8.S372/18.S996 Quantum Information Science III **Fall 2020**

Lecture 11: October 6, 2020

Lecturer: Aram Harrow *Scribe: Adam Bene Watts, Matthew Kearney*

In this section, we discuss some of the applications of relative entropy to show that it is a useful measure.

11.1 Application 1: Channel Coding

Consider a CQ channel $\{p(x), \sigma_x\}$, where message x is sent with probability p_x and σ_x is the resulting signal state. Define $\sigma = \sum_x p(x)\sigma_x$, making σ the average over the input states.

Recall that the relative entropy is given by

$$D(\sigma_x||\sigma) = \mathrm{tr}[\sigma_x(\log(\sigma_x) - \log(\sigma)] = -S(\sigma_x) - \mathrm{tr}[\sigma_x(\log(\sigma))]$$

If we then take the average over these relative entropies, we get the familiar Holevo χ, which describes the difference between the entropy of the average state and the average of the entropies of each of the states.

$$\begin{aligned}\sum_x p(x)D(\sigma_x||\sigma) &= -\sum_x p(x)S(\sigma_x) - \mathrm{tr}\left[\sum_x p(x)\sigma_x \log(\sigma)\right] \\ &= S(\sigma) - \sum_x p(x)S(\sigma_x) = \chi.\end{aligned}$$

This is an interesting result and leads us to ask Why should the relative entropy have anything to do with the channel capacity?

We can think of this as saying that the ability of the ensemble to carry information is related to the how surprising each message σ_x is compared to the average state σ, which is given by $D(\sigma_x||\sigma)$. To give a classical example, if we imagine that it rains 10% of the time and is sunny the other 90%, then the relative entropy between the state rainy and the average state will be low, meaning we are less surprised about it being sunny when the average state is our prior.

Along this line of thinking, we can imagine a hypothesis testing scenario where we try to distinguish a typical message $\sigma_{x^n} = \sigma_{x_1} \otimes \sigma_{x_2} \otimes ... \otimes \sigma_{x_n}$ from the average state

$\sigma^{\otimes n}$, which serves as our prior of the messages we will receive. Stein's Lemma tells us we mistakenly identify the message as the average state with probability $2^{-n\chi}$. This quantity is important for hypothesis testing realizations like sequential decoding where we may need to test against exponentially many possible states before testing against the correct state and we want to be very sure that we are not accepting the wrong messages.

While this discussion is suggestive of a strong link between hypothesis testing and channel coding, Ogawa and Nagaoka formalized this link by showing that you can prove the HSW theorem using hypothesis testing with carefully chosen states.

11.2 Application 2: Thermal States

Let H be a Hamiltonian and define

$$\gamma_T = \frac{e^{-H/T}}{\operatorname{tr}[e^{-H/T}]} \qquad\qquad F(\rho) = E(\rho) - TS(\rho) = \operatorname{tr}[H\rho] - TS(\rho).$$

Where $F(\rho)$ is the free energy and the thermal state γ_T is the state that minimizes free energy. Recall from PSET 4 that We derived an expression for a measure of how close a state's free energy is to the minimum free energy given by

$$\frac{D(\rho||\gamma_T)}{\ln(2)} = \frac{F(\rho) - F(\gamma_T)}{T} =: \frac{\Delta F}{T}.$$

Where ΔF is excess free energy. This shows that if the free energy of a state is small, that state is close to the thermal state.

To see why this is the case, we ask the following question: What is the probability that you measure a thermal state and get a state that looks like ρ? That's sort of like asking what the probability is that ρ arises from fluctuation which, by Crooks fluctuation theorem, is given by $e^{-\Delta F/T} = 2^{D(\rho||\gamma_T)}$. So the relative entropy is saying something about how surprised you should be to see ρ when you look at γ_T.

There is also another interpretation of $D(\rho||\gamma_T)$ in this case related to information removal and storage. Recall briefly Maxwell's Demon:

In this thought experiment, there is a box of gas particles with a partition in the middle, separating the left half from the right half. Further, there is a small hatch in the middle of this partition that can be open and shut by a demon in such as way as to not use any energy. If the demon opens that latch whenever a gas particle from the left side of the box is headed for it and closes it whenever a gas particle from the right is headed for it, eventually the gas particles will all end up on the right side of the box. This will have reduced the entropy and can therefore be used to perform work by opening the hatch and making the leftward motion of the gas particles do work. This however seems to violate the second law of thermodynamics.

The Landauer resolution to this paradox says that in fact this is not a violation because although the particles may be loosing entropy, the Demon is gaining information about which side of the box the gas is on and therefore is gaining entropy. In the act of gaining information, the Demon must also erase old information to make room for the new information. This erasure increased the entropy by at least as much as it is decreased by the collecting of the gas, giving us the minimal amount of work it costs to erase a bit, which by Landauer's erasure principle is $K_B T \ln(2)$. T

Along these lines, we $D(\rho||\gamma_T)$ as telling us how much space the state ρ has to store information. The amount of work that can be extracted from state ρ is given by $\Delta F = TD(\rho||\gamma_T)/\ln(2)$. Then, if we extract all the work we can from the state ρ and use it to erase bits we can erase a total of

$$\frac{TD(\rho||\gamma_T)/\ln(2)}{K_B T \ln(2)} = \frac{D(\rho||\gamma_T)}{K_B}$$

bits. We can alternatively think of this operation as storing $D(\rho||\gamma_T)/K_B$ bits inside ρ.

Second Law We can also state a strong version of the second law of thermodynamics. For any channel $\mathcal{N}$ satisfying $\mathcal{N}(\gamma_T) = \gamma_T$ and any state ρ we have

$$F(\mathcal{N}(\rho)) \leq F(\rho)$$

This is because any quantum channel can only decrease the relative entropy between ρ and γ_T, so

$$F(\rho) - F(\gamma_T) = \frac{D(\rho||\gamma_T)}{\ln(2)} \geq \frac{D(\mathcal{N}(\rho)||\mathcal{N}(\gamma_T))}{\ln(2)} = F(\mathcal{N}(\rho)) - F(\gamma_T).$$

11.3 Application 3: Quantifying Entanglement

In this section we seek principled ways of quantifying entanglement between subsystems. We can first ask, what properties of this quantification migth make sense or be useful? To answer this questions, it will be useful to draw analogy to the case of trying to quantify someone's wealth. You can imagine that it might be easy to quantify the wealth of two people whose money is all in US dollars, we can simply count who has more. But what about comparing someone whose money is in US dollars to someone whose money is in Euros? Or someone whose wealth is in diamonds compared to someone whose wealth is in gold? It would be useful to have a single metric (such as a "gold standard") to compare these values on, such as converting them all to US dollars first. We also want this to be a fair comparison. If we suppose that in the process of converting from Euros to US dollars someone loses an excess amount of wealth or somehow gains extra wealth such that when they convert back to Euros, they end with significantly more or less money than they started with, then this hardly seems like a fair comparison. Therefore we want to be able to convert between currencies with a minimal "exchange fee" so as to not significantly changing our wealth. Lastly, we would like it to be true that If we have our wealth in two different bank accounts, if we convert this wealth to US dollars, it doesn't matter if we convert it together or separately, we want to end up with the same amount of total US dollars at the end. This gives us the following properties:

1. Convertability
2. Small Conversion Fee
3. Additivity

These will be some of the properties that we may find useful when trying to judge a quantification scheme for entanglement.

11.3.1 Pure State Entanglement

We will begin by talking about quantifying the entanglement of pure states. Given a pure state $|\psi\rangle_{AB}$ the entanglement is quantified by the **entropy of entanglement**

$$S(A)_\psi = S(B)_\psi =: E$$

Explicitly, if $|\psi\rangle = \sum_i \sqrt{\lambda_i}\,|a_i\rangle \otimes |b_i\rangle$ then $E = H(\lambda)$.

Now we can ask Why is the entropy of entanglement a good measure of bipartite entanglement? We can imagine a conversion scheme, where different entangled states can be converted to the same "currency" (in our case, this will be EPR pairs) through some set of operations. Further, we don't want to allow these operations to create new entanglement, just as we didn't want our conversions in the wealth example to create new wealth. Therefore, if we allow only local operators and classical communication

(LOCC), Bennett, Bernstein, Popescu, and Schumacher (arXiv 9511030) showed that we can transform our state ψ as

$$\psi^{\otimes n} \to \Phi^{\otimes n(E-\delta)} \text{ (Entanglement Distillation)}$$
$$\Phi^{\otimes n(E+\delta)} \to \psi^{\otimes n} \text{ (Entanglement Dilution)}$$

where $\Phi = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. This gives us our "exchange rate" between any entangled state and the EPR state as $E = S(A) = S(B)$. Further, there is only a small exchange fee of δ.Therefore, asymptotically up to LOCC we can think of ψ as equalling E copies of Φ.

11.3.2 Mixed State Entanglement

What about the theory of bipartite entanglement for mixed states? We can try and do something analogous. Define the **distillable entanglement** $E_D(\rho)$ and **entanglement cost** $E_C(\rho)$ to be the max and min real numbers respectively such that

$$\rho^{\otimes n} \approx^{LOCC} \Phi^{E_D(\rho)} \text{ and} \tag{11.1}$$
$$\Phi^{E_C(\rho)} \approx^{LOCC} \rho^{\otimes n}. \tag{11.2}$$

Unfortunately, these definitions do not lead to the properties we discussed in the beginning. Some properties of these measures of entanglement include:

- They are not additive, so sometimes $E_D(\rho_1 \otimes \rho_2) > E_D(\rho_1) + E_D(\rho_2)$.
- There is no single letter formula known for these quantities. To get around this we can define the **entanglement of formation** $E_F(\rho)$ to be the minimum value of $\sum_i p_i S(\psi_i^A)$, taken over pairs of mixtures of pure states (p_i, ψ_i) satisfying $\sum_i p_i \psi_i = \rho$. We have $E_C \leq E_F$, but sometimes this inequality is strict.
- Sometimes we have entangled states with $E_D = 0$ ("bound entanglement"), but we have no general theory of these states or why they occur.
- On the other hand, for any entangled state $E_C > 0$.
- $E_D \leq E_C$, and sometimes this inequality is strict, meaning we could lose "entangledness" in converting back and forth between EPR pairs and certain states.

Let's try and come up with a nicer measure of bipartite entanglement for mixed states.

11.3.2.1 Relative Entropy of Entanglement

First define the **separable states** to be states in the set

$$\mathrm{Sep}(d_A, d_B) = \mathrm{conv}\left\{\alpha \otimes \beta : \alpha \in D_A, \beta \in D_B\right\}$$

where D_A are $d_A \times d_A$ density matrices D_B are $d_B \times d_B$ density matrices and

$$\mathrm{conv}(X) = \left\{\sum_{x\in X} p_x x : p_x > 0, \sum_x p_x = 1\right\}$$

is the convex hull of the points in X (the smallest convex set that contains all of the points in X). These are our unentangled states. Unfortunately, it is NP hard to determine if a given state is separable.

Now, we can define the **relative entropy of entanglement**

$$E_R(\rho) = \min_{\sigma\in\mathrm{Sep}} D(\rho||\sigma)$$

This measure may be non-additive, so we also define the **regularized relative entropy of entanglement**

$$E_R^{\infty}(\rho) = \lim_{n\to\infty} \frac{1}{n} E_R(\rho^{\otimes n}) \leq E_R(\rho)$$

(and this inequality is sometimes strict).

Why is the relative entropy of entanglement nice? Define the **asymptotically non-entangling operations** (a family of operations that includes LOCC) to be channels $\Lambda_1, \Lambda_2, ...\Lambda_n : (A'\otimes B')^{\otimes n} \to (A\otimes B)^{\otimes n}$ with Λ_n approximately sending separable states to separable states.

To make this definition precise define the **Rèyni divergences**

$$S_\alpha(A||B) = \frac{1}{\alpha - 1}\log(\mathrm{tr}[A^\alpha B^{1-\alpha}])$$

where as a few examples we have

$$\begin{aligned} S_1(A||B) &= S(A||B) \\ S_{1/2}(A||B) &= -2\log(F(A,B)) \\ S_\infty(A||B) &= \log \parallel B^{-1/2}AB^{-1/2} \parallel_\infty = \inf\{\lambda : A \leq 2^\lambda B\} \end{aligned}$$

Now channels $\Lambda_1, \Lambda_2, ...\Lambda_n$ are asymptotically non-entangling if

$$\forall \rho, \sigma \in \mathrm{Sep} : S_\infty(\Lambda_n(\rho^{\otimes n})||\sigma) \leq \epsilon_n$$

with $\epsilon_n \to 0$ as $n \to \infty$. This gives us our precise definition of these operations that we will now use to examine the regularized relative entropy of entanglement.

Theorem 7 (Brandao-Plenio arXiv:0710.5827) *Up to asymptotically non-entangling operations*

$$\rho^{\otimes n} \leftrightarrow \Phi^{\otimes n E_R^\infty(\rho)}$$

This is exactly the type of conversion we were looking for in our metric of entanglement.

We note that a similar result holds in thermodynamics. Define **thermal operations** to be channels

$$\mathcal{N}(\rho) = \mathrm{tr}_E \left[V \left(\rho_S \otimes \frac{\exp(-\beta H_E)}{\mathrm{tr}[\exp(-\beta H_E)]} \right) V^\dagger \right]$$

with $[V, H_S \otimes I + I \otimes H_E] = 0$. H_S is the system Hamiltonian while H_E is the environment (or bath) Hamiltonian. These are the operations that are free if the thermal states are free. In other words, they do not create any free energy. Let $\gamma_T = e^{-\beta H_S}/\mathrm{tr}[e^{-\beta H_S}]$ be the thermal state of the system. Under thermal operations, we can transform a state ρ into a state σ at a rate $D(\rho||\gamma_T)/D(\sigma||\gamma_T)$. In this way, we can think of entanglement and non-thermal states as resources.

We will now conclude by sketching the proof of Brandao-Plenio.

First, we compute $E_R(\Phi^{\otimes n})$. We do this via Stein's Lemma and ote that the optimal measurement to distinguish any state from the EPR state is $M = \Phi^{\otimes n}$. Taking $\sigma \in \mathrm{Sep}$ we have

$$\begin{aligned} \max_\sigma \mathrm{tr}[M\sigma] &= \max_{|\alpha\rangle, |\beta\rangle} \left| \langle \Phi |^{\otimes n} |\alpha\rangle |\beta\rangle \right|^2 \\ &= \max_{|\alpha\rangle, |\beta\rangle} \left| \langle \alpha | |\beta\rangle \right|^2 / 2^n = 2^{-n} \end{aligned}$$

which gives $E_R(\Phi^{\otimes n}) = n$. This is obviously the dsired result, since it tells us that n EPR states are worth n EPR states worth of entanglement.

Now, for any state ρ, $S_\infty(\rho||\mathrm{Sep}) = \lambda$ implies that there exists a $\sigma \in \mathrm{Sep}$ with $\rho^{\otimes n} \leq 2^\lambda \sigma$. Equivalently, $2^{-\lambda}\rho^{\otimes n} \leq \sigma$. Then we can write

$$\sigma = 2^{-\lambda}\rho^{\otimes n} + (I - 2^{-\lambda})\gamma$$

for some density matrix γ. Define an asymptotically non-entangling operation where Λ_n is the either the measurement $\Phi^{\otimes nR}$ and outputs $\rho^{\otimes n}$ or the measurement $I - \Phi^{\otimes nR}$ and outputs γ. The outcome of this measurement on $\Phi^{\otimes nR}$ is $\phi^{\otimes n}$ and the outcome of the measurement on any separable state is $2^{-\lambda}\rho^{\otimes n} + (I - 2^{-\lambda})\gamma = \sigma$. So the measurement is asymptotically non-entangling and maps $\Phi^{\otimes nR}$ to $\rho^{\otimes n}$.

To map in the other direction we use the optimal test distinguishing $\rho^{\otimes n}$ from Sep.