Homework Set #4
Differential Entropy, Gaussian Channel, Lagrange multipliers (KKT conditions)

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 $\sigma_i^2,\ \ i = 1, 2$.

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

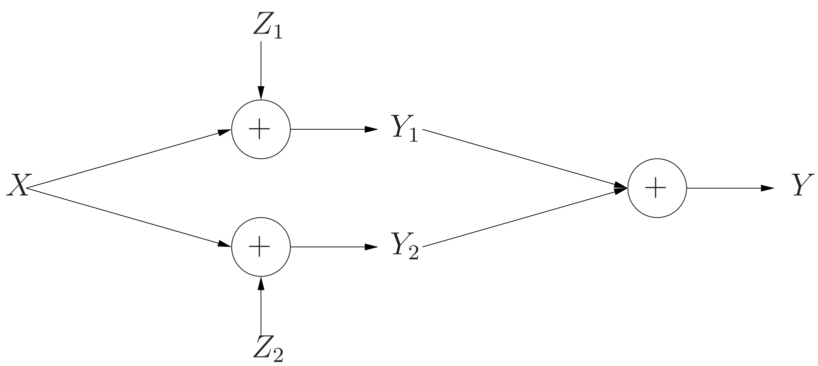
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right).$$

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

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 $I(X;Z)$.

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 $EY^2 \leq P$. Assume $\sigma^2 < P$. Find the channel capacity.

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.



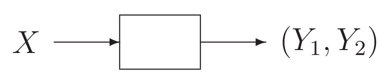
Let $Y = Y_1 + Y_2$ and $EX^2 \leq P$.

(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 ?

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) $\rho = 1$.

(b) $\rho = 0$.

(c) $\rho =$ -1.

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 $W \in \{1, 2, \dots, 2^{nR}\}$ 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 $\frac{1}{n}E[\sum_{i=1}^n X_{1,i}^2] \le P_1$ and $\frac{1}{n}E[\sum_{i=1}^n X_{2,i}^2] \le P_2$. The channel has a multiplying effect on X_1, X_2 by factor h_1, h_2, respectively, i.e., $Y = h_1 X_1 + h_2 X_2 + Z$, 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 $P_{X_1,X_2} I(X_1, X_2; Y)$.)

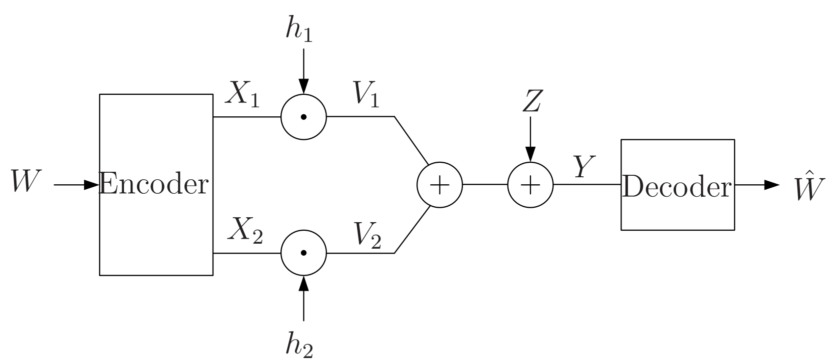


Figure 1: The communication model

(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. $h_1 = 0$, $h_2 = 0$?

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.

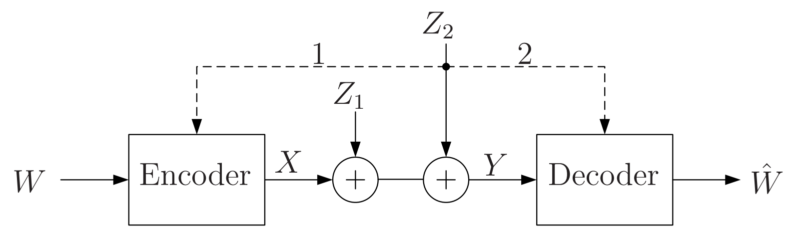


Figure 2: Two noise sources

(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.

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?

10. **Blahut-Arimoto's algorithm and KKT conditions (Lagrange multiplier)** Recall, that the capacity of a memoryless channel is given by

$$C = \max_{p(x)} I(X;Y).$$

Solving this optimization problem is a difficult task for the general channel. In this question we develop an iterative algorithm for finding the solution for a <u>fixed</u> channel $p(y|x)$.

   (a) Prove that the mutual information as a function of $p(x)$ and $p(x|y)$ may be written as

$$I(X;Y) = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)}.$$

   (b) Show that $I(X;Y)$ as written above is concave in both $p(x)$, $p(x|y)$ (Hint. You may use the log-sum-inequality).

   (c) Find an expression for $p(x)$ that maximizes $I(X;Y)$ when $p(x|y)$ is fixed (Hint. You may use the Lagrange multipliers method with the constraint $\sum_x p(x) = 1$. No need to take into account that $p(x) \geq 0$ since it will obtained anyway.)

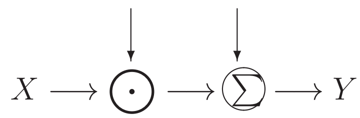
   (d) Find an expression for $p(x|y)$ that maximizes $I(X;Y)$ when $p(x)$ is fixed (Hint. You may use the Lagrange multipliers method with constraints $\sum_x p(x|y) = 1$ for all y. No need to take into account that $p(x|y) \geq 0$ since it will obtained anyway.).

   (e) Using (d), conclude that $C = \max_{p(x),p(x|y)} I(X;Y)$.

The Blahut-Arimoto's algorithm is performed by maximizing in each iteration over another variable; first over $p(x)$ when $p(x|y)$ is fixed, then over $p(x|y)$ when $p(x)$ is fixed, and so on. This iterative algorithm converges, and hence one can find the capacity of any DMC $p(y|x)$ with reasonable alphabet size.

11. **Fading channel.**
Consider an additive noise fading channel

V Z



$$Y = XV + Z,$$

where Z is additive noise, V is a random variable representing fading, and Z and V are independent of each other and of X.

(a) Argue that knowledge of the fading factor V improves capacity by showing

$$I(X;Y|V) \geq I(X;Y).$$

(b) Incidentally, conditioning does not always increase mutual information. Give an example of $p(u,r,s)$ such that $I(U;R|S) < I(U;R)$.

12. **Additive Gaussian channel where the noise might be a relay**

In this question we consider a channel with additive Gaussian noise as seen in class.

Consider the channel presented in Fig. 3.

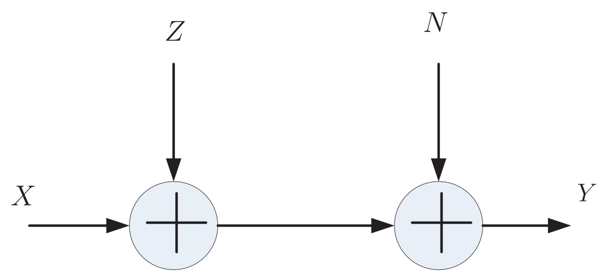


Figure 3: Additive Gaussian noise channel.

$$Y = X + Z + N,$$

where $Z \sim \mathcal{N}(0, \sigma_1^2)$ and $N \sim \mathcal{N}(0, \sigma_2^2)$ are additive noises and the input, X, is with power constraint P. N, Z and X are independent.

(a) Calculate the capacity of the channel assuming that the noise is independent of the message that the encoder uses for determining X_i.

(b) Now it is given that Z_i is an output of a relay-encoder which has access to the same message M that the channel encoder has. Hence X and Z are no longer independent. It is also given that Z has a power constraint P, namely $\frac{1}{n}\sum_{i=1}^{n} Z_i \leq P$ with high probability. Find the capacity of the channel and the probability density function $f(x,z)$ for which it is achieved.

13. **True or False**
Copy the following to your notebook and write **true** or **false**. Then, if it's true, prove it. Otherwise, if it's false, give a counter example or prove that the opposite is true.

- Let X be a continuous random variable. Then the following holds

$$I(X;X) = h(X).$$

14. **Two antennas with Gaussian noise**
In this question we consider a point-to-point discrete memoryless channel (DMC) in which the transmitter and the receiver both have two antennas, illustrated in Fig. 4. This channel is defined by two input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a channel transition matrix $P_{Y_1Y_2|X_1X_2}$. A message M is randomly and uniformly chosen from the message set $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$ and is to be transmitted from the encoder to the decoder in a lossless manner (as defined in class).

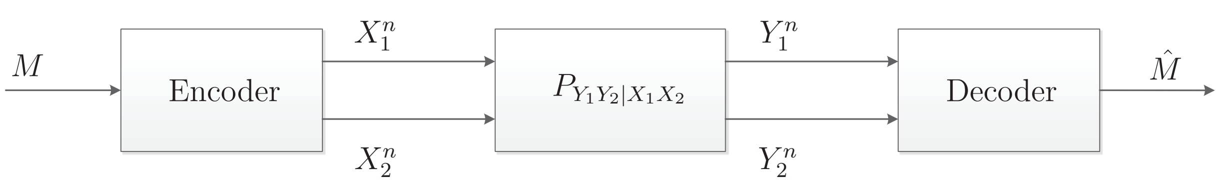


Figure 4: Two antenna point-to-point DMC.

(a) What is the capacity of the channel?

Now, consider the following Gaussian two antenna point-to-point DMC illustrated in Fig. 5

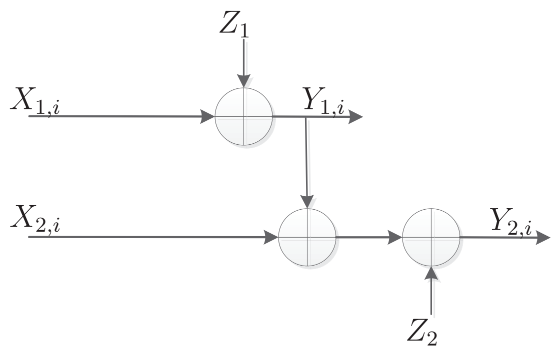


Figure 5: A Gaussian two antenna point-to-point DMC.

The outputs of the channel for every time $i \in \{1, \ldots, n\}$ are give by,

$$Y_{1,i} = X_{1,i} + Z_1, \tag{1}$$
$$Y_{2,i} = X_{1,i} + X_{2,i} + Z_1 + Z_2, \tag{2}$$

where (Z_1, Z_2) are two independent (of each other and of everything else) Gaussian random variable distributed according to $Z_1 \sim \mathcal{N}(0, N_1)$ and $Z_1 \sim \mathcal{N}(0, N_2)$. The input signals are bound to an average power constraints,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2\right] \leq P_1 \quad ; \quad \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_{2,i}^2\right] \leq P_2. \tag{3}$$

(b) Find the capacity of the Gaussian channel in terms of the provided parameters and state the joint distribution of (X_1, X_2) that achieves it.

15. **Complex Gaussian Channel.** The following question focuses on the complex Gaussian point-to-point communication channel.

Let $Z = U + iV$ be a complex Gaussian RV in the sense that U and V are independent and identically distributed real Gaussian RVs. In the following sections $Z \sim \mathcal{CN}(0, \gamma)$, where,

$$0 = \mathbb{E}[Z] \quad ; \quad \gamma = \mathbb{E}[|Z|^2], \tag{4}$$

and γ is a given positive parameter.

(a) Find the distribution of the random vector $(\Re\{Z\}, \Im\{Z\})^T = (U, V)^T$.

(b) Is it true that $h(Z) = h(U, V)$? Justify your answer.

(c) Calculate $h(Z)$.

(d) What is the maximum of the differential entropy over all centered complex RVs $Z = U + iV$ with $\mathbb{E}[U^2] + \mathbb{E}[V^2] \leq \gamma$? Which distribution of Z achieves this maximum?

Finally, consider the complex Gaussian channel illustrated in Fig. 6.

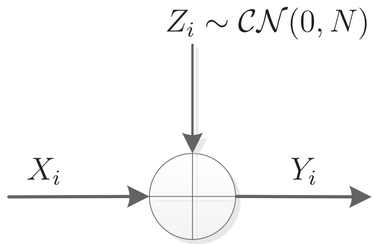


Figure 6: A complex Gaussian point-to-point channel.

The output of the channel for every time $i \in \{1, \ldots, n\}$ is give by,

$$Y_i = X_i + Z_i, \tag{5}$$

where X_i, for $i \in \{1, \ldots, n\}$, is a complex channel input and Z_i is distributed i.i.d according to $Z_i \sim \mathcal{CN}(0, N)$. The input signal is bound to an average power constraint,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} |X_i|^2\right] \leq P. \tag{6}$$

The capacity of the complex Gaussian channel is given by,

$$C = \max_{f_X:\ \mathbb{E}[|X|^2] \leq P} I(X; Y). \tag{7}$$

(e) Express the capacity in (7) in terms of the parameters of the problem (i.e., as a function of P and N) and state the distribution of the complex input RV X that achieves the maximum.

(f) Compare the result to the capacity of the real point-to-point channel. Explain the difference.

16. **Fast fading Gaussian channel**:
Consider a Gaussian channel given by $Y_i = G_i X_i + Z_i$, where $Z_i \overset{i.i.d}{\sim} \mathcal{N}(0, N)$

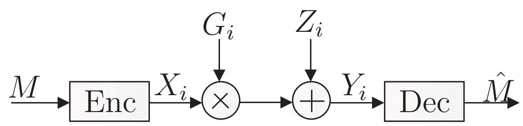


Figure 7: Fast fading Gaussian channel

and $G_i \overset{i.i.d}{\sim} P_G(g)$.
The gains and noise are independent, i.e., $\{Z_i\} \perp\!\!\!\perp \{G_i\}$, and

$$P_G(g) = \begin{cases} 0.5 & \text{if } g = 1 \\ 0.5 & \text{if } g = 2 \end{cases}$$

(a) Assume that the states are known at the decoder only, and there is an input constraint P.

i. What is the capacity formula?
ii. Find the optimal inputs distribution in the formula you gave.
iii. Compute the capacity as a function of N and P.

(b) Now the states are known both to the encoder and decoder, and the input constraint is P.

i. What is the capacity formula?
ii. Compute the capacity as a function of N and P.
You can write your answer as an optimization problem.

(c) Assume

$$P_G(g) = \begin{cases} 0.5 & \text{if } g = 0 \\ 0.5 & \text{if } g = 1 \end{cases}.$$

Repeat 16b.