CSCI5370 Quantum Computing December 2, 2013

Lecture 12: Quantum Information IV - Channel Coding

Lecturer: Shengyu Zhang *Scribe: Hing Yin Tsang*

12.1 Shannon's channel coding theorem

A classical (discrete memoryless) channel is described by the transition matrix $p(y|x)$. For such a channel, if the encoder sends a message $x^n \in \mathcal{X}^n$, the decoder will then receive the sequence $y^n \in \mathcal{Y}^n$ with probability $\prod_{i=1}^n p(y_i|x_i)$. A rate R is achievable if there exists a sequence of encoding and decoding functions $C_n : [2^{nR}] \to \mathcal{X}^n$ and $D_n : \mathcal{Y}^n \to [2^{nR}]$ such that the error probability for any message M tends to 0 as $n \to \infty$. The channel capacity is defined as the maximum achievable rate. A celebrated result of Shannon gives a complete characterization of channel capacity in terms of mutual information:

Theorem 12.1 (Shannon's channel coding theorem)**.** *Let $p(y|x)$ be a discrete memoryless channel. Then*

$$C = \max_{p(x)} I(X;Y).$$

Proof sketch of achievability. The proof is by probabilistic method. Fixing $p(x)$, we generate 2^{nR} codewords independently according to the distribution $p(x^n) = \prod_{i=1}^n p(x_i)$. Both the encoder and decoder knows the codebook. Let y_M^n be the messaged received by the decoder when x_M^n is sent. The decoder output any M' such that $(x_{M'}^n, y_M^n)$ are *jointly typical*. The error occurs only when (x_M^n, y_M^n) are not jointly typical, or there are other $M' \neq M$ such that $(x_{M'}^n, y_M^n)$ are jointly typical. The first event happens with very small probability by the law of large number. The second event also happens with small probability. It can be shown that for independent $\tilde{X}^n$ and $\tilde{Y}^n$ with the same marginals as $p(x^n, y^n)$, the probability that $(\tilde{X}^n, \tilde{Y}^n)$ are jointly typical is roughly $2^{-nI(X;Y)}$. Thus a simple union bound argument works when the number of codeword is smaller than $2^{nI(X;Y)}$. It follows that there exists codebook with small average error probability.

By removing the worst half of the codeword, the maximum error probability would be small as well and hence the proof is completed. □

Proof of converse. Let M be a uniformly random message, let $\hat{M}$ be the output of the decoder. Let $p = \Pr[M \neq \hat{M}]$ be the average error probability. We have

$$nR = H(M) = H(M|\hat{M}) + I(M;\hat{M}) \leq H(M|\hat{M}) + I(X^n;Y^n) \leq H(M|\hat{M}) + nC \leq H(p) + npR + nC,$$

where the first inequality is data processing inequality, the second inequality follows by the fact that the channel is memoryless and by the definition of C, and the last inequality follows by Fano's inequality. Hence if $R > C$, then $p \geq (R-C)/R - H(p)/n$ which is bounded away from 0 as $n \to \infty$. □

12.2 Quantum channel coding theorem

For the quantum analogue, we are given a quantum channel instead of a classical channel so that the codeword can be a quantum state. The quantum channel is now described by a trace-preserving quantum operator, namely $\mathcal{E}$ defined over some set H. In this case, if the encoder sends $\rho = \rho_1 \otimes \cdots \otimes \rho_n \in H^{\otimes n}$, the decoder will receive a state $\sigma = \sigma_1 \otimes \cdots \otimes \sigma_n = \mathcal{E}(\rho_1) \otimes \cdots \otimes \mathcal{E}(\rho_n)$. By abusing notation, we also write $\sigma = \mathcal{E}(\rho)$. In the quantum case, the decoding is done by performing measurement. A rate R is achievable if there exists a sequence of product state $\{\rho_M\}_{M=1,\ldots,2^{nR}}$ and a measurement described by POVM elements $\{E_M\}_{M=1,\ldots,2^{nR}}$ such that

$$\lim_{n\to\infty} \max_M (1 - \mathrm{Tr}(\mathcal{E}(\rho_M E_M)) = 0.$$

The product state capacity of $\mathcal{E}$, denoted by $C^{(1)}(\mathcal{E})$, is defined as the maximum achievable rate of $\mathcal{E}$. The following theorem is the quantum analogue of Shannon's channel coding theorem:

Theorem 12.2 (Holevo-Schumacher-Westmoreland theorem)**.**

$$C^{(1)}(\mathcal{E}) = \max_{\{p_j,\rho_j\}} \left[S\left(\mathcal{E}\left(\sum_j p_j\rho_j\right)\right) - \sum_j p_j S(\mathcal{E}(\rho_j))\right],$$

where the maximum is over all ensembles $\{p_j, \rho_j\}$ of possible input state ρ_j to the channel $\mathcal{E}$.

Proof sketch of acheviability. The proof is to argue randomly generate codeword (note that it is a quantum state now) can achieve small *average* error probability, then we use the trick in Shannon's proof to turn average error probability to maximum error probability by removing bad codes. Note that the rate is not affected by much in this step.

Specifically, let $\{p_j, \rho_j\}$ be an ensemble. For each message M, we generate a state $\rho_M = \rho_{M_1} \otimes \cdots \otimes \rho_{M_n}$ where M_i are independent and $M_i = j$ with probability p_j. It is the codebook we need.

Before defining the measurement, we need to definite several things first. Let $\sigma_{M_i} = \mathcal{E}(\rho_{M_i})$, $\sigma_M = \sigma_{M_1} \otimes \cdots \otimes \sigma_{M_n}$ and $\bar{\sigma} = \sum_j p_j\sigma_j$. Let P be the ε-typical subspace of $\bar{\sigma}^{\otimes n}$. We know that by the typical subspace theorem, for any $\delta > 0$ and n sufficiently large,

$$\mathrm{Tr}(\bar{\sigma}^{\otimes n}(I - P)) \le \delta.$$

Let $\sigma_j = \sum_k \lambda_k^j |e_k^j\rangle\langle e_k^j|$ be the spectral decomposition. Thus we have

$$\sigma_M = \sum_K \lambda_K^M |E_K^M\rangle\langle E_K^M|,$$

where $K = (K_1, \ldots, K_n)$, $\lambda_K^M = \prod_{i=1}^n \lambda_{K_i}^{M_i}$ and $|E_K^M\rangle = |e_{K_1}^{M_1}\rangle \cdots |e_{K_n}^{M_n}\rangle$. Let $\bar{S} = \sum_j p_j S(\sigma_j)$,

$$T_M = \left\{ K = (K_1, \ldots, K_n) : \left| \frac{1}{n} \log \frac{1}{\lambda_K^M} - \bar{S} \right| \right\},$$

and P_M be the projector to the subspace spanned by $\{|E_K^M\rangle : K \in T_M\}$. Note that P_M are random as $|E_K^M\rangle = |e_{K_1}^{M_1}\rangle \cdots |e_{K_n}^{M_n}\rangle$ are random. It can be shown (by law of large number again) that for any $\delta > 0$ and n large enough, $\mathbb{E}(\mathrm{Tr}(\sigma_M P_M)) \ge 1 - \delta$. Moreover, $\mathbb{E}(\mathrm{Tr}(P_M)) \le 2^{n(\bar{S}+\varepsilon)}$.

Now we define the measurement E_M. Let

$$E_M = \left(\sum_{M'} PP_{M'}P\right)^{-1/2} PP_MP \left(\sum_{M'} PP_{M'}P\right)^{-1/2}$$

and $E_0 = I - \sum_M E_M$.

For such measurement $\{E_M\}$, it is possible to show that the average error probability p_{ave}, defined by

$$p_{ave} = \frac{\sum_M 1 - \mathrm{Tr}(\sigma_M E_M)}{2^{nR}},$$

can be upper bounded by

$$p_{ave} \le \frac{1}{2^{nR}} \sum_M \left[3\mathrm{Tr}(\sigma_M(I - P)) + \sum_{M' \ne M} \mathrm{Tr}(P\sigma_M PP_{M'}) + \mathrm{Tr}(\sigma_M(I - P_M)) \right].$$

Taking expectation over all random codes, since $\mathbb{E}(\sigma_M) = \bar{\sigma}$ and σ_M and $P_{M'}$ are independent for $M \ne M'$, we have

$$\begin{aligned} \mathbb{E}(p_{ave}) &\le 3\mathrm{Tr}(\bar{\sigma}(I - P)) + (2^{nR} - 1)\mathrm{Tr}(P\bar{\sigma}P\mathbb{E}(P_1)) + \mathbb{E}(\mathrm{Tr}(\sigma_M(I - P_M))) \\ &\le 4\delta + (2^{nR} - 1)\mathrm{Tr}(P\bar{\sigma}P\mathbb{E}(P_1)). \end{aligned}$$

Finally, observe that $P\bar{\sigma}P \preceq 2^{-n(S(\bar{\sigma})-\varepsilon)}I$. Thus we have

$$\mathrm{Tr}(P\bar{\sigma}P\mathbb{E}(P_1)) \leq 2^{-n(S(\bar{\sigma})-\varepsilon)}\mathrm{Tr}(\mathbb{E}(P_1)) \leq 2^{-n(S(\bar{\sigma})-\bar{S}-2\varepsilon)}.$$

Whenever $R < S(\bar{\sigma}) - \bar{S}$, we can choose ε small enough so that $\mathbb{E}(p_{ave}) \to 0$ as $n \to \infty$. □

Proof of converse. Denote $\chi(\mathcal{E})$ the quantity at the right-hand side. Let $\rho_M = \rho_1^M \otimes \cdots \otimes \rho_n^M$ be the set of codewords and $\sigma_M = \mathcal{E}(\rho_M) = \sigma_1^M \otimes \cdots \otimes \sigma_n^M$. Let $\{E_M\}$ be the POVM measurement. We will show that even the average error probability is bounded away from 0 when $R > \chi(\mathcal{E})$, which will complete the proof. Note that

$$p_{ave} = \frac{\sum_M (1 - \mathrm{Tr}(\sigma_M E_M))}{2^{nR}} = \Pr[M \neq Y].$$

Now, consider the ensemble $\{1/2^{nR}, \sigma_M\}$, let $\bar{\sigma} = \sum_M \sigma_M / 2^{nR}$. By Holevo's bound, we have

$$\begin{aligned} S(\bar{\sigma}) - \sum_M \frac{S(\sigma_M)}{2^{nR}} &\geq I(M;Y) \\ &= H(M) - H(M|Y) \\ &= nR - H(M|Y) \\ &\geq nR - H(p_{ave}) - p_{ave}\log(2^{nR}-1) \\ &\geq nR - H(p_{ave}) - np_{ave}R. \end{aligned}$$

Since σ_M are product states, we have $S(\sigma_M) = \sum_{j=1}^n S(\sigma_j^M)$. By the subadditivity of entropy, $S(\bar{\sigma}) \leq \sum_{j=1}^n S(\bar{\sigma}^j)$ where $\bar{\sigma}^j = \sum_M \sigma_j^M / 2^{nR}$. By definition of $\chi(\mathcal{E})$, $S(\bar{\sigma}) - \sum_M S(\sigma_j^M) \leq \chi(\mathcal{E})$ for all i and hence we have

$$n\chi(\mathcal{E}) \geq nR - H(p_{ave}) - np_{ave}R.$$

As $n \to \infty$,

$$p_{ave} \geq \frac{R - \chi(\mathcal{E})}{R},$$

which is bounded away from 0 when $R > \chi(\mathcal{E})$ and the proof is completed. □