Information Theory for Wireless Communication

Lecture 9: Converse to Channel Coding Theorem for Discrete-time Continuous Memoryless Channels

Lecture by Dr. Saif K. Mohammed

Scribe by T. V. K. Chaitanya

In this lecture notes, we first prove the converse of the channel coding theorem proved in Lecture 7 for discrete-time continuous memoryless channels. In the later part, we prove some information theory inequalities.

I. Channel Coding Theorem: Achievability

We defined the capacity of the discrete-time continuous memoryless channel as [Lecture 7]:

$$C(s) = \max_{f_X(x)} I(X;Y)$$
$$\text{subject to} \quad E_{f_X}(g(x)) \leq s \tag{1}$$

We proved that ***for any*** $0 \leq R < C(s)$***, there exists a sequence of*** $\left(n, 2^{nR}, \lambda_n\right)$ ***codes such that*** $\lambda_n \to 0$ ***as*** $n \to \infty$, where a (n, N, λ) code $C(s)$ is defined as a set of two-tuples containing codewords and their decoding regions, $C(s) = \{(u_1, A_1), \cdots, (u_N, A_N)\}$ with the following properties

1) $u_i \in \mathcal{R}^n$, $A_i \cap A_j = \phi, \forall i \neq j$, and $A_i \subseteq \mathcal{R}^n, \forall i$.
2) $\frac{\sum_{k=1}^{n} g(u_i(k))}{n} \leq s, \forall i$.
3) Average probability of error for the code is $\bar{\lambda}$ and the maximum probability of error of the code is $\lambda_M \leq \lambda$.

II. Converse to the Channel Coding Theorem

Converse: ***If there exists a sequence of codes*** $\left\{\left(n, 2^{nR}, \lambda_n\right)\right\}$ ***such that*** $\lambda_n \to 0$ ***as*** $n \to \infty$***, then*** $R \leq C(s)$. Before proving the converse, we prove some Lemmas.

Lemma 1. *Consider two probability density functions* $f_X(x)$ *and* $g_X(x)$*, then*

$$\int f_X(x) \log_2 \frac{f_X(x)}{g_X(x)} dx \geq 0$$

with equality if and only if $f_X(x) = g_X(x)$*(in the region* Θ *where* $f_X(x) > 0, \forall x \in \Theta$ *and* $\int_\Theta f_X(x)\, dx = 1$*).*

Proof: For $x \geq 0$, we have $\ln x \leq x - 1$. So we have

$$\ln \frac{g_X(x)}{f_X(x)} \leq \frac{g_X(x)}{f_X(x)} - 1,$$

from which, we have

$$\int f_X(x) \ln \frac{g_X(x)}{f_X(x)} dx \leq \int f_X(x) \left(\frac{g_X(x)}{f_X(x)} - 1\right) = 0.$$

■

Lemma 2. $I(X;Y)$ *is a concave function of* $f_X(.)$ *for a fixed* $f_{Y|X}(.)$.

Proof: Consider two input probability density functions (PDFs) $g_X(x)$ and $h_X(x)$. Let

$$f_X(x) \triangleq \lambda g_X(x) + (1-\lambda) h_X(x), \quad 0 \leq \lambda \leq 1. \tag{2}$$

be another PDF. We need to show that

$$I_f(X;Y) \geq \lambda I_g(X;Y) + (1-\lambda) I_h(X;Y), \tag{3}$$

where $I_f(X;Y)$ is the mutual information between X and Y when the input PDF is $f_X(x)$,

and is given by:

$$I_f(X;Y) = \int\int f_X(x)\, f_{Y|X}(y|x) \log_2 \frac{f_X(x)\, f_{Y|X}(y|x)}{f_X(x) \int f_X(t)\, f_{Y|X}(y|t)\, dt} dx\, dy. \tag{4}$$

We can similarly define $I_g(X;Y)$ and $I_h(X;Y)$. Now using (2) in (4), we have

$$\begin{aligned} I_f(X;Y) = \lambda &\int\int g_X(x)\, f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{\int f_X(t)\, f_{Y|X}(y|t)\, dt} dx\, dy \\ &+ (1-\lambda) \int\int h_X(x)\, f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{\int f_X(t)\, f_{Y|X}(y|t)\, dt} dx\, dy \end{aligned} \tag{5}$$

Now considering

$$\begin{aligned} & I_f(X;Y) - \lambda I_g(X;Y) - (1-\lambda) I_h(X;Y) \\ = & \quad \lambda \left\{ \int\int g_X(x)\, f_{Y|X}(y|x) \log_2 \frac{\overbrace{\int g_X(t)\, f_{Y|X}(y|t)\, dt}^{\triangleq f_g(y)}}{\underbrace{\int f_X(t)\, f_{Y|X}(y|t)\, dt}_{\triangleq f_f(y)}} dx\, dy \right\} \\ & + (1-\lambda) \left\{ \int\int h_X(x)\, f_{Y|X}(y|x) \log_2 \frac{\overbrace{\int h_X(t)\, f_{Y|X}(y|t)\, dt}^{\triangleq f_h(y)}}{\int f_X(t)\, f_{Y|X}(y|t)\, dt} dx\, dy \right\} \\ = & \quad \lambda \left\{ \int \log_2 \frac{f_g(y)}{f_f(y)} \left[\int g_X(x)\, f_{Y|X}(y|x)\, dx \right] dy \right\} \\ & + (1-\lambda) \left\{ \int \log_2 \frac{f_h(y)}{f_f(y)} \left[\int h_X(x)\, f_{Y|X}(y|x)\, dx \right] dy \right\} \\ = & \quad \lambda \left\{ \int f_g(y) \log_2 \frac{f_g(y)}{f_f(y)} dy \right\} + (1-\lambda) \left\{ \int f_h(y) \log_2 \frac{f_h(y)}{f_f(y)} dy \right\} \overset{(*)}{\geq} 0 \end{aligned} \tag{6}$$

where (*) follows from Lemma 1. ■

Lemma 3. *With the $C(s)$ definition in (1), we have the following results*

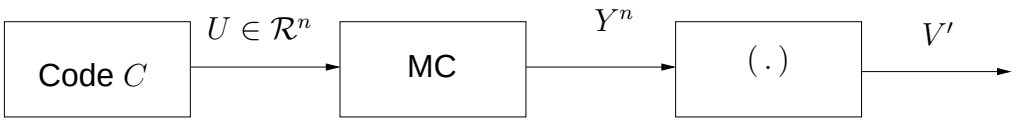

Figure 1. Communication over a discrete-time continuous memoryless channel.

1) $C(s)$ is a monotonically non-decreasing function of s.
2) $C(s)$ is concave in s. i.e., for $s_3 = \lambda s_1 + (1-\lambda) s_2$, and any $0 \leq \lambda \leq 1$, we have $C(s_3) \geq \lambda C(s_1) + (1-\lambda) C(s_2)$.

A. Proof of the Converse

To prove the converse, assuming that there exists a sequence of codes $\left(n, 2^{nR}, \lambda_n\right)$ with $\lambda_n \to 0$ as $n \to \infty$, we need to show that $R \leq C(s)$. This is equivalent to showing that for a code with $R = C(s) + \epsilon, \epsilon > 0$, $\lambda_n \nrightarrow 0$ as $n \to \infty$.

Consider a (n, N, λ) code $C = \{(u_1, A_1), \cdots, (u_N, A_N)\}$ with $A_1 \cup A_2 \cup \cdots \cup A_N = \mathcal{R}^n$ used for communication over a discrete-time continuous memoryless channel as shown in Fig. 1. We assume that all the codewords used for transmission on the MC are equally likely. i.e.,

$$\Pr(U = x^n) = \begin{cases} \frac{1}{N} & \text{if } x^n \in \{u_1, u_2, \cdots, u_N\} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The probability of error for the codeword u_i is given by

$$\lambda(u_i) = 1 - \int_{A_i} f_{Y^n|U}(y^n|u_i)\, dy^n \tag{8}$$

From the definition of the code C, we have

$$\max_i \lambda(u_i) \leq \lambda \tag{9}$$

and the average probability of error

$$\overline{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \lambda(u_i) \leq \lambda. \tag{10}$$

The decision metric at the receiver is given by:

$$V' = j, \quad \text{if } Y^n \in A_j, \forall 1 \leq j \leq N. \tag{11}$$

In Lecture 4, as a part of proof for weak converse to the channel coding theorem [1], we have proved the Fano's inequality

$$\begin{aligned} H\left(U|V'\right) &\leq \overline{\lambda}\log_2\left(N-1\right) - \overline{\lambda}\log_2\overline{\lambda} - \left(1-\overline{\lambda}\right)\log_2\left(1-\overline{\lambda}\right) \\ &< \lambda\log_2 N + 1 \end{aligned} \tag{12}$$

Therefore,

$$\begin{aligned} H\left(U\right) =& H\left(U\right) - H\left(U|V'\right) + H\left(U|V'\right) \\ =& I\left(U;V'\right) + H\left(U|V'\right) \\ <& I\left(U;V'\right) + \lambda\log_2 N + 1. \end{aligned} \tag{13}$$

Before proceeding further, we prove the following Lemma.

Lemma 4. $I\left(U;V'\right) \leq I\left(U;Y^n\right)$

Proof:

$$\begin{aligned} I\left(U;V'\right) &= \sum_i \sum_j p_{U,V'}\left(u_i,j\right)\log_2\frac{p_{V'|U}\left(j|u_i\right)}{p_{V'}\left(j\right)} \\ &= \sum_i \sum_j \frac{1}{N}\underbrace{\left\{\int_{A_j} f_{Y^n|U}\left(y^n|u_i\right)dy^n\right\}}_{=p_{V'|U}(j|u_i)}\log_2\frac{p_{V'|U}\left(j|u_i\right)}{p_{V'}\left(j\right)} \end{aligned} \tag{14}$$

and

$$\begin{aligned} I\left(U;Y^n\right) &= \sum_i p_U\left(u_i\right)\int_{\mathcal{R}^n} f_{Y^n|U}\left(y^n|u_i\right)\log_2\frac{f_{Y^n|U}\left(y^n|u_i\right)}{f_{Y^n}\left(y^n\right)}dy^n \\ &= \sum_i p_U\left(u_i\right)\sum_j\int_{A_j} f_{Y^n|U}\left(y^n|u_i\right)\log_2\frac{f_{Y^n|U}\left(y^n|u_i\right)}{f_{Y^n}\left(y^n\right)}dy^n \end{aligned} \tag{15}$$

Using (14) and (15), we have

$$I\left(U;Y^n\right)-I\left(U;V'\right)=\sum_i\sum_j\frac{1}{N}\int_{A_j}f_{Y^n|U}\left(y^n|u_i\right)\log_2\left[\frac{\frac{f_{Y^n|U}(y^n|u_i)}{p_{V'|U}(j|u_i)}}{\frac{f_{Y^n}(y^n)}{p_{V'}(j)}}\right]dy^n \tag{16}$$

For any given $i,j\in\{1,2,\cdots,N\}$, we define new probability density functions

$$g_{Y^n}^{(i,j)}\left(y^n\right)\triangleq\begin{cases}\frac{f_{Y^n|U}(y^n|u_i)}{p_{V'|U}(j|u_i)} & \text{if } y^n\in A_j\\ 0 & \text{if } y^n\notin A_j\end{cases}$$

and

$$h_{Y^n}^{j}\left(y^y\right)\triangleq\begin{cases}\frac{f_{Y^n}(y^n)}{p_{V'}(j)} & \text{if } y^n\in A_j\\ 0 & \text{if } y^n\notin A_j\end{cases}$$

we can easily see that

$$\int_{\mathcal{R}^n}g_{Y^n}^{(i,j)}\left(y^n\right)=\int_{A_j}\frac{f_{Y^n|U}\left(y^n|u_i\right)}{p_{V'|U}\left(j|u_i\right)}dy^n=\frac{1}{p_{V'|U}\left(j|u_i\right)}\int_{A_j}f_{Y^n|U}\left(y^n|u_i\right)dy^n=1,$$

and

$$\int_{\mathcal{R}^n}h_{Y^n}^{j}\left(y^y\right)=\int_{A_j}\frac{f_{Y^n}\left(y^n\right)}{p_{V'}\left(j\right)}dy^n=\frac{1}{p_{V'}\left(j\right)}\int_{A_j}f_{Y^n}\left(y^n\right)dy^n=1.$$

Now from (16), we have

$$I\left(U;Y^n\right)-I\left(U;V'\right)=\sum_i\sum_j\frac{1}{N}p_{V'|U}\left(j|u_i\right)\underbrace{\int_{A_j}\frac{f_{Y^n|U}\left(y^n|u_i\right)}{p_{V'|U}\left(j|u_i\right)}\log_2\left[\frac{\frac{f_{Y^n|U}(y^n|u_i)}{p_{V'|U}(j|u_i)}}{\frac{f_{Y^n}(y^n)}{p_{V'}(j)}}\right]dy^n}_{\geq 0\text{ (from Lemma 1)}}\geq 0 \tag{17}$$

■

Now using the result of Lemma 4 in (13), we have

$$\begin{aligned}H\left(U\right)&<I\left(U;V'\right)+\lambda\log_2 N+1\\&\leq I\left(U;Y^n\right)+\lambda\log_2 N+1\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{=} h\left(Y^n\right) - h\left(Y^n|U^1U^2\cdots U^n\right) + \lambda\log_2 N + 1 \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} h\left(Y_i\right) - \sum_{i=1}^{n} h\left(Y_i|U^i\right) + \lambda\log_2 N + 1 \\
&= \sum_{i=1}^{n} I\left(Y_i;U^i\right) + \lambda\log_2 N + 1
\end{aligned} \tag{18}$$

in (a), we have used the definition of $I\left(U;Y^n\right)$, where $U = \left[U\left(1\right)U\left(2\right)\cdots U\left(n\right)\right]$ with $U\left(i\right)$ denoting the random variable corresponding to the ith component of U. In (b), we have used $h\left(Y^n\right) \leq \sum_{i=1}^{n} h\left(Y_i\right)$ (follows from the chain rule and the conditional entropy properties) and that $h\left(Y^n|U\left(1\right)U\left(2\right)\cdots U\left(n\right)\right) = \sum_{i=1}^{n} h\left(Y_i|U\left(i\right)\right)$ (follows from the chain rule and the memoryless channel property). Now defining

$$s_i \triangleq E_{U^i}\left[g\left(U\left(i\right)\right)\right] = \frac{1}{N}\sum_{j=1}^{N} g\left(u_j\left(i\right)\right), \forall 1 \leq i \leq n \tag{19}$$

where $u_j\left(i\right)$ denotes the ith component of the jth codeword. In other words, the average in (19) is taken over ith component of all the N codewords $u_1, u_2, \cdots, u_N$. let

$$\begin{aligned}
&C\left(s_i\right) = \max_{f_X(x)} I\left(X;Y\right) \\
&\text{subject to } E_{f_X}\left(g\left(X\right)\right) \leq s_i
\end{aligned} \tag{20}$$

We can now easily see that

$$I\left(Y_i;U\left(i\right)\right) \leq C\left(s_i\right) \tag{21}$$

From (18) and (21), we can write

$$H\left(U\right) < \sum_{i=1}^{n} C\left(s_i\right) + \lambda\log_2 N + 1 \tag{22}$$

Now considering the summation term in (22), we can write

$$\begin{aligned}
\sum_{i=1}^{n} C\left(s_i\right) &= n\left[\frac{1}{n}\sum_{i=1}^{n} C\left(s_i\right)\right] \\
&\overset{(c)}{\leq} nC\left(\frac{\sum_i s_i}{n}\right)
\end{aligned}$$

$$\overset{(d)}{\leq} \quad nC\left(s\right). \tag{23}$$

In which, (c) and (d) follow from the concavity property and the monotonicity of $C\left(s\right)$ in Lemma 3[1]. Using (23) in (22), we have

$$H\left(U\right) \overset{\text{from (7)}}{=} \log_2 N \quad < \quad nC\left(s\right) + \lambda \log_2 N + 1 \tag{24}$$

Now assuming that we have $R = C\left(s\right) + \epsilon, \epsilon > 0$. Then the number of codewords in the code is

$$N = 2^{nR} = 2^{n(C(s)+\epsilon)}. \tag{25}$$

From (24) and (25), we have

$$\lambda > \frac{\epsilon}{C\left(s\right)+\epsilon} - \frac{1}{n\left(C\left(s\right)+\epsilon\right)} \tag{26}$$

from which it follows that $\lambda \nrightarrow 0$ as $n \to \infty$.

III. Miscellaneous Results

A. Data Processing Inequality

If X, Y and Z are three continuous random variables ,and that X,Z are conditionally independent given Y, i.e.,

$$f_{X,Z|Y}\left(xz|y\right) = f_{X|Y}\left(x|y\right) f_{Z|Y}\left(z|y\right)$$

then $I\left(X;Y\right) \geq I\left(X;Z\right)$.

Proof:

$$\begin{aligned} I\left(X;Y,Z\right) &= I\left(X;Z\right) + I\left(X;Y|Z\right) \\ &= I\left(X;Y\right) + I\left(X;Z|Y\right) \end{aligned} \tag{27}$$

[1] $\frac{\sum_{i=1}^{n} s_i}{n} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N}\frac{g\left(u_j(i)\right)}{n} = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}\frac{g\left(u_j(i)\right)}{n}$. For each $j = 1, 2, \cdots, N$, from the cost constraint for each codeword, we have $\sum_{i=1}^{n}\frac{g\left(u_j(i)\right)}{n} \leq s$. Hence $\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}\frac{g\left(u_j(i)\right)}{n} \leq s$.

however since X and Z are conditionally independent given Y, we have $I(X;Z|Y)=0$, which implies

$$I(X;Y)=I(X;Z)+I(X;Y|Z) \tag{28}$$

Since mutual information is non-negative, $I(X;Y|Z)\geq 0$. From which it follows that $I(X;Y)\geq I(X;Z)$. ■

B. Estimation Error and Differential Entropy

Given a random variable X and its estimate $\hat{X}$, we have

$$E\left[\left(X-\hat{X}\right)^2\right]\geq \frac{2^{2h(X)}}{2\pi e} \tag{29}$$

Proof: We have

$$E\left[\left(X-\hat{X}\right)^2\right]\geq \text{Var}(X) \tag{30}$$

and we know that

$$h(X)\leq \frac{1}{2}\log_2 2\pi e\text{Var}(X) \tag{31}$$

From (30) and (31), (29) follows. We have equality in (29) if and only if X is Gaussian (results in equality in (31)) and $\hat{X}=E(X)$ (results in equality in (30)). ■

REFERENCES

[1] A. Pitarokoilis, "Information Theory for Wireless Communication, Lecture 4: Converse to Channel Coding Theorems".