CS 410/510 Introduction to Quantum Computing

Lecture 13

Portland State U, Spring 2017 — *May 16, 2017*

Lecturer: Fang Song — *Scribe: Enis K. Inan*

VERSION: OCTOBER 9, 2017

Last lecture we introduced the density operator, which made it easier for us to describe general quantum states. We will briefly review it before proceeding any further. In the old notation, we only had pure states of the form $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. With our new framework, a pure state is expressed as $|\psi\rangle\langle\psi|$ where for our example,

$$|\psi\rangle\langle\psi| = \begin{pmatrix} \alpha\overline{\alpha} & \alpha\overline{\beta} \\ \overline{\alpha}\beta & \beta\overline{\beta} \end{pmatrix} = \begin{pmatrix} |\alpha|^2 & \alpha\overline{\beta} \\ \overline{\alpha}\beta & |\beta|^2 \end{pmatrix}$$

The expressive power of this notation is apparent when we have general quantum states. Specifically, imagine we have a register X that could be in one of the quantum states $|\psi_1\rangle, \ldots, |\psi_k\rangle$ with a probability distribution $p_1, \ldots, p_k$ where p_i is the probability that X is in state $|\psi_i\rangle$. With our new notation, we expressed our "knowledge" of X as a density matrix ρ where

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

Intuitively, ρ corresponds to a weighted average of each state. There were also several important properties associated with density matrices:

1. $\mathrm{Tr}(\rho) = 1$
2. ρ is positive semidefinite.

Property 2 will be important later in this lecture, so keep it in the back of your head for now.

In our old notation, the physically allowable operations on $|\psi\rangle$ were unitary operations U – the state $|\psi'\rangle = U|\psi\rangle$ is the resulting state after applying the unitary U on $|\psi\rangle$. In our new notation, $\rho' = U\rho U^\dagger$ corresponds to applying the unitary U on the general quantum state described by ρ – the result is ρ'.

1 General (Physically Admissible) Quantum Operations

The density matrix formalism does not limit us to just unitary operations. Specifically, any *physically admissible operation* Φ is a series of matrices $A_1, A_2, \ldots, A_k$ where $A_{i_{jk}} \in \mathbb{C}$ and

$$\sum_i A_i^\dagger A_i = I \tag{1}$$

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger \tag{2}$$

Note that each A_i in Φ does *not* have to be a square matrix, and that unitary matrices U are a special case of Φ with $k = 1$.
Applying Φ to ρ yields another density matrix, just like how applying unitary U to $|\psi\rangle$ resulted in another valid quantum state. In fact, Φ is also known as a completely positive trace preserving (CPTP) operator, which is just jargon for an operation that preserves a matrix's trace and semi-positive definitiveness. Another way to think of Φ is as follows. Let $\mathcal{X}$ and $\mathcal{Y}$ denote spaces describing m and n qubits, respectively, and let $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ be the set of all linear mappings from $\mathcal{X}$ to $\mathcal{Y}$. Let $\mathcal{D}(\mathcal{X})$ be the set of all density matrices in $\mathcal{X}$; define $\mathcal{D}(\mathcal{Y})$ in a similar manner. Then $\Phi : \mathcal{D}(\mathcal{X}) \to \mathcal{D}(\mathcal{Y})$ or, informally, Φ is the set of all density matrix mappings from $\mathcal{X}$ to $\mathcal{Y}$.

Example 1: Here we investigate decoherence as a physically admissible operation, which is when a quantum system interacts with some environmental noise and loses its quantumness (i.e. gets measured). Note that the operation used in this example is equivalent to measurement in the standard basis $\{|0\rangle, |1\rangle\}$. Here we have $A_0 = |0\rangle\langle 0|$ and $A_1 = |1\rangle\langle 1|$ so that

$$\Phi(\rho) = A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger$$

Firs we check if A_0 and A_1 are valid, i.e. that they satisfy Eqn. 1:

$$\begin{aligned}
\sum_i A_i^\dagger A_i &= A_0^\dagger A_0 + A_1^\dagger A_1 \\
&= |0\rangle\langle 0|0\rangle\langle 0| + |1\rangle\langle 1|1\rangle\langle 1| \\
&= |0\rangle\langle 0| + |1\rangle\langle 1| \\
&= I
\end{aligned}$$

which they do. Now assume that $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, i.e. that ρ corresponds to a pure state. We have,

$$\begin{aligned}
\Phi(|\psi\rangle\langle\psi|) &= A_0|\psi\rangle\langle\psi|A_0^\dagger + A_1|\psi\rangle\langle\psi|A_1^\dagger \\
&= |0\rangle\langle 0||\psi\rangle\langle\psi||0\rangle\langle 0| + |1\rangle\langle 1||\psi\rangle\langle\psi||1\rangle\langle 1| \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} |\alpha|^2 & \alpha\overline{\beta} \\ \overline{\alpha}\beta & |\beta|^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} |\alpha|^2 & \alpha\overline{\beta} \\ \overline{\alpha}\beta & |\beta|^2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix} \\
&= |\alpha|^2 |0\rangle\langle 0| + |\beta|^2 |1\rangle\langle 1|
\end{aligned}$$

which is exactly what we get by doing a measurement on the standard basis! To see why, the last line shows that we can think of $\Phi(|\psi\rangle\langle\psi|)$ as representing the general quantum state where we have an $|\alpha|^2$ chance of being in $|0\rangle$ and a $|\beta|^2$ chance of being in $|1\rangle$. This is the definition of measurement in the standard basis.

Another important operation is the partial trace. Imagine that we have an $m + n$ qubit-state, e.g. $|\psi\rangle = \sum_{x\in\{0,1\}^m} \sum_{y\in\{0,1\}^n} \alpha_{xy}|x\rangle|y\rangle$. We want to examine the resulting state after we discard either x or y. For our old notation, this is easy – for $|\psi\rangle$, we can just remove $|x\rangle$ or $|y\rangle$. But in our new notation of density matrices, it is not as obvious. Let $\mathcal{X}$ and $\mathcal{Y}$ be the spaces describing $|x\rangle$ and

$|y\rangle$, respectively, and assume without any loss of generality that we want to remove $|y\rangle$. Then what we want is an operation $\Phi : \mathcal{D}(\mathcal{X} \otimes \mathcal{Y}) \to \mathcal{D}(\mathcal{X})$ – this is the partial trace, denoted as Tr_y for our specific case to explicitly indicate that we want to discard $|y\rangle$.

Let's examine what Tr_y would look like when $\mathcal{X}$ and $\mathcal{Y}$ both describe 1-qubit each – i.e., when we have a two-qubit state and we want to discard the second qubit. We have Φ be the matrices:

$$A_0 = I_x \otimes \langle 0|_y = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$
$$A_1 = I_x \otimes \langle 1|_y = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Note that A_0 and A_1 are 2×4 matrices – I_x is the 2×2 identity matrix. We first check that A_0 and A_1 satisfy Eqn. 1:

$$\begin{aligned} \sum_i A_i^\dagger A_i &= A_0^\dagger A_0 + A_1^\dagger A_1 \\ &= (I_x \otimes |0\rangle_y)(I_x \otimes \langle 0|_y) + (I_x \otimes |1\rangle_y)(I_x \otimes \langle 1|_y) \\ &= (I_x \otimes |0\rangle\langle 0|_y) + (I_x \otimes |1\rangle\langle 1|_y) \\ &= I_x \otimes (|0\rangle\langle 0|_y + |1\rangle\langle 1|_y) \\ &= I_x \otimes I_y = I_{xy} \end{aligned}$$

Now let's do an example calculation using Φ.

Example 2: Let $\rho = |0\rangle\langle 0| \otimes |1\rangle\langle 1|$. Intuitively, we should get $|0\rangle\langle 0|$ back after applying Φ. Let's see if this is indeed the case.

$$\begin{aligned} \Phi(\rho) &= A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger \\ &= (I_x \otimes \langle 0|_y)(|0\rangle\langle 0| \otimes |1\rangle\langle 1|)(I_x \otimes |0\rangle_y) + (I_x \otimes \langle 1|_y)(|0\rangle\langle 0| \otimes |1\rangle\langle 1|)(I_x \otimes |1\rangle_y) \\ &= (|0\rangle\langle 0| \otimes \cancelto{0}{\langle 0|1\rangle}\langle 1|)(I_x \otimes |0\rangle_y) + (|0\rangle\langle 0| \otimes \langle 1|1\rangle\langle 1|)(I_x \otimes |1\rangle_y) \\ &= (|0\rangle\langle 0| \otimes \langle 1|)(I_x \otimes |1\rangle_y) \\ &= |0\rangle\langle 0| \end{aligned}$$

which is exactly what we expected. In the general case, e.g. when we have $\rho' = \rho \otimes |0\rangle$, then $\Phi(\rho') = \rho$ which is consistent with the physical meaning of Tr_y – take away y and only look at x. We can think of it as "tracing out y", which is where the name "partial trace" comes from.

Now what if the qubits x and y are entangled? What do we get after we apply Tr_y? We will use $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{xy}$ as our example.

Example 3: $\text{Tr}_y(|\phi^+\rangle\langle\phi^+|) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}I$ which can be verified by brute-force computation using Φ, like we did in Example 2. So we see that we get a *maximally mixed state*, i.e. a density matrix where the pure states form an orthonormal basis and each state has a probability $1/n$ of occurring, where n is the dimension of the matrix. Here, $n = 2$. In fact, it can be shown

that Tr_y for *any* Bell state is $\frac{1}{2}I$. Here's why this makes sense. In a Bell state, when we measure the second register, we *collapse* the resulting state to either just $|0\rangle$ with probability $1/2$, or just $|1\rangle$ with probability $1/2$ – exactly our maximally mixed state. For example in $|\phi^+\rangle$, if we measure a 1 in the second register, then the state collapses to $|1\rangle_x$; otherwise, it collapses to $|0\rangle_x$.

The next operation we will consider is an extension of Example 1. Instead of constraining the measurement to a specific basis, e.g. the standard basis, let's generalize it. Imagine that Γ represents the set of all possible outcomes resulting from our measurement and, for an $a \in \Gamma$, M_a is the matrix "capturing" an outcome a. For example in Example 1, $\Gamma = \{0, 1\}$, and $M_0 = |0\rangle\langle 0|$, $M_1 = |1\rangle\langle 1|$. Then if

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

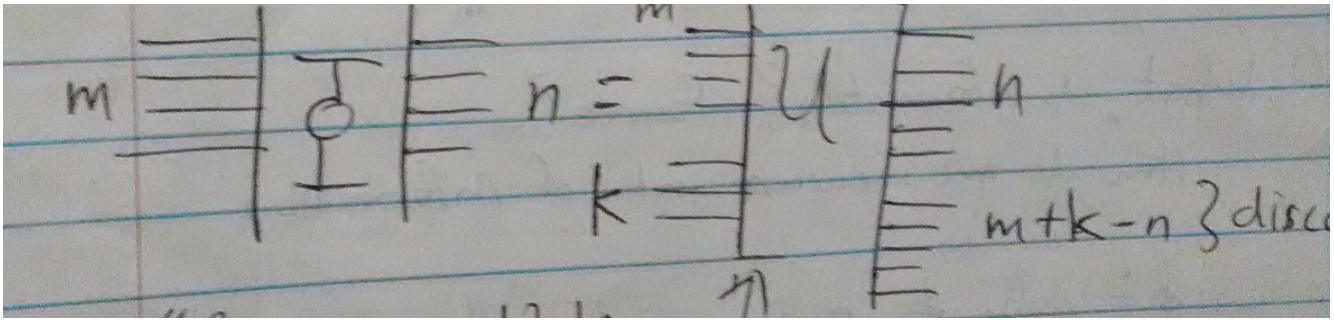
then the set of matrices $M = \{M_a \mid a \in \Gamma\}$ form a physically admissible operation Φ, i.e. Φ is a valid measurement. Let's take an outcome a and its measurement matrix M_a where M_a is a projecter ($M_a^2 = M_a$). Then $\text{Tr}(M_a \rho M_a^\dagger)$ is the probability that we will observe outcome a when we measure the quantum system described by ρ. In Example 1, $\text{Tr}(M_0 \rho M_0^\dagger) = |\alpha|^2$, which is exactly the probability of measuring a 0. Now assume that we did measure a. Then our state would collapse to

$$\frac{M_a \rho M_a^\dagger}{\text{Tr}(M_a \rho M_a^\dagger)}$$

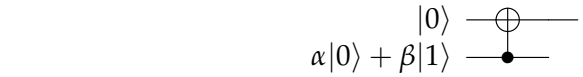
which makes sense. If M_a is the measurement matrix "capturing" an outcome a then when we see a, we should only have $M_a \rho M_a^\dagger$ left in our summation for $\Phi(\rho)$. However $M_a \rho M_a^\dagger$ may not be a valid density matrix, so we need to normalize its trace back down to 1 – this is what the $\text{Tr}(M_a \rho M_a^\dagger)$ in the denominator does. This corresponds to what we do in the pure state case, which is to keep only parts of the state that contain the measured outcome in its qubits ($M_a \rho M_a^\dagger$ here), then normalize that part (divide by $\text{Tr}(M_a \rho M_a^\dagger)$).

The above is known as a Von-Neumann measurement. Another way to think about measurement is through positive operator valued measurements (POVMs) which are the most general class of quantum measurements. In a POVM, we do not care about the resulting state after the measurement; only the probability of a specific outcome a. Above, we determined this to be $\text{Tr}(M_a \rho M_a^\dagger)$. Using a nice property of the trace, $\text{Tr}(AB) = \text{Tr}(BA)$, we see that $\text{Tr}(M_a \rho M_a^\dagger) = \text{Tr}(M_a^\dagger M_a \rho) = \text{Tr}(E_a \rho)$ where $E_a = M_a^\dagger M_a$. Given that $\text{Tr}(E_a \rho)$ is the probability that we will see outcome a, we can think of E_a as a "probability" matrix for outcome a that we multiply with ρ to obtain the probability that we measure a. Then note that the completeness condition, i.e. Eqn. 1, can be restated as $\sum_a E_a = 1$ – intuitively, this is like $\sum_i p_i = 1$ for a probability distribution. So POVM is a generalization of measurement, but in a different flavor.

It would appear on the surface that a physically admissible operator Φ is more powerful than a unitary U because U is just Φ with $k = 1$, i.e. Φ has more matrices that we're applying on ρ. Surprisingly according to Stinespring's Dilation Theorem, this is not the case - unitaries and physically admissible operators are *equally* powerful. Proving this result is beyond the scope of this class, but the basic idea is that we can simulate any circuit implementing Φ with a corresponding unitary circuit using some ancilla bits as input to the latter. At a high-level, the diagram looks like the following:



Example 4: As an example, let's simulate Φ as described in Example 1. The following circuit will do the trick:



so we see that we've introduced an extra ancillary bit and applied the CNOT gate to do the measurement, discarding the ancilla at the end.

2 Quantum Information Theory

So far we talked about the no-cloning theorem (1), distinguishing between two quantum states (2), entanglement (3), and the density matrix formalistm (4). An example of (2) was given in HW. 1 Problem 4. Here Alice and Bob are physically separated, each given one of the qubits of some 2-qubit state; the 2-qubit state is either State I or State II. The problem asked us to devise a strategy they could use to distinguish whether they're sharing State I or State II given that they are allowed to make only local measurements and communicate across a classical channel. An example of (3) was quantum teleportation where we showed that through entanglement, Alice can transmit a 1-qubit state to Bob.

In both classical and quantum information theory, we have three important issues that we would like to consider. These are:

0. What is the information comprised of? What is its source? How do we measure the *amount* of information that's being transmitted?
    - In the classical world, the source is a random variable X. We measure the amount of information using the Shannon entropy, $\mathrm{H}(X)$.
    - In the quantum world, the source is a mixed state ρ. We measure the amount of information using the Von Neumann entropy, $\mathrm{S}(\rho)$.
1. How do we transmit information over a noiseless channel? To put another way, if we store the information on a physical device, then what is the optimal compression that's possible for storing this information source?
    - Given an m-bit source, if we want to represent it using only $\alpha \bullet m$ bits where $\alpha < 1$, then Shannon's noiseless source coding theorem says that we can only do so iff $\alpha \geq H(X)$
    - Given a density matrix ρ on m-qubits, if we want to represent it using only $\alpha \bullet m$ qubits where $\alpha < 1$, then Schumacher's noiseless quantum source coding theorem says that we can only do so iff $\alpha \geq S(\rho)$.

2. How do we transmit information over a noisy channel?
    - We can do so using error correction codes, ECC.
    - We can do so using quantum error correction codes, QECC.

Imagine that we have a classical source churning out m-bit binary strings that are being transmitted on a quantum channel. The information transfer can be summarized by the following steps for a given n-bit binary string x:

1. The classical source churns out x
2. Some kind of process encodes x as a $\alpha \bullet n$ qubit quantum source ρ_x, where $\alpha \leq 1$.
3. ρ_x is transmitted across the quantum channel.
4. ρ_x is decoded at the end to get back x.

The question here is whether we can use a shorter quantum message to transmit our classical message, i.e. if we can find an $\alpha < 1$. Unfortunately, we're stuck with $\alpha = 1$ due to Holevo's theorem, which states that it is impossible to communicate more than n bits of classical information when transmitting n qubits.

3 Entropy

3.1 Shannon Entropy

We think of information as some kind of a message but, as counter-intuitive as it might be, a message may not necessarily contain any "information". For example, if all participating parties know the exact contents of the message that's being transmitted, then anybody who opens to read the message does not gain any new "information" – they already know what to expect! So really, it is better to think of "information" as the *amount of uncertainty* that's present in a message. The more uncertainty we have in our message, the more "information" it contains. For example, an e-mailed job offer would tell you a lot of information because, in general, there are a lot of factors that determine whether you will get the job or not and hence, a lot of uncertainty on whether you will get said job.

In physics, entropy is a quantity used to capture the amount of uncertainty in a given system. Similarly, the *Shannon Entropy* quantifies the amount of uncertainty in a given piece of information – that is why we use the term entropy. Mathematically it is defined as follows. Say we have a source that outputs a random variable $X \in \{0,1\}^m$ with probability p_x. Then we define the Shannon Entropy $\mathrm{H}(x)$ as:

$$\mathrm{H}(X) = -\sum_x p_x \lg p_x \tag{3}$$

where we define $0 \lg 0 = 1$.

Example 5: Let $X \in \{0,1\}$ where $p_0 = 0$ and $p_1 = 1$. Note that there is no uncertainty involved in X – we know that our source will always transmit a 1. Thus we expect that $\mathrm{H}(X) = 0$. Doing the

calculation:

$$\begin{aligned} \mathrm{H}(X) &= -\sum_x p_x \lg p_x \\ &= -1 \lg 1 \\ &= 0 \end{aligned}$$

Example 6: Let $X \in \{0,1\}$ where $p_0 = p_1 = 1/2$, i.e. X corresponds to a fair coin. Then we see that:

$$\begin{aligned} \mathrm{H}(X) &= -\sum_x p_x \lg p_x \\ &= -\left(\frac{1}{2}\lg\frac{1}{2} + \frac{1}{2}\lg\frac{1}{2}\right) \\ &= -(-\frac{1}{2} + -\frac{1}{2}) \\ &= 1 \end{aligned}$$

which is exactly 1-bit of information. This is the mathematical definition of a bit.

Generalizing Example 6, if we have $X \in \{0,1\}^m$ and $H(X) = m$, then this corresponds to m fair coins each with 1-bit of information for a total of m-bits of information. Note that the Shannon Entropy is non-negative, bounded by the size of It also works well in a syntactic sense in that we can take many independent and identically distributed random variables

3.2 Von Neumann Entropy

Imagine that we have an m-qubit space $\mathcal{X}$, and a source that outputs k pure states with probability p_k for each. Then our source is summarized by the density matrix $\rho_x = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. First, note that the quantum source subsumes the classical source as any classical source with $X \in \{0,1\}^m$ and associated probability p_x is just the densitry matrix $\rho' = \sum_{x\in\{0,1\}^m} p_x |x\rangle\langle x|$. Let us try to come up with a way to describe the amount of information contained in the density matrix, i.e. its entropy. At a glance, it might seem like we could use H defined above. Letting $\mathrm{E}(\rho_x)$ denote the entropy quantity of ρ_x, we get:

$$\mathrm{E}(\rho_x) = -\sum_i p_i \lg p_i$$

However this definition does not work because two different quantum sources can have the same entropy – entropy is meant to be unique. For example, the quantum source S_1 that outputs $|0\rangle$ with probability $1/2$ and $|1\rangle$ with probability $1/2$ is certainly different than the quantum source S_2 that outputs $|0\rangle$ with probability $1/2$ and $|+\rangle$ with probability $1/2$ – the density matrices are not the same! However the probabilities of all the possible pure states in each source are the same, so $\mathrm{E}(S_1) = \mathrm{E}(S_2)$. The intuition is that in the quantum world, we have different bases that we can output our information in such as the standard and Hadamard bases. Further, there is no requirement that the pure states outputted by our source be orthogonal or even form an orthonormal basis – S_2 is an example of one such source. In the classical world, however, everything is in binary – all our possible sources will output information that's the same

representation, so it is enough to just examine the probabilities. But the Shannon Entropy is too limited in the quantum world because we have sources that output different kinds of pure states together.

Fortunately, there is a way we can resolve our problem and that is by using the spectral decomposition. Because ρ is positive semi-definite, it has a unique decomposition under the eigenbasis which is:

$$\rho = \sum_{i=0}^{d-1} \lambda_i |\psi_i\rangle \langle \psi_i|$$

where d is the dimension of ρ's space, λ_i is an eigenvalue, and $|\psi_i\rangle$ is an eigenvector such that $\rho|\psi_i\rangle = \lambda_i|\psi_i\rangle$ and $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. Note that $\lambda_i \geq 0$. Because $\mathrm{Tr}(\rho) = 1$, we see that $\sum_i \lambda_i = 1$ so the eigenvalues form a valid probability distribution. Then we define the Von-Neumann entropy S as:

$$\mathrm{S}(\rho) = -\sum_{i=0}^{d-1} \lambda_i \lg \lambda_i \tag{4}$$

Example 7: Consider the source summarized by $\rho = |0\rangle\langle 0|$, i.e. a source that always outputs the pure state $|0\rangle$. We would expect that $\mathrm{S}(\rho) = 0$ since there is no uncertainty in what state we get. This is indeed the case, as $\mathrm{S}(\rho) = -1 \lg 1 = 0$, i.e. $\lambda = 1$ is the only eigenvalue of ρ.

Example 8: Consider a fair coin, i.e. the source where $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. Here, $\mathrm{S}(\rho) = 1$ which matches the corresponding classical case. In fact, if we have a mixed state on n-qubits where $\rho = \frac{1}{2^n} I$, then $\mathrm{S}(\rho) = n$.