Introduction to Information and Coding Theory September 23, 2014
(Prof. Permuter Haim, Mr. Tal Kopetz, and Mr. Oron Sabag)

Final Exam - Moed Bet
Total time for the exam: 3 hours!

1) **Parallel Gaussian channels (25 Points)** Consider a channel consisting of 2 parallel Gaussian channels, with inputs X_1 and X_2 and outputs given by

$$Y_1 = X_1 + Z_1,$$
$$Y_2 = X_2 + Z_2.$$

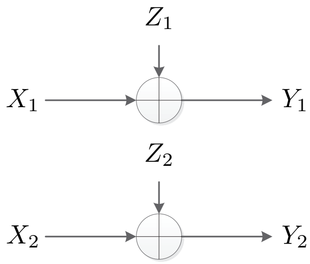

Fig. 1: Parallel Gaussian channels.

The random variables Z_1 and Z_2 are independent of each other and of the inputs, and have the variances σ_1^2 and σ_2^2 respectively, with $\sigma_1^2 < \sigma_2^2$.

a) Suppose $X_1 = X_2 = X$ and we have the power constraint $E[X^2] \le P$. At the receiver, an output $Y = Y_1 + Y_2$ is generated. What is the capacity C_a of the resulting channel with X as the input and Y as the output?
b) Suppose that we still have to transmit the same signal on both channels, but we can now choose how to distribute the power between the channels, i.e. $X_1 = aX$ and $X_2 = bX$. The new constraint is $E[X_1^2] + E[X_2^2] \le 2P$. What is the capacity, C_b, of this channel with X as the input and (Y_1, Y_2) as the output? Which a and b achieve that capacity?
c) We now assume that Z_1 and Z_2 are dependent, specifically, $Z_2 = 2Z_1$. As in subsection b, we can choose how to distribute the power between the channels, i.e. $X_1 = aX$ and $X_2 = bX$ under the power constraint $E[X_1^2] + E[X_2^2] \le 2P$. The outputs of the channels are given by

$$Y_1 = aX + Z_1,$$
$$Y_2 = bX + 2Z_1.$$

What is the capacity, C_c, of this channel with X as the input and (Y_1, Y_2) as the output? Which a and b achieve that capacity?

Solution

a) This channel has an input X and output Y and as we learned in class, the capacity of the Gaussian channel is given by

$$C = \frac{1}{2}\log(1 + \text{SNR}). \tag{1}$$

In our case,

$$\begin{aligned} \text{SNR} &= \frac{E[(X_1 + X_2)^2]}{E[(Z_2 + Z_2)^2]} \\ &\le \frac{4P}{\sigma_1^2 + \sigma_2^2}. \end{aligned} \tag{2}$$

So, the capacity of this channel is given by

$$C = \frac{1}{2}\log\left(1 + \frac{4P}{\sigma_1^2 + \sigma_2^2}\right). \tag{3}$$

b) Let

$$\begin{aligned} Y_1 &= aX + Z_1 \\ Y_2 &= bX + Z_2, \end{aligned} \tag{4}$$

where Z_1 and Z_2 are independent of each other and have the variances σ_1^2 and σ_2^2 respectively, with $\sigma_1^2 < \sigma_2^2$. We seek the values of a, b that maximize

$$
\begin{aligned}
I(X;Y_1,Y_2) &= h(Y_1,Y_2) - h(Y_1,Y_2|X) \\
&= h(Y_1,Y_2) - h(Z_1,Z_2) \\
&= h(Y_1,Y_2) - \frac{1}{2}\log 2\pi e\sigma_1^2\sigma_2^2, \qquad (5)
\end{aligned}
$$

under the constraint $a^2 + b^2 \leq 2$. In order to find $h(Y_1, Y_2)$ we need to find the covariance matrix of Y_1, Y_2, which is given by

$$
\Sigma_{Y_1,Y_2} = \begin{pmatrix} a^2P + \sigma_1^2 & abP \\ abP & b^2P + \sigma_2^2 \end{pmatrix}. \qquad (6)
$$

Then,

$$
\begin{aligned}
|\Sigma_{Y_1,Y_2}| &= (a^2P + \sigma_1^2)(b^2P + \sigma_2^2) - a^2b^2P^2 \\
&= P(a^2\sigma_2^2 + b^2\sigma_1^2) + \sigma_1^2\sigma_2^2 \\
&\leq P(a^2\sigma_2^2 + (2 - a^2)\sigma_1^2) + \sigma_1^2\sigma_2^2 \\
&= a^2P(\sigma_2^2 - \sigma_1^2) + (2P + \sigma_2^2)\sigma_1^2, \qquad (7)
\end{aligned}
$$

and

$$
h(Y_1,Y_2) \leq \log 2\pi e + \frac{1}{2}\log[a^2P(\sigma_2^2 - \sigma_1^2) + (2P + \sigma_2^2)\sigma_1^2]. \qquad (8)
$$

We can now see that, since $\sigma_1^2 < \sigma_2^2$, the expression in (8) achieves its maximum value when a achieves its maximal value, namely, for $a = \sqrt{2}$. We conclude that the optimal strategy in this case is to use only X_1 to transmit the data, and the capacity is thus

$$
C_b = \frac{1}{2}\log\left(1 + \frac{2P}{\sigma_1^2}\right) \qquad (9)
$$

c) In this case, we can set $X_1 = 0$, $X_2 = X$ and $Y = Y_2 - 2Y_1$. Substituting the equations for Y_1, Y_2 and Z_2 we see that

$$
Y = X. \qquad (10)
$$

Thus, the capacity is infinite.

2) **Erasure Channel with Feedback (25 Points)**

Let X be a random variable that is uniformly distributed in the interval $[0, 1]$.

a) Is it possible to generate from one realization of X a binary random variable that is distributed Bernoulli(p)? If yes, prove it.

Consider the erasure channel with feedback as depicted in Fig. 2.

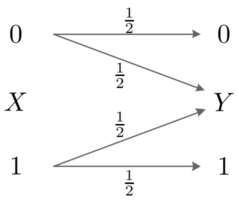

Fig. 2: Erasure Channel with erasure parameter $\epsilon = \frac{1}{2}$.

A student provided the following coding scheme for the erasure channel: The message M has a finite alphabet of size 2^{nR} and the points of the alphabet are distributed uniformly in the interval $[0, 1]$, i.e. $m \in \{k \cdot \frac{1}{2^{nR}}\}_{k=0}^{2^{nR}-1}$. Fix a parameter $p \in [0, 1]$. The interval $[0, 1]$ is divided into two parts, $[0, p)$ and $[p, 1)$. In the first transmission, if $m \in [0, p)$ the encoder transmits $'0'$ and if $m \in [p, 1)$ the encoder transmits $'1'$.

Upon a successful transmission, the decoder knows the interval where the message falls and this interval is divided again with the same parameter p. If the transmission failed, the encoder repeats the transmitted bit until a successful transmission is established.

b) What is the rate of the proposed coding scheme.

c) Can this coding scheme achieve the capacity of the erasure channel? If yes, prove it.

Solution

a) Yes. We construct the RV $Y \sim$Bernoulli(p) in the following way

$$
Y = \left\{ \begin{array}{ccc} 0 & , & x \in [0, p] \\ 1 & , & x \in (p, 1] \end{array} \right\} \qquad (11)
$$

b) The capacity for an erasure channel is $C = \max_{p(x)} H(X)(1 - \epsilon)$ where in our case $p(x)$ is set and $\epsilon = \frac{1}{2}$ and thus $C = \frac{1}{2}H_b(p)$.

c) It can be seen that when $p = \frac{1}{2}$ we obtain the capacity for the erasure channel which is $C = \frac{1}{2}$.

3) **Secure Network Coding (25 Points)** Consider the network depicted in Fig. 3.
The source S would like to transmit a message W to the terminal T. The message, W, is a random binary vector of length k, i.e. $W = [w_1, w_2, \dots, w_k]$, where each element w_i is distributed $w_i \sim Bern(0.5)$. Each link in the network can carry only one bit, the bit b_1 is transmitted at the upper link and b_2 through the lower link. A spy acquires, E, which is a random observation of one of the links. We know that $E = b_1$ with probability p and $E = b_2$ with probability $1 - p$.
Our goal is to maximize the amount of information that is transmitted to the terminal, while preserving that $I(E; W) = 0$ which means zero information available to the spy. All codebooks are known to the encoder, decoder, and to the spy.

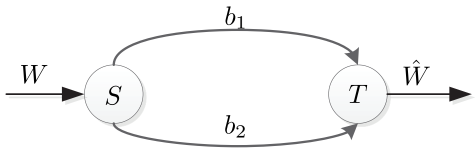

Fig. 3: Network with one source and one terminal.

a) Find $I(A; A \oplus B)$ for $A \sim Bern(\alpha)$ and $B \sim Bern(0.5)$.
b) What is the maximum number of bits (maximum k) that the source S can send to node T in one transmission assuming that the spy is NOT listening, i.e., $I(E; W)$ is NOT necessarily 0?
Provide an achievability scheme and a converse.
c) What is the maximum number of bits (maximum k) that the source S can send to node T in one transmission while preserving $I(E; W) = 0$ for any value of p?
Provide an achievability scheme and a converse. For the achievablitiy, you may use an additional RV which is distributed uniformly in the interval $[0, 1]$ and is drawn at the encoder S.
d) Is there a specific value of p which will allow us to send more bits? If yes, prove and if no give a counter example.

Solution

a) $I(A; A \oplus B) = 0$. Since $B \sim Bern(0.5)$, we obtain a new RV $C = A \oplus B$ that is distributed by $C \sim Bern(0.5)$ and is independent of A and thus the mutual information is zero.
b) We can send maximum 2 bits.
Achievability: Each bit from each link.
Converse: Cut-set bound.
c) We can send maximum 1 bit.
Achievability: We send a random bit d where $D \sim Bern(0.5)$ through link b_1 and the bit $w \oplus d$ through link b_2. If $e = b_1 = d$, then $I(E; W) = 0$ since d does not carry any information regarding w. If $e = b_1 = w \oplus d$, as in section a, $W \oplus D$ is independent of W and thus $I(E; W) = 0$.
Converse: We will prove this by contradiction. Assuming we can send 2 bits of information w_1, w_2. This means that $b_1 = f(w_1, w_2)$ and $b_2 = f(w_1, w_2)$. Now we must make sure that $I(E; W) = 0$, but $I(E; W) = pI(b_1; w_1, w_2) + \bar{p}I(b_2; w_1, w_2) = pH(b_1) - pH(b_1|w_1, w_2) + \bar{p}H(b_2) - \bar{p}H(b_2|w_1, w_2) = pH(b_1) + \bar{p}H(b_2)$ and this equals to 0 only if b_1 and b_2 are constants. In this case no information regarding w_1, w_2 will be transmitted and thus sending 2 bits of information is impossible and we have a contradiction.
d) Since we have proven the converse in section c for any p, no more than 1 bit of information is possible.

4) **Bhattacharyya distance (25 Points)** For two probability density functions, $f(x)$ and $g(x)$, define the *Bhattacharyya distance* between f and g as

$$D_b(f, g) = -\log \left(\int_{-\infty}^{\infty} \sqrt{f(x)g(x)} dx \right) \tag{12}$$

The Bhattacharyya distance is widely used in various fields such as machine learning, statistics, and more.
For this question, the base of the logarithm is 2.

a) Prove that $0 \leq D_b(f, g) \leq \infty$.
When does $D_b(f, g) = 0$? When does $D_b(f, g) = \infty$?
Hint: You can use the Cauchy Schwarz inequality: for any two real valued functions $f_1(x), f_2(x)$, we have:

$$\left| \int_{-\infty}^{\infty} f_1(x) f_2(x) dx \right|^2 \leq \int_{-\infty}^{\infty} |f_1(x)|^2 dx \int_{-\infty}^{\infty} |f_2(x)|^2 dx. \tag{13}$$

b) We define the differential divergence as follows:

$$D(f||g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx. \tag{14}$$

Let $h(x)$ be a third probability density function. Show that

$$D_b(f, g) \leq \frac{1}{2} \left(D(h||f) + D(h||g) \right). \tag{15}$$

c) Assume that $D_b(f,g) < \infty$. For what $h(x)$, there is an equality in Eq.(15)?
d) Does the following inequality holds?

$$2D_b(f,g) \leq \min\{D(g||f), D(f||g)\} \tag{16}$$

If yes, prove it, if not, give a counter example.

Solution

a) By Cauchy Schwarz inequality we have:

$$\begin{aligned}
\int_{-\infty}^{\infty} \sqrt{f(x)g(x)}dx &\leq \int_{-\infty}^{\infty} \left|\sqrt{f(x)}\right|^2 dx \int_{-\infty}^{\infty} \left|\sqrt{g(x)}\right|^2 dx &(17)\\
&= \int_{-\infty}^{\infty} f(x)dx \int_{-\infty}^{\infty} g(x)dx &(18)\\
&= 1 \cdot 1 = 1. &(19)
\end{aligned}$$

Since $-\log$ is a monotonically decreasing function, we have:

$$\begin{aligned}
D_b(f,g) &= -\log\left(\int_{-\infty}^{\infty} \sqrt{f(x)g(x)}dx\right) &(20)\\
&\geq -\log(1) = 0. &(21)
\end{aligned}$$

The last equality holds only when there is equality in the Cauchy Schwarz inequality, which happens only if $f = \alpha \cdot g$. Since both f and g have an integral equal to one:

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} g(x)dx = 1, \tag{22}$$

then $\alpha = 1$, which means $f = g$. The other equality $D_b(f,g) = \infty$ happens when

$$\int_{-\infty}^{\infty} \sqrt{f(x)g(x)}dx = 0, \tag{23}$$

and since $f(x)g(x) \geq 0$, we have that $f(x)g(x) = 0$ for almost every x. That means that f and g have different supports.

b)

$$\begin{aligned}
\frac{1}{2}\left(D(h||f) + D(h||g)\right) &= \frac{1}{2}\left(\mathbb{E}_h\left[\log\left(\frac{h(X)}{f(X)}\right)\right] + \mathbb{E}_h\left[\log\left(\frac{h(X)}{g(X)}\right)\right]\right) &(24)\\
&= \frac{1}{2}\mathbb{E}_h\left[\log\left(\frac{h(X)}{f(X)}\right) + \log\left(\frac{h(X)}{g(X)}\right)\right] &(25)\\
&= \frac{1}{2}\mathbb{E}_h\left[\log\left(\frac{h^2(X)}{f(X)g(X)}\right)\right] &(26)\\
&= \frac{1}{2}\mathbb{E}_h\left[-\log\left(\frac{f(X)g(X)}{h^2(X)}\right)\right] &(27)\\
&= \mathbb{E}_h\left[-\log\left(\frac{\sqrt{f(X)g(X)}}{h(X)}\right)\right] &(28)\\
&\overset{(a)}{\geq} -\log\left(\mathbb{E}_h\left[\frac{\sqrt{f(X)g(X)}}{h(X)}\right]\right) &(29)\\
&= \log\left(\int_{-\infty}^{\infty} \frac{\sqrt{f(x)g(g)}}{h(x)} h(x)dx\right) &(30)\\
&= D_b(f,g) &(31)
\end{aligned}$$

where (a) follows from Jensen's inequality.

c) There is an equality in Eq.(15) if and only if there is an equality in Jensen's inequality. Since $-\log$ is a strictly convex function, there is an equality iff $\frac{\sqrt{f(X)g(X)}}{h(X)}$ is deterministic (equals to a constant). That means:

$$\alpha\sqrt{(f(X)g(X)} = h(X), \quad \text{with probability 1.} \tag{32}$$

In order to find the constant we integrate both sides:

$$\alpha \int_{-\infty}^{\infty} \sqrt{f(x)g(x)}dx = 1 \tag{33}$$

$$\alpha 2^{-D_b(f,g)} = 1 \tag{34}$$

$$\alpha = 2^{D_b(f,g)} \tag{35}$$

$$h(x) = 2^{D_b(f,g)}\sqrt{f(x)g(x)}. \tag{36}$$

d) Take once $h = f$ and once $h = g$ to get:
If $h = f$:

$$D_b(f,g) \leq \frac{1}{2}\left(D(f||f) + D(f||g)\right) \tag{37}$$

$$= \frac{1}{2}\left(0 + D(f||g)\right). \tag{38}$$

If $h = g$:

$$D_b(f,g) \leq \frac{1}{2}\left(D(g||f) + D(g||g)\right) \tag{39}$$

$$= \frac{1}{2}\left(D(g||f) + 0\right). \tag{40}$$

Good Luck!