Electrical Engineering 229A Lecture 17 Notes

Daniel Raban

October 21, 2021

1 Upper Bound for Channel Capacity, Perfect Noiseless Feedback, and Joint Source Channel Coding

1.1 Upper bound for Shannon's channel coding theorem

Last time, we were proving Shannon's channel coding theorem for discrete memoryless channels. A DMC is given by a probability transition matrix $[p(y \mid x)]_{x \in \mathscr{X}, y \in \mathscr{Y}}$, where $\mathscr{X}, \mathscr{Y}$ are finite. Shannon's formulation uses block codes $((e_n, d_n), n \geq 1)$, where

$$e_n : [M_n] \to \mathscr{X}^n, \qquad d_n : \mathscr{Y}^n \to [M_n],$$

where M_n is exponentially growing in n. The "memoryless" part means

$$p(y_1^n \mid x_1^n) = \prod_{i=1}^n p(x_i, y_i).$$

Definition 1.1. We say **communication is possible at rate** R if there exist $((e_n, d_n), n \geq 1)$ such that

$$\mathbb{P}(d_n(e_n(W_n)) \neq W_n) \xrightarrow{n \to \infty} 0$$

and

$$\liminf_n \frac{1}{n} \log M_n \geq R.$$

Let

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

Theorem 1.1 (Shannon's channel coding theorem)**.**

$$\sup\{R : \textit{can communicate at rate } R\} = C.$$

C is called the **Shannon capacity** of the channel. Let's finish the proof.

Proof. We have proved achievability: For $\varepsilon > 0$, we can communicate at rate $C - \varepsilon$. Now we prove the converse. Consider any $((e_n, d_n), n \geq 1)$. We have the Markov chain $W_n - X_1^n - \widehat{Y}_1^n - \widehat{W}_n$, where $W_n \sim \mathrm{Unif}([M_n])$, and $\widehat{W}_n = d_n(Y_1^n)$. In this notation, the error probability is $p_e^{(n)} = \mathbb{P}(\widehat{W}_n \neq W_n)$; let's assume $p_e^{(n)} \to 0$. We will prove that this implies $\limsup_n \frac{1}{n} H(W_n) \leq C$ as follows.

$$\begin{aligned} H(W_n) &= H(W_n \mid Y_1^n) + I(W_n; Y_1^n) \\ &\leq H(W_n \mid Y_1^n) + I(X_1^n; Y_1^n) \\ &\leq H(W_n \mid \widehat{W}_n) + I(X_1^n; Y_1^n) \end{aligned}$$

Fano's inequality says that $H(W_n \mid \widehat{W}_n) \leq h(p_e^{(n)}) + p_e^{(n)} \log(M_n - 1)$.

$$\leq h(p_e^{(n)}) + p_e^{(n)} \log M_n + I(X_1^n; Y_1^n)$$

To deal with the last term, use the chain rule to write

$$\begin{aligned} I(X_1^n; Y_1^n) &= H(Y_1^n) - H(Y_1^n X_1^n) \\ &\leq \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n \underbrace{H(Y_i \mid X_1^n, Y_1^n)}_{=H(Y_i \mid X_i)} \\ &= \sum_{i=1}^n H(Y_i \mid X_i) \\ &\leq nC. \end{aligned}$$

Our issue is now that $\log M_n$ looks like n. We can deal with this by noting that $H(W_n) = \log M_n$ on the left. So far, we have that

$$\log M_n \leq h(p_e^{(n)}) + p_e^{(n)} \log m_n + nC.$$

Hence,

$$(1 - p_e^{(n)}) \frac{\log M_n}{n} \leq \frac{h(p_e^{(n)})}{n} + C.$$

If $p_e^{(n)} \to 0$ as $n \to \infty$, this implies that $\limsup_n \frac{1}{n} \log M_n \leq C$. □

1.2 Communication with perfect noiseless feedback

Earlier, we had $X_i = e_{n,i}(m)$, where $e_n = (e_{n,1}, \ldots, e_{n,n})$ and $m \in [M_n]$ is a message.

Definition 1.2. Perfect noiseless feedback is when we have $X_i = e_{n,i}(m, Y_1, \ldots, Y_{i-1})$.

Theorem 1.2. *Perfect noiseless feedback cannot increase the rates at which communication is possible over a DMC.*

Proof. The achievability at rate $C - \varepsilon$ is the same as Shannon's coding theorem, since the encoder can ignore the feedback. But for the converse, we do not have a Markov chain. As before, write

$$\begin{aligned}\log M_n &= H(W_n) \\ &= H(W_n \mid Y_1^n) + I(W; Y_1^n) \\ &\le h(p_e^{(n)}) + p_e^{(n)} \log M_n,\end{aligned}$$

where $p_e^{(n)} := P(d_n(Y_1^n) \neq W_n)$. Note that we can still use Fano's inequality because we have the Markov chain $W_n - Y_1^n - \widehat{W}_n$. Here, Y_i conditioned on $(X_1^{i-1}, Y_1^{i-1}, X_i = x_i)$ has the law $p(y_i \mid x_i)$. Observe that $p(m, x_1^n, y_1^n) = \frac{1}{M_n} \prod_{i=1}^n \mathbb{1}_{\{x_i = e_i(m, y_1^{i-1})\}} p(y_i \mid x_i)$.

The chain rule gives

$$\begin{aligned}I(W_n, Y_1^n) &= H(Y_1^n) - H(Y_1^n \mid W_n) \\ &\le \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i \mid W_n, Y_1^{i-1})\end{aligned}$$

But $X_i = e_{n,i}(W_n, Y_1^{i-1})$, so $H(Y_i \mid W_n, Y^{i-1}) = H(Y_i \mid X_n, Y_1^{i-1}, X_i) = H(Y_i, X_i)$. So

$$I(W_n, Y_1^n) \le nC,$$

and the rest of the proof proceeds as before. □

1.3 Joint source channel coding

Model a source as a random sequence $(V_k, k \in \mathbb{Z})$ (think stationary and ergodic) with $V_k \in \mathscr{V}$, where $\mathscr{V}$ is finite.

Definition 1.3. A **source channel code at block length** n is an encoding map

$$e_n : \mathscr{V}^{\ell_n} \to \mathscr{X}^n$$

and a decoding map

$$d_n : \mathscr{Y}^n \to \mathscr{V}^{\ell_n}.$$

Note that ℓ_n might be different from n. Here, Y_1^n results from X_1^n over a DMC.

$$V_1^{\ell_n} \xrightarrow{e_n} \mathscr{X}_1^n \xrightarrow{\text{DMC}} \mathscr{Y}_1^n \xrightarrow{d_n} \mathscr{V}_1^{\ell_n}.$$

Theorem 1.3 (Joint source channel coding theorem)**.** *If the source has entropy rate $H(V)$, then there exists $((e_n, d_n), n \ge 1)$ with $\mathbb{P}(d_n(e_n(V_1^{\ell_n})) \neq V_1^{\ell_n}) \to 0$ if and only if*

$$\limsup_n \frac{\ell_n H(V)}{nC} \le 1.$$

Proof. Achievability: The idea is to compress the source and then use Shannon's channel coding theorem. Take $\ell_n = n$. If $H(V)/C \leq 1 - \delta$, we can compress V_1^n to $n(H(V) + \delta/2)$ bits with probability going to 0 as $n \to \infty$. Then send those bits over a DMC with error probability going to 0.

Converse: We have the Markov chain

$$V_1^{\ell_n} - X_1^n - Y_1^n - \widehat{V}_1^{\ell_n},$$

so Fano's inequality gives

$$H(V_1^{\ell_n} \mid \widehat{V}_1^{\ell_n}) \leq 1 + P_e^{(n)}(\ell_n \log |\mathscr{V}|).$$

We then have

$$\begin{aligned} H(V_1^n) &= H(V_1^{\ell_n} \mid \widehat{V}_1^{\ell_n}) + I(V_1^{\ell_n}; \widehat{V}_1^{\ell_n}) \\ &\leq 1 + p_e^{(n)}(\ell_n \log |\mathscr{V}|) + I(X_1^n; Y_1^n) \\ &\leq 1 + p_e^{(n)}(\ell n \log |\mathscr{V}|) + nC. \end{aligned}$$

Divide by ℓ_n and let $n \to \infty$ (we can assume without loss of generality that $\ell_n \to \infty$, otherwise we automatically have the limsup bounded by 1). We get

$$\frac{H(V_1^{\ell_n})}{\ell_n} \leq \frac{1}{\ell_n} + p_e^{(n)} \log |\mathscr{V}| + \frac{nC}{\ell_n}.$$

The left hand side converges to $H(V)$. The first term on the right goes to 0 because $\ell_n \to \infty$. The second term on the right goes to 0 because $p_e^{(n)} \to 0$ by assumption. $\square$