CIE6020/MAT3350
Selected Topics in Information Theory

Lecture 14: Converse of Channel Coding Theorem

---

15 March 2019

The Chinese University of Hong Kong, Shenzhen

Converse for Zero-Error Codes

Zero-Error codes

- Suppose we have a $(n, 2^{nR})$ code (f, ϕ) with $\lambda_{\max} = 0$.
- Let U be the uniform distribution on the message set.
- Note that $U \to \mathbf{X} \to \mathbf{Y} \to \hat{U}$ forms a markov chain, where $\mathbf{X} = f(U)$, $\hat{U} = U = \phi(\mathbf{Y})$.
- We can write

$$\begin{aligned} nR = H(U) &= H(U|\mathbf{Y}) + I(U;\mathbf{Y}) \\ &= I(U;\mathbf{Y}) \\ &\leq I(\mathbf{X};\mathbf{Y}) \\ &\leq \sum_{i=1}^{n} I(X_i;Y_i) \\ &\leq nC. \end{aligned}$$

Fano's Inequality

Fano's Inequality

Lemma
For random variables X and Y with the same alphabet $\mathcal{X}$,

$$H(X|Y) \leq P_e \log(|\mathcal{X}| - 1) + H(P_e),$$

where $P_e = \Pr\{X \neq Y\}$.

- If X is a function of Y, which is equivalent to $H(X|Y) = 0$, the guessing has zero error.
- We hope to estimate X with low probability of error only if the conditional entropy $H(X|Y)$ is small.

Proof of Fano's Inequality

- Define random variable Z with $Z = 0$ if $X = Y$ and $Z = 1$ otherwise.
- Then,

$$\begin{aligned} H(X|Y) &= H(X|Y) + H(Z|X,Y) \\ &= H(X,Z|Y) \\ &= H(Z|Y) + H(X|Z,Y) \\ &\leq H(Z) + H(X|Z,Y) \\ &= H(P_e) + H(X|Z,Y). \end{aligned}$$

- $H(X|Y, Z = 0) = 0$, and $H(X|Y, Z = 1) \leq \log(|\mathcal{X}| - 1)$.

Converse

A Channel Code

- Let R be an achievable rate.
- Consider an n-length code (f, φ) such that $\frac{1}{n} \log M > R - \epsilon$ and $\lambda_{\max} < \epsilon$.
- Let U be the uniform distributed random variable over the message set $\{1, 2, \ldots, M\}$.
- The codeword we transmit for U is the random variable $\mathbf{X} = f(U)$.
- Let $\mathbf{Y}$ be the output of the channel for input $\mathbf{X}$, i.e., $(\mathbf{X}, \mathbf{Y}) \sim p_{\mathbf{X}}(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x})$.
- Let $\hat{U} = \varphi(\mathbf{Y})$.
- We have a Markov chain

$$U \to \mathbf{X} \to \mathbf{Y} \to \hat{U}.$$

Bound on $I(\mathbf{X};\mathbf{Y})$

- Since the channel is memoryless,

$$\begin{aligned} H(\mathbf{Y}|\mathbf{X}) &= \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) H(\mathbf{Y}|\mathbf{X}=\mathbf{x}) \\ &= \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \sum_{i=1}^{n} H(Y_i|X_i=x_i) = \sum_{i=1}^{n} H(Y_i|X_i), \end{aligned}$$

- Hence,

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) &= H(\mathbf{Y}) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &= \sum_{i=1}^{n} I(X_i;Y_i) \leq nC. \end{aligned}$$

Converse

- Due to Fano's inequality and the bound on $I(\mathbf{X};\mathbf{Y})$,

$$\begin{aligned}\log M = H(U) &= H(U|\hat{U}) + I(U;\hat{U}) \\ &\leq H(U|\hat{U}) + I(\mathbf{X};\mathbf{Y}) \\ &\leq 1 + P_e \log(M-1) + nC \\ &< 1 + \epsilon \log M + nC,\end{aligned}$$

- which implies

$$R - \epsilon < \frac{1}{n}\log M < \frac{1}{1-\epsilon}\left(\frac{1}{n} + C\right).$$

- Since for an achievable rate R, we require the above inequality holds for any $\epsilon > 0$ and all sufficiently large n, we conclude that

$$R \leq C.$$

Feedback capacity

Definition
An (n, M) code for a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with feedback consists of a sequence of encoding functions

$$f_i : \{1, 2, \ldots, M\} \times \mathcal{Y}^{i-1} \to \mathcal{X}^n$$
$$(u, y_1, y_2, \cdots, y_{i-1}) \mapsto x_i$$

where y_i, $i = 1, \ldots, i-1$ are the first $i-1$ output symbols of the DMC, and a decoding function

$$\varphi : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$$
$$(y_1, y_2, \ldots, y_n) \mapsto \hat{u}.$$

Definition
The feedback capacity of a DMC C_{FB} is the supremum of all the achievable rates.

Theorem

For a DMC, $C_{FB} = C$.

Proof of Achievability.
$C_{\mathsf{FB}} \geq C$. $\square$

Proof outline of converse

- Let U be the uniform distributed random variable over the message set.
- For $i = 1, \dots, n$, define $X_i = f_i(U, Y^{i-1})$.
- Let Y_i be the output of the channel with X_i as the input.
- We do not have the markov chain $U \to \mathbf{X} \to \mathbf{Y} \to \hat{U}$.
- Using $U \to \mathbf{Y} \to \hat{U}$, we write

$$\begin{aligned} \log M &= H(U) \\ &= H(U|\hat{U}) + I(U;\hat{U}) \\ &\le H(U|\hat{U}) + I(U;\mathbf{Y}). \end{aligned}$$

- By the chain rule for entropy

$$\begin{aligned}
H(\mathbf{Y}|U) &= \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}) \\
&= \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}, X_i) \\
&= \sum_{i=1}^{n} H(Y_i|X_i).
\end{aligned}$$

- By the chain rule for entropy

$$
\begin{aligned}
H(\mathbf{Y}|U) &= \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}) \\
&= \sum_{i=1}^{n} H(Y_i|U, Y^{i-1}, X_i) \\
&= \sum_{i=1}^{n} H(Y_i|X_i).
\end{aligned}
$$

- Then,

$$
\begin{aligned}
I(U;\mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y}|U) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I(X_i;Y_i).
\end{aligned}
$$