Solutions to Homework Set #2
Broadcast channel, degraded message set, Csiszar Sum Equality

1. **Convexity of capacity region of broadcast channel.** Let $\mathbf{C} \subseteq \mathbf{R}^2$ be the capacity region of all achievable rate pairs $\mathbf{R} = (R_1, R_2)$ for the broadcast channel. Show that $\mathbf{C}$ is a convex set by using a timesharing argument.

   Specifically, show that if $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$ are achievable, then $\lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$ is achievable for $0 \leq \lambda \leq 1$.

   Solution to Convexity of Capacity Regions.

   Let $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$ be two achievable rate pairs. Then there exist a sequence of $((2^{nR_1^{(1)}}, 2^{nR_2^{(1)}}), n)$ codes and a sequence of $((2^{nR_1^{(2)}}, 2^{nR_2^{(2)}}), n)$ codes for the channel with $P_e^{(n)}(1) \to 0$ and $P_e^{(n)}(2) \to 0$. We will now construct a code of rate $\lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$.

   For a code length n, use the concatenation of the codebook of length λn and rate $\mathbf{R}^{(1)}$ and the code of length $(1-\lambda)n$ and rate $\mathbf{R}^{(2)}$. The new codebook consists of all pairs of codewords and hence the number of X_1 codewords is $2^{\lambda n R_1^{(1)}} 2^{(1-\lambda)nR_1^{(2)}}$, and hence the rate is $\lambda R_1^{(1)} + (1-\lambda)R_1^{(2)}$. Similarly the rate of the X_2 codeword is $\lambda R_2^{(1)} + (1-\lambda)R_2^{(2)}$.

   We will now show that the probability of error for this sequence of codes goes to zero. The decoding rule for the concatenated code is just the combination of the decoding rule for the parts of the code. Hence the probability of error for the combined codeword is less than the sum of the probabilities for each part. For the combined code,

   $$P_e^{(n)} \leq P_e^{(\lambda n)}(1) + P_e^{((1-\lambda)n)}(2) \tag{1}$$

   which goes to 0 as $n \to \infty$. Hence the overall probability of error goes to 0, which implies the $\lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$ is achievable.

2. **Joint typicality** Let x^n, y^n be jointly strong-typical i.e., $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$, and let Z^n be distributed according to $\prod_{i=1}^n p_{Z|X}(z_i|x_i)$ (instead of $p_{Z|X,Y}(z_i|x_i,y_i)$). Then, $P\{(x^n, y^n, Z^n) \in T_\epsilon^{(n)}(X,Y,Z)\} \le 2^{-n(I(Y;Z|X)-\delta(\epsilon))}$, where $\delta(\epsilon) \to 0$ when $\epsilon \to 0$.

   Solution to Joint typicality.

   Since $(x^n, x^n) \in T_\epsilon^{(n)}(X,Y)$ and $Z^n \sim \prod_{i=1}^n p_{Z|X}(z_i|x_i)$ then if $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$ results in $p(x^n, y^n, z^n) = p\{(x^n)p(z^n|x^n)p(y^n|x^n)$.

$$
\begin{aligned}
p[(x^n, y^n, z^n) &\in T_\epsilon^{(n)}(X,Y,Z)] && (2)\\
&= \sum_{(x^n,y^n,z^n)\in T_\epsilon^{(n)}(X,Y,Z)} p(x^n, y^n, z^n) && (3)\\
&= \sum_{(x^n,y^n,z^n)\in T_\epsilon^{(n)}(X,Y,Z)} p(x^n)p(z^n|x^n)p(y^n|x^n) && (4)\\
&\le \sum_{(x^n,y^n,z^n)\in T_\epsilon^{(n)}(X,Y,Z)} 2^{-n(H(X)-\epsilon)}2^{-n(H(Y|X)-\epsilon)}2^{-n(H(Z|X)-\epsilon)} && (5)\\
&\le 2^{n(H(X,Y,Z)+\epsilon)}2^{-n(H(X)+H(Z|X)+H(Y|X)-3\epsilon)} && (6)\\
&= 2^{n(H(Y,Z|X)-H(Y|X)-H(Z|X)+4\epsilon)} && (7)\\
&= 2^{n(H(Y,Z|X)-H(Y|X)+4\epsilon)} && (8)\\
&= 2^{-n(I(Y;Z|X)-\delta(\epsilon))} && (9)
\end{aligned}
$$

   Note: $p(z^n|x^n) \le 2^{-n(H(Z|X)-\epsilon)}$. (see Cover, page 522.)

3. **Broadcast capacity depends only on the conditional marginals.** Consider the general broadcast channel $(X, Y_1 \times Y_2, p(y_1, y_2 \mid x))$. Show that the capacity region depends only on $p(y_1 \mid x)$ and $p(y_2 \mid x)$. To do this, for any given $((2^{nR_1}, 2^{nR_2}), n)$ code, let

$$
\begin{aligned}
P_1^{(n)} &= P\{\hat{W}_1(\mathbf{Y}_1) \neq W_1\}, && (10)\\
P_2^{(n)} &= P\{\hat{W}_2(\mathbf{Y}_2) \neq W_2\}, && (11)\\
P^{(n)} &= P\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\}. && (12)
\end{aligned}
$$

   Then show

$$\max\{P_1^{(n)}, P_2^{(n)}\} \le P^{(n)} \le P_1^{(n)} + P_2^{(n)}.$$

The result now follows by a simple argument.

Remark: The probability of error $P^{(n)}$ *does* depend on the conditional joint distribution $p(y_1, y_2 \mid x)$. But whether or not $P^{(n)}$ can be driven to zero (at rates (R_1, R_2)) *does not* (except through the conditional marginals $p(y_1 \mid x), p(y_2 \mid x)$).

Solution to Broadcast channel capacity depends only on conditional marginals

$$P_1^{(n)} = P(\hat{W}_1(\mathbf{Y}_1) \neq W_1) \tag{13}$$

$$P_2^{(n)} = P(\hat{W}_2(\mathbf{Y}_2) \neq W_2) \tag{14}$$

$$P^{(n)} = P((\hat{W}_1(\mathbf{Y}_1), \hat{W}_2(\mathbf{Y}_2)) \neq (W_1, W_2)) \tag{15}$$

Then by the union of events bound, it is obvious that

$$P^{(n)} \leq P_1^{(n)} + P_2^{(n)}. \tag{16}$$

Also since $(\hat{W}_1(\mathbf{Y}_1) \neq W_1)$ or $(\hat{W}_2(\mathbf{Y}_2) \neq W_2)$ implies $((\hat{W}_1(\mathbf{Y}_1), \hat{W}_2(\mathbf{Y}_2)) \neq (W_1, W_2))$, we have

$$P^{(n)} \geq \max\{P_1^{(n)}, P_2^{(n)}\}. \tag{17}$$

Hence $P^{(n)} \to 0$ iff $P_1^{(n)} \to 0$ and $P_2^{(n)} \to 0$.

The probability of error, $P^{(n)}$, for a broadcast channel does depend on the joint conditional distribution. However, the individual probabilities of error $P_1^{(n)}$ and $P_2^{(n)}$ however depend only on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$ respectively. Hence if we have a sequence of codes for a particular broadcast channel with $P^{(n)} \to 0$, so that $P_1^{(n)} \to 0$ and $P_2^{(n)} \to 0$, then using the same codes for another broadcast channel with the same conditional marginals will ensure that $P^{(n)}$ for that channel as well, and the corresponding rate pair is achievable for the second channel. Hence the capacity region for a broadcast channel depends only on the conditional marginals.

4. **Degraded broadcast channel.** Find the capacity region for the degraded broadcast channel in Figure 1.

   Degraded broadcast channel. From the expression for the capacity region, it is clear that the only on trivial possibility for the auxiliary random variable U is that it be binary. From the symmetry of the problem,

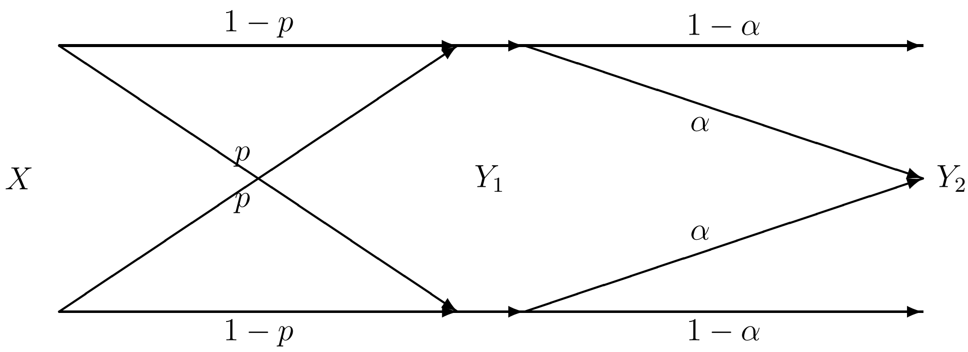

Figure 1: Broadcast channel with a binary symmetric channel and an erasure channel

we see that the auxiliary random variable should be connected to X by a binary symmetric channel with parameter β.

Hence we have the setup as shown in Figure 2.

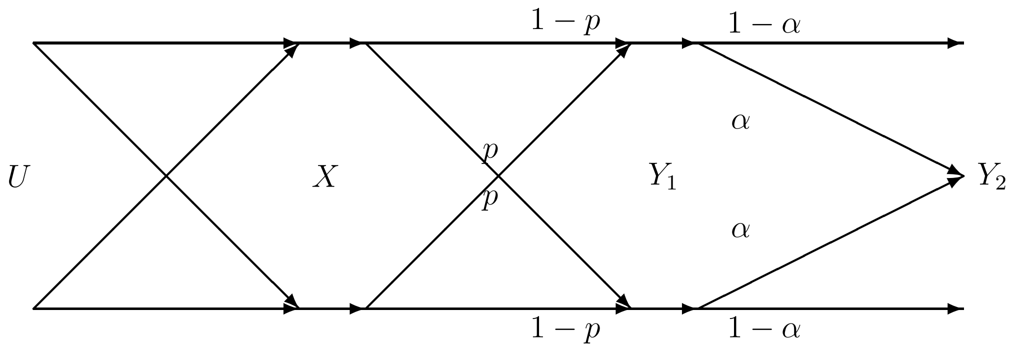

Figure 2: Broadcast channel with auxiliary random variable

We can now evaluate the capacity region for this choice of auxiliary random variable. By symmetry, the best distribution for U is the uniform.

Hence

$$\begin{aligned}
R_2 &= I(U;Y_2) && (18)\\
&= H(Y_2) - H(Y_2|U) && (19)\\
&= H\left(\frac{\overline{\alpha}}{2}, \alpha, \frac{\overline{\alpha}}{2}\right) - H((\overline{\beta}\overline{p} + \beta p)\overline{\alpha}, \alpha, (\overline{\beta}p + \beta\overline{p})\overline{\alpha}) && (20)\\
&= H(\alpha) + \overline{\alpha}H\left(\frac{1}{2}\right) - H(\alpha) - \overline{\alpha}H(\overline{\beta}p + \beta\overline{p}) && (21)\\
&= \overline{\alpha}(1 - H(\overline{\beta}p + \beta\overline{p})). && (22)
\end{aligned}$$

Also

$$\begin{aligned}
R_1 &= I(X;Y_1|U) && (23)\\
&= H(Y_1|U) - H(Y_1|U,X) && (24)\\
&= H(\beta\overline{p} + \overline{\beta}p) - H(p). && (25)
\end{aligned}$$

These two equations characterize the boundary of the capacity region as β varies. When $\beta = 0$, then $R_1 = 0$ and $R_2 = \overline{\alpha}(1 - H(p))$. When $\beta = \frac{1}{2}$, we have $R_1 = 1 - H(p)$ and $R_2 = 0$.

5. **Csiszar Sum Equality**. Let X^n and Y^n be two random vectors with arbitrary joint probability distribution. Show that:

$$\sum_{i=1}^{n} I(X_{i+1}^n; Y_i|Y^{i-1}) = \sum_{i=1}^{n} I(Y^{i-1}; X_i|X_{i+1}^n) \qquad (26)$$

As we shall see this inequality is useful in proving converses to several multiple user channels. (Hint: You can prove this by induction or by expanding the terms on both sides using the chain rule.)

Solution: Csiszar Sum Equality.

$$
\begin{aligned}
\sum_{i=1}^{n} I(X_{i+1}^n; Y_i|Y^{i-1}) &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} I(X_j; Y_i|Y^{i-1}, X_{j+1}^n) \\
&= \sum_{j=2}^{n}\sum_{i=1}^{j-1} I(X_j; Y_i|Y^{i-1}, X_{j+1}^n) \\
&= \sum_{j=2}^{n} I(X_j; Y^{j-1}|X_{j+1}^n) \\
&= \sum_{j=1}^{n} I(X_j; Y^{j-1}|X_{j+1}^n) \\
&= \sum_{i=1}^{n} I(Y^{i-1}; X_i|X_{i+1}^n)
\end{aligned}
$$

where the first and third equalities follow from chain rule, and the second equality follows from switching of the summations.

6. **Broadcast Channel with Degraded Message Sets.** Consider a general DM broadcast channel $(\mathcal{X}; p(y_1, y_2|x); \mathcal{Y}_1; \mathcal{Y}_2)$. The sender X encodes two messages $(W_0; W_1)$ uniformly distributed over $\{1, 2..., 2^{nR_0}\}$ and $\{1, 2, ...2^{nR_1}\}$. Message W_0 is to be sent to both receivers, while message W_1 is only intended for receiver Y_1.

   The capacity region is given by the set $\mathcal{C}$ of all $(R_0; R_1)$ such that:

$$
\begin{aligned}
R_0 &\leq I(U; Y_2) && (27) \\
R_1 &\leq I(X; Y_1|U) && (28) \\
R_0 + R_1 &\leq I(X; Y_1) && (29)
\end{aligned}
$$

   for some $p(u)p(x|u)$.

   (a) Show that the set $\mathcal{C}$ is convex.

   (b) Provide the achievability proof

   (c) Provide a converse proof. You may derive your own converse or use the steps below.

- An alternative characterization of the capacity region is the set $\mathcal{C}'$ of all (R_0, R_1) such that:

$$R_0 \leq \min\{I(U;Y_1); I(U;Y_2)\} \tag{30}$$
$$R_0 + R_1 \leq \min\{I(X;Y_1); I(X;Y_1|U) + I(U;Y_2)\} \tag{31}$$

  for some $p(u)p(x|u)$. Show that $\mathcal{C} = \mathcal{C}'$.

- Define $U_i = (W_0; Y_1^{i-1}, Y_{2,i+1}^n)$. Show that

$$n(R_0 + R_1) \leq \sum_{i=1}^{n} (I(X_i; Y_{1,i}|U_i) + I(U_i; Y_{2,i})) + n\epsilon_n \tag{32}$$

  using the steps

$$\begin{aligned} n(R_0 + R_1) &\leq I(W_1; Y_1^n|W_0) + I(W_0; Y_2^n) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} I(X_i; Y_{1,i}|U_i) + I(Y_{2,i+1}^n; Y_{1,i}|W_0.Y_1^{i-1}) \\ &\quad + I(U_i; Y_{2,i}) - I(Y_1^{i-1}; Y_{2,i}|W_0, Y_{2,(i+1)}^n) + n\epsilon_n. \end{aligned} \tag{33}$$

  Then use the identity from previous exercise to cancel the second and fourth terms.

Solution: Broadcast Channel with Degraded Message Sets.

1. Convex proof.

   To show that the set $\mathcal{C}$ is convex is suffices to show that for any random variable Q s.t. $Q \to (U, X) \to (Y_1, Y_2)$ form a Markov chain, any rate pair (R_0, R_1) s.t.:

   $$R_0 \leq I(U; Y_2|Q), \tag{34}$$

   $$R_1 \leq I(X; Y_1|U, Q), \tag{35}$$

   $$R_0 + R_1 \leq I(X; Y_1|Q), \tag{36}$$

   is in C. To show this, define $U' = (Q, U)$. Then $U' \to X \to (Y_1, Y_2)$ form a Markov chain, and

   $$R_0 \leq I(U; Y_2|Q) \leq I(U'; Y_2), \tag{37}$$

   $$R_1 \leq I(X; Y_1|U, Q) \leq I(X; Y_1; U'), \tag{38}$$

   $$R_0 + R_1 \leq I(X; Y_1|Q) \leq I(X; Y_1), \tag{39}$$

   This completes the proof of convexity

2. The achieveability proof

   The proof of achievability of $\mathcal{C}$ uses superposition coding and is identical to the proof of achievability for degraded broadcast channel.

   First generate 2^{nR_0} codewords $u^n(w_0)$ according to $p(u^n) = \prod_{i=1}^n p(u_1)$. For each codeword $u^n(w_0)$, generate 2^{nR_1} codewords $x^n(w_0, w_1)$ according to $\prod_{i=1}^n p(x_i|u_i(w_0))$. To send the pair (w_0, w_1), send the corresponding codeword $x^n(w_0, w_1)$.

   At the receiver side, receiver 2 decodes W_0 with arbitrarily small probability of error if $R_0 \leq I(U; Y_2)$ and receiver 1 decodes W_0, W_1 with arbitrarily small probability of error if $R_0 + R_1 < I(X;Y_1)$ and $R_1 < I(X;Y_1|U)$. Thus, $P_e^n \to 0$ if

$$R_0 \leq I(U;Y_2), \tag{40}$$

$$R_1 \leq I(X;Y_1|U), \tag{41}$$

$$R_0 + R_1 \leq I(X;Y_1), \tag{42}$$

3. Converse proof.

   Step 1

   The converse in the following section proves that $\mathcal{C} \subseteq \mathcal{C}'$. To show that $\mathcal{C}' \subseteq \mathcal{C}$ it suffices to show that any point on the boundary of $\mathcal{C}'$ is in $\mathcal{C}$. Consider any point on the boundary of $\mathcal{C}'$, i.e., any point s.t.:

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_1); I(U;Y_2)\} && (43)\\ R_0 + R_1 &\leq \min\{I(X;Y_1), I(X;Y_1|U) + I(U;Y_2)\} && (44) \end{aligned}$$

   Clearly $R_0 \leq I(U;Y_2)$ and $R_0 + R_1 \leq I(X;Y_1)$. Now, if $R_0 = I(U;Y_2)$, then $R_1 = \min\left\{I\left(X;Y_1\right) - I\left(U;Y_2\right), I(X;Y_1|U)\right\} \leq I(X;Y_1|U)$. If on the other hand $R_0 = I(U;Y_1)$, then
   $R_1 = \min\left\{I\left(X;Y_1|U\right), I(X;Y_1|U) + I(U;Y_2) - I(U,Y_1)\right\} \leq I(X;Y_1|U)$. Thus, $\mathcal{C}' = \mathcal{C}$.

   Step 1

$$\begin{aligned} nR_0 &= H(W_0) && (45)\\ &\leq I(W_0;Y_1^n) + n\epsilon_n && (46)\\ &= H(Y_1^n) - H(Y_1^n|W_0) + n\epsilon_n && (47)\\ &\leq \sum_{i=1}^{n} \big(H(Y_{1i}) - H(Y_{1i}|W_0, Y_1^{i-1})\big) + n\epsilon_n && (48)\\ &= \sum_{i=1}^{n} \big(H(Y_{1i}) - H(Y_{1i}|W_0, Y_1^{i-1}, Y_{2(i+1)}^n)\big) + n\epsilon_n && (49)\\ &\leq \sum_{i=1}^{n} (H(Y_{1i}) - H(Y_{1i}|U_1)) + n\epsilon_n && (50)\\ &= \sum_{i=1}^{n} I(U_i;Y_{1i}) + n\epsilon && (51) \end{aligned}$$

Similarly, it can be shown that $nR_0 \leq \sum_{i=1}^{n} I(U_i;Y_{2i}) + n\epsilon_n$. Now, consider

$$\begin{aligned}
n(R_0+R_1) &= H(W_0,W_1) && (52)\\
&\leq I(W_0,W_1;Y_1^n) + n\epsilon_n && (53)\\
&\leq I(X^n;Y_1^n) + n\epsilon_n && (54)\\
&= H\left(Y_1^n\right) - H(Y_1^n|X^n) + n\epsilon_n && (55)\\
&\leq \sum_{i=1}^{n} \left(H(Y_{1i}) - H(Y_{1i}|X_i)\right) + n\epsilon_n && (56)\\
&= \sum_{i=1}^{n} I(X_i;Y_{1i}) + n\epsilon_n. && (57)
\end{aligned}$$

Finally, consider

$$
\begin{aligned}
n(R_0 + R_1) &= H(W_0) + H(W_1) && (58)\\
&= H(W_1|W_0) + H(W_0) && (59)\\
&\leq I(W_1; Y_1^n|W_0) + I(W_0; Y_2^n) + 2n\epsilon_n && (60)\\
&= \sum_{i=1}^{n} I(W_1; Y_{1i}|W_0, Y_1^{i-1}) + I(W_0; Y_{2i}|Y_{2(i+1)}^n) + 2n\epsilon_n && (61)\\
&\leq \sum_{i=1}^{n} I(W_1, Y_{2(i+1)}^n; Y_{1i}|W_0, Y_1^{i-1}) + I(W_0, Y_{2(i+1)}^n; Y_{2i}) + 2n\epsilon_n && (62)\\
&= \sum_{i=1}^{n} I(Y_{2(i+1)}^n; Y_{1i}|W_0, Y_1^{i-1}) + I(W_1; Y_{1i}|W_0, Y_1^{i-1}, Y_{2(i+1)}^n) && (63)\\
&+ I(W_0; Y_{2(i+1)}^n; Y_{2i}) && (64)\\
&+ I(Y_1^{i-1}; Y_{2i}|W_0, Y_{2(i+1)}^n) - I(Y_1^{i-1}; Y_{2i}|W_0, Y_{2(i+1)}^n) + 2n\epsilon_n && (65)\\
&= \sum_{i=1}^{n} I(Y_{2(i+1)}^n; Y_{1i}|W_0, Y_1^{i-1}) + I(W_1; Y_{1i}|U_i) && (66)\\
&+ I(W_0, Y_1^{i-1}, Y_{2(i+1)}^n; Y_{2i}) - I(Y_1^{i-1}; Y_{2i}|W_0, Y_{2(i+1)}^n) + 2n\epsilon_n && (67)\\
&= \sum_{i=1}^{n} I(Y_{2(i+1)}^n; Y_{1i}|W_0, Y_1^{i-1}) + I(X_i; Y_{1i}|U_i) && (68)\\
&+ I(U_i; Y_{2i}) - I(Y_i^{i-1}; Y_{2i}|W_o, Y_{2(i+1)}^n) + 2n\epsilon_n && (69)\\
&= \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i) + I(U_i; Y_{2i}) + 2n\epsilon_n && (70)
\end{aligned}
$$

Using the standard 'time-sharing' argument, the converse follows.