CS294-4 Quantum Computing **Fall 2002**

Lecture 9: November 1

Lecturer: Umesh Vazirani *Scribes: Yozo Hida, Adam Megacz*

Edited by: Ashwin Nayak, May 2003

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, trace distance between mixed quantum states, and bit commitment.

9.1 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 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 reason we consider mixed states is the following. Quantum systems are hard to isolate, and hence are often entangled to the environment. It turns out that the behaviour of the quantum system is then more appropriately described by a mixed state derived from the joint state of the system and the environment.

9.2 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 $\mathbb{C}^{2^n}$, and thus we can associate the outer product $|\psi_i\rangle\langle\psi_i| = \psi_i\psi_i^*$, which is a $2^n \times 2^n$ 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.

We now show that two mixed states can be distinguished if and only if the density matrix ρ are different:

Theorem 9.1 *Suppose we measure a mixed state* $\{p_j, |\psi_j\rangle\}$ *in an orthonormal basis* $\{|\beta_k\rangle\}$. *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

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

■

Corollary 9.2 *If we measure the mixed state* $\{p_j, |\psi_j\rangle\}$ *in the standard basis, we have* $\Pr[k] = \rho_{k,k}$, *the diagonal entry of the density matrix* ρ.

We list several more properties of the density matrix:

1. $\mathrm{tr}\rho = 1$. This follows immediately from Corollary 9.2, since the probabilities $\Pr[k]$ must add up to 1.

2. ρ is Hermitian. This follows from the fact that ρ is a convex sum of Hermitian outer products $((\psi\psi^*)^* = \psi\psi^*)$.

3. Eigenvalues of ρ are non-negative. First of all, eigenvalues of a Hermitian matrix are real. Suppose that λ and $|e\rangle$ are an eigenvalue-eigenvector pair. Then if we measure in the eigenbasis, we have

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

   Since the probability must be non-negative, we see that $\lambda \geq 0$.

Now 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 orthonormal 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:

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

We can define the distance between two probability distributions (with respect to a basis $F = \{|f_i\rangle\}$) as

$$|\mathcal{D}_A - \mathcal{D}_B|_F = \sum |\Pr{}_A[i] - \Pr{}_B[i]|.$$

If F is the eigenbasis given by E, then

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

which is called the trace distance between A and B. Here, $|M|$ is defined as $\sqrt{M^*M}$, for any Hermitian matrix M.

Claim 9.3 *Measuring with respect to the eigenbasis E (of the matrix $A - B$) is optimal in the sense that it maximizes the distance $|\mathcal{D}_A - \mathcal{D}_B|_F$ between the two probability distributions.*

Proof: Let $A - B = E\Lambda E^* = \sum_i \lambda_i |e_i\rangle\langle e_i|$, in diagonal form. Let $F = \{|f_j\rangle\}$ be any orthonormal basis. Then,

$$\begin{aligned}
|\mathcal{D}_A - \mathcal{D}_B|_F &= \sum_j |\langle f_j|A|f_j\rangle - \langle f_j|B|f_j\rangle| \\
&= \sum_j |\langle f_j|(A - B)|f_j\rangle| \\
&= \sum_j \left| \langle f_j| \left(\sum_i \lambda_i |e_i\rangle\langle e_i| \right) |f_j\rangle \right| \\
&\leq \sum_i |\lambda_i| \sum_j \langle f_j|e_i\rangle\langle e_i|f_j\rangle \\
&\leq \sum_i |\lambda_i| \;=\; \|A - B\|_{\mathrm{tr}}.
\end{aligned}$$

We may verify easily that for any incomplete von Neumann measurement there is a complete measurement which gives rise to probability distributions that are at least as far apart as the original distributions:

Let $F = \{P_j\}$ be an incomplete von Neumann measurement given by orthogonal projection operators $P_j = \sum_k |f_{jk}\rangle\langle f_{jk}|$. Then, the difference in the probability of outcome j is

$$\begin{aligned}
|\mathrm{tr}(P_j A) - \mathrm{tr}(P_j B)| &= |\mathrm{tr}[P_j(A-B)]| \\
&= \left|\sum_k \langle f_{jk}|(A-B)|f_{jk}\rangle\right| \\
&\leq \sum_k |\langle f_{jk}|(A-B)|f_{jk}\rangle| .
\end{aligned}$$

So if we measure in the basis $\{|f_{jk}\rangle\}_{j,k}$, the distance between the resulting distributions is at least as large.

The above argument also applies to more general measurements (with the inclusion of some ancillary qubits in a fixed state). Without loss in generality, we may assume that the ancilla in such a measurement is $|\bar{0}\rangle\langle\bar{0}|$. Our task then reduces to distinguishing density matrices $A \otimes |\bar{0}\rangle\langle\bar{0}|$ and $B \otimes |\bar{0}\rangle\langle\bar{0}|$ via a von Neumann measurement. The maximum distance achieved then is $\| A \otimes |\bar{0}\rangle\langle\bar{0}| - B \otimes |\bar{0}\rangle\langle\bar{0}| \|_{\mathrm{tr}} = \| A - B \|_{\mathrm{tr}}$. ■

The above proof may directly be written in terms of general measurments, i.e., in terms of the corresponding POVM's (positive operator valued measurements). Let $F = \{Q_j\}$ be a POVM, so that Q_j are positive semidefinite operators such that $\sum_j Q_j = I$. Then, we may bound the distance $|\mathcal{D}_A - \mathcal{D}_B|_F$ as

$$\begin{aligned}
|\mathcal{D}_A - \mathcal{D}_B|_F &= \sum_j |\mathrm{tr}(Q_j A) - \mathrm{tr}(Q_j B)| \\
&= \sum_j |\mathrm{tr}(Q_j(A-B))| \\
&= \sum_j \left|\mathrm{tr}\left(Q_j \left(\sum_i \lambda_i |e_i\rangle\langle e_i|\right)\right)\right| \\
&= \sum_j \left|\sum_i \lambda_i \langle e_i|Q_j|e_i\rangle\right| \\
&\leq \sum_i |\lambda_i| \sum_j \langle e_i|Q_j|e_i\rangle \\
&\leq \sum_i |\lambda_i| \langle e_i| \left(\sum_j Q_j\right) |e_i\rangle \\
&= \sum_i |\lambda_i| \;=\; \| A - B \|_{\mathrm{tr}} .
\end{aligned}$$

9.3 Bit Commitment

Bit commitment is a cryptographic communication protocol in which one party, Alice, wishes to commit to some value $b \in \{0,1\}$ by exchanging some qubits with another party, Bob. At some later time, she reveals b to Bob. At the time of revelation, she also provides some extra information that allows Bob to verify that her commitment was consistent with the bit she claims to have sent. Bob must gain no knowledge of b before Alice reveals it, and Alice must not be able to change b after committing without being detected.

In 1984, Bennett and Brassard published the first paper on quantum cryptography, outlining algorithms for bit commitment and quantum key distribution. Their algorithm involved two ensembles:

$$I : \begin{cases} \text{with probability } 1/2 \text{ , } |0\rangle \\ \text{with probability } 1/2 \text{ , } |1\rangle \end{cases}$$

$$II : \begin{cases} \text{with probability } 1/2 \text{ , } |+\rangle \\ \text{with probability } 1/2 \text{ , } |-\rangle \end{cases}$$

Alice sends a qubit in mixed state I if $b = 0$, and in state II if $b = 1$. Since the density matrices of the two ensembles are the same, $\rho_I = \rho_{II}$, the two mixed states are indistinguishable. So no information is leaked to Bob. When she reveals her bit, she has to indicate which state she used for encoding her bit (e.g., if $b = 1$, whether she sent a $|+\rangle$ or a $|-\rangle$). If Alice changes her bit in the revelation stage, she will fail Bob's test with probability at least 1/2.

Unfortunately, this algorithm is flawed; Alice can perform an "entanglement attack". She does so by preparing 2 qubits in Bell a state and sending half of the pair as the commitment. I.e., Alice sets her qubit pair in state $(1/\sqrt{2})(|00\rangle + |11\rangle)$, and sends one of the two entangled qubits to Bob. To reveal the bit $b = 0$, Alice measures her half of the Bell pair in the $|0\rangle, |1\rangle$ basis; and to commit to $b = 1$, Alice measures in the $|+\rangle, |-\rangle$ basis. It is easily verified that Bob's state collapses to the state Alice observes, and so Alice can pass Bob's test with probability 1! In fact, no quantum protocol for ideal bit commitment (i.e., as stated above) is possible.

Claim 9.4 *Quantum bit committment is impossible. In fact, these two properties cannot be achieved simultaneously:*

- *Alice stays committed, and*
- *Bob gets no information.*

The probability of commitment and the information leaked are intertwined; increasing one decreases the other.