CMSC 657: Introduction to Quantum Information Processing
Lecture 26

Instructor: Daniel Gottesman

Fall 2024

1 Entanglement and Channel Capacities

1.1 Negativity

Another measure of entanglement is the *negativity.* Negativity doesn't have an operational interpretation (or at least no standard one), but unlike the previous measures, it is easy to calculate. The partial transpose ρ^{T_A} is a map that is positive but not completely positive. It is thus not a physical operation, but we can do it mathematically. In particular, since it is positive, when you apply partial transpose to a tensor product state, you get a positive state. However, since it is not completely positive, if you apply it to some entangled states, you get "states" that have negative eigenvalues for the "density matrix". The sum of the negative eigenvalues is the negativity of the state. It is an entanglement monotone, and moreover it is additive and easy to calculate. Unfortunately, though, the negativity is not faithful. There are some states that have 0 negativity but are not separable. Actually, such states must be bound entangled states (and that is how we know that bound entangled states exist), because a successful distillation procedure produces maximally entangled states, which have non-zero negativity. Since negativity cannot increase under LOCC, no such protocol is possible.

1.2 Distillable Entanglement and Quantum Channel Capacity

Entanglement distillation protocols where Alice sends information to Bob but not vice-versa are related to QECCs: We can think of the mixed state as the result of sending a maximally entangled state over a noisy channel. If Alice and Bob succeed in producing maximally entangled states, they can then use them to transmit qubits. On the other hand, Alice could try to teleport her state (encoded in a QECC) to Bob immediately via these noisy entangled states, but those with noise will cause errors in the qubits. The channel produced is not necessarily identical to the original noisy channel, but it is true that in many cases, the one-way distillable entanglement (produced by an LOCC protocol with one-way communication) is equal to the quantum channel capacity. However, in general the *two-way* distillable entanglement (given via general LOCC protocols) is strictly higher than the one-way distillable entanglement.

1.3 Two-Way Capacity and Entanglement Distillation

One can define a capacity for distilling entanglement similar to the channel capacity. That is, Alice and Bob share some joint state $\rho^{\otimes n}$ and want to get something close to $(|00\rangle + |11\rangle)^{\otimes k}$. What is the maximum k they can achieve with high probability in the limit of $n \to \infty$? We can think of this as a kind of channel capacity by letting $\rho = (I \otimes \mathcal{E})(|00\rangle + |11\rangle)$.

What they can achieve depends on the rules and what they can do. If they can't communicate, then they are very limited. If Alice can send classical information freely to Bob, but not vice-versa, this is very closely related to the quantum capacity (the "one-way" capacity), since Alice can then send quantum information to Bob by first distilling entanglement, then teleporting. If Alice and Bob can communicate in both directions classically, then this is the “two-way” capacity, which is higher in general than the one-way capacity. You can also consider tradeoffs where the classical communications is not free and unlimited.

1.4 Classical Capacity of a Quantum Channel

Another interesting question is how much classical information can we send using a noisy quantum channel? We know that using superdense coding, with prior shared entanglement, we can send two bits per qubit sent through a noiseless channel. In this scenario, however, we will assume no prior shared entanglement and a noisy quantum channel. (You can also consider the case of entanglement + noisy channel.)

Let us consider a single-shot encoding protocol, i.e., one message to one use of the channel. The encoding procedure can then be considered as a set of states ρ_i: When Alice wishes to send i, she prepares the state ρ_i. The channel $\mathcal{N}$ then transforms these states into $\sigma_i = \mathcal{N}(\rho_i)$. Bob receives these and we want to know how much information he can extract about i. Bob then makes some measurement, getting a classical random variable. If Alice's classical input distribution is X and Bob's output distribution is Y, the mutual information between X and Y is bounded by the *Holevo chi quantity*:

$$I(X:Y) \leq \chi(\{p_i, \sigma_i\}) = S(\sigma) - \sum_i p_i S(\sigma_i). \tag{1}$$

Here, p_i is the probability that the message is i (i.e., the probability that distribution X produces i) and $\sigma = \sum_i p_i \sigma_i$ is Bob's average density matrix. Basically, this says that the information Bob can get about Alice's message is limited by the difference between the entropy of the average density matrix and the average of the entropies.

In fact, this rate is achievable for the best possible ensemble of mixed states. That is, n uses of an i.i.d. quantum channel $\mathcal{N}$ can send about Cn bits of information, where

$$C \geq \max_{\{p_i, \rho_i\}} \chi(\{p_i, \sigma_i\}). \tag{2}$$

But the χ function is an upper bound to Bob's information, so why doesn't this equal the capacity? The answer is once again super-additivity. If Alice sends tensor product states, then the best Alice and Bob can do is given by the RHS. However, if Alice sends entangled states, there are some cases where they can do better. Thus, the actual capacity is given by the regularized χ capacity:

$$C = \lim_{n\to\infty} \max_{\{p_i^{(n)}, \rho_i^{(n)}\}} \frac{1}{n} \chi(\{p_i^{(n)}, \sigma_i^{(n)}\}). \tag{3}$$

(Note that entangled measurements without sending entangled states doesn't help.)

2 Bell Inequalities

One stand-out feature that entangled states have that separates them from classical states is the ability to violate Bell inequalities. Quantum mechanics is famously unintuitive and surprising. Many people have tried to resolve these weird aspects of quantum mechanics by coming up a “local hidden variable theory” to describe quantum systems. The “hidden variable” part means that the theory is actually deterministic, unlike standard quantum mechanics, which has some randomness in the measurement results. A hidden variable theory instead has additional degrees of freedom which are not directly accessible to experiment (they are “hidden”) but determine the outcome of any possible measurement. The quantum theory would then only appear to be random because we don't know the value of the hidden variables and different copies of the system have different values of the hidden variables, so when we repeat the same experiment, we can get different outcomes.

It is possible to make hidden variable theories that agree with standard quantum mechanics, but Bell's theorem says that any such theory must be *non-local*, meaning that measurement results at one location have to be affected by actions at distant locations.

The specific form of Bell's theorem we will discuss is the CHSH inequality, in the form of something called a non-local box.

Definition 1. *A perfect* non-local box *is a operation which takes two single-bit inputs a and b and gives two single-bit outputs x and y such that x and y each separately have a probability distribution of* 50% 0 *and* 50% 1 *for all inputs but $x \oplus y = ab$. (That is, x and y agree unless $a = b = 1$, in which case they disagree.)*

These conditions imply that for a perfect non-local box,

$$P((x,y) = (0,0)|(a,b) \neq (1,1)) = P((x,y) = (1,1)|(a,b) \neq (1,1)) = 1/2 \tag{4}$$

$$P((x,y) = (0,1)|(a,b) = (1,1)) = P((x,y) = (1,0)|(a,b) = (1,1)) = 1/2. \tag{5}$$

We imagine that the non-local box acts on a system shared by Alice and Bob who are far apart from each other. Alice enters the input a into the box and receives the output x. At the same time, Bob enters the input b and receives the output y. Because Alice and Bob are far apart, neither Alice nor Bob knows the input received by the other. A non-local box therefore cannot be implemented perfectly in either classical or quantum mechanics, but quantum mechanics can do better than any hidden variable theory.

What precisely do we mean by a hidden variable theory? Imagine this is a game played by Alice and Bob as a team, who have to try to reproduce the effect of the non-local box. Alice and Bob can pre-agree on their strategy before learning and some additional numbers λ. λ can be used, for instance, by Alice and Bob to generate x and y; this counts as satisfying the uniform distribution of x and y separately because they can pick the appropriate bits of λ to be random bits. The strategy must be decided before Alice and Bob know a and b. They only receive the inputs a and b once they are far apart and can no longer communicate (perhaps because of the finite speed of light). They must therefore be able to generate x and y locally with the aid of the pre-arranged strategy and λ.

The full probability distribution of outputs conditioned on the inputs and the hidden variable λ is thus $P(x, y|a, b, \lambda)$. We can let there be additional randomness in Alice's and Bob's measurement outcomes, but we could also let that additional randomness be absorbed into extra bits of λ. We can therefore assume that $P(x, y|a, b, \lambda)$ is either 0 or 1 for any specific values of x, y, a, b, and λ. Now, λ is shared by Alice and Bob, but only Alice knows a and only Bob knows b. They must act independently except for their shared knowledge of λ. That means that

$$P(x, y|a, b, \lambda) = P(x|a, \lambda)P(y|b, \lambda). \tag{6}$$

And again, $P(x|a, \lambda)$ and $P(y|b, \lambda)$ are either 0 or 1.

Consider some possibilities for strategy for Alice and Bob. One option is for them to say, λ is one random bit, and just let your output (x or y, respectively) be equal to λ. Then when a and b are not both 1, their answers are the same, which is correct, but when $a = b = 1$, the answers are still the same but they are supposed to be different. Thus, if the pair (a, b) is uniformly random, they "win" the game 3/4 of the time. Or they could do something that actually uses a and b. For instance, they could have Alice always use the bit λ but Bob uses λ when $b = 0$ and $1 - \lambda$ when $b = 1$. This works when $(a, b) = (0, 0)$, $(a, b) = (1, 0)$, and $(a, b) = (1, 1)$, but not when $(a, b) = (0, 1)$. Again, they win with probability 3/4.

Theorem 1. *For any strategy and distribution on λ, Alice and Bob can never win with probability greater than* 3/4.

Theorem 2. *If Alice and Bob share entanglement, they can win the non-local box game with probability* $\frac{1+\sqrt{2}}{2\sqrt{2}} = 0.85\ldots$