Lecture 15: Quantum information revisited (continued)

March 16, 2006

More on the partial trace

In the previous lecture we were discussing an important admissible operation called the partial trace. We will continue to discuss this operation in the first part of this lecture.

Recall that if we have registers X and Y with corresponding spaces $\mathcal{X}$ and $\mathcal{Y}$, then the admissible operation that describes the action of throwing away Y and leaving X alone is the partial trace (over the space $\mathcal{Y}$). Assuming that $\mathcal{Y} = \mathbb{C}(\Gamma)$ for some arbitrary classical state set Γ, we can express the partial trace over $\mathcal{Y}$ as follows:

$$\mathrm{Tr}_{\mathcal{Y}}\, \rho = \sum_{a\in\Gamma}(I_{\mathcal{X}} \otimes \langle a|)\rho(I_{\mathcal{X}} \otimes |a\rangle).$$

You could replace the vectors $\{|a\rangle \,:\, a \in \Gamma\}$ (and their corresponding bra vectors) by any other orthonormal basis of $\mathcal{Y}$ without changing the operation. Another way of expressing the partial trace, which makes this fact apparent, is $\mathrm{Tr}_{\mathcal{Y}} = I_{\mathrm{L}(\mathcal{X})} \otimes \mathrm{Tr}$ (where $I_{\mathrm{L}(\mathcal{X})}$ refers to the identity operator on $\mathrm{L}(\mathcal{X})$, which is the admissible operation that corresponds to doing nothing at all).

One can define $\mathrm{Tr}_{\mathcal{X}}$, the partial trace over the space $\mathcal{X}$, similarly to $\mathrm{Tr}_{\mathcal{Y}}$. More generally, we could imagine any number of registers and consider the operation that corresponds to discarding some sub-collection of them.

Let us now go back to the simple case of two qubits, and consider what happens when the partial trace is applied to an entangled state. Suppose that the pair of qubits (X, Y) is in the state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|11\rangle\,.$$

The corresponding density matrix is therefore $|\phi^+\rangle\,\langle\phi^+|$. The effect of discarding the second qubit is computed as follows. By the definition of the partial trace, we have

$$\mathrm{Tr}_{\mathcal{Y}}\,|\phi^+\rangle\,\langle\phi^+| = (I \otimes \langle 0|)\,|\phi^+\rangle\,\langle\phi^+|\,(I \otimes |0\rangle) + (I \otimes \langle 1|)\,|\phi^+\rangle\,\langle\phi^+|\,(I \otimes |1\rangle).$$

In order to simplify this, note that

$$(I \otimes \langle 0|)\,|\phi^+\rangle = (I \otimes \langle 0|)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}\,|0\rangle\,,$$

$$(I \otimes \langle 1|)\,|\phi^+\rangle = (I \otimes \langle 1|)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}\,|1\rangle\,,$$

and so

$$\mathrm{Tr}_{\mathcal{Y}}\,|\phi^+\rangle\,\langle\phi^+| = \frac{1}{2}\,|0\rangle\,\langle 0| + \frac{1}{2}\,|1\rangle\,\langle 1|\,.$$

By the way, we have seen this density matrix a couple of times now—it is sometimes called the *totally mixed state*, and essentially represents complete ignorance about the qubit in question. Try repeating this sort of calculation for initial states $|\phi^-\rangle$, $|\psi^+\rangle$, and $|\psi^-\rangle$. You will find that in each of these cases the resulting state is also the totally mixed state.

Of course it is not necessary to view the partial trace as representing the *destruction* of one or more quantum registers—we can use the partial trace when we simply wish to consider some part of a larger system in isolation. For instance, if Alice and Bob share some state, and we want to consider the outcomes of any measurements Bob can perform on his part of the shared system alone, then we would consider the state that results from tracing out Alice's part of the system. When this is done, the resulting state is called the *reduced state* of Bob's registers.

Example 1. Alice and Bob each have one qubit, and the two qubits together are known to be in one of the four Bell states:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle\,, \quad |\psi^+\rangle = \frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle\,,$$
$$|\phi^-\rangle = \frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{2}}|11\rangle\,, \quad |\psi^-\rangle = \frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|10\rangle\,.$$

Let $\mathcal{A}$ and $\mathcal{B}$ be the spaces corresponding to Alice's and Bob's qubit, respectively. Bob wants to determine which of the four Bell states they share without Alice's help. It is impossible for him to gain any information whatsoever about which of the four states they have, however, because

$$\mathrm{Tr}_{\mathcal{A}}|\phi^+\rangle\langle\phi^+| = \mathrm{Tr}_{\mathcal{A}}|\phi^-\rangle\langle\phi^-| = \mathrm{Tr}_{\mathcal{A}}|\psi^+\rangle\langle\psi^+| = \mathrm{Tr}_{\mathcal{A}}|\psi^-\rangle\langle\psi^-| = \frac{1}{2}I.$$

In all four cases, Bob's reduced state is the same, so there will be no difference in the distribution of outcomes Bob would get using any measurement whatsoever.

On the other hand, if Alice and Bob shared one of two states for which Bob's reduced state was *different* for the two states, then he could gain some information about which state they share by making some well-chosen measurement.

Example 2. Alice and Bob again each have one qubit. This time they share one of the two states

$$|\psi_1\rangle = \frac{3}{5}|00\rangle + \frac{4}{5}|11\rangle \quad \text{or} \quad |\psi_2\rangle = \frac{4}{5}|00\rangle - \frac{3}{5}|11\rangle\,.$$

The corresponding reduced states for Bob are

$$\rho_1 = \mathrm{Tr}_{\mathcal{A}}|\psi_1\rangle\langle\psi_1| = \frac{9}{25}|0\rangle\langle 0| + \frac{16}{25}|1\rangle\langle 1| \quad \text{and} \quad \rho_2 = \mathrm{Tr}_{\mathcal{A}}|\psi_2\rangle\langle\psi_2| = \frac{16}{25}|0\rangle\langle 0| + \frac{9}{25}|1\rangle\langle 1|\,.$$

These density matrices are different, so there is some measurement that Bob could make that would give him some information about which state they share. If Bob measures with respect to the standard basis, he could guess that they shared $|\psi_1\rangle$ if the measurement result was 1 and guess $|\psi_2\rangle$ if the measurement result was 0. He would be right 64% of the time, implying a gain of partial information about which state it was.

Purifications

Suppose that X is a register with corresponding space $\mathcal{X} = \mathbb{C}(\Sigma)$, and let $\rho \in \mathrm{D}(\mathcal{X})$ be a mixed state of X. Sometimes it is helpful to view ρ as being the reduced state of some pure state of the pair (X, Y), where Y is some other register. It is always possible to do this so long as the space $\mathcal{Y}$ corresponding to the register Y has dimension at least that of $\mathcal{X}$. (Really, as long as the dimension of $\mathcal{Y}$ is as large as the rank of ρ, this will be possible.) One way to see this is to consider a *spectral decomposition* of ρ:

$$\rho = \sum_{i=1}^{n} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|,$$

where $n = |\Sigma|$, $(p_1, \ldots, p_n)$ is a probability vector, and $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ is an orthonormal basis of $\mathcal{X}$. Every density matrix has such a decomposition. For $\mathcal{Y} = \mathcal{X}$, define a unit vector $|\phi\rangle \in \mathcal{X} \otimes \mathcal{Y}$ as follows:

$$|\phi\rangle = \sum_{i=1}^{n} \sqrt{p_i} \left|\psi_i\right\rangle \left|\psi_i\right\rangle.$$

Then $\mathrm{Tr}_{\mathcal{Y}} \left|\phi\right\rangle \left\langle\phi\right| = \rho$ (which you should check for yourself). The vector $|\phi\rangle$ (or the state $|\phi\rangle\langle\phi|$) is called a *purification* of ρ. Another purification is

$$|\phi\rangle = \sum_{i=1}^{n} \sqrt{p_i} \left|\psi_i\right\rangle |i\rangle$$

for $\mathcal{Y} = \mathbb{C}^n$, which has the same dimension as $\mathcal{X}$.

The Schmidt decomposition

By considering reduced states of bipartite systems, we can learn something interesting about pure states of those systems.

Suppose that $|\phi\rangle \in \mathcal{X} \otimes \mathcal{Y}$ is a pure quantum state of registers (X, Y), and let $\rho \in \mathrm{D}(\mathcal{X})$ denote the reduced state of X:

$$\rho = \mathrm{Tr}_{\mathcal{Y}} \left|\psi\right\rangle \left\langle\psi\right| \in \mathrm{D}(\mathcal{X}).$$

Like all density matrices, ρ is positive semidefinite and has trace equal to 1. It is therefore possible to write ρ as

$$\rho = \sum_{j=1}^{n} p_j \left|\mu_j\right\rangle \left\langle\mu_j\right|$$

for some orthonormal basis $\{|\mu_1\rangle, \ldots, |\mu_n\rangle\}$ of $\mathcal{X}$ and some probability vector $(p_1, \ldots, p_n)$. The vectors $|\mu_1\rangle, \ldots, |\mu_n\rangle$ are eigenvectors of ρ, and $p_1, \ldots, p_n$ are the associated eigenvalues. (We know the eigenvalues are all nonnegative because ρ is positive semidefinite, and they sum to 1 and therefore form a probability distribution because ρ has trace equal to 1.)

Suppose now that we write

$$|\psi\rangle = \sum_{j=1}^{n} \left|\mu_j\right\rangle \left|\nu_j\right\rangle$$

for some choice of vectors $|\nu_1\rangle, \ldots, |\nu_n\rangle \in \mathcal{Y}$. This is always possible, following from the fact that $\{|\mu_1\rangle, \ldots, |\mu_n\rangle\}$ is a basis of $\mathcal{X}$. The vectors $|\nu_1\rangle, \ldots, |\nu_n\rangle$ are uniquely determined given that the basis $\{|\mu_1\rangle, \ldots, |\mu_n\rangle\}$ has been fixed: it must be that $|\nu_j\rangle = (\langle\mu_j| \otimes I)|\psi\rangle$. A quick calculation now shows that because

$$\operatorname{Tr}_{\mathcal{Y}} |\psi\rangle\langle\psi| = \rho = \sum_{j=1}^{n} p_j |\mu_j\rangle\langle\mu_j|,$$

it must be the case that

$$\langle\nu_i|\nu_j\rangle = \begin{cases} p_i & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Let us throw away all of the vectors for which $p_j = 0$ and re-number the remaining ones so that

$$\rho = \sum_{j=1}^{k} p_j |\mu_j\rangle\langle\mu_j|$$

for $p_1, \ldots, p_k \neq 0$. We now see that

$$|\psi\rangle = \sum_{j=1}^{k} |\mu_j\rangle|\nu_j\rangle$$

for $\{|\nu_1\rangle, \ldots, |\nu_k\rangle\}$ an orthogonal set of nonzero vectors. By normalizing these vectors, i.e., letting $|\gamma_j\rangle = |\nu_j\rangle / \sqrt{p_j}$ for each $j = 1, \ldots, k$, we have that

$$|\psi\rangle = \sum_{j=1}^{k} \sqrt{p_j}\,|\mu_j\rangle|\gamma_j\rangle$$

for both $\{|\mu_1\rangle, \ldots, |\mu_k\rangle\}$ and $\{|\gamma_1\rangle, \ldots, |\gamma_k\rangle\}$ *orthonormal* sets.

An expression of $|\psi\rangle$ in this form is known as a *Schmidt decomposition*. (It is really only a cosmetic variant of the *singular value decomposition* of matrices, which was not discovered by Schmidt. Nevertheless, people use the term Schmidt decomposition, so that is what we will use.)

You may wonder why this is interesting. It turns out to be an incredibly useful fact, and we will see later some uses for it. Here is one consequence of the above proof.

Fact. Suppose $|\phi\rangle, |\psi\rangle \in \mathcal{X} \otimes \mathcal{Y}$ satisfy $\operatorname{Tr}_{\mathcal{Y}} |\phi\rangle\langle\phi| = \operatorname{Tr}_{\mathcal{Y}} |\psi\rangle\langle\psi|$. Then there exists a unitary operator $U \in \mathrm{L}(\mathcal{Y})$ such that $(I_{\mathcal{X}} \otimes U)|\phi\rangle = |\psi\rangle$.

To see that this fact holds, consider performing the previous construction of the Schmidt decomposition for two pure states $|\phi\rangle$ and $|\psi\rangle$ satisfying $\operatorname{Tr}_{\mathcal{Y}} |\phi\rangle\langle\phi| = \operatorname{Tr}_{\mathcal{Y}} |\psi\rangle\langle\psi|$. Because the reduced state of register X is the same in both cases, you could take $\{|\mu_1\rangle, \ldots, |\mu_n\rangle\}$ and $(p_1, \ldots, p_n)$ to be identical for the two cases. Following through with the construction, you would then have

$$|\phi\rangle = \sum_{j=1}^{k} \sqrt{p_j}\,|\mu_j\rangle|\gamma_j\rangle,$$
$$|\psi\rangle = \sum_{j=1}^{k} \sqrt{p_j}\,|\mu_j\rangle|\gamma_j'\rangle,$$

for two (possibly different) orthonormal sets $\{|\gamma_1\rangle, \ldots, |\gamma_k\rangle\}$ and $\{|\gamma'_1\rangle, \ldots, |\gamma'_k\rangle\}$. There is always at least one unitary transformation U mapping the first set to the second, given that both sets are orthonormal.

It happens to be the case that there are also approximate versions of the above fact, but because we have not discussed meaningful distance measures for quantum states it will not be possible to go into greater detail about this. Expressed informally, the approximate versions are of this form: if

$$\mathrm{Tr}_{\mathcal{Y}} |\phi\rangle \langle\phi| \approx \mathrm{Tr}_{\mathcal{Y}} |\psi\rangle \langle\psi|$$

then there exists a unitary operator $U \in \mathrm{L}(\mathcal{Y})$ such that $(I \otimes U) |\phi\rangle \approx |\psi\rangle$.

Example 3. It was claimed above that all four Bell states result in the totally mixed state when one of the two qubits is traced out:

$$\mathrm{Tr}_{\mathcal{A}} |\phi^+\rangle \langle\phi^+| = \mathrm{Tr}_{\mathcal{A}} |\phi^-\rangle \langle\phi^-| = \mathrm{Tr}_{\mathcal{A}} |\psi^+\rangle \langle\psi^+| = \mathrm{Tr}_{\mathcal{A}} |\psi^-\rangle \langle\psi^-| = \frac{1}{2} I$$

if the spaces are $\mathcal{A}$ and $\mathcal{B}$. This means that they are all purifications of the totally mixed state. It follows that for any two of them, there is a unitary operator acting just on Alice's qubit that maps the first to the second. This is essentially what is happening in super-dense coding—Alice decides to transform $|\phi^+\rangle$ to one of the four Bell states depending on her two classical input bits, sends her qubit to Bob, and he performs a measurement to determine which of the four states it is to determine Alice's classical bits.

Measurements

Finally, the last remaining piece of the general formalism of quantum information is measurements. There are more general types of measurements than the simple type we have discussed previously. Mathematically speaking, they are described in a manner similar to admissible operations.

Suppose that X is a quantum register with classical state set Σ and corresponding vector space $\mathcal{X} = \mathbb{C}(\Sigma)$. A *measurement* on X can have any finite, nonempty set Γ of possible outcomes. Formally, a measurement is described by a collection

$$\{M_a \,:\, a \in \Gamma\} \subset \mathrm{L}(\mathcal{X})$$

of matrices that satisfies

$$\sum_{a \in \Gamma} M_a^\dagger M_a = I.$$

If the register X is measured with respect to this measurement while it is in the state $\rho \in \mathrm{D}(\mathcal{X})$, the outcome must be some element of Γ. As before, the outcome may be random: for each $a \in \Gamma$, the outcome will be a with probability

$$\mathrm{Tr}\left(M_a \rho M_a^\dagger\right).$$

Conditioned on any given outcome a, the density matrix describing the state of X then becomes

$$\frac{M_a \rho M_a^\dagger}{\mathrm{Tr}\left(M_a \rho M_a^\dagger\right)}.$$

Example 4. Suppose X is an n qubit register, so that $\mathcal{X} = \mathbb{C}(\{0,1\}^n)$. Measuring the first qubit with respect to the "old" notion of measurement is now described by the set $\{M_0, M_1\}$ where

$$M_0 = |0\rangle\langle 0| \otimes I \quad \text{and} \quad M_1 = |1\rangle\langle 1| \otimes I.$$

In the special case where each of the matrices M_a of a measurement $\{M_a : a \in \Gamma\}$ is a *projection*, the measurement is said to be a *projective measurement* (or *von Neumann measurement*). (A projection is a matrix M that satisfies $M = M^\dagger$ and $M^2 = M$.) In this case, the number of possible outcomes $|\Gamma|$ can be at most $|\Sigma|$ (the dimension of $\mathcal{X}$). If we have that $M_a = |\psi_a\rangle\langle\psi_a|$ for $\{|\psi_a\rangle : a \in \Gamma\}$ an orthonormal basis of $\mathcal{X}$ (which requires $|\Gamma| = |\Sigma|$), then the measurement is a *complete projective measurement*. In this case we sometimes say that the measurement is *with respect to the basis* $\{|\psi_a\rangle : a \in \Gamma\}$.

When one doesn't care about the state of the register X after the measurement, it is sufficient to know just the operators $\{M_a^\dagger M_a : a \in \Gamma\}$ to determine the probabilities of the various outcomes:

$$\Pr[\text{outcome is } a] = \operatorname{Tr}\left(\left(M_a^\dagger M_a\right)\rho\right).$$

This follows from the fact that $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$ for any choice of matrices A and B. Sometimes these operators are called *POVM elements*, and the measurement is called a POVM (short for positive operator valued measure).