2nd Semester 2010

Solutions to Homework Set #5
Differential Entropy and Gaussian Channel

1. **Differential entropy.**
   Evaluate the differential entropy $h(X) = -\int f \ln f$ for the following:

   (a) Find the entropy of the exponential density $\lambda e^{-\lambda x},\ x \geq 0$.

   (b) The sum of X_1 and X_2, where X_1 and X_2 are independent normal random variables with means μ_i and variances $\sigma_i^2,\ \ i = 1, 2$.

   Solution: Differential entropy.

   (a)

$$h(f) \quad = \quad \log \frac{e}{\lambda} \text{ bits.} \tag{1}$$

   (b) Sum of two normal distributions.
   The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2} \log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.} \tag{2}$$

2. **Mutual information for correlated normals.** Find the mutual information $I(X;Y)$, where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right).$$

   Evaluate $I(X;Y)$ for $\rho = 1, \rho = 0$, and $\rho = -1$, and comment.

   Mutual information for correlated normals.

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right) \tag{3}$$

Using the expression for the entropy of a multivariate normal derived in class

$$h(X,Y) = \frac{1}{2}\log(2\pi e)^2|K| = \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2). \tag{4}$$

Since X and Y are individually normal with variance σ^2,

$$h(X) = h(Y) = \frac{1}{2}\log 2\pi e\sigma^2. \tag{5}$$

Hence

$$I(X;Y) = h(X) + h(Y) - h(X,Y) = -\frac{1}{2}\log(1-\rho^2). \tag{6}$$

(a) $\rho = 1$. In this case, $X = Y$, and knowing X implies perfect knowledge about Y. Hence the mutual information is infinite, which agrees with the formula.

(b) $\rho = 0$. In this case, X and Y are independent, and hence $I(X;Y) = 0$, which agrees with the formula.

(c) $\rho = -1$. In this case, $X = -Y$, and again the mutual information is infinite as in the case when $\rho = 1$.

3. **Markov Gaussian mutual information.**
Suppose that (X,Y,Z) are jointly Gaussian and that $X \to Y \to Z$ forms a Markov chain. Let X and Y have correlation coefficient ρ_1 and let Y and Z have correlation coefficient ρ_2. Find $I(X;Z)$.

Solution: Markov Gaussian mutual information.
First note that we may without any loss of generality assume that the means of X, Y and Z are zero. If in fact the means are not zero one can subtract the vector of means without affecting the mutual information or the conditional independence of X, Z given Y. Similary we can also assume the variances of X, Y, and Z to be 1. (The scaling may change the differential entropy, but not the mutual information.)

Let

$$\Sigma = \begin{pmatrix} 1 & \rho_{xz} \\ \rho_{xz} & 1 \end{pmatrix},$$

be the covariance matrix of X and Z. From Eqs. (9.93) and (9.94)

$$\begin{aligned} I(X;Z) &= h(X) + h(Z) - h(X,Z) \\ &= \frac{1}{2}\log\left(2\pi e\right) + \frac{1}{2}\log\left(2\pi e\right) - \frac{1}{2}\log\left(2\pi e \det(\Sigma)\right) \\ &= -\frac{1}{2}\log(1-\rho_{xz}^2) \end{aligned}$$

Now from the conditional independence of X and Z given Y, we have

$$\begin{aligned} \rho_{xz} &= \mathbf{E}[XZ] \\ &= \mathbf{E}\left[\mathbf{E}[XZ|Y]\right] \\ &= \mathbf{E}\left[\mathbf{E}[X|Y]\cdot\mathbf{E}[Z|Y]\right] \\ &= \mathbf{E}[\rho_1 Y \cdot \rho_2 Y] \\ &= \rho_1\rho_2. \end{aligned}$$

We can thus conclude that

$$I(X;Z) = -\frac{1}{2}\log(1-\rho_1^2\rho_2^2)$$

4. **Output power constraint.**
Consider an additive white Gaussian noise channel with an expected output power constraint P. (We might want to protect the eardrums of the listener.) Thus $Y = X + Z$, $Z \sim N(0,\sigma^2)$, Z is independent of X, and $EY^2 \le P$. Assume $\sigma^2 < P$. Find the channel capacity.

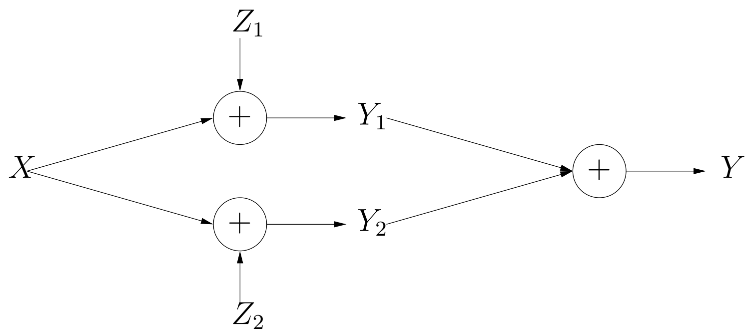
Solution: Output power constraint.
The output power constraint $EY^2 \le P$ is equivalent to the input power constraint

$$E(X+Z)^2 = EX^2 + EZ^2 = EX^2 + \sigma^2 \le P,$$

that is, $EX^2 \le P - \sigma^2$. Thus, we reduce the problem to a previously known one and get

$$C = \frac{1}{2}\log\left(\frac{P}{\sigma^2}\right).$$

5. **Multipath Gaussian channel.**
Consider a Gaussian noise channel of power constraint P, where the signal takes two different paths and the received noisy signals are added together at the antenna.



Let $Y = Y_1 + Y_2$ and $EX^2 \leq P$.

(a) Find the capacity of this channel if Z_1 and Z_2 are jointly normal with covariance matrix

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}.$$

(b) What is the capacity for $\rho = 0, -1$, and 1 ?

Solution: Multipath Gaussian channel.

(a) Since

$$\begin{aligned} Y &= Y_1 + Y_2 \\ &= X + Z_1 + X + Z_2 \\ &= 2X + (Z_1 + Z_2), \end{aligned}$$

and $Z_1 + Z_2$ is $\sim N(0, 2N(1+\rho))$, the capacity is given by

$$C = \frac{1}{2} \log \left(1 + \frac{4P}{2N(1+\rho)} \right) = \frac{1}{2} \log \left(1 + \frac{2P}{N(1+\rho)} \right).$$

(b) When $\rho = 0$,

$$C = \frac{1}{2} \log \left(1 + \frac{2P}{N} \right).$$

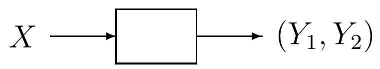
When $\rho = 1$,

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right),$$

which makes sense since $Y_1 = Y_2$ and $Y = 2Y_1$. (Scaling the output does not change the mutual information.)

When $\rho = -1$, we have $C = \infty$. Since $Z_1 + Z_2 = 0$, the channel is given by $Y = 2X$ without any additive noise. Hence we can transmit unbounded amount of information (any real number satisfying the power constraint) over the channel without any error.

6. **The two-look Gaussian channel.**



Consider the ordinary additive noise Gaussian channel with two correlated looks at X, i.e., $Y = (Y_1, Y_2)$, where

$$Y_1 = X + Z_1$$

$$Y_2 = X + Z_2$$

with a power constraint P on X, and $(Z_1, Z_2) \sim \mathcal{N}_2(\mathbf{0}, K)$, where

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}.$$

Find the capacity C for

(a) ρ = 1.

(b) ρ = 0.

(c) ρ = -1.

Solution: The two-look Gaussian channel

It is clear that the input distribution that maximizes the capacity is $X \sim \mathcal{N}(0, P)$. Evaluating the mutual information for this distribution,

$$\begin{aligned} C_2 &= \max I(X; Y_1, Y_2) \\ &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2) \end{aligned}$$

Now since

$$(Z_1, Z_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}\right),$$

we have

$$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2|K_Z| = \frac{1}{2}\log(2\pi e)^2 N^2(1-\rho^2).$$

Since $Y_1 = X + Z_1$, and $Y_2 = X + Z_2$, we have

$$(Y_1, Y_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} P+N & P+\rho N \\ P+\rho N & P+N \end{bmatrix}\right),$$

and

$$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2|K_Y| = \frac{1}{2}\log(2\pi e)^2(N^2(1-\rho^2) + 2PN(1-\rho)).$$

Hence the capacity is

$$\begin{aligned} C_2 &= h(Y_1, Y_2) - h(Z_1, Z_2) \\ &= \frac{1}{2}\log\left(1 + \frac{2P}{N(1+\rho)}\right). \end{aligned}$$

(a) $\rho = 1$. In this case, $C = \frac{1}{2}\log(1 + \frac{P}{N})$, which is the capacity of a single look channel. This is not surprising, since in this case $Y_1 = Y_2$.

(b) $\rho = 0$. In this case,

$$C = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right),$$

which corresponds to using twice the power in a single look. The capacity is the same as the capacity of the channel $X \to (Y_1 + Y_2)$.

(c) $\rho = -1$. In this case, $C = \infty$, which is not surprising since if we add Y_1 and Y_2, we can recover X exactly.

Note that the capacity of the above channel in all cases is the same as the capacity of the channel $X \to Y_1 + Y_2$.

7. **Diversity System**

For the following system, a message $W \in \{1, 2, \ldots, 2^{nR}\}$ is encoded into two symbol blocks $X_1^n = (X_{1,1}, X_{1,2}, ..., X_{1,n})$ and $X_2^n = (X_{2,1}, X_{2,2}, ..., X_{2,n})$ that are being transmitted over a channel. The average power constrain on the inputs are $\frac{1}{n}E[\sum_{i=1}^{n} X_{1,i}^2] \leq P_1$ and $\frac{1}{n}E[\sum_{i=1}^{n} X_{2,i}^2] \leq P_2$. The channel has a multiplying effect on X_1, X_2 by factor h_1, h_2, respectively, i.e., $Y = h_1X_1 + h_2X_2 + Z$, where Z is a white Gaussian noise $Z \sim N(0, \sigma^2)$.

(a) Find the joint distribution of X_1 and X_2 that bring the mutual information $I(Y; X_1, X_2)$ to a maximum? (You need to find $\arg\max P_{X_1,X_2} I(X_1, X_2; Y)$.)

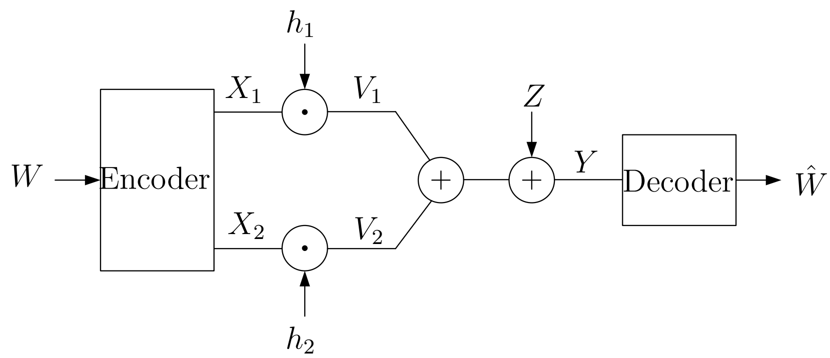


Figure 1: The communication model

(b) What is the capacity of the system ?

(c) Express the capacity for the following cases:

i. $h_1 = 1,\ h_2 = 1$?

ii. $h_1 = 1,\ h_2 = 0$?

iii. $h_1 = 0,\ h_2 = 0$?

Solution: Diversity System

(a)

$$Y = h_1 X_1 + h_2 X_2 + Z$$

The mutual information is:

$$\begin{aligned} I(X_1, X_2; Y) &= h(Y) - h(Y|X_1, X_2) \\ &= h(Y) - h(Z) \end{aligned}$$

Since $h(z)$ is constant, we need to find the maximum of $h(Y)$, the second moment of Y is:

$$\begin{aligned} E[Y^2] &= E[(h_1X_1 + h_2X_2 + Z)^2] \\ &\overset{(i)}{=} E[(h_1X_1 + h_2X_2)^2] + E[Z^2] \\ &= h_1^2[X_1^2] + h_2^2[X_2^2] + 2h_1h_2E[X_1X_2] + \sigma_Z^2 \\ &\leq h_1^2P_1 + h_2^2P_2 + 2h_1h_2E[X_1X_2] + \sigma_Z^2 \\ &\overset{(ii)}{\leq} h_1^2P_1 + h_2^2P_2 + 2h_1h_2\sqrt{E[X_1^2]E[X_2^2]} + \sigma_Z^2 \\ &\leq h_1^2P_1 + h_2^2P_2 + 2h_1h_2\sqrt{P_1P_2} + \sigma_Z^2 \\ &= (h_1\sqrt{P_1} + h_2\sqrt{P_2})^2 + \sigma_Z^2 \end{aligned}$$

(i) - Z is independent of X_1, X_2.
(ii) - Cauchy-Schwarz inequality. Where $X_1 = \alpha X_2$, $\binom{X_1}{X_2} \sim N(0, K)$ and $K = \begin{pmatrix} P_1 & \sqrt{P_1P_2} \\ \sqrt{P_1P_2} & P_2 \end{pmatrix}$ will result with equality and bring the mutual information to a maximum.

Therefore, the mutual information is bounded by:

$$I(X_1, X_2; Y) \leq \frac{1}{2}\log\left(1 + \frac{(h_1\sqrt{P_1} + h_2\sqrt{P_2})^2}{\sigma_Z^2}\right)$$

(b) The capacity of the system is:

$$C = \max_{P_{x_1,x_2}} I(X_1, X_2; Y) = \frac{1}{2}\log\left(1 + \frac{(h_1\sqrt{P_1} + h_2\sqrt{P_2})^2}{\sigma_Z^2}\right)$$

(c) For $h_1 = 1$ and $h_2 = 1$ the capacity of the system would be:

$$\begin{aligned} C &= \frac{1}{2}\log\left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{\sigma_Z^2}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{P_1 + 2\sqrt{P_1P_2} + P_2}{\sigma_Z^2}\right) \end{aligned}$$

For $h_1 = 1$ and $h_2 = 0$ the capacity of the system would be:

$$C = \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_Z^2}\right)$$

For $h_1 = 0$ and $h_2 = 0$ the capacity of the system would be:

$$C = \frac{1}{2}\log(1) = 0$$

We can see that having 2 Gaussian channels with one message, it is the best to transmit the signals coherently.

8. **AWGN with two noises**

Figure 2 depicts a communication system with an AWGN (Additive white noise Gaussian) channel whith two i.i.d. noises $Z_1 \sim N(0, \sigma_1^2)$, $Z_2 \sim N(0, \sigma_2^2)$ that are independent of each other and are added to the signal X, i.e., $Y = X + Z_1 + Z_2$. The average power constrain on the input is P, i.e., $\frac{1}{n}E[\sum_{i=1}^{n} X_i^2] \leq P$. In the sub-questions below we consider the cases where the noise Z_2 may or may not be known to the encoder and decoder.

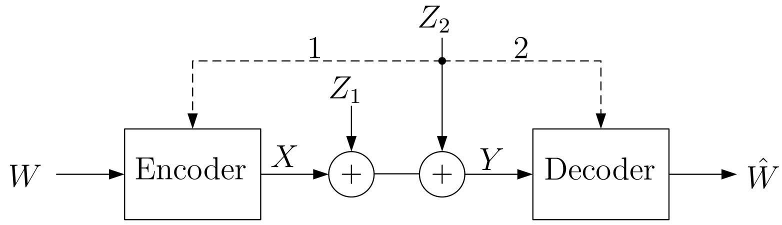


Figure 2: Two noise sources

(a) Find the channel capacity for the case in which the noise in not known to either sides (lines 1 and 2 are <u>disconnected</u> from the encoder and the decoder).

(b) Find the capacity for the case that the noise Z_2 is known to the encoder and decoder (lines 1 and 2 are <u>connected</u> to both the encoder and decoder). This means that the codeword X^n may depend on the message W and the noise Z_2^n and the decoder decision $\hat{W}$ may depend on the output Y^n and the noise Z_2^n. (**Hint**: Could the capacity be lager than $\frac{1}{2}\log(1 + \frac{P}{\sigma_1^2})$?)

(c) Find the capacity for the case that the noise Z_2 is known only to the decoder. (line 1 is <u>disconnected</u> from the encoder and line 2 is <u>connected</u> to the decoder). This means that the codewords X^n may depend only on the message W and the decoder decision $\hat{W}$ may depend on the output Y^n and the noise Z_2^n.

Solution: AWGN with two noises

(a) Since the noise is not know to both sides, the total noise is $\sigma_1^2 + \sigma_2^2$ and the capacity is:

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

(b) Once Z_2 is known to the receiver, we can add a subtraction unit in the decoder that subtract Z_2 and therefore the noise is only Z_1. And the capacity is:

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

9. **Parallel channels and waterfilling**
Consider a pair of parallel Gaussian channels, i.e.,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

where

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right),$$

and there is a power constraint $E(X_1^2 + X_2^2) \leq P$. Assume that $\sigma_1^2 > \sigma_2^2$. At what power does the channel stop behaving like a single channel with noise variance σ_2^2, and begin behaving like a pair of channels, ie., at what power does the worst channel become useful?

Solution: Parallel channels and waterfilling
By the result of water filling taught in the class , it follows that we will put all the signal power into the channel with less noise until the total power of noise + signal in that channel equals the noise power in the other channel. After that, we will split any additional power evenly between the two channels.

Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e., when $P = \sigma_1^2 - \sigma_2^2$.