Multi-user Information Theory 2 December 20th, 2012

Lecture 7

Lecturer: Haim Permuter *Scribe: Oron Sabag*

I. Degraded Message Set Broadcast Channel

We define a new setup as an extension for the BC channel. In the degraded message set BC the encoder transmits 2 messages $\{M_0, M_1\}$ over the channel. M_0 is decoded in both decoders, i.e, common message and M_1 is decoded only at the first decoder. A figure of this channel is described in Fig.1.

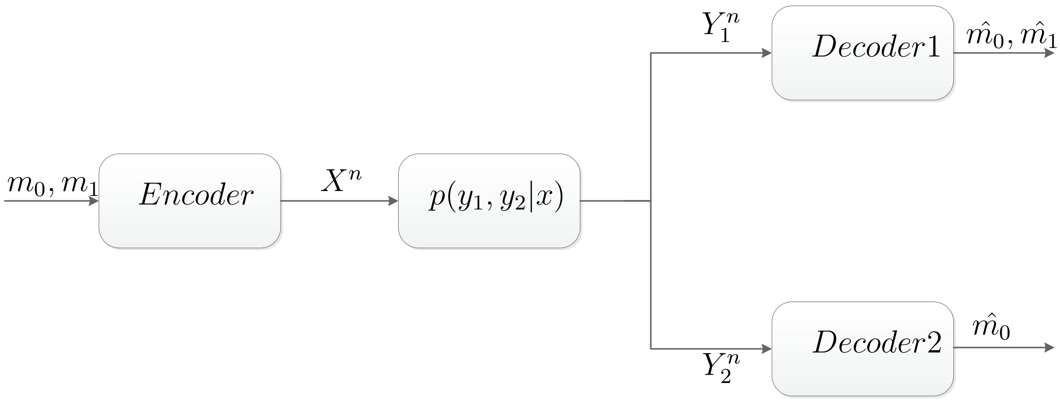


Fig. 1. Broadcast channel with degraded message set.

We begin with some basic definitions related to this setting:

Definition 1 (Memoryless BC) A broadcast channel consists of an input alphabet $\mathcal{X}$, two outputs alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a probability channel function $p(y_1, y_2|x)$. The channel is called a memoryless channel if

$$p(y_{1,i}, y_{2,i}|x^i, y_1^{i-1}, y_2^{i-1}) = p(y_{1,i}, y_{2,i}|x_i). \tag{1}$$

Definition 2 (Code for the message set degraded BC) A $((2^{nR_0}, 2^{nR_1}), n)$ code for the degraded message set BC consists of two sets of integers $\mathcal{M}_0 = \{1, 2, \cdots, 2^{nR_0}\}$ and

$\mathcal{M}_1 = \{1, 2, \cdots, 2^{nR_1}\}$, called the message sets. There is an encoding function:

$$f : \mathcal{M}_0 \times \mathcal{M}_1 \to \mathcal{X}^n \tag{2}$$

and two decoding functions:

$$g_1 : \mathcal{Y}_1^n \to \mathcal{M}_0 \times \mathcal{M}_1 \tag{3}$$

$$g_2 : \mathcal{Y}_2^n \to \mathcal{M}_0. \tag{4}$$

Definition 3 (Average probability of error) The average probability of error is the probability that one of the decoded messages is not equal to the transmitted messages. That is, $P_e^{(n)} = \Pr\left(\{M_1 \neq \hat{M}_1\} \cup \{M_2 \neq \hat{M}_2\}\right) = \Pr\left((M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)\right)$. We assume that the messages (M_1, M_2) are distributed uniformly and independent of each other over $2^{nR_0} \times 2^{nR_1}$.

Definition 4 (Achievable rate pair) A rate pair (R_1, R_2) is said to be achievable for the degraded message set BC if there exists a sequence of $((2^{nR_0}, 2^{nR_1}), n)$ codes s.t $P_e^{(n)} \to 0$ as $n \to \infty$.

Definition 5 (Capacity region) The capacity region is the closure of the union of all achievable rate pairs.

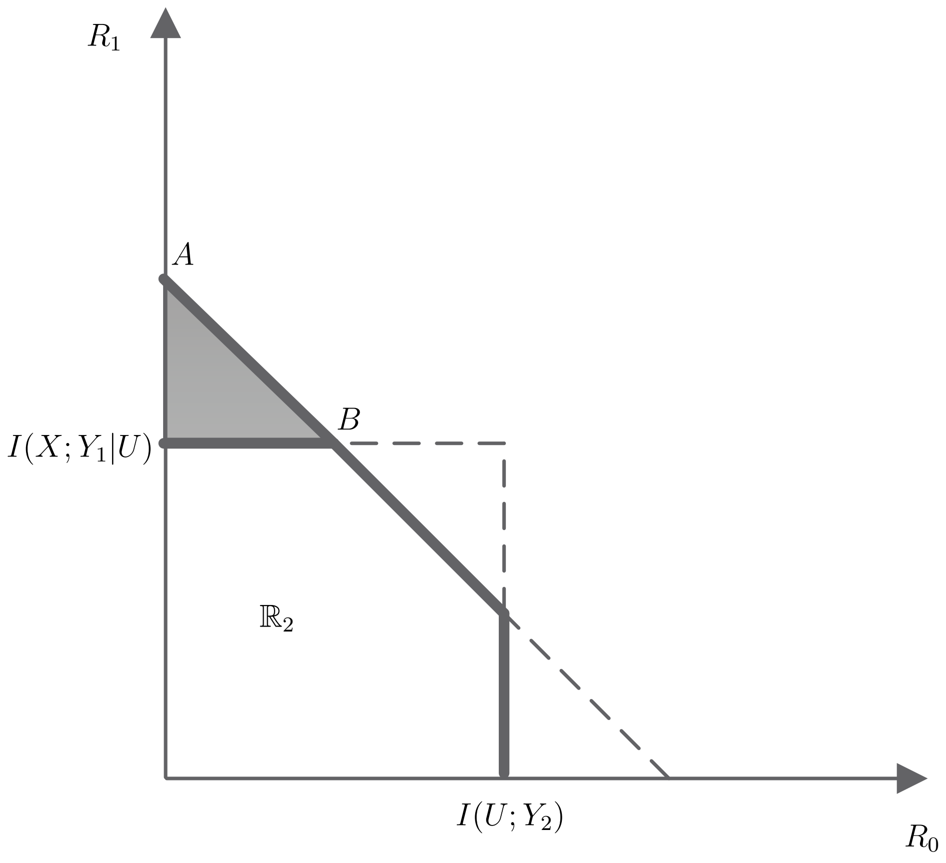
Theorem 1 (Capacity region for the degraded message set BC) The capacity region for the degraded message set BC is :

$$\mathbb{R}_1 = \bigcup_{p(u,x)p(y_1,y_2|x)} \begin{pmatrix} R_0 \leq I(U;Y_2) \\ R_0 + R_1 \leq I(X;Y_1|U) + I(U;Y_2) \\ R_0 + R_1 \leq I(X;Y_1) \end{pmatrix}. \tag{5}$$

Lemma 1 (Equivalent achievabile rate region for degraded message set BC) For the degraded message set BC the region $\mathbb{R}_1$ is an achievable region iff $\mathbb{R}_2$ is achievable region.

$$\mathbb{R}_2 = \bigcup_{p(u,x)p(y_1,y_2|x)} \begin{pmatrix} R_0 \leq I(U;Y_2) \\ R_1 \leq I(X;Y_1|U) \\ R_0 + R_1 \leq I(X;Y_1) \end{pmatrix}. \tag{6}$$

Proof: (Equivalent achievable region) The first side of the proof is obvious since $\mathbb{R}_2 \subseteq \mathbb{R}_1$. In order to prove the second direction of proof we can look at graphical description of the regions in fig. I.



The gray triangle is the gap between the regions i.e, $\mathbb{R}_1 \backslash \mathbb{R}_2$. The point A is achievable if the point B is achievable since we can use time sharing to transfer rate from R_0 to R_1. The convexity of $\mathbb{R}_2$ will be proven in appendix 1. ■

Proof: **(Achievability for theorem 1)** We will prove the achievability for the region $\mathbb{R}_1$ and conclude that $\mathbb{R}_2$ is also achievable region by Lemma 1. Fix a joint distribution $P_{X,U,Y_1,Y_2} = P_{X,U} P_{Y_1,Y_2|X}$, where $P_{Y_1,Y_2|X}$ is given by the channel.

Codebook generation:

1) Generate 2^{nR_0} independent codewords of length n, $U(i)$,$i \in \{1, 2, \ldots, 2^{nR_0}\}$ according to P_U.
2) For each codeword U^n, generate 2^{nR_1} independent codewords, $X^n(s), s \in \{1, 2, \ldots, 2^{nR_1}\}$ according to $P_{x|u}$.

Encoding: The encoder sends sequence $X^n(u^n(m_0), m_1)$.

Decoding:

1) Decoder 1 looks for $(\hat{m}_0, \hat{m}_1)$ such that $(y_1^n, u^n(\hat{m}_0), x^n(u^n(\hat{m}_0), \hat{m}_1)) \in A_\epsilon^*(X, U, Y_1)$.
2) Decoder 2 looks for $\hat{m}_0$ such that $(y_2^n, u^n(\hat{m}_0)) \in A_\epsilon^*(U, Y_2)$.

Error analysis: Let us assume W.L.O.G that the messages $(m_0, m_1) = (1, 1)$ were sent. The error events are as follows:

1) Define E_1 for the event that the sequences were generated are not in the typical set, i.e $(y_1^n, y_2^n, x^n(u^n(1), 1)) \notin A_\epsilon^*(X, U, Y_1, Y_2)$. By using law of large numbers and the fact that all sequences were generated i.i.d the probability of this event is small if n is large enough.
2) For the second decoder let's define the event $E_2 : \exists j \neq 1 s.t (y_2^n, u^n(j)) \in A_\epsilon^*(U, Y_2)$. We know by covering lemma that $P(E_2) \to 0$ for $n \to \infty$ if $R_0 \leq I(U; Y_2) - \epsilon$.
3) For the first decoder there are 3 different events for decoding the wrong pair of messages, for convenience we show each event and its constraint according to the covering lemma in a table.

Event	$\hat{m}_0$	$\hat{m}_1$	Constraint
E_{31}	1	$\hat{m}_1 \neq 1$	$R_1 \leq I(X; Y_1 \vert U) - \epsilon$
E_{32}	$\hat{m}_0 \neq 1$	1	$R_0 \leq I(X; Y_1) - \epsilon$
E_{33}	$\hat{m}_0 \neq 1$	$\hat{m}_1 \neq 1$	$R_0 + R_1 \leq I(U, X; Y_1) - \epsilon$

One can see in the third constraint that $I(U, X; Y_1) = I(X; Y_1)$ since $Y_1 - X - U$ is a markov chain. Therefore, we can remove the second constraint since its region contained in the third constraint region. By using the union bound we can see that:

$$P_e^{(n)} \leq \Pr(E_1) + \Pr(E_2) + \Pr(E_{31}) + \Pr(E_{32}) + \Pr(E_{33}) \to 0 \tag{7}$$

if

$$R_0 \leq I(U;Y_2) \quad (8)$$

$$R_1 \leq I(X;Y_1|U) \quad (9)$$

$$R_0 + R_1 \leq I(U,X;Y_1) = I(X;Y_1). \quad (10)$$

■

(Proof of the converse for Theorem 1) We prove the converse for region $\mathbb{R}_1$.

$$\begin{aligned}
nR_0 &\overset{(a)}{=} H(M_0) && (11)\\
&\overset{(b)}{\leq} I(Y_2^n;M_0) + n\epsilon_n && (12)\\
&= \sum_{i=1}^{n} I(Y_{2i};M_0|Y_2^{i-1}) + n\epsilon_n && (13)\\
&\overset{(c)}{=} \sum_{i=1}^{n} I(Y_{2i};M_0,Y_2^{i-1}) + n\epsilon_n && (14)\\
&\leq \sum_{i=1}^{n} I(Y_{2i};M_0,Y_2^{i-1},Y_{i+1}^n) + n\epsilon_n && (15)\\
&\overset{(d)}{=} \sum_{i=1}^{n} I(Y_{2i};U_i) + n\epsilon_n && (16)
\end{aligned}$$

Where (a) follows from the uniform distribution of $M_0 \in [1,...,2^{nR_0}]$ as defined in Def.(3).

Where (b) follows from Fano's inequality where $H(M_0|Y_2^n) \leq n\epsilon_n$.

Where (c) follows from independence of $\{Y_{2i}\}_{i=1}^n$ over i which applies $H(Y_{2i}|M_o,Y_2^{i-1}) = H(Y_{2i}|M_o)$.

Where (d) follows from substituting $U_i = (M_0,Y_2^{i-1},Y_{i+1}^n)$.

$$\begin{aligned}
n(R_0+R_1) &\overset{(a)}{=} H(M_0,M_1) && (17)\\
&\overset{(b)}{\leq} I(M_0,M_1;Y_1^n) + n\epsilon_n && (18)\\
&\overset{(c)}{=} I(M_0,M_1,X^n;Y_1^n) + n\epsilon_n && (19)\\
&\overset{(d)}{=} I(Y_1^n;X^n) + n\epsilon_n && (20)
\end{aligned}$$

$$\leq \sum_{i=1}^{n} H(Y_{1i}|Y_1^{i-1}) - H(Y_{1i}|X^n, Y_1^{i-1}) + n\epsilon_n \tag{21}$$

$$\overset{(e)}{\leq} \sum_{i=1}^{n} H(Y_{1i}) - H(Y_{1i}|X_i) + n\epsilon_n \tag{22}$$

$$= \sum_{i=1}^{n} I(Y_{1i}; X_i) + n\epsilon_n \tag{23}$$

Where (a) follows from the uniform distribution of $M_0 \in [1, ..., 2^{nR_0}]$ and $M_1 \in [1, ..., 2^{nR_1}]$ as defined in Def.(3).

Where (b) follows from Fano's inequality where $H(M_0, M_1|Y_1^n) \leq n\epsilon_n$.

Where (c) follows the fact that X^n is a function of (M_0, M_1).

Where (d) follows from Markov chain $(M_0, M_1) - X^n - Y_1^n$.

Where (e) follows from the given memoryless channel as defined in Def.(1).

$$n(R_0 + R_1) \overset{(a)}{=} H(M_0, M_1) \tag{24}$$

$$= H(M_1|M_0) + H(M_0) \tag{25}$$

$$\overset{(b)}{\leq} I(M_1; Y_1^n|M_0) + n\hat{\epsilon}_n + I(Y_2^n; M_0) + n\tilde{\epsilon} \tag{26}$$

$$\overset{(c)}{\leq} \sum_{i=1}^{n} I(M_1; Y_{1i}|M_0, Y_{1_{i+1}}^n) + I(Y_{2i}; M_0|Y_2^{i-1}) + 2n\epsilon_n \tag{27}$$

$$\overset{(d)}{\leq} \sum_{i=1}^{n} I(M_1, Y_2^{i-1}; Y_{1i}|M_0, Y_{1_{i+1}}^n) + I(Y_{2i}; M_0, Y_2^{i-1}) + 2n\epsilon_n \tag{28}$$

$$\overset{(e)}{=} \sum_{i=1}^{n} I(Y_2^{i-1}; Y_{1i}|M_0, Y_{1_{i+1}}^n) + I(M_1; Y_{1i}|M_0, Y_{1_{i+1}}^n, Y_2^{i-1}) + I(Y_{2i}; M_0, Y_2^{i-1}) + 2n\epsilon_n$$

$$\overset{(f)}{=} \sum_{i=1}^{n} I(Y_2^{i-1}; Y_{1i}|M_0, Y_{1_{i+1}}^n) + I(X_1; Y_{1i}|M_0, Y_{1_{i+1}}^n, Y_2^{i-1}) + I(Y_{2i}; M_0, Y_2^{i-1}) + 2n\epsilon_n \tag{29}$$

$$\overset{(g)}{=} \sum_{i=1}^{n} I(Y_{1_{i+1}}^n; Y_{2i}|M_0, Y_2^{i-1}) + I(X_i; Y_{1i}|U_i) + I(Y_{2i}; M_0, Y_2^{i-1}) + 2n\epsilon_n \tag{30}$$

$$\overset{(h)}{=} \sum_{i=1}^{n} I(M_0, Y_2^{i-1}, Y_{1_{i+1}}^n; Y_{2i}) + I(X_i; Y_{1i}|U_i) + 2n\epsilon_n \tag{31}$$

$$\overset{(i)}{=} \sum_{i=1}^{n} I(U_i; Y_{2i}) + I(X_i; Y_{1i}|U_i) + 2n\epsilon_n \tag{32}$$

Where (a) follows from the uniform distribution of $M_0 \in [1, ..., 2^{nR_0}]$ and $M_1 \in [1, ..., 2^{nR_1}]$ as defined in Def.(3).

Where (b) follows from Fano's inequality where $H(M_1|Y_1^n, M_0) \leq H(M_0, M_1|Y_1^n) \leq n\hat{\epsilon}_n$ and $H(M_0|Y_2^n) \leq n\tilde{\epsilon}_n$.

Where (c) follows from the chain rule and substituting $n\epsilon_n = max\{\hat{\epsilon}_n, \tilde{\epsilon}_n\}$.

Where (d) follows from independence of $\{Y_{2i}\}_{i=1}^n$ over i and the fact that mutual information has a non-negative value.

Where (e) follows from the chain rule of mutual information.

Where (f) follows from the fact that X_i is a deterministic function of (M_0, M_1).

Where (g) follows from Csiszar sum identity and substituting $U_i = (M_0, Y_2^{i-1}, Y_{i+1}^n)$.

Where (h) follows from the chain rule of mutual information.

Where (i) follows from substituting $U_i = (M_0, Y_2^{i-1}, Y_{i+1}^n)$.

II. General upper bound for the Broadcast Channel

Theorem 2 (General upper bound for the Broadcast Channel) We want to proceed and develop an upper bound for the general broadcast channel. The upper bound for the general BC is as follows:

$$
\begin{aligned}
R_1 &\leq I(U_1; Y_1) && (33)\\
R_2 &\leq I(U_2; Y_2) && (34)\\
R_1 + R_2 &\leq I(U_2; Y_2) + I(X; Y_1|U_2) && (35)\\
R_1 + R_2 &\leq I(U_1; Y_1) + I(X; Y_2|U_1) && (36)
\end{aligned}
$$

Proof: In this proof we will use the converse we proved for the message set degraded BC. First, we define 2 auxiliary r.v's

$$
\begin{aligned}
U_{1i} &\triangleq (M_1, Y_2^{i-1}, Y_{i+1}^n) && (37)\\
U_{2i} &\triangleq (M_2, Y_2^{i-1}, Y_{i+1}^n) && (38)
\end{aligned}
$$

Inequality (33) can be derived by replacing (M_0, U_i) with (M_1, U_{1i}) in the converse of the message set degraded in (11)-(16). Inequality (34) can be derived by replacing (M_0, U_i)

with (M_2, U_{2i}) in the converse of the message set degraded in (11)-(16). Inequality (35) can be derived by replacing (M_0, U_i) with (M_2, U_{2i}) in the converse of the message set degraded in (24)-(32). The last inequality, i.e (36), is proved by following a similar steps as the converse in (24)-(32). For some convenience we add the proof but with no arguments.

$$
\begin{aligned}
n(R_1+R_2) &\overset{(a)}{=} H(M_1, M_2) && (39)\\
&= H(M_2|M_1) + H(M_1) && (40)\\
&\overset{(b)}{\leq} I(M_2; Y_2^n|M_1) + n\hat{\epsilon}_n + I(M_1; Y_1^n) + n\tilde{\epsilon} && (41)\\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(M_2; Y_{2i}|M_0, Y_2^{i-1}) + I(M_1; Y_{1i}|Y_{1_{i+1}}^n) + 2n\epsilon_n && (42)\\
&\overset{(d)}{\leq} \sum_{i=1}^{n} I(M_2, Y_{1_{i+1}}^n; Y_{2i}|M_1, Y_2^{i-1}) + I(M_1, Y_{1_{i+1}}^n; Y_{1i}) + 2n\epsilon_n && (43)\\
&\overset{(e)}{=} \sum_{i=1}^{n} I(Y_{1_{i+1}}^n; Y_{2i}|M_1, Y_2^{i-1}) + I(M_2; Y_{2i}|M_1, Y_2^{i-1}, Y_{1_{i+1}}^n) + I(M_1, Y_{1_{i+1}}^n; Y_{1i}) + 2n\epsilon_n\\
&\overset{(f)}{=} \sum_{i=1}^{n} I(Y_2^{i-1}; Y_{1i}|M_1, Y_{1_{i+1}}^n) + I(X_i; Y_{2i}|U_{1i}) + I(M_1, Y_{1_{i+1}}^n; Y_{1i}) + 2n\epsilon_n && (44)\\
&\overset{(g)}{=} \sum_{i=1}^{n} I(Y_{1i}; M_1, Y_2^{i-1}, Y_{1_{i+1}}^n) + I(X_i; Y_{1i}|U_{1i}) + 2n\epsilon_n && (45)\\
&\overset{(h)}{=} \sum_{i=1}^{n} I(Y_{1i}; U_{1i}) + I(X_i; Y_{1i}|U_{1i}) + 2n\epsilon_n && (46)
\end{aligned}
$$

■

III. Semi Deterministic Broadcast Channel

The semi-deterministic channel is a broadcast channel where Y_1 is a deterministic function of X, therefore its joint distribution can be written as: $P(x, y_1, y_2) = p(x)p(y_2|x)\mathbb{1}_{Y_1=g(X)}$. This channel is an example for overlapping between the inner and outer bound we have developed. The capacity region will be found using Marton inner bound as proved in Lec .6 and the general upper bound discussed previously.

Definition 6 Define the region,

$$\mathbb{R}_{Semi} = \bigcup_{p(u)p(x|u)\mathbf{1}_{Y_1=g(X)}p(y_2|x)} \begin{pmatrix} R_1 \leq H(Y_1) \\ R_2 \leq I(U;Y_2) \\ R_1 + R_2 \leq H(Y_1|U) + I(U;Y_2) \end{pmatrix}. \tag{47}$$

Theorem 3 The capacity region for the semi-deterministic BC channel is $\mathbb{R}_{Semi}$.

Proof: First we show that this region is achievable, substituting $U_1 = Y_1$ and $U_2 = U$ into Marton's inner bound yields the following region:

$$R_1 \; = \; \leq I(Y_1;Y_1) = H(Y_1) \tag{48}$$

$$R_2 \; \leq \; I(U;Y_2) \tag{49}$$

$$R_1 + R_2 \; \leq \; H(Y_1) + I(U;Y_2) - I(U;Y_1) = H(Y_1|U) + I(U;Y_2). \tag{50}$$

Now we need to prove that the outer bound overlaps with the inner bound,

$$R_1 \; \leq \; I(U_1;Y_1) \leq H(Y_1) \tag{51}$$

$$R_2 \; \leq \; I(U_2;Y_2) \tag{52}$$

$$R_1 + R_2 \; \leq \; I(U_2;Y_2) + I(X;Y_1|U_2) = I(U_2;Y_2) + H(Y_1|U_2) \tag{53}$$

Where the last inequality holds since Y_1 is a function of X, therefore $H(Y_1|X, U_2) = 0$.

■

For further reading attached are references for the original papers, the general broadcast channel was first introduced in 1972 by cover in [5]. Korner and Marton found the capacity region of the degraded message set BC in [2]. The general upper bound which we showed is a direct consequence from [3] by El Gamal. The capacity region of the semi deterministic BC was solved by Marton in [4].

APPENDIX A
CONVEXITY OF THE REGION $\mathbb{R}_2$

In appendix A we complete the proof of Lemma.1 by showing the region convexity of $\mathbb{R}_2$. The proof begins with introducing a new region and show its convexity and then we proceed to show that that the new region equals to the required region $\mathbb{R}_2$. Let us define the region:

$$\mathbb{C}' = \bigcup_{p(q)p(u,x|q)p(y_1,y_2|x)} \begin{pmatrix} R_0 \leq I(U;Y_2|Q) \\ R_1 \leq I(X;Y_1|U,Q) \\ R_0 + R_1 \leq I(X;Y_1|Q) \end{pmatrix}. \tag{54}$$

We define two rate pairs $(R_0^A, R_1^A), (R_0^B, R_1^B) \in C'$ which corresponds to $p_A(u,x)$ and $p_B(u,x)$ respectively. The region is convex if a convex combination ,i.e

$$(R_0^C, R_1^C) = \alpha(R_0^A, R_1^A) + \bar{\alpha}(R_0^b, R_1^B) \ \ \forall \alpha \in [0,1] \tag{55}$$

holds

$$(R_0^C, R_1^C) \in \mathbb{C}'. \tag{56}$$

Let us substitute into the the rate pairs into the region inequalities so we get:

$$R_0^A \leq I(U;Y_2|Q=A) \qquad R_0^B \leq I(U;Y_2|Q=B) \tag{57}$$

$$R_1^A \leq I(X;Y_1|U,Q=A) \qquad R_1^B \leq I(X;Y_1|U,Q=B) \tag{58}$$

$$R_0^A + R_1^A \leq I(X;Y_1|Q=A) \qquad R_0^B + R_1^B \leq I(X;Y_1|Q=B) \tag{59}$$

One can see that the terms of region $\mathbb{C}'$ can be written as:

$$I(U;Y_2|Q) = P(Q=A)I(U;Y_2|Q=A) + P(Q=B)I(U;Y_2|Q=B) \tag{60}$$

$$I(X;Y_1|U,Q) = P(Q=A)I(X;Y_1|U,Q=A) + P(Q=B)I(X;Y_1|U,Q=B) \tag{61}$$

$$I(X;Y_1|Q) = P(Q=A)I(X;Y_1|Q=A) + P(Q=B)I(X;Y_1|Q=B) \tag{62}$$

By choosing distribution of Q with $Q = \{A,B\}, P(Q=A) = \alpha$ and $P(Q=B) = \bar{\alpha}$ and using the inequalities we developed we get:

$$\begin{aligned} I(U;Y_2|Q) &= \alpha I(U;Y_2|Q=A) + \bar{\alpha} I(U;Y_2|Q=B) & (63) \\ &\leq \alpha R_0^A + \bar{\alpha} R_0^B & (64) \\ &= R_0^C. & (65) \end{aligned}$$

$$I(X;Y_1|U,Q) = \alpha I(X;Y_1|U,Q=A) + \bar{\alpha} I(X;Y_1|U,Q=B) \tag{66}$$

$$
\begin{aligned}
&\leq \alpha R_1^A + \bar{\alpha} R_1^B && (67)\\
&= R_1^C. && (68)
\end{aligned}
$$

$$
\begin{aligned}
I(X;Y_1|Q) &= \alpha I(X;Y_1|Q=A) + \bar{\alpha} I(X;Y_1|Q=B) && (69)\\
&\leq \alpha(R_0^A + R_1^A) + \bar{\alpha}(R_0^B + R_1^B) && (70)\\
&= (R_0^C + R_1^C). && (71)
\end{aligned}
$$

Thus we showed that $(R_0^C, R_1^C) \in \mathbb{C}'$ and the region is convex. In order to complete the proof that $\mathbb{R}_2$ is convex we show that $\mathbb{R}_2 = \mathbb{C}'$. First we notice that $\mathbb{C}' \supseteq \mathbb{R}_2$ by substituting $Q = \emptyset$. Then we show that $\mathbb{R}_2 \supseteq \mathbb{C}'$, consider

$$
\begin{aligned}
I(U;Y_2|Q) &\leq I(U,Q;Y_2) \overset{(a)}{=} I(\tilde{U};Y_2) && (72)\\
I(X;Y_1|U,Q) &\overset{(a)}{=} I(X;Y_1|\tilde{U}) && (73)\\
I(X;Y_1|Q) &\overset{(b)}{=} H(Y_1|Q) - H(Y_1|X) \leq H(Y_1) - H(Y_1|X) = I(X;Y_1) && (74)
\end{aligned}
$$

(a) follows from choosing new auxiliary r.v $\tilde{U} = (U,Q)$.

(b) follows from the markov chain $Q - X - Y_1$.

APPENDIX B
ALTERNATIVE CONVERSE FOR GELFAND PINSKER MODEL

Appendix B provides a proof for the ”Telescoping Identity” which was introduced in [1] and its use for an alternative proof of the converse for Gelfand Pinsker model. Then we proceed and prove an extended version of Telescoping Identity which is used to prove Csiszar sum identity.

Lemma 2 (Telescoping Identity) Let A^n and B^n be any sequences of random variables, then

$$\sum_{i=1}^{n} I(A^i;B_{i+1}^n) - I(A^{i-1};B_i^n) = 0 \tag{75}$$

Proof: We expand the series as follows:

$$\sum_{i=1}^{n} I(A^i;B_{i+1}^n) - I(A^{i-1};B_i^n) \quad = \tag{76}$$

$$
\begin{aligned}
i=1 \quad I(A^1;B_2^n)-I(A^0;B_1^n)+ &\ \cdots && (77)\\
i=2 \quad I(A^2;B_3^n)-I(A^1;B_2^n)+ &\ \cdots && (78)\\
&\ \vdots \\
i=n \quad I(A^n;B_{n+1}^n)-I(A^{n-1};B_n^n) &= && (79)\\
-I(A^0;B_1^n)+I(A^n;B_{n+1}^n) &= 0. && (80)
\end{aligned}
$$

■

Alternative proof of converse for Gelfand-Pinsker model:

Fix a code $(n,2^{nR})$ with an average probability of error $P_\epsilon^{(n)}$. Now consider,

$$
\begin{aligned}
nR &= H(M) && (81)\\
&\overset{(a)}{\leq} I(M;Y^n)+n\epsilon_n && (82)\\
&\overset{(b)}{=} I(M;Y^n)-I(M;S^n)+n\epsilon_n && (83)\\
&\overset{(c)}{=} \sum_{i=1}^{n} I(M,S_{i+1}^n;Y^i)-I(M,S_i^n;Y^{i-1})+n\epsilon_n && (84)\\
&\overset{(d)}{=} \sum_{i=1}^{n} I(M,S_{i+1}^n;Y^{i-1})+I(M,S_{i+1}^n;Y^i|Y^{i-1})-I(M,S_{i+1}^n;Y^{i-1})-I(S_i;Y^{i-1}|S_{i+1}^n,M)+n\epsilon_n \\
&= \sum_{i=1}^{n} I(M,S_{i+1}^n;Y_i|Y^{i-1})-I(S_i;Y^{i-1}|S_{i+1}^n,M)+n\epsilon_n && (85)\\
&= \sum_{i=1}^{n} H(Y_i|Y^{i-1})-H(Y^i|Y^{i-1},M,S_{i+1}^n)-(H(S_i|S_{i+1}^n,M)+H(S_i|Y^{i-1}S_{i+1}^n,M))+n\epsilon_n && (86)\\
&\overset{(e)}{\leq} \sum_{i=1}^{n} H(Y_i)-H(Y^i|Y^{i-1},M,S_{i+1}^n)-(H(S_i)+H(S_i|Y^{i-1}S_{i+1}^n,M))+n\epsilon_n && (87)\\
&= \sum_{i=1}^{n} I(Y_i;Y^{i-1},M,S_{i+1}^n)-I(S_i;Y^{i-1}S_{i+1}^n,M)+n\epsilon_n && (88)\\
&\overset{(f)}{=} \sum_{i=1}^{n} I(Y_i;U_i)-I(S_i;U_i)+n\epsilon_n && (89)
\end{aligned}
$$

Where (a) follows from Fano's inequality.

Where (b) follows from independence of side information S^n and message M. item[]

Where (c) follows from substituting $A^n=Y^n$ and $B^n=(M,S^n)$ in lemma.2, one can see that the only arguments are not equal to zero correspond to indices $i = 1, n$.

Where (d) follows from chain rule of mutual information.

Where (e) follows from conditioning reduces entropy.

Where (f) follows from substituting $U_i = (Y^{i-1}, M, S_{i+1}^n)$.

■

Lemma 3 Let A^n, B^n, C^n and D^n be any sequences of random variables, then

$$\sum_{i=1}^{n} I(A^i; B_{i+1}^n | C_{i+1}^n, D^{i-1}) - I(A^{i-1}; B_i^n | C_{i+1}^n, D^{i-1}) = 0 \tag{90}$$

We don't provide an explicit proof since it follows from expansion of the series again.

Alternative proof for Csiszar sum identity:

The proof follows from the chain rule of mutual information and substituting the results into lemma.2.

$$\begin{aligned} I(A^i; B_{i+1}^n) &= I(A^{i-1}; B_{i+1}^n) + I(A_i; B_{i+1}^n | A^{i-1}) & (91) \\ I(A^{i-1}; B_i^n) &= I(A^{i-1}; B_{i+1}^n) + I(B_i; A_{i-1}^n | B_{i+1}^n) & (92) \end{aligned}$$

■

REFERENCES

[1] G. Kramer, Teaching IT: an identity for the Gelfand-Pinsker converse, IEEE Inform. Theory Society Newsletter, vol. 61, no. 4, pp. 46, 2012.

[2] J. Korner and K. Marton. ”General broadcast channels with degraded message sets”. IEEE Trans. IT, 23(1):6064, January 1977.

[3] A. A. El Gamal, The capacity of a class of broadcast channels, IEEE Trans. Inform. Theory, vol. 25, no. 2, pp. 166169, Mar. 1979.

[4] K. Marton, The capacity region of deterministic broadcast channels,Trans. Int. Symp. Inform. Theory, 1977.

[5] T. M. Cover, Broadcast channels, IEEE Trans. Inform. Theory, vol. IT-18, pp. 214, Jan. 1972.