CMSC 657: Introduction to Quantum Information Processing
Lecture 24

Instructor: Daniel Gottesman

Fall 2024

1 Quantum Data Compression

Now suppose we want to send qubits. We can consider a source producing pure state messages which are given to Alice. Alice does not know what the individual messages are, but she does know what the distribution is: She gets $|\psi_i\rangle$ with probability p_i. We can describe this source with a density matrix

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1}$$

Again, we imagine a situation where an i.i.d. source of this type produces n such messages that Alice must send to Bob. In this case, our goal is that the final state decoded by Bob should have a high fidelity to the initial state. This should be true in the most general case. In particular, suppose the source is actually handing Alice not pure states but parts of entangled states. The part of the state given to Alice still has density matrix ρ, and each such density matrix is purified using a reference system R. Given this purification, someone holding R can prepare any ensemble $\{p_i, |\psi_i\rangle\}$ for Alice by making an appropriate measurement on R. This already tells us that whatever Alice is doing can only depend on ρ and not on the precise decomposition.

The goal, then, is that after compression and decoding, the density matrix of Bob and R should have high fidelity to the initial state of Alice and R. In particular, as $n \to \infty$, the entanglement fidelity $\to 1$. If $|\Psi\rangle_{AR}$ is the initial state of Alice and R, $\mathcal{E}$ is Alice's encoding map, and $\mathcal{D}$ is Bob's decoding map, then the entanglement fidelity is $F(|\Psi\rangle_{AR}\langle\Psi|, (\mathcal{D}_A \otimes I_R) \circ (\mathcal{E}_A \otimes I_R)(|\Psi\rangle_{AR}\langle\Psi|))$.

We can achieve an asyptotically good entanglement fidelity via *Schumacher compression*. For this protocol, Alice simply works in a basis in which ρ is diagonal, with diagonal entries q_j (which don't have to match the p_i's). Then she performs a reversible version of the classical compression protocol for $\{q_j\}$, and makes sure to do it coherently and with no left-over scratch bits. Bob works in the same basis to uncompress the state. The typical strings of basis states are decoded correctly, and there is only a small chance for an atypical string of basis states, so the final state does indeed have high fidelity to the correct state.

There are also universal compression schemes where Alice and Bob do not even need to know ρ. Basically, Alice makes some gentle measurements on the states she is getting and eventually gets enough information to learn ρ to whatever accuracy is desired. Then she can run Schumacher compression along with a few bits that tell Bob what ρ is. There is substantial extra overhead, but it is negligible in the limit $n \to \infty$.

2 Channel Capacity

Shannon's other major theorem of information theory is called the Channel Coding Theorem, and it talks about error correction:

Theorem 1 (Shannon's Channel Coding Theorem)**.** *Suppose we have a scheme to send n bits through a noisy channel with i.i.d. noise and with asymptotic logical error rate $\to 0$ as $n \to \infty$. Then the scheme uses*

at least $n/C + o(n)$ bits transmitted through the channel. C is the channel capacity *of the noisy channel and is given by* $\max_{input\ distributions\ X} I(X : Y)$*, with Y the probability distribution given by applying the noisy channel to X and $I(X : Y) = H(X) + H(Y) - H(X, Y)$ the* mutual information *of X and Y. Moreover, there exist encoding schemes using only $n/C + o(n)$.*

Basically, we can send C bits per use of the channel. The way to do this, of course, is error-correcting codes, and the Channel Coding Theorem tells us there exist codes that correct a constant fraction of typical errors and encode data at a constant rate. The subsequent 75 years of work on channel capacities has in large part focused on finding efficient codes that come close to the channel capacity for various channels of interest, with the complexity of encoding and decoding being another important factor.

We can similarly define the *quantum channel capacity* as the highest rate with which we can send quantum information through a noisy quantum channel via quantum error-correcting codes and still come out with asymptotically high fidelity. We need to find something that will play the role of mutual information. The obvious definition of quantum mutual information (using von Neumann entropies instead of Shannon entropies) is not well-behaved. For instance, suppose the state of system XY is a maximally entangled state $|00\rangle + |11\rangle$. Then $S(X, Y) = 0$ since it is a pure state, but $S(X) = S(X) = 1$ since $\rho_X = \rho_Y = I/2$. The mutual information is then 2, but we don't expect that we can transmit 2 qubits of quantum information by sending only 1 qubit through the channel.

The correct quantity is something called the coherent information, but there is a major complication in the quantum case, meaning we don't have as nice a result as the classical one.

Definition 1. *Let $\mathcal{E}$ be a noisy quantum channel. Let ρ be a quantum state. The* coherent information *$I(\rho, \mathcal{E})$ for ρ and $\mathcal{E}$ is given as follows: Let $|\psi\rangle_{AR}$ be a purification of ρ with a reference system R. Then*

$$I(\rho, \mathcal{E}) = S(\mathcal{E}(\rho)) - S((\mathcal{E} \otimes I)(|\psi\rangle_{AR}\langle\psi|)). \tag{2}$$

The second term is $S(AR)$ after the application of the channel. Another way of formulating this expression is to note that we can purify the channel as well with a system E. Then after the channel, the whole system ARE is in a pure state, so $S(AR) = S(E)$ after the channel. Thus, $I(\rho, \mathcal{E}) = S(\mathcal{E}(\rho)) - S(E)$.

Then the quantum channel capacity is given by the following formula:

$$\lim_{n\to\infty} \sup_{\rho} \frac{1}{n} I(\rho, \mathcal{E}^{\otimes n}). \tag{3}$$

The supremum is taken over states ρ on the n-qubit Hilbert space. The big difference between this and the classical formula (known as a "single letter formula") is that the quantum formula requires us to consider arbitrarily large systems, making it impossible to compute the channel capacity in pratice for most channels. The need for this comes out of the possibility of degenerate quantum codes — basically, it is possible to get better rates in some cases by using degenerate codes. In the context of the above formula, it is sometimes better to use entangled ρ rather than tensor product ρ. The upshot is that the coherent information is *superadditive*, namely $\max_\rho I(\rho, \mathcal{E} \otimes \mathcal{F}) \geq \max_{\rho_A} I(\rho_A, \mathcal{E}) + \max_{\rho_B} I(\rho_B, \mathcal{F})$, and there are cases (even when $\mathcal{E} = \mathcal{F}$) that the inequality is strict.