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

Lecture 5: Sep 15, 2020

Lecturer: Aram Harrow *Scribe: Changnan Peng, Annie Wei*

5.1 Information Theory: Classical and Quantum

5.1.1 Noiseless Coding Theorem

5.1.1.1 Review

Today we will continue talking about information theory.

Recall that last time we defined the **Shannon entropy** as a measure of uncertainty for the probability distribution $p(x)$:

$$H(p) = -\sum_x p(x) \log p(x). \tag{5.1}$$

Then we discussed an application of Shannon entropies to the problem of compression. Recall that we started by defining a **typical set** (also known as the "asymptotic equipartition property"), which is a set of strings with probability δ-close to a probability distribution p:

$$T^n_{p,\delta} = \left\{ x^n = (x_1, ..., x_n) \text{ s.t. } \left| -\frac{1}{n} \log p^{\otimes n}(x^n) - H(x) \right| \leq \delta \right\}. \tag{5.2}$$

The probability of being non-typical, characterized by $\epsilon = 1 - p^{\otimes n}(T^n_{p,\delta})$, goes to zero, as $n \to \infty$. In fact, $\epsilon \leq n^{O(1)} 2^{-n\delta}$.

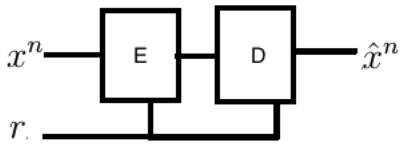
5.1.1.2 Converse

Now, continuing from last time, let's talk about the **converse** to the Shannon coding theorem: Let's say that we compress to k bits, and we assume that there exists a set S that is decoded correctly. Note that by definition $|S| \leq 2^k$. Then the probability of correctly decoding is

$$p^n(S) \leq p^n(T^n_{p,\delta} \cap S) + p^n(\overline{T^n_{p,\delta}}) \leq 2^k 2^{-nH(p)+n\delta} + \epsilon \tag{5.3}$$

Note that the right hand side is a small number if $k \leq n(H(p) - \delta)$. In this argument we assumed deterministic coding and encoding, but what if Alice and Bob share randomness? We'll claim that even with the shared randomness, the bound still holds.

Note that with shared randomness, we have a diagram that looks something like the following:



Note that we can condition on the shared randomness r, reducing it to the deterministic case, and then sum over r,

$$P(x^n = \hat{(}x)^n | r) \leq \epsilon', \tag{5.4}$$

so shared randomness should not change our results.

5.1.2 Quantum Entropy and Compression

5.1.2.1 Quantum Entropies

In the quantum case, we can define the von Neumann entropy,

$$S(\rho) = H(\text{eig}(\rho)) = -tr[\rho \log \rho]. \tag{5.5}$$

It is zero if and only if ρ is a pure state:

$$S(\rho) = 0 \Leftrightarrow eig(\rho) = (1, 0, ..., 0) \Leftrightarrow \rho = |\psi><\psi|. \tag{5.6}$$

It attains its maximum when ρ is maximally mixed. Letting $d = \dim(\rho)$,

$$S(\rho) \leq \log d \tag{5.7}$$

$$S(\rho) = \log d \Leftrightarrow \rho = I/d. \tag{5.8}$$

We can generalize all of our classical entropies to the quantum case described by

density matrix ρ_{XY}:

$$S(X) = S(\rho_X) \tag{5.9}$$
$$S(X|Y) = S(XY) - S(Y) \tag{5.10}$$
$$I(X:Y) = S(X) + S(Y) - S(XY) \tag{5.11}$$
$$D(\rho||\sigma) = \mathrm{tr}\rho[\log\rho - \log\sigma] \tag{5.12}$$
$$D(p||q) = \sum_x p(x)\log\frac{p(x)}{q(x)} \tag{5.13}$$

Note that again $S(X|Y) \leq S(X)$, which allows us to derive concavity of entropy. A good question to ask is how we should actually interpret the quantity $S(X|Y)$. For example, for the state

$$|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2},$$

how would we actually condition on the state of the second system? This isn't so clear! Note, also that the conditional entropy in the quantum case can be negative. An example is again the Bell state, where $S(XY)_\psi = 0$, $S(Y)_\psi = 1$, $S(X|Y) = -1$.

Now let's extend the notion of typical sets. Say we have the state

$$\rho = \sum_x \lambda_x |v_x\rangle\langle v_x|,$$

where the $|v_x\rangle$'s are orthonormal. Then we can define a typical projector

$$\Pi^n_{p,\delta} = \sum_{x^n \in T_{\lambda,\delta}} |v_{x^n}\rangle\langle v_{x^n}|.$$

Here

$$|v_{x^n}\rangle = |v_{x_1}\rangle \otimes ... \otimes |v_{x_n}\rangle.$$

This results in

$$\rho^{\otimes n} = \sum_{x^n} \lambda_{x_1}...\lambda_{x_n} |v_{x^n}\rangle\langle v_{x^n}|$$

Note that this projector projects to the typical subspace. The projection measurement $\{\Pi^n_{p,\delta}, I - \Pi^n_{p,\delta}\}$ has resulting probability

$$\mathrm{tr}[\rho^{\otimes n}\Pi^n_{p,\delta}] = \sum_{x^n} \lambda_{x_1}...\lambda_{x_n} 1_{x^n \in T^n_{\lambda,\delta}} = \lambda^n(T^n_{\lambda,\delta}) \tag{5.14}$$

Note that this approaches 1 as $n \to \infty$.

Note that we would like to discuss compressing unknown ρ for qubits with known $S(\rho)$. If ρ is known, then we can use a classical compression scheme working in the eigenbasis $|v_1\rangle$,...,$|v_d\rangle$.

5.1.2.2 Efficient Classical Compression

Now let's look at an example of an **efficient classical compression scheme**, specifically Huffman encoding.

For an example case, suppose we have 4 symbols, with probabilities given in the second column below. Then we can assign a code using the third column below:

A	$1/2$	0
B	$1/4$	10
C	$1/8$	110
D	$1/8$	111

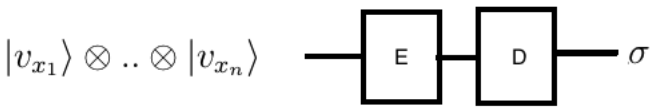
Note that this encoding encodes x with $\lceil \log 1/p(x) \rceil$ bits, and that this is always possible. Note also that this encoding is prefix-free. If the probabilities are not powers of $1/2$, we can use block coding, i.e. by expanding our code to include multiple bits in a block. In our example, we would take the codewords to be blocks AA, AB, BB, etc. Then this allows us to assign probabilities to each block that are closer to powers of $1/2$.

5.1.2.3 Quantum compression

We start off by asking the question, what does it mean to compress ρ? Here are some possible answers:

1. 

   with $F(\rho^{\otimes n}, \sigma) \approx 1$.

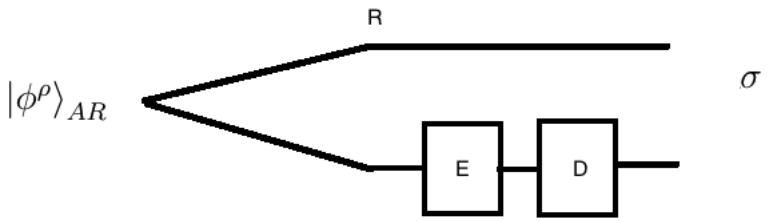
2. 

with $E_{x^n \sim \lambda}[F(|v_{x^n}\rangle \langle v_{x^n}|, \sigma)] \approx 1$.

3. Letting $\rho = \sum_i p_i |w_i\rangle \langle w_i|$, where the $|w_i\rangle$ are not necessarily orthonormal,

with $E_{i^n \sim \rho}[F(|w_{i^n}\rangle, \sigma) \approx 1]$.

4. Letting $|\phi^\rho\rangle_{AR}$ be a purification such that $\phi_A^\rho = \rho$,



with $F(|\phi^\rho\rangle^{\otimes n}, \sigma) \approx 1$.

Note that the first scheme doesn't work because it allows for the possibility where you input the maximally mixed state and produce the output by just throwing away the input state and always outputting the maximally mixed state. The classical equivalent would be a source that always emits the uniform distribution, and an encoding scheme that throws away the acutal message and just returns the uniform distribution.

The second to fourth options are roughly the same, and give us **Schumacher Jozsa compression**. Formally, the way this works is the following: Say we have state $\rho^{\otimes n}$. Apply $\{\Pi_{p,\delta}^n, I - \Pi_{p,\delta}^n\}$, where the first result is successful and the second is a failure. If this is successful, the state is contained in a subspace of dimension $\text{tr}\Pi_{p,\delta}^n \leq \exp(n(S(\rho) + \delta) = \exp(nH(\lambda) + \delta)$. Thus it fits into $n(S(\rho) + \delta)$ qubits.

An application of this is to algorithmic cooling, where our states (representing, for example, nuclear spins in large magnetic fields) are of the form $\rho^{\otimes n}$ with

$$\rho = \begin{pmatrix} \frac{1+\epsilon}{2} & 0 \\ 0 & \frac{1-\epsilon}{2} \end{pmatrix}.$$