CS 294-2 Spring 2007

Density Matrices, von Neumann Entropy

3/7/07 Lecture 13

In this lecture, we will discuss the basics of quantum information theory. In particular, we will discuss mixed quantum states, density matrices, von Neumann entropy and the trace distance between mixed quantum states.

1 Mixed Quantum State

So far we have dealt with *pure* quantum states

$$|\psi\rangle = \sum_x \alpha_x |x\rangle.$$

This is not the most general state we can think of. We can consider a probability distribution of pure states, such as $|0\rangle$ with probability $1/2$ and $|1\rangle$ with probability $1/2$. Another possibility is the state

$$\begin{cases} |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) & \text{with probability } 1/2 \\ |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) & \text{with probability } 1/2 \end{cases}$$

In general, we can think of *mixed* state $\{p_i, |\psi_i\rangle\}$ as a collection of pure states $|\psi_i\rangle$, each with associated probability p_i, with the conditions $0 \le p_i \le 1$ and $\sum_i p_i = 1$. One context in which mixed states arise naturally is in quantum protocols, where two players share an entangled (pure) quantum state. Each player's view of their quantum register is then a probability distribution over pure states (achieved when the other player measures their register). Another reason we consider such mixed states is because the quantum states are hard to isolate, and hence often entangled to the environment.

2 Density Matrix

Now we consider the result of measuring a mixed quantum state. Suppose we have a mixture of quantum states $|\psi_i\rangle$ with probability p_i. Each $|\psi_i\rangle$ can be represented by a vector in $\mathscr{C}^{2^n}$, and thus we can associate the outer product $|\psi_i\rangle\langle\psi_i| = \psi_i \psi_i^*$, which is an $2^n \times 2^n$ matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \cdots & \bar{a}_N \end{pmatrix} = \begin{pmatrix} a_1\bar{a}_1 & a_1\bar{a}_2 & \cdots & a_1\bar{a}_N \\ a_2\bar{a}_2 & a_1\bar{a}_2 & \cdots & a_2\bar{a}_N \\ \vdots & & & \vdots \\ a_N\bar{a}_1 & a_N\bar{a}_2 & \cdots & a_N\bar{a}_N \end{pmatrix}.$$

We can now take the average of these matrices, and obtain the *density matrix* of the mixture $\{p_i, |\psi_i\rangle\}$:

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

We give some examples. Consider the mixed state $|0\rangle$ with probability of $1/2$ and $|1\rangle$ with probablity $1/2$. Then

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$|1\rangle\langle 1| = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus in this case

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Now consider another mixed state, this time consisting of $|+\rangle$ with probability $1/2$ and $|-\rangle$ with probability $1/2$. This time we have

$$|+\rangle\langle +| = (1/2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and

$$|-\rangle\langle -| = (1/2)\begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Thus in this case the offdiagonals cancel, and we get

$$\rho = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Note that the two density matrices we computed are identical, even though the mixed state we started out was different. Hence we see that it is possible for two different mixed states to have the same density matrix.

Nonetheless, the density matrix of a mixture completely determines the effects of making a measurement on the system:

Theorem 13.1: *Suppose we measure a mixed state $\{p_j, |\psi_j\rangle\}$ in an orthonormal bases $|\beta_k\rangle$. Then the outcome is $|\beta_k\rangle$ with probability $\langle\beta_k|\rho|\beta_k\rangle$.*

Proof: We denote the probability of measuring $|\beta_k\rangle$ by $\Pr[k]$. Then

$$\begin{aligned}
\Pr[k] &= \sum_j p_j |\langle\psi_j|\beta_k\rangle|^2 \\
&= \sum_j p_j \langle\beta_k|\psi_j\rangle\langle\psi_j|\beta_k\rangle \\
&= \left\langle \beta_k \middle| \sum_j p_j |\psi_j\rangle\langle\psi_j| \middle| \beta_k \right\rangle \\
&= \langle\beta_k|\rho|\beta_k\rangle.
\end{aligned}$$

□

We list several properties of the density matrix:

1. ρ is Hermitian, so the eigenvalues are real and the eigenvectors orthogonal.

2. If we measure in the standard basis the probability we measure i, $P[i] = \rho_{i,i}$. Also, the eigenvalues of ρ are non-negative. Suppose that λ and $|e\rangle$ are corresponding eigenvalue and eigenvector. Then if we measure in the eigenbasis, we have

$$\Pr[e] = \langle e|\rho|e\rangle = \lambda\langle e|e\rangle = \lambda.$$

3. $\mathrm{tr}\rho = 1$. This is because if we measure in the standard basis $\rho_{i,i} = Pr[i]$ but also $\sum_i Pr[i] = 1$ so that $\sum_i \rho_{i,i} = \sum_i Pr[i] = 1$.

Consider the following two mixtures and their density matrices:

$$\left.\begin{array}{lll} \cos\theta|0\rangle + \sin\theta|1\rangle \quad \text{w.p. } 1/2 & = \frac{1}{2}\begin{pmatrix} c\theta \\ s\theta \end{pmatrix}\begin{pmatrix} c\theta & s\theta \end{pmatrix} & = \frac{1}{2}\begin{pmatrix} c^2\theta & c\theta s\theta \\ c\theta s\theta & s^2\theta \end{pmatrix} \\ \cos\theta|0\rangle - \sin\theta|1\rangle \quad \text{w.p. } 1/2 & = \frac{1}{2}\begin{pmatrix} c\theta \\ -s\theta \end{pmatrix}\begin{pmatrix} c\theta & -s\theta \end{pmatrix} & = \frac{1}{2}\begin{pmatrix} c^2\theta & -c\theta s\theta \\ -c\theta s\theta & s^2\theta \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

$$\left.\begin{array}{lll} |0\rangle \quad \text{w.p.} \cos^2\theta & = \cos^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} & = \cos^2\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ |1\rangle \quad \text{w.p. } \sin^2\theta & = \sin^2\theta\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix} & = \sin^2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

Thus, since the mixtures have identical density matrices, they are indistinguishable.

3 Von Neumann Entropy

We will now show that if two mixed states are represented by different density measurements, then there is a measurement that distinguishes them. Suppose we have two mixed states, with density matrices A and B such that $A \neq B$. We can ask, what is a good measurement to distinguish the two states? We can diagonalize the difference $A - B$ to get $A - B = E\Lambda E^*$, where E is the matrix of orthogonal eigenvectors. Then if e_i is an eigenvector with eigenvalue λ_i, then λ_i is the difference in the probability of measuring e_i:

$$\mathrm{Pr}_A[i] - \mathrm{Pr}_B[i] = \lambda_i.$$

We can define the distance between two probability distributions (with respect to a basis E) as

$$|\mathscr{D}_A - \mathscr{D}_B|_E = \sum (\mathrm{Pr}_A[i] - \mathrm{Pr}_B[i]).$$

If E is the eigenbasis, then

$$|\mathscr{D}_A - \mathscr{D}_B|_E = \sum_i |\lambda_i| = \mathrm{tr}|A - B| = \|A - B\|_{\mathrm{tr}},$$

which is called the trace distance between A and B.

Claim Measuring with respect to the eigenbasis E (of the matrix $A - B$) is optimal in the sense that it maximizes the distance $|\mathscr{D}_A - \mathscr{D}_B|_E$ between the two probability distributions.

Before we prove this claim, we introduce the following definition and lemma without proof.

Definition Let $\{a_i\}_{i=1}^N$ and $\{b_i\}_{i=1}^N$ be two non-increasing sequences such that $\sum_i a_i = \sum_i b_i$. Then the sequence $\{a_i\}$ is said to majorize $\{b_i\}$ if for all k,

$$\sum_{i=1}^k a_i \geq \sum_{i=1}^k b_i.$$

Lemma[Schur] Eigenvalues of any Hermitian matrix majorizes the diagonal entries (if both are sorted in nonincreasing order).

Now we can prove claim 3.

Proof Since we can reorder the eigenvectors, we can assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Note that $\mathrm{tr}(A-B) = 0$, so we must have $\sum_i \lambda_i = 0$. We can split the λ_i's into two groups: positive ones and negative ones, we must have

$$\sum_{\lambda_i > 0} = \frac{1}{2}\|A-B\|_{\mathrm{tr}} \qquad \sum_{\lambda_i < 0} = -\frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Thus

$$\max_k \sum_{i=1}^{k} \lambda_i = \frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Now consider measuring in another basis. Then the matrix $A-B$ is represented as $H = F(A-B)F^*$, and let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be the diagonal entries of H. Similar argument shows that

$$\max_k \sum_{i=1}^{k} \mu_i = \frac{1}{2}\sum_{i=1}^{n} |\mu_i| = \frac{|\mathscr{D}_A - \mathscr{D}_B|_F}{2}.$$

But by Schur's lemma the λ_i's majorizes μ_i's, so we must have

$$|\mathscr{D}_A - \mathscr{D}_B|_F \leq |\mathscr{D}_A - \mathscr{D}_B|_E = \|A-B\|_{\mathrm{tr}}.$$

Let $H(X)$ be the *Shannon Entropy* of a random variable X which can take on states $p_1 \ldots p_n$.

$$H(\{p_i\}) = \sum_i p_i \log \frac{1}{p_i}$$

In the quantum world, we define an analogous quantity, $S(\rho)$, the *Von Neumann entropy* of a quantum ensemble with density matrix ρ with eigenvalues $\lambda_1, \ldots, \lambda_n$:

$$S(\rho) = H\{\lambda_1, \ldots, \lambda_n\} = \sum_i \lambda_i \log \frac{1}{\lambda_i}$$

4 Two open questions related to NRM

Liquid NMR (Nuclear Magnetic Resonance) quantum computers have successfully implemented 7 qubits and performed a stripped down version of quantum factoring on the number 15. In liquid NMR, the quantum register is composed of the nuclear spins in a suitably chosen molecule - the number of qubits is equal to the number of atoms in the molecule. We can think of the computer as consisting of about 10^{16} such molecules (a macroscopic amount of liquid), each controlled by the same operations simultaneously. Thus we will have 10^{16} copies of our state, each consisting of say 7 qubits. We assume that we can address the qubits individually, so that for example, we could preform an operation such as *CNOT* on the 2nd and 4th qubit (simultaneously on each copy).

The catch in liquid NMR quantum computing is that initializing the register is hard. Each qubit starts out in state $|0\rangle$ with probability $1/2+\varepsilon$ and in state $|1\rangle$ with probability $1/2-\varepsilon$. Here ε depends upon the strength of the magnetic field that the liquid sample is placed in. Using very strong magnets in the NMR apparatus, the polarization ε is still about 10^{-5}.

If $\varepsilon = 0$ then the density matrix describing the quantum state of the register is $\rho = \frac{1}{2^n}I$. This means that if we apply a unitary transformation U, the density matrix of the resulting state is $I \rightarrow_U UIU^\dagger = I$. So you cannot perform any meaningful computation.

The way NMR quantum computation works is this: the initial mixed state (with $\varepsilon = 10^{-5}$) is preprocessed (through a sequence of quantum gates) to obtain a new mixed state which is maximally mixed ($\frac{1}{2^n}I$) with probability $1-\delta$ and $|0000000\rangle$ with probability δ. Now, if we apply a unitary transformation to this state, we get $\frac{1}{2^n}I$ with probability $1-\delta$ and $U|0000000\rangle$ with probability δ. Thus if we measure the state, we obtain a coin flip with bias δ^2 towards the correct answer. Another way of thinking about this is that the $\frac{1}{2^n}I$ gives no net signal in the measurement, while the δ^2 signal gets amplified by the 10^{16} copies of the computation being carried out simultaneously. The problem is that δ is exponentially small in n the number of qubits. Therefore liquid NMR quantum computation cannot scale beyond 10-20 qubits.

Question 1: Say we have a single clean bit (and n maximally mixed qubits), what can we do with this?

We can do at least one quantum computation, phase estimation to approximate the trace of a unitary matrix. Use the single clean qubit as the control bit and apply the (controlled) unitary to the n maximally mixed qubits. We can think of the n qubits as being a uniform mixture over the eigenvectors of the unitary! (see previous lecture for details).

Is there anything else that we can do with just one qubit? Can you prove limits on what can be done with one clean qubit.

Question 2

Let ρ be a density matrix for a mixed state which takes on $|x\rangle$ with probability $(1/2+\varepsilon)^{\#0}(1/2-\varepsilon)^{\#1}$ where #0 and #1 are the number of 0s and 1s in $|x\rangle$ respectively. Therefore, ρ can be written as a probability distribution over unentangled states. Such mixed states are called separable mixed states. It turns out, that if the density matrix ρ' is sufficiently close to the identity matrix I then there always exists some such decomposition into a distribution over unentangled states. Using the bound 10^{-5} for the polarization of qubits in liquid NMR, it turns out that for $n \leq 10$ the initial state of an n-qubit NMR register is a separable mixed state.

So if we apply a unitary transformation U to ρ, we get a new density matrix $\rho' = U\rho U^\dagger$ which is also close to I and thus is also a separable mixed state.

Given that there is no entanglement, can we simulate such a quantum computation efficiently? We don't know how, since to write the state as a separable mixture we might have to perform a change of basis after each quantum gate. On the other hand, we don't know how to obtain any quantum speedup in this model either.