CS 294 Quantum Random Access Codes and Applications 4/1/09
Spring 2009 Lecture 17

1 Quantum Information Theory

1.1 Density Matrices

We saw that the state of an isolated quantum system of n particles is described by a unit vector $\left|\Psi\right\rangle$ in a 2^n-dimensional Hilbert space. However, in general a system that is not isolated (i.e., it could be entangled with the environment) is called a *mixed state*: it is a probability distribution $\{p_i\}$ over pure states $\{\left|Psi_i\right\rangle\}$, described by its *density matrix* $\rho = \sum_i p_i|\Psi_i\rangle\langle\Psi_i|$. A pure state simply has $p_1 = 1$ and $p_i = 0$ for $i > 1$. In general, a density matrix is a positive matrix with trace 1. One can always diagonalize it as $\rho = \sum_i \lambda_i|\Psi_i\rangle\langle\Psi_i|$ where the λ_i are positive eigenvalues that sum to 1 because of the trace condition. Note that a given density matrix can have different diagonalization bases, and thus different interpretations as a mixture of pure states.

1.2 Quantum Channels

Suppose we have a quantum channel over which we can send n entangled qubits. How much information can be sent over this quantum channel?

Let $X = x_1x_2\cdots x_m$ be the m-bit random variable which we want to send across the channel. Our general procedure will be to encode X into the n-dimensional quantum state ρ_X and send this across the channel. The person on the other side adds an ancilla to get $\rho_X \otimes \left|0^{m-n}\right\rangle\langle 0^{m-n}|$, and then measures in some basis to get an m-bit random variable Y. We would like for Y to be close to X in some way. One measure of how close X and Y are is the mutual information, denoted by $I(X:Y)$, defined as

$$I(X:Y) = H(X) + H(Y) - H((X,Y)) \tag{1}$$

Let us now see what we can prove about the mutual information between X and Y for the quantum channel above. More generally, suppose that each X appears with probability p_X and is encoded into a mixed quantum state ρ_X. Then, $\rho = \sum_X p_X\rho_X$ is the density matrix of the state sent across the channel.

Then we have the following,

Theorem 17.1: **(Holevo)**

$$I(X:Y) \leq S(\rho) - \sum_x p_x S(\rho_x) \tag{2}$$

where S indicates the Von Neumann entropy: $S(\rho) = \sum_i \lambda_i \log(1/\lambda_i)$, for λ_i the eigenvalues of ρ.

Given that the Von Neumann entropy is a nonnegative quantity, it immediately follows that $I(X:Y) \leq S(\rho)$. But $S(\rho) \leq n$ as ρ describes a state on n qubits. Hence, $I(X:Y) \leq n$. So if X is uniform and $Y = X$, then $m \leq n$. Of course, this is no better than we can achieve classically.

The proof of Holevo's bound is slightly technical. The simple intuition is that, by a quantum equivalent of the data processing inequality, the mutual information between X and ρ should be an upper bound to $I(X:Y)$. Then, if we are willing to believe that the Von Neumann entropy behaves similarly to its classical counterpart:

$$I(X:Y) \leq S(\rho) - S(\rho|X) = S(\rho) - \sum_x p_x S(\rho_x)$$

1.3 Nayak's bound

In this section we prove a weaker statement than Holevo's bound, which is much simpler to prove fully:

Theorem 17.2: **(Nayak)** *If X is a m bit binary string, we send it using n qubits, and decode it via some mechanism back to an m bit string Y, then our probability of correct decoding is given by*

$$Pr[X = Y] \leq \frac{2^n}{2^m} \tag{3}$$

This shows that any encoding using a number of qubits much smaller than m will be "really bad".

Proof: Say x gets mapped to $|\phi_x\rangle$. Consider the message space as being a $C_2^{\otimes n}$ subspace of the full decoding space. So each $|\phi_{x_0}\rangle$ is mapped to some $|e_{x_0}\rangle$. Then, since the decoder is measuring in an orthonormal state

$$\begin{aligned} Pr[X = Y] \quad &= \tfrac{1}{2^m} \textstyle\sum_x ||P_x|\phi_x\rangle||^2 \\ &= \tfrac{1}{2^m} \textstyle\sum_{x,j} ||\langle \phi_x | e_{x,j} \rangle||^2 \\ &\leq \tfrac{1}{2^m} \textstyle\sum_{x_j} ||Q|e_{x,j}\rangle||^2 \end{aligned}$$

where Q is some projection back into the message space, $e \longrightarrow \phi$. The reverse projection will be at least as long.

Now pick an orthonormal basis $|f_1\rangle ... |f_{2^n}\rangle$ of the message space. The projection equals

$$\begin{aligned} \frac{1}{2^m} \sum_{x_j} ||Q|e_{x,j}\rangle||^2 \quad &= \tfrac{1}{2^m} \textstyle\sum_{x,j} \sum_i^{2^n} ||\langle e_{x,j} | f_i \rangle||^2 \\ &= \tfrac{1}{2^m} \textstyle\sum_i^{2^n} \sum_{x,j} ||\langle e_{x,j} | f_i \rangle||^2 \end{aligned}$$

But $\sum_{x,j} ||\langle e_{x,j} | f_i \rangle||^2 = 1$ because the f_i were unit vectors, which leaves us with

$$\begin{aligned} Pr[X = Y] &\leq \frac{1}{2^m} \sum_i^{2^n} 1 \\ &= \frac{2^n}{2^m} \end{aligned}$$

□

2 Random Access Codes

We have just seen that quantum does not buy us much in communication as transmitting m bits of information requires a quantum channel on $O(m)$ qubits. Consider however the scenario where the decoding party is not interested in the whole string $X_1, \cdots, X_m$ but in an arbitrary bit X_i, with i unknown to the encoder. The question is: can we recover X_i with probability $p \geq 1/2 + \varepsilon$ for any i and still use very few qubits, i.e. $n = o(m)$?

Figure 1: A quantum random access code with m=2, n=1

Notice that the previous bound does not apply to this case. In particular, the question above does not pose itself in the classical case: in fact, if we are able to recover every single bit, we must be able to recover the whole string. However, in the quantum case, a single measurement to decode one bit will collapse the state and make it impossible to read off other bits.

2.1 Construction for m=2, n=1

Figure 2.1 shows how to encode each x into a quantum state $f(x)$, such that we can retrieve x_1 by measuring along the basis u and x_2 by measuring along the basis v. Notice that the angle between each vector in the image is $\pi/8$, so that the probability of success is $\cos^2 \pi/8 \approx .85$ for both bits.

2.2 Lower bound for random access codes

Is it possible to extend this construction to higher values of m? Notice that it is possible to show that there exists a family of c^M almost orthogonal states (with inner product $< \varepsilon$) in $\mathscr{C}^M$ (where $c > 1$ is a universal constant). For $M = 2^m$, this might suggest that we could get an exponential compression in the quantum encoding. However, it turns out that mutual orthogonality does not suffice to ensure the robust decoding of every bit required by random access codes. Indeed, we are going to prove the following lower bound:

$$n \geq m(1 - H(p))$$

The following lemma is a consequence of Holevo's bound:

Lemma 17.1: *Let σ_0 and σ_1 be different mixed states. Suppose that there exists a measurement yielding $b \in \{0,1\}$ on σ_b with probability at least p. Moreover, let $\sigma = \frac{1}{2}(\sigma_0 + \sigma_1)$. Then:*

$$S(\sigma) \geq \frac{1}{2}(S(\sigma_0) + S(\sigma_1)) + (1 - H(p))$$

Proof: Let the random variable Y be the output of the measurement and take b to be uniform over $\{0,1\}$. By Holevo's bound:

$$I(b:Y) \leq S(\sigma) - \frac{1}{2}(S(\sigma_0) + S(\sigma_1))$$

But we know that $Pr[b = Y] \geq p$, which implies that $I(b:Y) \geq 1 - H(p)$. By combining the two inequalities and rearranging terms we obtain the theorem. □

Now we are ready to prove the lower bound. Let

$$\rho = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \rho_x$$

For $y \in \{0,1\}^k$, $k < n$, let:

$$\rho_y = \frac{1}{2^{n-k}} \sum_{x \in \{0,1\}^{n-k}} \rho_{xy}$$

We use induction on k to prove that $S(\rho_y) \geq (1 - H(p))(m - k)$. This is trivial for $k = m$, as $S(\rho_y) \geq 0$, by definition of Von Neumann entropy. Assuming the hypothesis for $k+1$, notice that $\rho_y = \frac{1}{2}(\rho_{0y} + \rho_{1y})$. By the lemma and the inductive hypothesis:

$$S(\rho_y) \geq \frac{1}{2}(S(\rho_{0y}) + S(\rho_{1y})) + (1 - H(p)) \geq$$

$$\frac{1}{2}(1 - H(p))2(m-k-1) + (1 - H(p)) = (1 - H(p))(m-k)$$

as required. Finally, the fact that $S(\rho) \leq n$ yields the lower bound.

3 Applications

3.1 Locally decodable codes

Assume that we want to encode n classical bits into m classical bits, such that any of the original bits can be recovered by looking up a constant number of bits of the encoding. A way to do this is given by the Hadamard code, which encodes $x \in \{0,1\}^n$ into $\{x \cdot y\}_{y \in \{0,1\}^n} \in \{0,1\}^{2^n}$. To recover x_i, simply pick a y at random, query $x \cdot y$ and $x \cdot (y + e_j)$ (where e_j is the indicator vector of position j), and add them mod 2. The drawback of this encoding is that it uses an encoding of exponential size $m = 2^n$. Is it possible to do better?

Katz and Trevisan proved that $m > n^{1+1/(q-1)}$, for any LDC that makes at most q queries. Kerenidis and de Wolf much improved this bound for the $q = 2$ case by showing, using a quantum argument, that any 2-query LDC must have $m = \Omega(exp(n))$. We sketch their proof below; it uses a reduction to random access codes before applying the bound on random access codes that we saw in the previous section.

Katz and Trevisan showed that any 2-query LDC can be transformed into a normal form (called a *smooth code*), in which for each $i \in [n]$ there exists a matching $M_i \subset [m] \times [m]$ of size $M_i = \Omega(m)$, such that the following protocol decodes successfully:

1. Pick a random edge $e = (u,v) \in M_i$,
2. Query u and v from the codeword, get y_u and y_v,
3. Output $f(y_u, y_v)$, where f is a known boolean function that depends on i, u, v.

We now show how to construct a quantum random access code from this normal form. Consider the encoding $x \mapsto |\Psi_x\rangle = \sum_{i \in \{0,1\}^m} (-1)^{y_i} |i\rangle$, where y is the encoding of x according to the LDC. Note that $|\Psi_x\rangle$ is a state on $\log m$ qubits. Assume we want to recover x_i from ρ_x. Using an ancillary register, compute for each j an identifier e_j for the edge in M_i that contains j; if j does not appear in M_i then write a special symbol. Measure the ancillary register; with constant probability we will get $1/\sqrt{2}((-1)^{y_u}|u\rangle + (-1)^{y_v}|v\rangle) \otimes |e_{(u,v)}\rangle$. Discard the second register, rename the first registers and assume for simplicity that we have an extra state in the superposition (this can be achieved by a simple modification of the original encoding), so that our state is now

$$\frac{1}{\sqrt{3}}(|0\rangle + (-1)^{y_u}|1\rangle + (-1)^{y_b}|2\rangle)$$

Now consider the four orthogonal states

$$|\Psi_{a,b}\rangle = \frac{1}{2}(|0\rangle + (-1)^a|1\rangle + (-1)^b|2\rangle + (-1)^{a+b}|3\rangle)$$

Measuring in that basis, one sees that the probability of obtaining the correct outcome $|\Psi_{y_u,y_v}\rangle$ is $3/4$, while each of the three remaining possibilities has equal probability $1/12$. We have thus constructed a random access code for x using $\log m$ qubits, which implies by Nayak's bound that $\log m = \Omega(n)$, yielding an exponential lower bound on 2-query LDCs.

3.2 Quantum state tomography

Assume we have a state $|\Psi\rangle = \sum_x \alpha_x |x\rangle$ over n qubits in mind. Suppose that we want a concise classical description of $|\Psi\rangle$ that enables us to predict the outcome of any measurement performed on $|\Psi\rangle$. In general, this would require a full description of all amplitudes, necessitating $\Omega(2^n)$ bits. However, imagine we are only concerned with a more modest goal: being able to estimate the outcome of *most* measurements on $|\Psi\rangle$, where most is defined with respect to an unknown, arbitrary distribution D. In this section we sketch the striking result due to Aaronson that such a learning task is possible using only $t = O(n/\operatorname{poly}(\varepsilon))$ samples obtained from measurements on $|\Psi\rangle$.

To give a simplified proof sketch, consider only 2-outcome measurements, and assume that we are only interested in being able to answer the question "given a measurement, is one of the two outcomes more likely than the other?". The main idea is to use the learning-theoretic concept of VC dimension, which is a measure of the complexity of the class that we are trying to learn: it is known that a class with VC dimension d requires a number of samples that is linear in d in order to be learned. So we only need to upper-bound the VC-dimension. Imagine that this dimension is d: this means that there exist d measurement operators $E_1, \ldots, E_d$ such that, for every set of outcomes $o_1, \ldots, o_d \in \{0,1\}^d$, there is a state $|\Psi_{o_1,\ldots,o_d}\rangle$ on n qubits that gives these outcomes. But this means that those states form a random access code for d-bit strings! Applying our bound on the dimension of random access codes, we get $n = \Omega(d)$, or equivalently $d = O(n)$, which means that a linear number of samples is sufficient.