Multi-User Information Theory November 11th, 2009

Lecture 7

Lecturer: Haim Permuter *Scribe: Avihay Shirazi*

I. Compound Channel

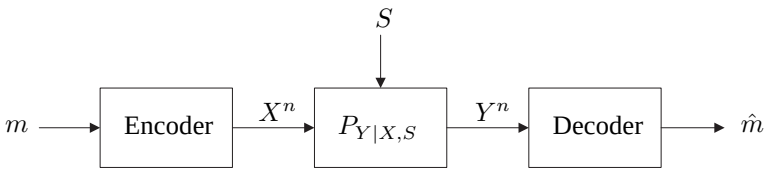

Fig. 1. The compound channel. Fixed state throughout the transmission block.

Definition 1 The *compound channel* is a channel with a state, where the state is constant over a whole transmission block.

Definition 2 A $(2^{nR}, n)$ code for a compound channel consists of the following:

1) An index set $\mathcal{M} = \{1, 2, \dots 2^{nR}\}$.
2) An encoding function $f: \mathcal{M} \longmapsto \mathcal{X}^n$.
3) A decoding function $g: \mathcal{Y}^n \longmapsto \mathcal{M}$.

Definition 3 The *average probability of error* $P_e^{(n)}$, is the probability $P_e^{(n)} \triangleq \Pr\{\hat{M} \neq M\}$, where M is chosen according to a uniform distribution over the set $\{1, 2, \dots 2^{nR}\}$.

Definition 4 A rate R is said to be *achievable* if there exists a sequence of codes $(2^{nR}, n)$ s.t. $P_e^{(n)} \xrightarrow[n\to\infty]{} 0$.

Definition 5 The *capacity* is the supremum of all achievable rates.

A. Compound channel with an unknown state

Theorem 1 (Compound channel capacity) [1] The *capacity of the compound channel* is given by:

$$C_{CC} = \max_{p(x)} \min_{s} I(X;Y|S=s), \tag{1}$$

Example 1 Suppose $\mathcal{S} = \{0, 1\}$, and consider the compound channel in Fig. 2.
We assume that the encoder does not know the state (we can assume, though, that the decoder is provided with the state information. If it is not the case initially, we can send a finite training sequence and estimate the state with arbitrarily small probability of error). Let us denote $\Pr\{X = 0\} = p$ and calculate the

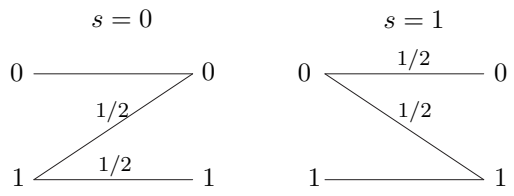

Fig. 2. Z-channel and S-channel.

capacity of the channel in the example:

For $S = 0$,

$$\begin{aligned} I(X;Y|S=0) &= H(Y|S=0) - H(Y|X,S=0) \\ &= H\Big(\frac{1}{2} + \frac{p}{2}\Big) - (1-p) \\ &= H\Big(\frac{1}{2} - \frac{p}{2}\Big) - (1-p). \end{aligned}$$

For $S = 1$,

$$\begin{aligned} I(X;Y|S=1) &= H(Y|S=1) - H(Y|X,S=1) \\ &= H\Big(\frac{p}{2}\Big) - p. \end{aligned}$$

We can now calculate the capacity of the channel using the given formula:

$$\begin{aligned} C &= \max_p \min_s \left\{ H\Big(\frac{1-p}{2}\Big) - (1-p) \ , \ H\Big(\frac{p}{2}\Big) - p \right\} \\ &= \max_p \begin{cases} H\Big(\frac{1-p}{2}\Big) - (1-p) & \text{if } p \le \frac{1}{2}, \\ H\Big(\frac{p}{2}\Big) - p & \text{if } p \ge \frac{1}{2} \end{cases} \\ &\overset{(a)}{=} H\Big(\frac{1}{4}\Big) - \frac{1}{2} \\ &= 0.3113 \text{ bit}, \end{aligned}$$

where we get (a) for $p = \frac{1}{2}$. (See Fig. 3).

Note that the capacity of each channel by itself is $C = H(\frac{1}{5}) - \frac{2}{5} = 0.3219$ bit, which is strictly higher than the capacity we found. Moreover, the compound channel is dictating capacity of the worst state for us, hence, even if we knew that one state is more likely than the other, still it would not change the capacity.

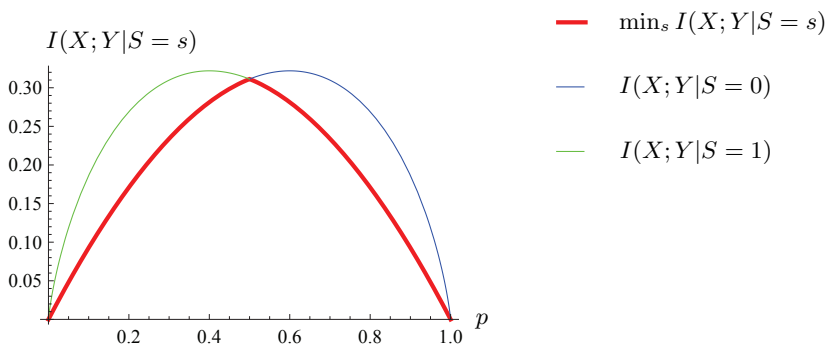

Fig. 3. The achievable rates scheme for *Example 1*

Proof of Theorem 1

Achievability in Theorem 1: Assume that the decoder is provided with full state information (CSIR, which is Channel State Information at the Receiver). Fix $p(x),\ p(s)$ and generate 2^{nR} codewords at random according to the distribution $p(x^n) = \Pi_{i=1}^{n} p(x_i)$, $X_i \sim\ p(x)\ i.i.d.$. Index the codewords $X^n(m), m \in 2^{nR}$, and reveal the content of this codebook ($\mathcal{C}$) to the sender and the reciever .

Encoding: Given the word m, the encoder sends the sequence $X^n(m)$ over the channel.

Decoding: Given $Y^n,\ S$, the decoder declares that $X^n(\hat{M})$ was sent, if there is one and only one sequence $X^n(\hat{M}) \in \mathcal{C}$ s.t.

$$\left(X^n(\hat{M}), Y^n\right) \in T_\epsilon^{(n)}\big(X, Y|S=s\big). \tag{2}$$

If there is no such sequence, or there is more than one, an error is declared.

Analysis of probability of error: Let us assume that the message $M = 1$ was sent. Define the following events:

$$E_1 = \left\{(X^n(1), Y^n) \notin T_\epsilon^{(n)}\big(X, Y|S=s\big)\right\}, \tag{3}$$

$$E_2 = \left\{\exists j \neq 1 : (X^n(j), Y^n) \in T_\epsilon^{(n)}\big(X, Y|S=s\big)\right\}. \tag{4}$$

Then, by union of events bound,

$$P_e^{(n)} = P(E_1 \cup E_2) \leq P(E_1) + P(E_2). \tag{5}$$

Because of the L.L.N. $P(E_1) \longrightarrow 0$.

Define $E_{2j} = \Big\{(X^n(j), Y^n) \in T_\epsilon^{(n)}\big(X, Y|S=s\big)\Big\}$, and note that

$$P(E_2) = P\left(\bigcup_{j=2}^{2^{nR}} E_{2j}\right) \tag{6}$$

$$\leq \sum_{j=2}^{2^{nR}} P(E_{2j}). \tag{7}$$

$$\tag{8}$$

Let $j \neq 1$, then

$$\Pr\Big\{\big(X^n(j), Y^n\big) \in T_\epsilon^{(n)}\big(X, Y|S=s\big)\Big\} \leq 2^{-n\left(I(X;Y|S=s)-\epsilon\right)}. \tag{9}$$

Therefore,

$$P(E_2) \leq 2^{nR-nI(X;Y|S=s)+n\epsilon}, \tag{10}$$

and this tends to 0 for $\quad R \leq I(X;Y|S=s),\ \ s \in \mathcal{S}$.

□

Converse

Fix a $(2^{nR}, n)$ code with probability of error $P_e^{(n)}$, then,

$$\begin{aligned}
nR &= H(M) && (11)\\
&= H(M|S=s) && (12)\\
&= H(M|S=s) - H(M|Y^n, S=s) + H(M|Y^n, S=s) && (13)\\
&= I(M; Y^n|S=s) + H(M|Y^n, S=s) && (14)\\
&\overset{(a)}{\leq} I(X^n; Y^n|S=s) + n\epsilon_n(s) && (15)\\
&= H(Y^n|S=s) - H(Y^n|, X^n, S=s) + n\epsilon_n && (16)\\
&\overset{(b)}{=} H(Y^n|S=s) - \sum_{i=1}^{n} H(Y_i|X_i, S=s) + n\epsilon_n(s) && (17)\\
&= \sum_{i=1}^{n} H(Y_i|Y_1^{i-1}, S=s) - \sum_{i=1}^{n} H(Y_i|X_i, S=s) + n\epsilon_n(s) && (18)\\
&\overset{(c)}{\leq} \sum_{i=1}^{n} \Big(H(Y_i|S=s) - H(Y_i|X_i, S=s)\Big) + n\epsilon_n(s) && (19)\\
&= \sum_{i=1}^{n} I(X_i; Y_i|S=s) + n\epsilon_n(s), && (20)\\
& && (21)
\end{aligned}$$

where

(a) follows from Fano's inequality and the data processing inequality

(b) follows from the fact that the channel is memoryless

(c) follows from the fact that conditioning decreases entropy

Let us now introduce the time sharing random variable Q, where $Q \sim Uniform[1, 2, \dots n]$ and independent of (M, X^n, Y^n, s). Let $X := X_Q,\ Y := Y_Q$. Then $Q - X - Y$ forms a Markov chain (recall that the channel is defined by $P_{Y|X,S}$), and hence,

$$\begin{aligned} nR &\leq n\sum_{i=1}^{n}\frac{1}{n}I(Y_Q; X_Q|S=s, Q=i) + n\epsilon_n(s) && (22)\\ &= nI(X_Q; Y_Q|S=s, Q) + n\epsilon_n(s) && (23)\\ &\overset{(d)}{\leq} nI(X_Q, Q; Y_Q|S=s) + n\epsilon_n(s) && (24)\\ &\overset{(e)}{=} nI(X; Y|S=s) + n\epsilon_n(s) && (25)\\ R &\leq I(X; Y|S=s) + \epsilon_n(s). && (26)\end{aligned}$$

This is true for every $s \in \mathcal{S}$, and in particular,

$$R \leq \min_s I(Y; X|S=s) + \epsilon_n(s). \quad (27)$$

Since $\epsilon_n(s) \xrightarrow[n\to\infty]{} 0$ for all $s \in \mathcal{S}$, we can conclude

$$R \leq \max_{p(x)} \min_s I(X; Y|S=s), \quad (28)$$

where

(d) follows from the mutual information chain rule

(e) follows from the Markov chain $Q - X - Y$ and the mutual information chain rule

This completes the proof of the compound channel capacity formula.

□

B. Compound channel with a state known to the encoder

Let us consider the same compound channel, but this time with full CSIT (Channel State Information at the Transmitter). For this discussion we will first note the following inequality:

Lemma 1

$$\max_b \min_a f(a, b) \leq \min_a \max_b f(a, b). \quad (29)$$

Proof:

$$\min_{b'} f(a, b') \leq f(a, b), \qquad \forall a, b. \quad (30)$$

Take $\max_a$ on both sides,

$$\max_a \min_{b'} f(a,b') \leq \max_a f(a,b), \qquad \forall b. \tag{31}$$

This is true for all b, and in particular for

$$b^* = \arg\min_b \max_a f(a,b). \tag{32}$$

Hence, we obtain

$$\max_a \min_{b'} f(a,b') \leq \max_a f(a,b^*). \tag{33}$$

■

Let us define a code for the compound channel with CSIT:

Definition 6 A $(2^{nR}, n)$ code for a compound channel with CSIT consists of the following:

1) An index set $\mathcal{M} = \{1, 2, \ldots 2^{nR}\}$.
2) An encoding function $f: \mathcal{M} \times \mathcal{S} \longmapsto \mathcal{X}^n$.
3) A decoding function $g: \mathcal{Y}^n \longmapsto \mathcal{M}$.

Theorem 2 (Capacity of compound channel with state information at the encoder) The *capacity of the compound channel with CSIT* is given by:

$$C_{CC-CSIT} = \max_{p(x|s)} \min_s I(X;Y|S=s) \tag{34}$$

$$= \min_s \max_{p(x)} I(X;Y|S=s). \tag{35}$$

The encoder can be informed in two ways: either by directly informing the transmitter, or by feedback (with a short training sequence). Note that the capacity of the compound channel with CSIT is greater or equals to the one without state information, as following from Theorem 1.

Remark 1 In this case, in contradiction to what we learned in *Lecture. 2.* ,the feedback **does** increase the capacity. The reason for this is that this channel is not memoryless. The state $S = s$ is constant over the whole block length, thus, at a given time within the block, the present state does depend on the state in the past. Therefor, the channel has memory.

Example 2 Let us modify the channel in Example 1 a little, and add CSIT. The capacity of the new channel is:

$$C_{CC-CSIT} = C_0 = C_1 = H\Big(\frac{1}{5}\Big) - \frac{2}{5}\text{bit}.$$

We can see that the capacity of the new channel equals the capacity of each of the sub-channels. This follows from:

$$\min_s \max_p \{I(X;Y|S=0), I(X;Y|S=1)\}$$

$$= \min_s \{\max_p I(X;Y|S=0) \ , \ \max_p I(X;Y|S=1)\}$$
$$= \min_s \{C_{S=0}, C_{S=1}\}$$
$$= H\left(\tfrac{1}{5}\right) - \tfrac{2}{5} \quad \text{bit},$$

which is strictly greater than the capacity of the channel in Example 1.

II. Channels with Causal State Information at the Encoder

In this section we consider a channel with a random state, where the state is known to the encoder in a causal way. Let the state $S_i \sim p(s) \quad i.i.d.$.

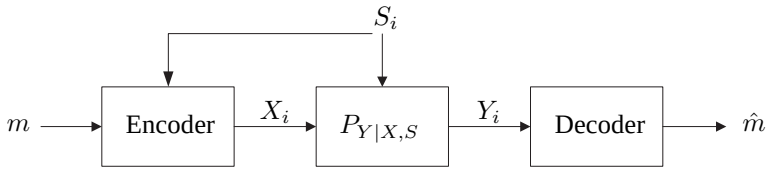

Fig. 4. State dependent channel with causal CSIT (Channel State Information at the Transmitter).

Definition 7 A $(2^{nR}, n)$ code for a channel with causal CSIT consists of the following:

1) An index set $\mathcal{M} = \{1, 2, \ldots 2^{nR}\}$
2) An encoding function $f_i : \ \mathcal{M} \times \mathcal{S}_1^i \longmapsto \mathcal{X}_i$
3) A decoding function $g : \ \mathcal{Y}^n \longmapsto \mathcal{M}$

A rate R is achievable if there exists a sequence of codes $(2^{nR}, n)$ s.t. $P_e^{(n)} \underset{n\to\infty}{\longrightarrow} 0$.
Capacity is the supremum of all achievable rates.

Theorem 3 (Capacity of channel with causal state information at the encoder) [2] The capacity of a channel with causal CSIT is given by:

$$C = \max_{\substack{p(u) \\ x=f(u,s)}} I(U;Y), \tag{36}$$

where $p(s,u,x,y) = p(s)p(u)p(x|s,u)p(y|x,s)$ and $p(x|s,u) \in \{0,1\}$, the letter which defines X as a deterministic function of U and S.

Proof of Theorem 3

Converse

Fix a $(2^{nR}, n)$ code with probability of error $P_e^{(n)}$, then,

$$nR \quad = \quad H(M) \tag{37}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} && I(M;Y^n) + n\epsilon_n && (38)\\
&= && H(Y^n) - H(Y^n|M) + n\epsilon_n && (39)\\
&= && \sum_{i=1}^{n} \Big(H(Y_i|Y_1^{i-1}) - H(Y_i|M, Y_1^{i-1}) \Big) + n\epsilon_n && (40)\\
&\overset{(b)}{\leq} && \sum_{i=1}^{n} \Big(H(Y_i) - H(Y_i|M, Y_1^{i-1}) \Big) + n\epsilon_n && (41)\\
&= && \sum_{i=1}^{n} I(M, Y_1^{i-1}; Y_i) + n\epsilon_n && (42)\\
&\overset{(c)}{\leq} && \sum_{i=1}^{n} I(M, Y_1^{i-1}, S_1^{i-1}; Y_i) + n\epsilon_n && (43)\\
&\overset{(d)}{=} && \sum_{i=1}^{n} I(M, Y_1^{i-1}, X_1^{i-1}, S_1^{i-1}; Y_i) + n\epsilon_n && (44)\\
&\overset{(e)}{=} && \sum_{i=1}^{n} I(M, X_1^{i-1}, S_1^{i-1}; Y_i) + n\epsilon_n && (45)\\
&\overset{(f)}{=} && \sum_{i=1}^{n} I(\underbrace{M, S_1^{i-1}}_{U_i}; Y_i) + n\epsilon_n && (46)\\
&\overset{(g)}{\leq} && n \max_{\substack{p(u)\\ x=f(u,s)}} I(U;Y) + n\epsilon_n, && (47)
\end{aligned}
$$

where

(a) follows from Fano's inequality

(b) follows from the fact that conditioning reduces entropy

(c) follows from the properties of mutual information

(d) follows from the fact that X_1^{i-1} is a function of M, S_1^{i-1}

(e) follows from the Markov chain $Y_1^{i-1} \longleftrightarrow (X_1^{i-1}, S_1^{i-1}) \longleftrightarrow Y_i$

(f) follows from the same reason presented in (d)

(g) follows from the following two facts:

(i) $X_i = f_i(U_i, S_i)$

(ii) U is independent of S

Let us show (i):

$$
\begin{aligned}
X_i &= f_i(M, S_1^i)\\
&= \tilde{f}_i(M, S_1^{i-1}, S_i, Y_1^{i-1})\\
&= \hat{f}_i(U_i, S_i).
\end{aligned}
$$

Let us show (ii):

$$p(m, s_1^i, y_1^{i-1}) = p(s_i)\, \underbrace{p(s_1^{i-1})p(m)p(y_1^{i-1}|\underbrace{m, s_1^{i-1}}_{x_1^{i-1}}, s_i)}_{p(y_1^{i-1}|x_1^{i-1}, s_1^{i-1})}\,.$$

□

REFERENCES

[1] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Stat., 30:1229-1241, 1959.*

[2] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Dev., vol. 2, pages 289-293, 1958*.