Lecture 4: September 17, 2024

Scribe: Aditi Venkatesh *Quantum Compression*

4.1 Classical Compression

Last class we talked about Shannon's noiseless coding theorem and data compression. We said if $R > H(p)$ compression is possible (direct) and if $R < H(p)$ compression is impossible (converse). Here we will prove the converse (the direct proof is in the previous lecture note). We can define a typical set as

$$T^n_{p,\delta} = \{x^n = (x_1, x_2, ..., x_n) : \left| -\frac{1}{n} \log p^{\otimes n}(x^n) - H(p) \right| < \delta\}$$

$$\epsilon = 1 - p^{\otimes n}(T^n_{p,\delta}) \leq 2^{-n\delta} \to 0 \quad \text{as } n \to \infty$$

This implies that the set of non-typical sequences becomes vanishingly small as n increases.

Now, suppose we want to compress the source X^n to k bits, where $k < nH(p)$. The goal is to show that this compression will lead to a vanishing probability of correctly decoding the original sequence.

Consider the compression process:

$$\boldsymbol{x}^n \xrightarrow{\text{Encoder (E)}} m \xrightarrow{\text{Decoder (D)}} \boldsymbol{x}^n$$

Here, $\boldsymbol{x}^n$ is the original sequence, m is the compressed version (with k bits), and $\boldsymbol{x}^n$ is the decoded sequence. The encoder uses r as a random seed to perform the compression. Pick r to maximize

$$\mathbb{P}\left(D(E(\boldsymbol{x}^n)) = \boldsymbol{x}^n \mid r\right)$$

Let $S \subseteq \Sigma^n$ be the set of sequences that can be decoded correctly. Since we're compressing to k bits, the size of this set is constrained by $|S| \leq 2^k$. We now want to bound the probability that a sequence $\boldsymbol{x}^n$ is decoded correctly.

$$p^n(S) \leq p^n(T^n_{p,\delta} \cap S) + p^n(T^{n^c}_{p,\delta})$$

The second term, $p^n(T^{n^c}_{p,\delta})$, is the probability of being in the non-typical set, which is upper-bounded by ϵ, a small quantity that goes to zero as $n \to \infty$. The first term can be bounded using the size of S and the fact that typical sequences occur with probability around $2^{-nH(p)}$:

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

Thus, the total probability is:

$$p^n(S) \leq 2^k 2^{-nH(p)+n\delta} + \epsilon$$
$$\to 0 \quad \text{if} \quad \frac{k}{n} < H(p)$$

This shows that if the compression rate k/n is less than $H(p)$, the probability of correctly decoding the sequence goes to zero as n increases. Therefore, it is not possible to compress below the entropy rate without losing information.

4.2 Quantum Compression (Entropy)

The entropy of a density matrix ρ, also known as the von Neumann entropy, is defined as:

$$S(\rho) = H(\text{eig}(\rho)) = -\text{Tr}(\rho \log \rho),$$

where $H(\text{eig}(\rho))$ is the Shannon entropy of the eigenvalues of ρ.

4.2.1 Bounds on Entropy

The von Neumann entropy satisfies the following bounds:

$$0 \leq S(\rho) \leq \log d,$$

where d is the dimension of the Hilbert space.

If $S(\rho) = 0$, then the eigenvalues of ρ are $(1, 0, 0, \ldots, 0)$, implying that $\rho = |\psi\rangle\langle\psi|$ for some pure state $|\psi\rangle$. If $S(\rho) = \log d$, then ρ is the maximally mixed state:

$$\rho = \frac{I}{d}.$$

4.2.2 Conditional Entropy

The entropy of a system X is given by:

$$S(X) = S(\rho_X).$$

The conditional entropy is defined as:

$$S(X|Y) = S(XY) - S(Y),$$

which can be negative. This definition carries over from classical information theory.

4.2.3 Mutual Information

The mutual information between two systems X and Y is defined as:

$$I(X : Y) = S(X) + S(Y) - S(XY),$$

which can also be expressed as:

$$I(X : Y) = S(X) - S(X|Y).$$

4.3 Typical Subspaces and Projectors

4.3.1 Definition of Typical Subspace

Let ρ be a density matrix acting on a Hilbert space $\mathcal{H}$, and let $\epsilon > 0$ be a small positive number. The typical subspace, denoted $\mathcal{T}_\epsilon^{(n)}$, is defined as the span of the eigenvectors of $\rho^{\otimes n}$ corresponding to eigenvalues close to $2^{-nS(\rho)}$, where $S(\rho)$ is the von Neumann entropy of ρ:

$$S(\rho) = -\text{Tr}(\rho \log \rho).$$

The typical subspace satisfies the following properties:

- $\mathbb{P}\left(\psi \in \mathcal{T}_\epsilon^{(n)}\right) \geq 1 - \epsilon$ for a random state ψ drawn from $\rho^{\otimes n}$.
- The dimension of the typical subspace is approximately $2^{nS(\rho)}$ for large n.

4.3.2 Properties of Typical Subspaces

1. *High Probability Support*: A quantum state ψ drawn according to $\rho^{\otimes n}$ has a high probability of being in the typical subspace. This is crucial for understanding the structure of quantum information over many copies.

2. *Dimensionality*: The typical subspace has dimension $d_{\text{typ}} \approx 2^{nS(\rho)}$, where $S(\rho)$ is the von Neumann entropy of ρ. This shows that the size of the subspace grows exponentially with the number of copies n.

4.4 Projectors onto Typical Subspaces

Given the typical subspace $\mathcal{T}_\epsilon^{(n)}$, we define a projector P_{typ} that projects any state onto this subspace.

4.4.1 Construction of the Projector

Let $\rho^{\otimes n}$ be the n-fold tensor product of the density matrix ρ. We diagonalize $\rho^{\otimes n}$ in its eigenbasis:

$$\rho^{\otimes n} = \sum_i \lambda_i |i\rangle\langle i|.$$

The typical subspace corresponds to the eigenvalues λ_i that satisfy

$$2^{-n(S(\rho)+\delta)} \leq \lambda_i \leq 2^{-n(S(\rho)-\delta)},$$

where $\delta > 0$ is a small positive number. The projector onto the typical subspace is given by

$$P_{\text{typ}} = \sum_{i:\lambda_i \text{ typical}} |i\rangle\langle i|.$$

4.4.2 Properties of the Projector

The projector P_{typ} has the following important properties:

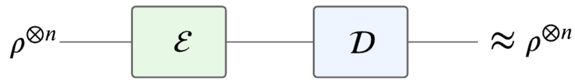
- *Approximate Preservation of Trace*: For large n, we have $\text{Tr}(P_{\text{typ}}\rho^{\otimes n}) \geq 1 - \epsilon$. This means that most of the probability mass of $\rho^{\otimes n}$ lies within the typical subspace.

- *Dimensionality*: The rank of P_{typ} (the dimension of the typical subspace) is approximately $2^{nS(\rho)}$.

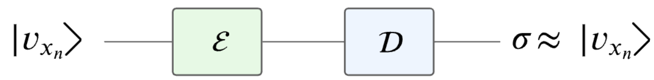
4.5 Quantum Compression

Below we give four possible schemes of quantum compression and decide whether or not they are valid.

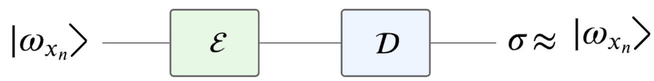
1. In the first scheme we have n copies of a density matrix ρ and we apply an encoder $\mathcal{E}$ and then a decoder $\mathcal{D}$ and check if the final density matrix matches the initial one.



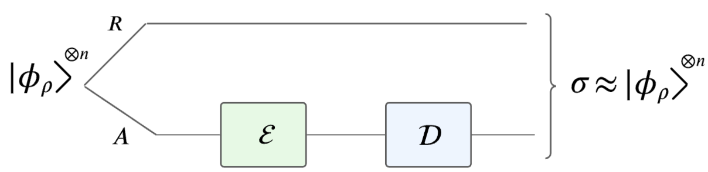
2. In the second scheme we can define x_n λ_n and check $\mathop{\mathbb{E}}_{x^n \sim \lambda^n} F(\sigma, |v_{x_n}\rangle) \approx 1$.



3. In the third scheme we define $\rho = \sum_i p_i |\omega_i\rangle \langle\omega_i|$ and check if the final state is close to the initial state.



4. In the fourth scheme we have n copies of a state $|\phi_\rho\rangle$ which we entangle with a reference. Then we apply the encoder and the decoder and check if the joint state is close to the initial state.



The first scheme does not work because it does not preserve correlations. We can see that the fourth, third, and second are equivalent ($4 \to 3 \to 2 \to 4$). It is easy to show that ($3 \to 2$) by choosing the eigenbasis. To show ($4 \to 3$), we use the fact that $\forall$ ensembles $\{p_i, |\omega_i\rangle\}$ such that $\rho = \sum_i p_i |\omega_i\rangle \langle\omega_i|$ there exists measurement operators $M_1, ..., M_l$ on R that induces this ensemble.