EE 376B
Information Theory
Prof. T. Cover

Handout #14
Tuesday, May 5, 2009

Solutions to Homework Set #4

1. **The cooperative capacity of a multiple access channel.**

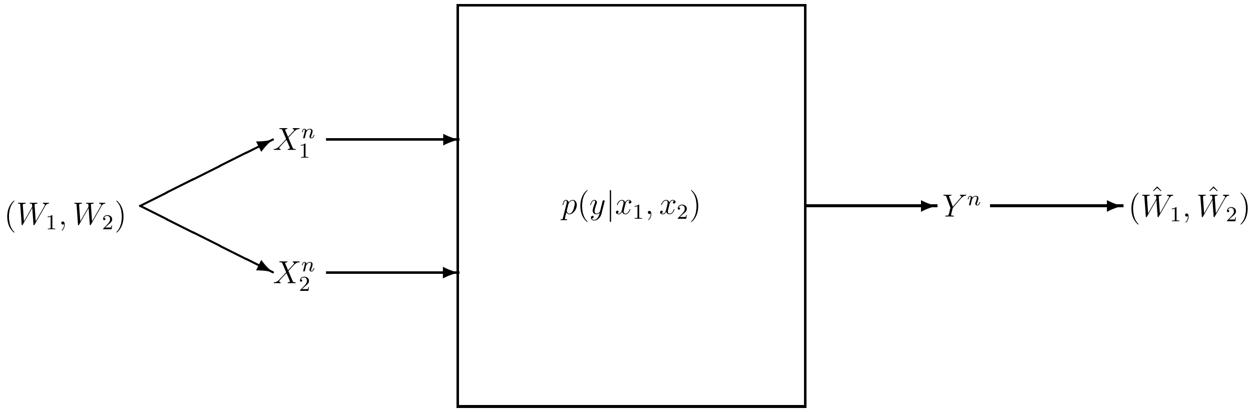


Figure 1: Multiple access channel with cooperating senders.

(a) Suppose X_1 and X_2 have access to *both* indices $W_1 \in \{1, 2^{nR_1}\}, W_2 \in \{1, 2^{nR_2}\}$. Thus the codewords $X_1^n(W_1, W_2), X_2^n(W_1, W_2)$ depend on both indices. Find the capacity region.

(b) Evaluate this region for the binary erasure multiple access channel $Y = X_1 + X_2$, $X_i \in \{0, 1\}$. Compare to the non-cooperative region.

Solution: The cooperative capacity of a multiple access channel

(a) When both senders have access to the pair of messages to be transmitted, they can act in concert. The channel is then equivalent to a single user channel with the input $X = (X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, and the message $W = (W_1, W_2)$. The capacity of this single user channel is $C = \max_{p(x)} I(X; Y) = \max_{p(x_1, x_2)} I(X_1, X_2; Y)$. The two senders can send at any combination of rates with the total rate

$$R_1 + R_2 \leq C.$$

(b) The capacity for the binary erasure multiple access channel was evaluated in class. When the two senders cooperate to send a common message, the capacity is

$$C = \max_{p(x_1,x_2)} I(X_1, X_2; Y) = \max H(Y) = \log 3,$$

achieved by (for example) a uniform distribution on the pairs, (0,0), (0,1) and (1,1). The cooperative and non-cooperative regions are illustrated in Figure 2.

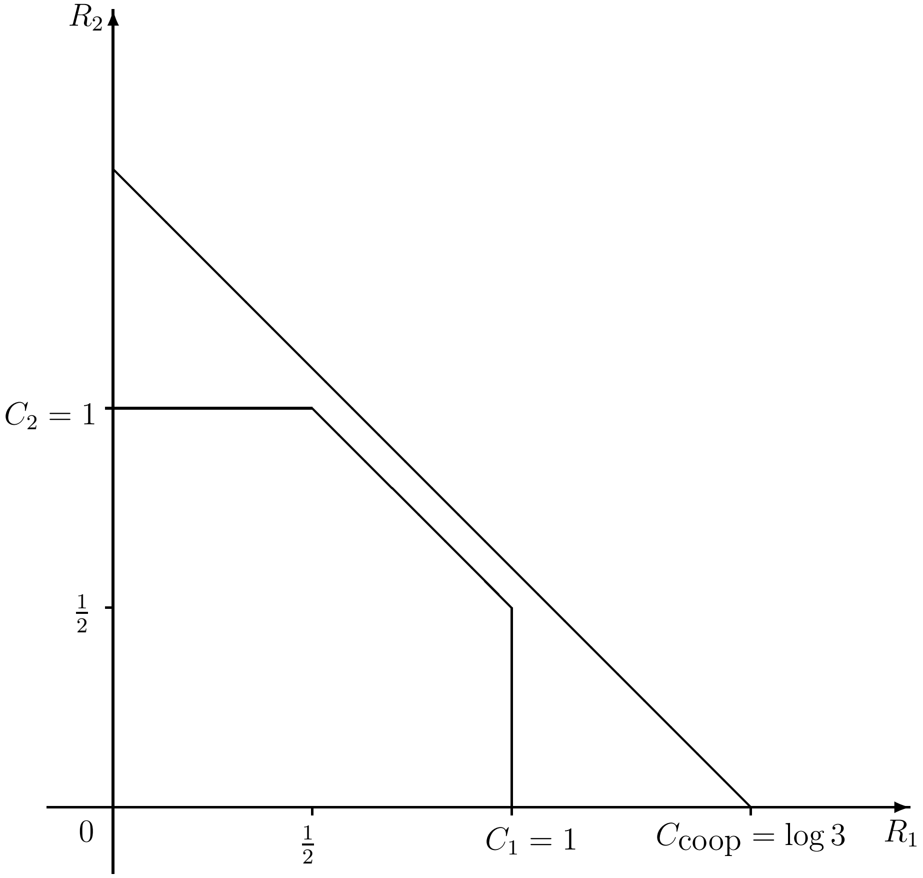


Figure 2: Cooperative and non-cooperative capacity for a binary erasure multiple access channel.

2. **Square channel.**
What is the capacity of the following multiple access channel?

$$\begin{aligned} X_1 &\in \{-1,0,1\} \\ X_2 &\in \{-1,0,1\} \\ Y &= X_1^2 + X_2^2 \end{aligned}$$

(a) Find the capacity region.

(b) Describe $p^*(x_1), p^*(x_2)$ achieving a point on the boundary of the capacity region.

(c) What is the capacity if $Y = X_1 X_2$?

Solution: Square channel.

(a) First of all note that $\{-1, 1\}$ are degenerate from the point of view of the receiver Y. Define $Z_1 = X_1^2$ and $Z_2 = X_2^2$ then $Y = Z_1 + Z_2$ and $Z_i \in \{0, 1\}$. Therefore, we are back to the case of binary erasure multiple access channel described in Example 14.3.3 of the text. The capacity region is as given in Figure 14.13.

(b) Note that a single distribution can achieve the whole region. (The union and convexification operations are not necessary for this specific channel.)

Take $p(x_1) = (\alpha, 1/2, 1/2-\alpha)$ and $p(x_2) = (\beta, 1/2, 1/2-\beta)$ for any $0 \le \alpha, \beta \le 1/2$. Then, we have

$$\begin{aligned} I(X_1; Y|X_2) &= H(Y|X_2) = 1 \\ I(X_2; Y|X_1) &= H(Y|X_1) = 1 \\ I(X_1, X_2; Y) &= H(Y) = 3/2. \end{aligned}$$

(Note: Putting positive probability masses on both $+1$ and -1 does no worse than sticking to either $+1$ or -1.)

(c) We can use the same argument as in Question 2. Since $Y \in \{-1, 0, +1\}$, $R_1 + R_2 \le \log(3)$ and $(R_1 = \log(3), R_2 = 0)$ is acheived by the distribution $p_1^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and p_2^* that puts all its mass on $\{+1\}$ or $\{-1\}$. Therefore, the capacity region is the same as the binary multiplier channel with the bound 1 replaced $\log(3)$.

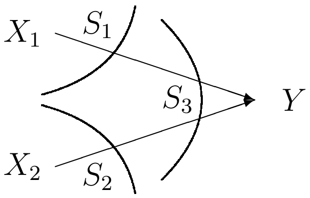
3. **Cut-set interpretation of capacity region of multiple access channel.** For the multiple access channel we know that (R_1, R_2) is achievable if

$$\begin{aligned} R_1 &< I(X_1; Y \mid X_2), && (1)\\ R_2 &< I(X_2; Y \mid X_1), && (2)\\ R_1 + R_2 &< I(X_1, X_2; Y), && (3) \end{aligned}$$

for X_1, X_2 independent. Show, for X_1, X_2 independent, that

$$I(X_1; Y \mid X_2) = I(X_1; Y, X_2).$$

Thus R_1 is less than the mutual information between X_1 and everything else.



Interpret the information bounds as bounds on the rate of flow across cutsets S_1, S_2 and S_3.

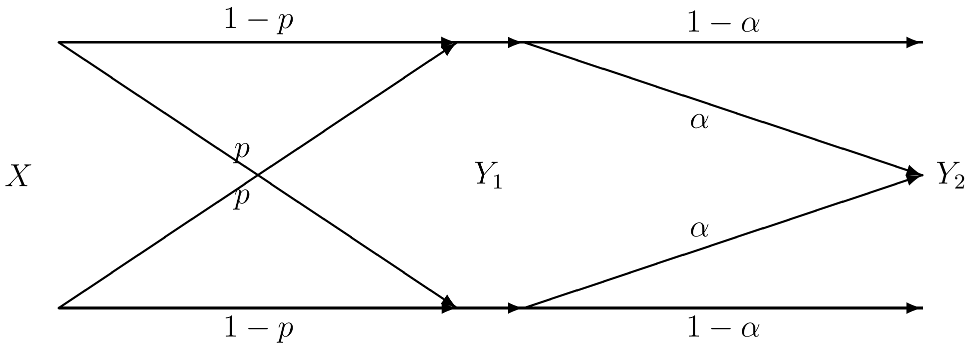
Solution: Cut-set interpretation of capacity region of multiple access channel.

By the chain rule for mutual information and the independence of X_1 and X_2,

$$I(X_1; Y, X_2) = I(X_1; X_2) + I(X_1; Y|X_2) = I(X_1; Y|X_2).$$

We can interpret $I(X_1; Y, X_2)$ as the maximum amount of information that could flow across the cutset S_1. This is an upper bound on the rate R_1. Similarly, we can interpret the other bounds.

4. **Degraded broadcast channel**
Find the capacity region for the degraded broadcast channel below.



Solution: Degraded broadcast channel
From the cardinality of X, Y_1, and Y_2, it suffices to choose U to be binary as well. From the symmetry of the problem, we see that both U and X should be Bern(1/2) and the only parameter we need to control is $\beta = \Pr(U \neq X)$. (The optimality of this form of joint distributions can be rigorously estabilished by the convexity of the capacity region in $p(u, x)$.)

Hence we have the setup as shown in Figure 3.

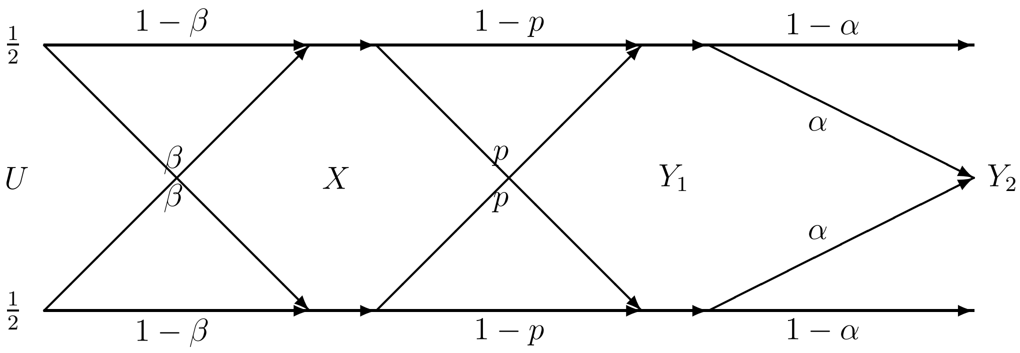


Figure 3: Broadcast channel with auxiliary random variable

We can now evaluate the capacity region for this choice of auxiliary random variable.

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

Also

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

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 $R_2 = 0$. The capacity region is sketched in Figure 4.

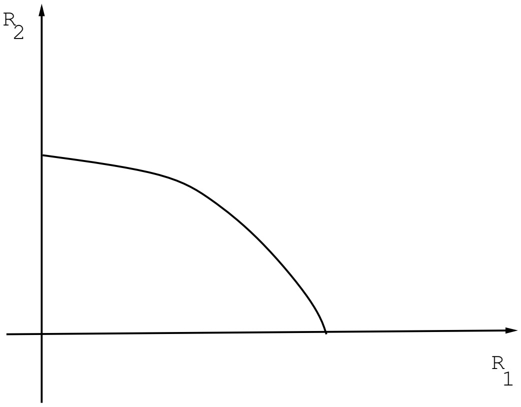


Figure 4: Capacity region of the broadcast channel.

5. **Slepian-Wolf for deterministically related sources**
Find and sketch the Slepian-Wolf rate region for the simultaneous data compression of (X, Y), where $y = f(x)$, and f is a given deterministic function.

Solution: Slepian-Wolf for deterministically related sources

The quantities defining the Slepian Wolf rate region are $H(X,Y) = H(X)$, $H(Y|X) = 0$ and $H(X|Y) \geq 0$. Hence the rate region is as shown in the Figure 5.

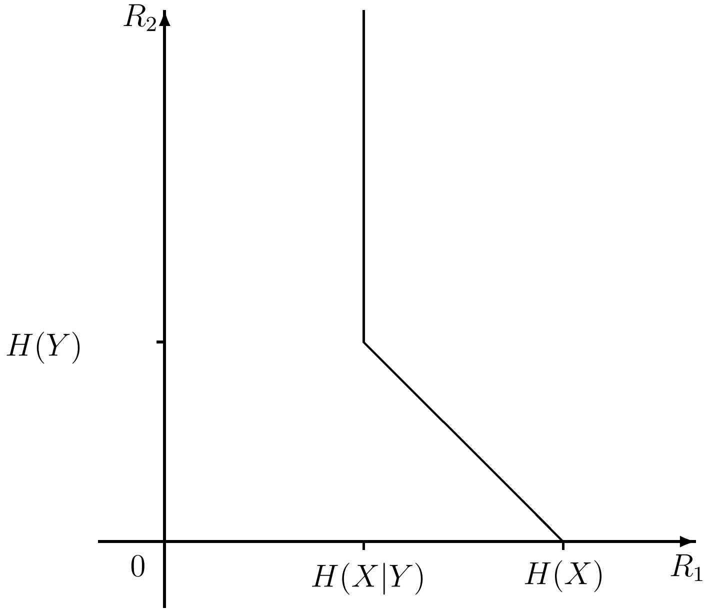


Figure 5: Slepian Wolf rate region for $Y = f(X)$.

6. **Converse for the degraded broadcast channel.**
The following chain of inequalities proves the converse for the degraded discrete memoryless broadcast channel. We are given a degraded broadcast channel $p(x)p(y|x)p(z|y)$. Suppose we have a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with (arithmetic average) probability of error tending to zero. Provide reasons for each of the labeled inequalities.

Setup for converse for degraded broadcast channel capacity:

Let W_1 and W_2 be independent and uniformly drawn over 2^{nR_1} and 2^{nR_2} respectively.

$$(W_1, W_2) \to X^n(W_1, W_2) \to Y^n \to Z^n$$

Encoding $f_n : 2^{nR_1} \times 2^{nR_2} \to X^n$

Decoding: $g_n : Y^n \to 2^{nR_1}, \qquad h_n : Z^n \to 2^{nR_2}$

Let $U_i = (W_2, Y^{i-1})$. Then

$$\begin{aligned}
nR_2 \quad &\overset{\cdot}{\leq}_{\text{Fano}} \quad I(W_2; Z^n) && (4)\\
&\overset{(a)}{=} \quad \sum_{i=1}^{n} I(W_2; Z_i \mid Z^{i-1}) && (5)\\
&\overset{(b)}{=} \quad \sum_{i} (H(Z_i \mid Z^{i-1}) - H(Z_i \mid W_2, Z^{i-1})) && (6)\\
&\overset{(c)}{\leq} \quad \sum_{i} (H(Z_i) - H(Z_i \mid W_2, Z^{i-1}, Y^{i-1})) && (7)\\
&\overset{(d)}{\leq} \quad \sum_{i} (H(Z_i) - H(Z_i \mid W_2, Y^{i-1})) && (8)\\
&\overset{(e)}{=} \quad \sum_{i=1}^{n} I(U_i; Z_i). && (9)
\end{aligned}$$

Continuation of converse. Give reasons for the labeled inequalities:

$$
\begin{aligned}
nR_1 \;&\overset{\cdot}{\leq}_{\text{Fano}}\; I(W_1;Y^n) && (10)\\
&\overset{(f)}{\leq}\; I(W_1;Y^n,W_2) && (11)\\
&\overset{(g)}{\leq}\; I(W_1;Y^n \mid W_2) && (12)\\
&\overset{(h)}{=}\; \sum_{i-1}^{n} I(W_1;Y_i \mid Y^{i-1},W_2) && (13)\\
&\overset{(i)}{\leq}\; \sum_{i=1}^{n} I(X_i;Y_i \mid U_i). && (14)
\end{aligned}
$$

(15)

Solution: Converse for the degraded broadcast channel

$$(W_1,W_2) \rightarrow \mathbf{X}(W_1,W_2) \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}, \qquad (16)$$

where W_1 and W_2 are chosen uniformly over the 2^{nR_1} and 2^{nR_2} respectively.

We also have

$$(W_1,W_2) \rightarrow X_i(W_1,W_2) \rightarrow Y_i \rightarrow Z_i. \qquad (17)$$

Let $U_i = (W_2, Y^{i-1})$.

By Fano's inequality,

$$H(W_2|Z^n) \leq P_2^{(n)} nR_2 + H(P_2^{(n)}) = n\epsilon_n \qquad (18)$$

where $\epsilon_n \rightarrow 0$ as $P_2^{(n)} \rightarrow 0$.

We then have the following chain of inequalities

$$\begin{aligned}
nR_2 &= H(W_2) &(19)\\
&= I(W_2;Z^n)+H(W_2|Z^n) &(20)\\
&\leq I(W_2;Z^n)+n\epsilon_n &(21)\\
&\overset{(a)}{=} \sum_i I(W_2;Z_i|Z^{i-1})+n\epsilon_n &(22)\\
&\overset{(b)}{=} \sum_i (H(Z_i|Z^{i-1})-H(Z_i|Z^{i-1},W_2))+n\epsilon_n &(23)\\
&\overset{(c)}{\leq} \sum_i (H(Z_i)-H(Z_i|Z^{i-1},W_2,Y^{i-1}))+n\epsilon_n &(24)\\
&\overset{(d)}{=} \sum_i (H(Z_i)-H(Z_i|W_2,Y^{i-1}))+n\epsilon_n &(25)\\
&\overset{(e)}{=} \sum_i I(U_i;Z_i)+n\epsilon_n &(26)
\end{aligned}$$

where (21) follows from Fano's inequality,
(a) from the chain rule,
(b) from the definition of conditional mutual information,
(c) from the fact that removing conditioning increases entropy and adding conditioning reduces it,
(d) from the fact that since the broadcast channel is degraded, Z^{i-1} depends only on Y^{i-1} and is conditionally independent of everything else, hence Z_i is conditionally independent of Z^{i-1} given Y^{i-1},
(e) follows from the definition of U_i.

Continuation Of Converse

Similarly by Fano's inequality,

$$H(W_1|Y^n) \leq P_1^{(n)} nR_1 + H(P_1^{(n)}) = n\epsilon_n \tag{27}$$

and we have the chain of inequalities,

$$\begin{aligned}
nR_1 &= H(W_1) && (28)\\
&= I(W_1;Y^n) + H(W_1|Y^n) && (29)\\
&\leq I(W_1;Y^n) + n\epsilon_n && (30)\\
&\overset{(f)}{\leq} I(W_1;W_2,Y^n) + n\epsilon_n && (31)\\
&\overset{(g)}{=} I(W_1;Y^n|W_2) + n\epsilon_n && (32)\\
&\overset{(h)}{\leq} I(W_1;Y_i|W_2,Y^{i-1}) + n\epsilon_n && (33)\\
&\leq I(W_1,X_i;Y_i|W_2,Y^{i-1}) + n\epsilon_n && (34)\\
&\overset{(i)}{\leq} I(X_i;Y_i|W_2,Y^{i-1}) + n\epsilon_n && (35)\\
&= I(X_i;Y_i|U_i) + n\epsilon_n && (36)
\end{aligned}$$

where (30) follows from Fano's inequality,
(f) follows from the fact that the difference, $I(W_1;W_2|Y^n) \geq 0$,
(g) follows from the chain rule for I and the fact that W_1 and W_2 are independent,
(h) from the chain rule for mutual information, and
(i) from the data processing inequality.

We can then use standard techniques like the introduction of a time-sharing random variable to complete the proof of the converse for the broadcast channel.

7. **Multiple access.**

(a) Find the capacity region for the multiple access channel

$$Y = X_1^{X_2}$$

where

$$X_1 \epsilon \{2, 4\} \ , \quad X_2 \epsilon \{1, 2\} \ .$$

(b) Suppose the range of X_1 is $\{1, 2\}$. Is the capacity region decreased?

Solution: Multiple access

(a) With $X_1 \in \{2, 4\}, X_2 \in \{1, 2\}$, the channel $Y = X_1^{X_2}$ behaves as:

X_1	X_2	Y
2	1	2
4	1	4
2	2	4
4	2	16

We compute

$$\begin{aligned}
R_1 &\leq I(X_1; Y|X_2) = I(X_1; X_1^{X_2}|X_2) = H(X_1) = 1 \text{ bit per trans} \\
R_2 &\leq I(X_2; Y|X_1) = I(X_2; X_1^{X_2}|X_1) = H(X_2) = 1 \text{ bit per trans} \\
R_1 + R_2 &\leq I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) = H(Y) = \frac{3}{2} \text{ bits per trans,}
\end{aligned}$$

where the bound on $R_1 + R_2$ is met at the corners in the picture below, where either sender 1 or 2 sends 1 bit per transmission and the other user treats the channel as a binary erasure channel with capacity $1 - p_{erasure} = 1 - \frac{1}{2} = \frac{1}{2}$ bits per use of the channel. Other points on the line are achieved by timesharing.

(b) With $X_1 \in \{1, 2\}, X_2 \in \{1, 2\}$, the channel $Y = X_1^{X_2}$ behaves as:

X_1	X_2	Y
1	1	1
2	1	2
1	2	1
2	2	4

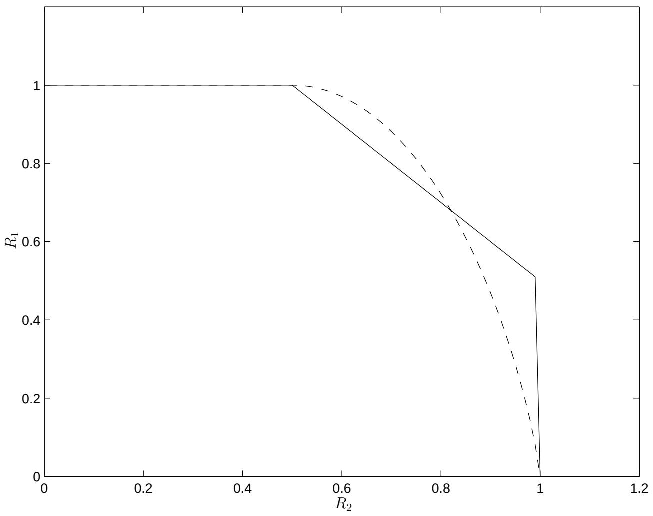


Figure 6: Rate regions for $X_1 \in \{2, 4\}$ and $X_1 \in \{1, 2\}$

Note when $X_1 = 1$, X_2 has no effect on Y and can not be recovered given X_1 and Y. If $X_1 \sim Br(\alpha)$ and $X_2 \sim Br(\beta)$ then:

$$\begin{aligned}
R_1 &\leq I(X_1; Y|X_2) = I(X_1; X_1^{X_2}|X_2) = H(\alpha) \\
R_2 &\leq I(X_2; Y|X_1) = H(Y|X_1) - H(Y|X_1, X_2) = H(Y|X_1) \\
&= p(X_1 = 1)H(Y|X_1 = 1) + p(X_1 = 2)H(Y|X_1 = 2) \\
&= \alpha H(\beta) \\
R_1 + R_2 &\leq I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) = H(Y) \\
&= H(\alpha\beta, \overline{\alpha}\overline{\beta}, 1 - \alpha\beta - \overline{\alpha}\overline{\beta}) = H(\alpha) + \alpha H(\beta)
\end{aligned}$$

We may choose $\beta = \frac{1}{2}$ to maximize the above bounds, giving

$$\begin{aligned}
R_1 &\leq H(\alpha) \\
R_2 &\leq \alpha \\
R_1 + R_2 &\leq H(\alpha) + \alpha
\end{aligned}$$

Above, we plot the region for $X_1 \in \{2, 4\}$ (solid line) against that when $X_1 \in \{1, 2\}$ (dotted). What we find is that, surprisingly, the rate region from the first case is not reduced in the second. In fact, neither region contains the other, so for each version of this channel, there are achievable rate pairs which are *not* achievable in the other.

8. **Broadcast channel.**

   (a) For the degraded broadcast channel $X \to Y_1 \to Y_2$, find the points a and b where the capacity region hits the R_1 and R_2 axes.

   (b) Show that $b \leq a$.

Solution: Broadcast channel

(a) The capacity region of the degraded broadcast channel $X \to Y_1 \to Y_2$ is the convex hull of regions of the form

$$R_1 \leq I(X;Y_1|U) \tag{37}$$

$$R_2 \leq I(U;Y_2) \tag{38}$$

over all choices of auxiliary random variable U and joint distribution of the form $p(u)p(x|u)p(y_1,y_2|x)$.

The region is of the form illustrated in Figure 7.

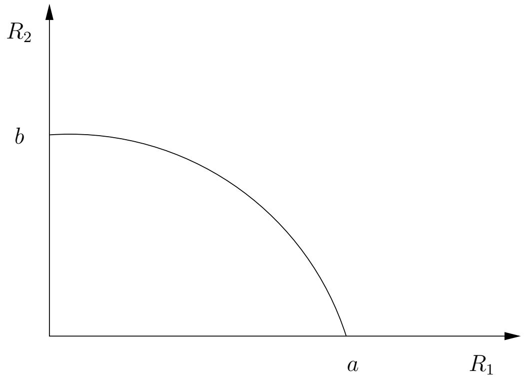


Figure 7: Capacity region of degraded broadcast channel

The point b on the figure corresponds to the maximum achievable rate from the sender to receiver 2. From the expression for the capacity region, it is the maximum value of $I(U;Y_2)$ for all auxiliary random variables U.

For any random variable U and $p(u)p(x|u)$, $U \to X \to Y_2$ forms a Markov chain, and hence $I(U;Y_2) \leq I(X;Y_2) \leq \max_{p(x)} I(X;Y_2)$. The maximum can be achieved by setting $U = X$ and choosing the distribution of X to be the one that maximizes $I(X;Y_2)$. Hence the point b corresponds to $R_2 = \max_{p(x)} I(X;Y_2)$, $R_1 = I(X;Y_1|U) = I(X;Y_1|X) = 0$.

The point a has a similar interpretation. The point a corresponds to the maximum rate of transmission to receiver 1. From the expression for the capacity region,

$$R_1 \le I(X;Y_1|U) = H(Y_1|U) - H(Y_1|X,U) = H(Y_1|U) - H(Y_1|X), \tag{39}$$

since $U \to X \to Y_1$ forms a Markov chain. Since $H(Y_1|U) \le H(Y_1)$, we have

$$R_1 \le H(Y_1) - H(Y_1|X) = I(X;Y_1) \le \max_{p(x)} I(X;Y_1), \tag{40}$$

and the maximum is attained when we set $U \equiv 0$ and choose $p(x) = p(x|u)$ to be the distribution that maximizes $I(X;Y_1)$. In this case, $R_2 \le I(U;Y_2) = 0$.

Hence the point a corresponds to the rates $R_1 = \max_{p(x)} I(X;Y_1), R_2 = 0$.

These results have a simple single user interpretation. If we not sending any information to receiver 1, then we can treat the channel to receiver 2 as a single user channel and send at capacity for this channel, i.e., $\max I(X;Y_2)$. Similarly, if we are not sending any information to receiver 2, we can send at capacity to receiver 1, which is $\max I(X;Y_1)$.

(b) Since $X \to Y_1 \to Y_2$ forms a Markov chain for all distributions $p(x)$, we have by the data processing inequality

$$\begin{aligned} b = \max_{p(x)} I(X;Y_2) &= I(X^*;Y_2) && (41)\\ &\le I(X^*;Y_1) && (42)\\ &= \max_{p(x)} I(X;Y_1) = a, && (43) \end{aligned}$$

where X^* has the distribution that maximizes $I(X;Y_2)$.