	y_3	$p_x(X)$
x_1	4/32	2/32	2/32	8/32
x_2	3/32	6/32	6/32	15/32
x_3	9/32	0	0	9/32
$p_y(Y)$	16/32	8/32	8/32	32/32

The ij th entry of this table is the probability of having $X = x_i$ and $Y = y_j$. Now, suppose we just want to know what the probability of $X = x_i$ is, for example $X = x_1$. Well, we know that $X = x_1$ when $X = x_1$ and $Y = y_1$, and also when $X = x_1$ and $Y = y_2$, and also when $X = x_1$ and $Y = y_3$. In order to get the probability that $X = x_1$, we must then take all of these cases into account by summing over their respective probabilities. As we see in the table above (summing across the top row in this table),

$$P(X = x_1) = (4/32) + (2/32) + (2/32) = 8/32. \quad (6.20)$$

The probability for any outcome of X is thus given by

$$p_x(X) = \sum_y p(X, Y). \tag{6.21}$$

Here, we are summing *part* of the total system to get information about the other part that we care about, in this case the X random variable. This underlying idea is the same for the partial trace of a quantum state.

If we only cared about particle A in our two particle system above, we would want to *sum over*, or take a partial trace, of particle B . In this situation, we would write the partial trace as

$$\text{Tr}_B[\rho_{AB}] = \sum_i \rho_A \otimes \langle i | \rho_B | i \rangle = \rho_A. \tag{6.22}$$

In the last step, we used the fact that the trace of a density operator is one.

In this case, since the state is separable, the resulting *marginal state* is just the state of the original A particle, ρ_A , ignoring the B particle. This is of course desirable for a definition of the partial trace.

In the general case, the state ρ_{AB} doesn't need to be separable for us to define the partial trace. In this case, we slightly generalize what we did above—leave the A particle alone and take the trace of the B particle by summing over its expectation values in some orthonormal basis. To generalize, we simply take “leave the A particle alone” to mean “act with the identity operator on the A subsystem,” which are of course equivalent. This allows us to formally define the partial trace for the state of a composite system AB consisting of two subsystems A and B below.

Note that we use the term “subsystem” instead of “particle.” A system can be one particle or many particles—this definition is valid for both cases. The “sub” prefix merely indicates that we are thinking of the system as a component of a larger system. We now state the definition of partial trace.

Definition 6.7 (Partial trace). Let A and B be two subsystems making up the composite system described by the density operator ρ_{AB} . The **partial trace over the B subsystem**, denoted Tr_B , is defined as

$$\text{Tr}_B[\rho_{AB}] := \sum_j (I_A \otimes \langle j |_B) \rho_{AB} (I_A \otimes |j \rangle_B) \tag{6.23}$$

where $\{|j\rangle\}$ is any orthonormal basis for the Hilbert space \mathcal{H}_B of subsystem B . We often write $\rho_A \equiv \text{Tr}_B[\rho_{AB}]$.

Similarly, the **partial trace over the A subsystem**, denoted Tr_A , is defined as

$$\text{Tr}_A[\rho_{AB}] := \sum_i (\langle i |_A \otimes I_B) \rho_{AB} (|i \rangle_A \otimes I_B) \tag{6.24}$$

where $\{|i\rangle\}$ is any orthonormal basis for the Hilbert space \mathcal{H}_A of subsystem A . We often write $\rho_B \equiv \text{Tr}_A[\rho_{AB}]$.

The partial trace can be generalized to composite systems of more than two parts, for example ρ_{ABC} or ρ_{ABCD} . The idea is the same, however—act with identity on the system(s) you care about, and trace out (sum over) the systems you don't care about. To build a partial trace, we need identity operators and orthonormal bases—the recipe is to put these in the right places.

Example 5: Partial trace of a maximally entangled state.

This is a classical example in quantum information theory^a. Consider the pure state

$$\rho_{AB} = |\Phi_{AB}^+\rangle \langle \Phi_{AB}^+| \tag{6.25}$$

where

$$|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}} [|00\rangle + |11\rangle] \tag{6.26}$$

is the maximally entangled Bell state. The density operator is then

$$\rho_{AB} = \frac{1}{2} [|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|]. \quad (6.27)$$

Let's now calculate the partial trace over the B subsystem. Doing so, following Definition 6.7 using the computational basis as our orthonormal basis $\{|j\rangle\}$, we have

$$\text{Tr}_B[\rho_{AB}] = \frac{1}{2} [(I_A \otimes \langle 0|_B)\rho_{AB}(I_A \otimes |0\rangle_B) + (I_A \otimes \langle 1|_B)\rho_{AB}(I_A \otimes |1\rangle_B)] \quad (6.28)$$

$$= \frac{1}{2} [|0\rangle\langle 0| + |1\rangle\langle 1|.] \quad (6.29)$$

Thus, the partial trace of the maximally entangled state is the maximally mixed state!

^aOr maybe...a *quantum* example?

Exercise 67: Write out the maximally entangled state (6.27) in matrix form. If you weren't thinking in terms of quantum information theory, you'd never know such an inconspicuous matrix could have so many theoretical implications!

Exercise 68: Write out the maximally mixed state (6.29) in matrix form.

Exercise 69: Take the partial trace over subsystem A in Example 5 and show that you get the same result. Do this first mathematically, then think physically in terms of symmetry and get the same result with much less algebra.

6.2.1 Expectation Values as Traces of Density Operators

Lastly, we note a useful way to write expectation operators in terms of traces of density operators. Here we transition back from the *partial* trace to the "full" trace, better known as just the trace.

First consider a pure state $\rho = |\psi\rangle\langle\psi|$, and examine what happens when we evaluate $\text{Tr}[\rho A]$ where A is some observable (hermitian operator):

$$\begin{aligned} \text{Tr}[\rho A] &= \sum_i \langle i|\rho A|i\rangle && \text{(by definition of trace)} \\ &= \sum_i \langle i||\psi\rangle\langle\psi|A|i\rangle && \text{(by definition of } \rho) \\ &= \sum_i \psi_i \langle\psi|A|i\rangle && \text{(let } \psi_i \text{ denote the } i\text{th coordinate of } |\psi\rangle) \\ &= \sum_i \langle\psi|A|(\psi_i|i)\rangle && \text{(group } \psi_i \text{ with } |i\rangle) \\ &= \langle\psi|A|\psi\rangle && \text{(since } |\psi\rangle = \sum_i \psi_i|i\rangle). \end{aligned}$$

This is precisely the expectation value of A with respect to $|\psi\rangle$! We have thus proved the following theorem:

Theorem 6.6 (Expectation value as trace over a pure state). Let $\rho = |\psi\rangle\langle\psi|$ be a pure state and A be an observable. Then,

$$\langle A \rangle_\psi := \langle \psi | A | \psi \rangle = \text{Tr}[\rho A]. \quad (6.30)$$

This theorem provides an alternative useful means of calculating expectation values.

A similar type of identity exists for the case of mixed states. In this case, we obtain a convex combination¹ of expectation values in each pure state present in the mixed state. We precisely state the theorem below.

Theorem 6.7. Consider the mixed state described by the density operator

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (6.31)$$

Then, for any observable A , we have

$$\text{Tr}[\rho A] = \sum_j p_j \langle \psi_j | A | \psi_j \rangle \quad (6.32)$$

Proof. We proceed as above in the case of pure states, using only definitions and simple algebra. Starting with the trace, we can write

$$\begin{aligned} \text{Tr}[\rho A] &= \sum_i \langle i | \rho A | i \rangle && \text{(by definition of trace)} \\ &= \sum_i \langle i | \left[\sum_j p_j |\psi_j\rangle\langle\psi_j| \right] A | i \rangle && \text{(by definition of } \rho \text{)} \\ &= \sum_j p_j \left[\sum_i \langle i | \psi_j \rangle \langle \psi_j | A | i \rangle \right] && \text{(rearrange summation order and terms)} \\ &= \sum_j p_j \langle \psi_j | A | \psi_j \rangle. && \text{(by the results for the pure state case)} \end{aligned}$$

Note that in the last step, we used the result of the proof for the similar derivation for pure states. Thus, we see that the trace of ρA is indeed a convex combination of pure state expectation values, as claimed. \square

Exercise 70: Explicitly write out all steps in the proof of Theorem 6.7 (instead of citing the result from the proof for pure states).

6.3 In Closing

In this seminar, we have formally introduced the notions of density operators, pure states, mixed states, trace, and partial trace. These are foundational concepts in quantum information theory and quantum algorithms. Mastering them is time well spent. In future seminars, we'll use these concepts to examine additional important topics in quantum information science, including:

- Distance between quantum states.
- Entanglement tests and entanglement measures.
- Noise and decoherence.

¹A *convex combination* is the fanciest way I know of saying a linear combination where all coefficients are non-negative and sum to one.