Ensembles and incomplete information

So far in this course, we have described quantum systems by states that are normalized vectors in a complex Hilbert space. This works so long as (a) the system is not entangled with anything else, and (b) we have complete information about the system. We will look at (a) in a moment; for now, let’s look at (b).

In the classical world, when we have incomplete information about a system, we specify a probability distribution over possible states. The probabilities reflect our degree of certainty or uncertainty about that state. We could imagine that the system was drawn from a large collection of systems in different states, where the proportion of systems in each state corresponds to the probability we assign to that state. Such a collection is called an ensemble.
We can try the same approach in quantum mechanics. Rather than specifying a unique state vector $|\psi\rangle$, we list a collection of state vectors $|\psi_1\rangle, \ldots, |\psi_N\rangle$ with probabilities $p_i$ such that $\sum_i p_i = 1, 0 \leq p_i \leq 1$. (For now, we assume that the number of states $N$ in the ensemble is finite.) We denote such an ensemble $\{p_i, |\psi_i\rangle\}$.

If there is a single state $|\psi\rangle$ with probability $p = 1$, our ensemble is just the same a state vector. We call such an ensemble a **pure state**. Any ensemble that is not pure we call a **mixed state**.

Suppose we have a system in a mixed state. How do our usual rules for dealing with quantum states change?
Unitary transformations of mixed states

Let our ensemble be \( \{ p_i, |\psi_i\rangle \} \). How does this evolve in time? If we knew we were in a definite pure state \( |\psi_i\rangle \), the state would just change by a unitary transformation \( |\psi_i\rangle \rightarrow \hat{U}|\psi_i\rangle \). In a mixed state, we just do the same thing term by term:

\[
\{ p_i, |\psi_i\rangle \} \rightarrow \{ p_i, \hat{U}|\psi_i\rangle \}.
\]

Note that this simple rule works because of the linearity of quantum evolution. The rule for time evolution has no explicit dependence on the state at the present time.

In unitary evolution, the probabilities \( p_i \) of the ensemble elements do not change.
Measurements of Ensembles

Let’s first restrict ourselves to projective measurements. Suppose we measure some observable whose eigenprojectors are \( \{ \hat{P}_j \} \). If the system were in state \( |\psi_i\rangle \), the probability of outcome \( \hat{P}_j \) would be

\[
p_{j|i} = \langle \psi_i | \hat{P}_j | \psi_i \rangle,
\]

and the state would become \( |\psi_i\rangle \rightarrow \hat{P}_j |\psi_i\rangle / \sqrt{p_{j|i}} \).

The overall probability of the outcome \( j \) is given by weighting the conditional probabilities \( p_{j|i} \) by the \textit{a priori} probabilities \( p_i \), to get

\[
p_j = \sum_i p_i p_{j|i} \equiv \sum_i p_i \& j.
\]
Conditioning the Probabilities

We could just replace the states $\psi_i$ with the updated states in the ensemble $\{p_i, \psi_i\}$; but that would not be quite right. We have gained information from the measurement about which of the states in the ensemble are likely; we need to update their weights $p_i$ as well. Since the uncertainty about the state is basically a classical uncertainty, we update the weights by the Bayes rule. Given that the measurement outcome was $j$,

$$p_i \rightarrow p_i|j = \frac{p_{i \& j}}{p_j} = \frac{p_{i|j} p_j}{p_j} = \frac{p_i \langle \psi_i | \hat{P}_j | \psi_i \rangle}{\sum_k p_k \langle \psi_k | \hat{P}_j | \psi_k \rangle}.$$ 

The ensemble then transforms to

$$\{p_i, \psi_i\} \rightarrow \{p_{i|j}, \hat{P}_j | \psi_j \}/\sqrt{p_{j|i}}.$$
Generalized Measurements

- Much the same applies to a generalized measurement, given by a set of measurements operators \( \{\hat{M}_j\} \). In this case, the probability of outcome \( j \) for a state \(|\psi_i\rangle\) is

\[
p_{j|i} = \langle \psi_i | \hat{M}_j^\dagger \hat{M}_j | \psi_i \rangle,
\]

and the state goes to

\[
|\psi_i\rangle \rightarrow \hat{M}_j |\psi_i\rangle / \sqrt{p_{j|i}}.
\]

- Just as for projective measurements, the ensemble then transforms to

\[
\{p_i, |\psi_i\rangle\} \rightarrow \left\{p_{i|j}, \hat{M}_j |\psi_j\rangle / \sqrt{p_{j|i}} \right\}.
\]
Note that the probabilities of outcomes are not changed in any way if each of the members of an ensemble are separately multiplied by an overall phase.

That is the ensembles \( \{ p_i, |\psi_i\rangle \} \) and \( \{ p_i, e^{i\theta_i} |\psi_i\rangle \} \) are equivalent for all practical purposes. We consider these two ensembles to be physically identical.

The rule for updating an ensemble after a measurement is rather complicated. We can simplify by changing the way we represent mixed states, using the idea of a \textit{density matrix}. 
Density matrices

Rather than represent a mixed state by an ensemble \( \{ p_i, |\psi_i\rangle \} \), we can instead use a density matrix (or density operator):

\[
\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|.
\]

Note that the operators \( |\psi_i\rangle \langle \psi_i| \) are one-dimensional projectors (hence positive operators with unit trace), and the weights \( p_i \) are positive numbers which sum to 1. (The \( |\psi_i\rangle \) need not be orthogonal.) This implies that \( \rho \) is a positive operator with unit trace:

\[
\text{Tr}\{\rho\} = 1, \quad \langle \phi | \rho | \phi \rangle \geq 0 \quad \forall |\phi\rangle, \quad \rho = \rho^\dagger.
\]
For a pure state $|\psi\rangle$, the corresponding density matrix is $\rho = |\psi\rangle\langle\psi|$. In this case only, $\rho^2 = \rho$ is a projector. This gives a simple test of a state’s purity.

A density matrix transforms under unitary evolution by $\rho \to \hat{U}\rho\hat{U}^\dagger$.

For a projective measurement with projectors $\{\hat{P}_j\}$, the probability of outcome $j$ is

$$p_j = \sum_i p_i \langle \psi_i | \hat{P}_j | \psi_i \rangle = \sum_i p_i \text{Tr} \left\{ \hat{P}_j | \psi_i \rangle \langle \psi_i | \right\} = \text{Tr} \left\{ \hat{P}_j \rho \right\}.$$  

This is a much simpler expression than for ensembles.
After getting measurement result $j$, the density operator is updated by

$$\rho \rightarrow \sum_i \frac{p_i|j}{p_j} \hat{P}_j |\psi_i\rangle \langle \psi_i| \hat{P}_j = \hat{P}_j \rho \hat{P}_j / p_j.$$

This is a far simpler rule than that for ensembles as well! For a generalized measurement, the expressions are similar:

$$p_j = \text{Tr} \left\{ \hat{M}_j^\dagger \hat{M}_j \rho \right\} \equiv \text{Tr} \left\{ \hat{E}_j \rho \right\} = \text{Tr} \left\{ \hat{M}_j \rho \hat{M}_j^\dagger \right\},$$

$$\rho \rightarrow \hat{M}_j \rho \hat{M}_j^\dagger / p_j.$$
The von Neumann equation

The unitary evolution of quantum states was derived from the Schrödinger equation:

\[
(i\hbar)\frac{d|\psi\rangle}{dt} = \hat{H}(t)|\psi\rangle,
\]

where \( \hat{H}(t) = \hat{H}^\dagger(t) \) is the Hamiltonian. This implies the unitary evolution \( |\psi(t_1)\rangle = \hat{U}(t_1, t_0)|\psi(t_0)\rangle \).

The corresponding equation for density matrices is

\[
d\rho/dt = -(i/\hbar) \left[ \hat{H}(t), \rho \right].
\]

This equation is sometimes referred to as the von Neumann equation. It implies the unitary evolution

\[
\rho(t_0) \rightarrow \rho(t_1) = \hat{U}(t_1, t_0)\rho(t_0)\hat{U}^\dagger(t_1, t_0).
\]
Ensemble equivalence

All observable results can be predicted from the density matrix, without needing to know the ensemble used to construct it. In ensembles that include many terms, this is a more compact representation. However, the relation between density matrices and ensembles is not unique. There can be different ensembles \( \{ p_i, |\psi_i\rangle \} , \{ p'_j, |\psi'_j\rangle \} \) with the same density matrix \( \rho \):

\[
\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i | = \sum_j p'_j |\psi'_j\rangle \langle \psi'_j |.
\]

The number of terms in the two ensembles need not be the same. In fact, except for pure states, any density operator can derive from an infinite number of possible ensembles. We call these possible decompositions of \( \rho \).
For example, consider the density matrix
\[ \rho = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1|. \]

This could arise from the ensemble containing $|0\rangle, |1\rangle$ with weights $p, 1 - p$. But it could also arise from the ensemble with states $\sqrt{p}|0\rangle \pm \sqrt{1 - p}|1\rangle$ with weights $1/2, 1/2$.

All physical consequences can be deduced from the density matrix. This means that two ensembles with the same density matrix make exactly the same predictions for any measurement. Two such ensembles are impossible to distinguish by any experimental test.
In fact, this indistinguishability is much stronger than the indistinguishability for nonorthogonal pure states. No measurement on a single copy of the system can reliably tell the two nonorthogonal states apart. Given many copies, however, one can collect statistics and distinguish the states.

This is not true for ensembles with the same density matrix. For any number of systems drawn from the two ensembles, all predictions are identical.

Because of this fact, we rarely describe mixed states in terms of ensembles. It is much more common just to represent them as density matrices. Indeed, in many cases we would have no idea what ensemble to choose; and the choice would add no useful information.
Orthogonal decomposition

Let us now take density matrices as our primary objects. A density matrix is a positive, Hermitian operator with unit trace. As such, it has an orthonormal basis of eigenstates $|\phi_j\rangle$, whose corresponding real eigenvalues $\lambda_j$ are $\geq 0$ and sum to 1. In diagonal form,

$$\rho = \sum_{j=1}^{D} \lambda_j |\phi_j\rangle \langle \phi_j|, \quad \lambda_j \geq 0, \quad \sum_j \lambda_j = 1.$$ 

We see that $\{\lambda_j, |\phi_j\rangle\}$ is an ensemble, and forms a decomposition of $\rho$; moreover, this decomposition is composed of orthogonal states (which is not true of most decompositions). It is sometimes useful to work in this orthogonal decomposition.
Composite systems and entanglement

For pure states, we represented the joint state of two subsystems, each in separate states $|\psi\rangle$ and $|\phi\rangle$, by a product state $|\psi\rangle \otimes |\phi\rangle$ in the tensor-product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. More generally, a joint state could be a superposition of product states; it could be entangled,

$$|\Psi\rangle = \sum_{i,j} t_{ij} |i\rangle \otimes |j\rangle.$$ 

What happens when we switch to mixed states? The density operator for a product state $|\psi\rangle \otimes |\phi\rangle$ is

$$\rho = (|\psi\rangle \otimes |\phi\rangle)(\langle\psi| \otimes \langle\phi|) = |\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|.$$ 

More generally, if the first system has a density matrix $\rho$ and the second has a density matrix $\chi$, the joint density matrix is a product $\rho \otimes \chi$. 
What if the two systems are in an entangled state $|\Psi\rangle$? Then the joint density matrix is $\rho = |\Psi\rangle\langle\Psi|$, and is a positive operator with unit trace on the joint Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. A mixed state would then take the form

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|.$$ 

In certain cases, this joint mixed state can be written in the form

$$\rho = \sum_i p_i \rho_i \otimes \chi_i,$$

where $0 \leq p_i \leq 1$, $\sum_i p_i = 1$, and $\rho_i$ and $\chi_i$ are density matrices. We call such a matrix $\rho$ separable.
If $\rho$ is separable, it can be written as a decomposition into pure product states:

$$\rho = \sum_k p_k |\psi_k\rangle \langle\psi_k| \otimes |\phi_k\rangle \langle\phi_k|.$$ 

We define a mixed state to be entangled if it is not separable; i.e., if it has no decomposition in terms of product states.

Unfortunately, this is not an easy thing to check! Every mixed-state density matrix has many decompositions. There is no efficient procedure to determine whether any of them uses only product states. (In fact, this problem has been shown to be NP-complete!)
Also, while the unentangled pure states form a set of measure zero, the separable states do not. The separable states form a convex set in the space of all density matrices.

All correlated pure states are entangled. But with mixed states, there are three levels: product states (uncorrelated); separable states (classically correlated); and nonseparable states (entangled).

Mixed state entanglement is a much more complicated subject than pure state entanglement. It is well beyond the scope of this class.
Density Operator Evolutions

- What is the most general linear transformation that takes density matrices to density matrices? A unitary transformation obviously works: \( \rho \rightarrow \hat{U}\rho\hat{U}^\dagger \). However, it is not the most general possibility.

- Note that we are now asking for linear transformations on \emph{operators}, not states. Such a linear transformation is sometimes called a \emph{superoperator}. The most general such transformation is given by pairs of operators \( \{\hat{A}_i, \hat{B}_i\} \):

\[
\hat{O} \rightarrow \sum_i \hat{A}_i \hat{O} \hat{B}_i.
\]

- It is easy to see that any linear transformation on operators can be written in this form.
CPTP Maps

- We then want to restrict ourselves to transformations which preserve both the trace and the positivity of operators: *trace-preserving positive superoperators*.

- In fact, we need to restrict things a little more than that. Suppose our system is just part of a composite system. It turns out that a transformation which is positive on the subsystem alone may *not* be positive on the joint system! The canonical example of this is *transposition*. Transposing a density matrix yields another density matrix. Doing a *partial transpose* on just one subsystem *doesn’t* necessarily yield a positive operator.

- A map which is positive even when applied only to part of a composite system is called *completely positive*. The allowed transformations are therefore *completely postitive trace-preserving (CPTP) maps*. 
It turns out that all such maps can be written in the form
\[
\rho \rightarrow \sum_j \hat{M}_j \rho \hat{M}_j^\dagger
\]
for some set of operators \( \{ \hat{M}_j \} \) that satisfy
\[
\sum_j \hat{M}_j^\dagger \hat{M}_j = \hat{I}.
\]

This is the Kraus representation, and the \( \{ \hat{M}_j \} \) are Kraus operators. It looks exactly like a generalized measurement on \( \rho \), but with the measurement outcome unknown. It also includes unitary transformations as a special case (with only 1 Kraus operator).
Note that just as the decompositions of $\rho$ are not unique, the Kraus representation of a completely positive map is not unique either. That is, there are different sets of operators $\{\hat{M}_j\}$ that correspond to the same completely positive map.

We call a trace-preserving completely positive map a *quantum operation*. Any quantum operation is, in principle, physically achievable. So if we allow mixed states, the set of physically allowed procedures are quantum operations and generalized measurements.

To compare this to what we have already done earlier in this course, for pure states the physically allowed procedures are unitary transformations and generalized measurements. Mixed states therefore have a larger class of physically allowed procedures.
Reduced density matrices

Suppose we have a composite system with density matrix \( \rho_{AB} \) on the tensor-product space \( \mathcal{H}_A \otimes \mathcal{H}_B \). We are only interested, however, in measurements and operations on subsystem A. (For example, subsystem B might be far away, and not accessible to us.) In that case, we can write down an effective density matrix for subsystem A alone by taking the partial trace over B:

\[
\rho_A = \text{Tr}_B \{ \rho_{AB} \}.
\]

The operator \( \rho_A \) is also a density matrix; we call it the reduced density matrix for subsystem A. (Obviously, we could also find a reduced density matrix for B.)
Recall that the partial trace of an operator is

$$\text{Tr}_B \left\{ \sum_j \hat{A}_j \otimes \hat{B}_j \right\} = \sum_j \hat{A}_j \text{Tr}\{\hat{B}_j\}. $$

If $A$ and $B$ are in a pure product state $|\psi\rangle \otimes |\phi\rangle$, then the reduced density matrix of $A$ is

$$\rho_A = \text{Tr}_B\{|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|\} = |\psi\rangle\langle\psi|,$$

which is a pure state.

If they are in an entangled pure state, however, the situation is quite different.
Consider the entangled state

\[ |\Psi\rangle = \sum_{i,j} t_{ij} |i\rangle \otimes |j\rangle, \]

where \(|i\rangle\) and \(|j\rangle\) are orthonormal bases, then the reduced density matrix will be a mixed state:

\[ \rho_A = \sum_{i,i',j} t_{ij} t_{i'j}^* |i\rangle \langle i'| \]

A mixed state, then, reflects either uncertainty about the state, or entanglement with another subsystem. No measurement on the subsystem alone can distinguish these possibilities.
The Schmidt decomposition

Given any pure state $|\Psi\rangle$ on a tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, we can always find bases $\{|i\rangle_A\}, \{|i\rangle_B\}$ on $\mathcal{H}_A$ and $\mathcal{H}_B$ and positive numbers $\{\lambda_i\}$ such that

$$|\Psi\rangle = \sum_i \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B, \quad \sum_i \lambda_i = 1.$$ 

(The construction is based on the singular value decomposition of the matrix $t_{ij}$ of amplitudes.)

If we write $|\Psi\rangle$ in Schmidt form, it is clear that

$$\rho_A = \sum_i \lambda_i |i\rangle_A \langle i|, \quad \rho_B = \sum_i \lambda_i |i\rangle_B \langle i|.$$
Purification

The Schmidt bases are the eigenbases of the reduced density matrices, and the matrices have the same eigenvalues. This makes it possible to *purify* a matrix.

Given a density matrix $\rho$ with orthogonal decomposition

$$\rho = \sum_i \lambda_i |i\rangle \langle i|,$$

we can introduce a fictional additional subsystem $B$ and a joint pure state $|\Psi\rangle$ on $A$ and $B$, such that $\rho$ is the reduced density matrix on $A$.

The problem is to find $|\Psi\rangle$ such that $\rho = \text{Tr}_B\{|\Psi\rangle \langle \Psi|\}$. 
This pure state $|\Psi\rangle$ is just

$$|\Psi\rangle = \sum_i \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B,$$

where the eigenbasis of $\rho$ now gives the Schmidt basis for $A$. We can use any orthonormal basis on $B$. Note that the dimension of $B$ must be at least equal to the number of nonzero eigenvalues of $\rho$.

In some cases, it is much easier to work with pure states than mixed states; in those cases, it often makes sense to purify the state $\rho$, do the calculation, and then take a partial trace at the end.
von Neumann entropy

Given an ensemble \( \{ p_i, |\psi_i\rangle \} \), we can associate with it a Shannon entropy

\[
S = - \sum_i p_i \log_2 p_i.
\]

We could try to identify this entropy with the uncertainty in the state \( \rho \). There is a problem, though: the different ensemble decompositions of \( \rho \) will not, in general, have the same Shannon entropy.

To get around this, we could define the entropy of the state \( \rho \) to be the minimum Shannon entropy over all decompositions of \( \rho \). This sounds very hard to calculate; but in fact, it is not. It turns out to be

\[
S(\rho) = - \text{Tr}\{\rho \log_2 \rho}\}.
\]
For a $D$-dimensional system, the maximum value $S$ can take is $\log_2 D$ (for $\rho = \hat{I}/D$).

The easiest way to calculate $S(\rho)$ is to write $\rho$ in diagonal form:

$$\rho = \sum_i \lambda_i |i\rangle\langle i|.$$ 

The von Neumann entropy is then

$$S(\rho) = -\sum_i \lambda_i \log_2 \lambda_i.$$ 

The decomposition of $\rho$ that minimizes the Shannon entropy is the orthogonal decomposition!
Entropy of Entanglement

If $\rho$ is a pure state, $S(\rho) = 0$. This lets us define a good measure of entanglement for entangled pure states.

Let $|\Psi\rangle$ be the joint state on $A$ and $B$, and $\rho_A$ be the reduced density matrix; then $S(\rho_A)$ is the *entropy of entanglement* for $|\Psi\rangle$. It vanishes if $|\Psi\rangle$ is a product state $|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle$, and attains the maximum value $\log_2 D$ for a *maximally entangled state* with Schmidt form

$$|\Psi\rangle = \sum_{j=1}^{D} \frac{1}{\sqrt{D}} |j\rangle_A \otimes |j\rangle_B.$$ 

For example, the Bell states are all maximally entangled, and have $S(\rho_A) = 1$. 