Mathematical methods in communication December 14th, 2009

Lecture 9

Lecturer: Haim Permuter *Scribe: Itzhak Tamo*

I. Non causal state information-Gelfand-Pinsker Theorem

We consider the channel coding problem depicted in Figure 1: Where the channel is DMC with s state

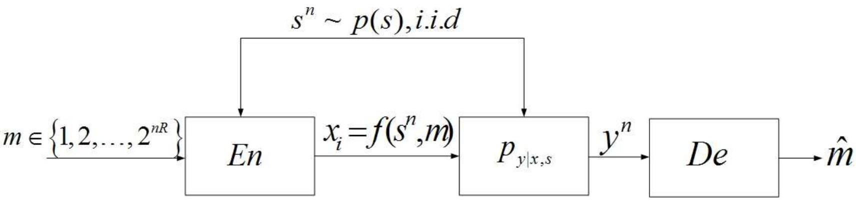

Fig. 1. Channel with state $s \sim p(s)$, distributed i.i.d and known causally at the encoder

$\mathcal{X} \times S$,$p(y|x,s)p(s), \mathcal{Y}$, and the state sequence, S^n, is i.i.d distributed according to $\sim p(s)$ and is known non causaly at the encoder. The defintions of *Achievability, Capacity* and *Error Probablity* are as before.

Definition 1 A code of rate R is function $f : \{1, 2, ..., 2^{nR}\} \times \mathcal{S}^n \to \mathcal{X}^n$, i.e. every codeword, x^n is a function of the message, $m \in \{1, 2, ..., 2^{nR}\}$ and the state sequence, s^n.

Theorem 1 [**Gelfand-Pinsker Theorem [1]]:** The *Capacity* of the DMC with state that is i.i.d distributed according to $\sim p(s)$ and is available noncausally only at the encoder is:

$$C = \max_{p(u|s), x=f(u,s)} (I(U;Y) - I(U;S)),$$

where $|\mathcal{S}| \le \min\{|\mathcal{X}||\mathcal{S}|, |\mathcal{Y}| + |\mathcal{S}| - 1\}$ and f is a deterministic function of u and s.

Example: First we deal with a binary case. Find the capacity of the channel depicted in Figure 2:

$$Y = X \oplus S \oplus Z,$$

Where $S \sim Bernoulli(p), Z \sim Bernoulli(q)$ and the state sequence is S^n known non causally at the encoder.

Solution: Answer: $C = 1 - H(q)$

- **Achievability:** Encode the message independently of S^n, and then do XOR to the codeword with the state vector, s^n. Thus the decoder gets $Y = X \oplus S \oplus S \oplus Z = X \oplus Z$, therefore

$$C = \max_{p(x)} I(X;Y) = 1 - H(q)$$

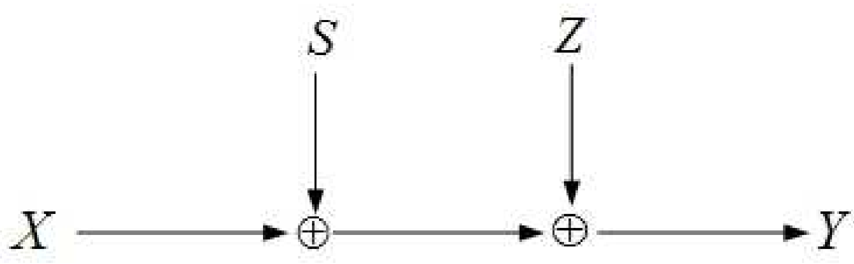

Fig. 2. Xor channel scheme

- **Achievability by Gelfand-Pinsker Theorem:** Let $U \sim Be(\frac{1}{2})$ independent of the state S, and let $X = U \oplus S$ then,

$$\begin{aligned} Y &= X \oplus S \oplus Z \\ &= U \oplus S \oplus S \oplus Z \\ &= U \oplus Z, \end{aligned}$$

and

$$\begin{aligned} C &= \max_{p_{u|s}, x=f(u,s)} I(U;Y) - I(U;S) \\ &\geq I(U;U \oplus Z) - I(U;S) \\ &= I(U;U \oplus Z) \\ &= H(U \oplus Z) - H(Z) \\ &= 1 - H(q) \end{aligned}$$

- **Converse:** Assume S^n is known to the encoder and decoder, thus the channel is reduced to an ordinary BSC with probability of error q, and $C \leq 1 - H(q)$

We now move on to prove Theorem 1:

Proof of Achievability:

- **Design of the code:** Fix $p(u|s)$ and $x = f(u,s)$. Generate randomly using $p(u)$, i.i.d, a $2^{n(I(U;Y)-\epsilon)}$ codewords, i.e. $(u^n(1), u^n(2), ..., u^n(R^{'}))$, where $R^{'} = (I(U;Y) - \epsilon)$. Generate 2^{nR} bins, one for each message $m \in \{1, 2, ..., 2^{nR}\}$, where $R = (I(U;Y) - I(U;S) - 2\epsilon)$. Distribute uniformly each of the codewords into one of the bins. Therefore for each message m we have generated a subcode $C(m)$ of size $\frac{2^{n(I(U;Y)-\epsilon)}}{2^{n(I(U;Y)-I(U;S)-2\epsilon)}} = 2^{n(I(U;S)+\epsilon)}$ which made of the codewords in bin m.

- **Encoding:** To send massage m, the encoder chooses a codeword u^n from bin m such that is jointly typical with the state sequence s^n, i.e.

$$(u^n, s^n) \in A^{\epsilon}(S, U).$$

The input to the channel at time i is $x_i = f(u_i, s_i)$

- **Decoding:** Looks for a codeword $\hat{u}^n$ that is jointly typical with the received codeword, y^n, i.e.

$$(\hat{u}^n, y^n) \in A^{\epsilon}(U, Y),$$

then declares the message $\hat{m}$ that is associated to the bin which contains $\hat{u}^n$.

- **Probability of error analysis:** An error occurs in the following cases:

1) There is no codeword in bin m that is associated to the given state sequence s^n:

$$E_1 = \{\forall u^n \in C(m), (u^n, s^n) \notin A^{\epsilon}(U, S)\}$$

2) we found a codeword, u^n, in bin m that is jointly typical with the state sequence s^n and sent $x^n = f(u^n, s^n)$, but the received codeword is not jointly typical with u^n :

$$E_2 = \{(u^n, y^n) \notin A^{\epsilon}(U, Y)\}$$

3) There exists $\hat{u}^n$ in bin $\hat{m}$ such that $\hat{m} \neq m$, that is jointly typical with the received codeword, y^n,i.e.

$$E_3^{\hat{m}} = \{\exists \hat{u}^n \in C(\hat{m}), \hat{m} \neq m, (\hat{u}^n, y^n) \in A^{\epsilon}(U, Y)\}$$

W.l.o.g we can assume that $m = 1$ therefore:

$$\begin{aligned} P_e^{(n)} &= P(\hat{m} \neq m | m = 1) \\ &= P(E_1 \cup E_2 \cup_{\hat{m}=2}^{2^{nR}} E_3^{\hat{m}}) \\ &\overset{(a)}{\leq} P(E_1) + P(E_2) + \sum_{\hat{m}=2}^{2^{nR}} P(E_3^{\hat{m}}), \end{aligned}$$

Where

(a) Union Bound.

We will see that each of the terms tends to zero as n tends to infinity.

1) In each bin there are $2^{n(I(U,S))+\epsilon}$ codewords, then according to the covering lemma (see lecture 10), with high probability, at least one codeword is jointly typical with s^n. In other words $P(E_1) \to 0$ as $n \to \infty$.

2) u^n and s^n are jointly typical by the choice of u^n. x is a function of (u, s) thus $(x^n, u^n, s^n) \in A_n^{\epsilon}(X, U, S)$. Therefore by the weak law of large number with high probability $(u^n, y^n) \in A_n^{\epsilon}(U, Y)$. i.e. $P(E_2) \to 0$ as $n \to \infty$

3) Let $\hat{m} \neq 1$, then $P((\hat{u}_n, y^n) \in A_n^\epsilon(U,Y)) \leq 2^{-n(I(U;Y)-3\epsilon)}$ for some $\hat{u}_n \in C(\hat{m})$ (see [2, Theorem 2.7.4 pp.33]). Thus,

$$\sum_{\hat{m}=2}^{2^{nR}} P(E_3^{\hat{m}}) = \sum_{\hat{m}=2}^{2^{nR}} \sum_{\hat{u}_n \in C(\hat{m})} P((\hat{u}_n, y^n) \in A_n^\epsilon(U,Y))$$

$$\leq \sum_{\hat{m}=1}^{2^{nR}} \sum_{\hat{u}_n \in C(\hat{m})} 2^{-n(I(U;Y)-3\epsilon)}$$

$$\leq 2^{n(I(U;Y)-I(U;S)-2\epsilon)} 2^{n(I(U;S)+\epsilon)} 2^{-n(I(U;Y)-3\epsilon)}$$

$$2^{-n\epsilon} \to 0.$$

Thus we have shown that under this encoding scheme the $P_e^{(n)} \to 0$ as $n \to \infty$, which means the rate R is achievable.

Proof of Converse:

Let R be an achievable rate, i.e. there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The trick is to find the auxilary random variable U_i that forms the markov chain $U_i \to (X_i, S_i) \to Y_i$. We can bound the rate R as

$$nR = H(M)$$

$$= H(M) - H(M|Y^n) + H(M|Y^n)$$

$$\stackrel{(a)}{\leq} I(M;Y^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, M) + n\epsilon_n$$

$$\stackrel{(b)}{\leq} \sum_{i=1}^{n} H(Y_i) - H(Y_i|Y^{i-1}, M) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(Y_i; Y^{i-1}, M) + n\epsilon_n$$

$$\stackrel{(c)}{\leq} \sum_{i=1}^{n} I(M, Y^{i-1}, S_{i+1}^n; Y_i) - I(Y_i; S_{i+1}^n|M, Y^{i-1}) + n\epsilon_n$$

$$\stackrel{(d)}{=} \sum_{i=1}^{n} I(M, Y^{i-1}, S_{i+1}^n; Y_i) - I(Y^{i-1}; S_i|S_{i+1}^n, M) + n\epsilon_n$$

$$\stackrel{(e)}{=} \sum_{i=1}^{n} I(M, Y^{i-1}, S_{i+1}^n; Y_i) - I(S_{i+1}^n, M, Y^{i-1}; S_i) + n\epsilon_n,$$

where

(a) Fano's inequality,

(b) coditioning reduces enropy,

(c) chain rule,

(d) Csiszar sum identity: $\sum_{i=1}^{n} I(X_{i+1}^n; Y_i|Y^{i-1}) = \sum_{i=1}^{n} I(Y^{i-1}; X_i|X_{i+1}^n)$ [3, HW 3, question 7]

(e) (M, S_{i+1}^n) is independent of S_i.

Now define $U_i \stackrel{def}{=} (M, Y^{i-1}, S_{i+1}^n)$ for $1 \leq i \leq n$, then we get:

$$
\begin{aligned}
nR &\leq \sum_{i=1}^{n} I(U_i; Y_i) - I(U_i, S_i) + n\epsilon_n \\
&\leq n \max_{p(u,x|s)} (I(U;Y) - I(U;S)) + n\epsilon_n
\end{aligned}
$$

We are almost done, we only have to show now that it suffices to maximize over $p(u|s)$ and a deterministic function $x = f(u,s)$, i.e. $p(u,x|s) = p(u|s)p(x|u,s)$ where $p(x|u,s) = 0, 1$. Note that $p(x|u,s) = 0, 1$. means that x is a deterministic functin of u, s. Fix $p(u|s)$ and note that the maximization in Gelfand-Pinsker formula is done only over $I(U;Y)$ because $I(U;S)$ is fixed by fixing $p(u|s)$. By [2, Theorem 2.7.4 pp.33] we know that mutual information $I(U;Y)$ is a convex function of $p(y|u)$ for a fixed $p(u|s)$. Noting that the Complete probability formula:

$$
p(y|u) = \sum_{x,s} p(s|u)p(x|u,s)p(y|x,s)
$$

is linear in $p(x|u,s)$ we conclude that $I(U;Y)$ is convex also in $p(x|u,s)$ for a fixed $p(u|s)$. This implies that the maximum of $I(U;Y)$ is achieved at the extreme points of the set of $P(x|u,s)$, that is $P(x|u,s) = 0, 1$. This completes the proof of the converse.

References

[1] Gel'fand, S. I. and Pinsker, M. S., '*Coding for Channel with Random Parameters*'. Problems of Control Theory, vol.9, no. 1, pp.19-31,1980

[2] T. M. Cover and J. A. Thomas, '*Elements of Information Theory.*'. Wiley, New York, 2nd edition 2006

[3] Multi-User Information Theory, http://www.ee.bgu.ac.il/ multi/HW3/hw3.pdf