

Solutions to Homework Set #4
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 σ_i^2 , $i = 1, 2$.

Solution: Differential entropy.

(a)

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

(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 $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

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

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.

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

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

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

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

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

Hence

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

- (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 $X \rightarrow Y \rightarrow Z$ 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 . Similarly 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)

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

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

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

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

Solution: Output power constraint.

The output power constraint $EY^2 \leq P$ is equivalent to the input power constraint

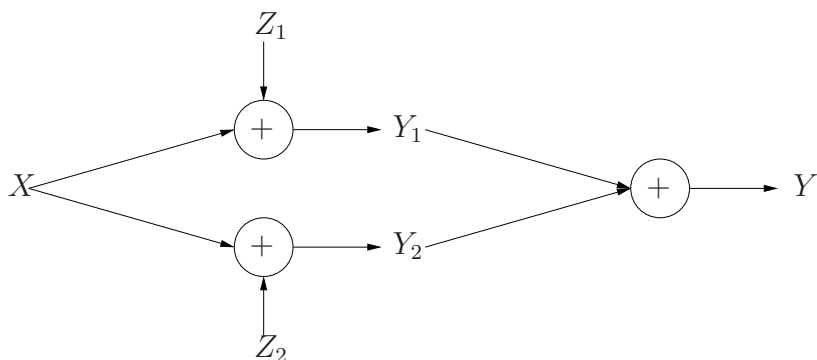
$$E(X + Z)^2 = EX^2 + EZ^2 = EX^2 + \sigma^2 \leq P,$$

that is, $EX^2 \leq P - \sigma^2$. 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 $\rho = 0$,

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

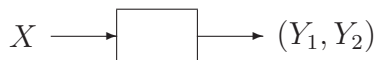
When $\rho = 1$,

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

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) $\rho = 1$.

(b) $\rho = 0$.

(c) $\rho = -1$.

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 $X \rightarrow (Y_1 + Y_2)$.

(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 $X \rightarrow Y_1 + Y_2$.

7. Diversity System

For the following system, a message $W \in \{1, 2, \dots, 2^{nR}\}$ is encoded into two symbol blocks $X_1^n = (X_{1,1}, X_{1,2}, \dots, X_{1,n})$ and $X_2^n = (X_{2,1}, X_{2,2}, \dots, X_{2,n})$ 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 $\arg \max P_{X_1, X_2} I(X_1, X_2; Y)$.)

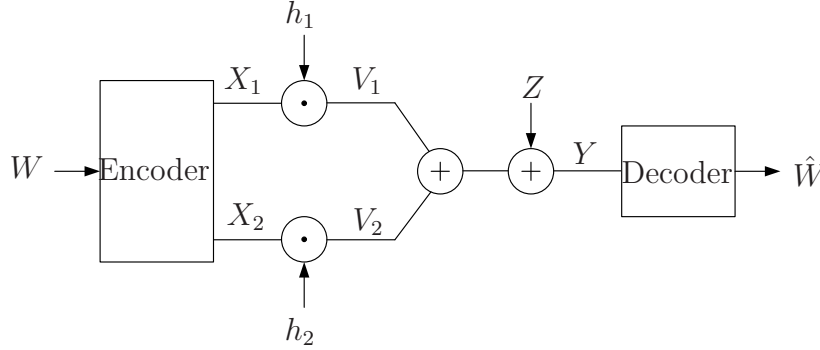


Figure 1: The communication model

- (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)

$$Y = h_1X_1 + h_2X_2 + Z$$

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:

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

(i) - Z is independent of X_1, X_2 .

(ii) - Cauchy-Schwarz inequality. Where $X_1 = \alpha X_2$, $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(0, K)$ 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 with 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 constraint 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.

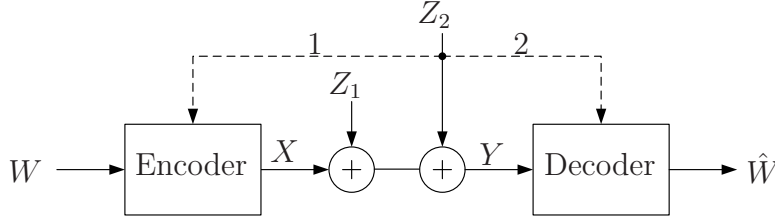


Figure 2: Two noise sources

- Find the channel capacity for the case in which the noise is not known to either side (lines 1 and 2 are disconnected from the encoder and the decoder).
- Find the capacity for the case that the noise Z_2 is known to the encoder and decoder (lines 1 and 2 are connected 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 larger than $\frac{1}{2} \log(1 + \frac{P}{\sigma_1^2})$?)
- Find the capacity for the case that the noise Z_2 is known only to the decoder. (line 1 is disconnected from the encoder and line 2 is connected 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

- Since the noise is not known 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:

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

- (c) Same as in (b), 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, i.e., 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 $P = \sigma_1^2 - \sigma_2^2$.

10. **Blahut-Arimoto's algorithm and KKT conditions** 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 fixed 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 be 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 be 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.

Solutions

- (a) Since $I(X; Y) = H(X) - H(X|Y)$, the answer is obvious.
- (b) Recall, that the Log-Sum inequality is

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}.$$

Hence

$$\begin{aligned} & (\lambda p_1(x) + (1 - \lambda)p_2(x)) \log \frac{\lambda p_1(x) + (1 - \lambda)p_2(x)}{\lambda p_1(x|y) + (1 - \lambda)p_2(x|y)} \\ & \leq \lambda p_1(x) \log \frac{p_1(x)}{p_1(x|y)} + (1 - \lambda)p_2(x) \log \frac{p_2(x)}{p_2(x|y)}. \end{aligned}$$

Taking the reciprocal of the logarithms yields

$$\begin{aligned} & (\lambda p_1(x) + (1 - \lambda)p_2(x)) \log \frac{\lambda p_1(x|y) + (1 - \lambda)p_2(x|y)}{\lambda p_1(x) + (1 - \lambda)p_2(x)} \\ & \geq \lambda p_1(x) \log \frac{p_1(x|y)}{p_1(x)} + (1 - \lambda)p_2(x) \log \frac{p_2(x|y)}{p_2(x)}. \end{aligned}$$

Multiplying by $p(y|x)$ and summing over all x, y , and letting $\mathcal{I}(p(x), p(x|y))$ be the mutual information as in (a), we obtain

$$\begin{aligned} & \mathcal{I}(\lambda p_1(x) + (1 - \lambda)p_2(x), \lambda p_1(x|y) + (1 - \lambda)p_2(x|y)) \\ & \geq \lambda \mathcal{I}(p_1(x), p_1(x|y)) + (1 - \lambda) \mathcal{I}(p_2(x), p_2(x|y)). \end{aligned}$$

(c) Define the lagrangian

$$L = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)} + \mu \left(\sum_x p(x) - 1 \right),$$

and differentiate over $p(x)$. Solving $\frac{\partial L}{\partial p(x)} = 0$ provides us with

$$p(x) = \frac{\prod_y p(x|y)^{p(y|x)}}{\sum_x \prod_y p(x|y)^{p(y|x)}}.$$

(d) Define the lagrangian

$$J = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)} + \mu(y) \left(\sum_x p(x|y) - 1 \right),$$

and differentiate over $p(x|y)$. Solving $\frac{\partial J}{\partial p(x|y)} = 0$ provides us with

$$p(x|y) = \frac{p(x)p(y|x)}{\sum_x p(x)p(y|x)}.$$

(e) The expression for $p(x|y)$ is the one that corresponds to $p(x)$, and hence maximizing over $p(x)$, $p(x|y)$ is the same as over $p(x)$ alone.

11. Fading channel.

Consider an additive noise fading channel

$$\begin{array}{c} \begin{array}{ccccccc} & & V & & Z & & \\ & & \downarrow & & \downarrow & & \\ X & \longrightarrow & \odot & \longrightarrow & \oplus & \longrightarrow & Y \end{array} \\ Y = XV + Z, \end{array}$$

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 $I(U; R|S) < I(U; R)$.

Solution: Fading channel

- (a) We may show the inequality as follows:

$$\begin{aligned} I(X; Y|V) &= h(X|V) - h(X|Y, V) \\ &= h(X) - h(X|Y, V) \end{aligned} \tag{7}$$

$$\begin{aligned} &\geq h(X) - h(X|Y) \\ &= I(X; Y) \end{aligned} \tag{8}$$

where (7) follows from the independence of X and V , and (8) is true because conditioning reduces entropy.

- (b) There are many examples for this case. For instance, if we consider two cascaded BSCs where the input to the channel is U , the output of the first channel is R and the output of the second channel is S , the Markov chain $U - R - S$ holds from the data processing inequality. Now we know that

$$I(U; R) \stackrel{(a)}{=} I(U; R, S) \tag{9}$$

$$\stackrel{(b)}{=} I(U; S) + I(U; R|S) \tag{10}$$

$$\geq I(U; R|S) \tag{11}$$

where (a) follows from the Markov chain and (b) follows since mutual information is always non-negative.

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.

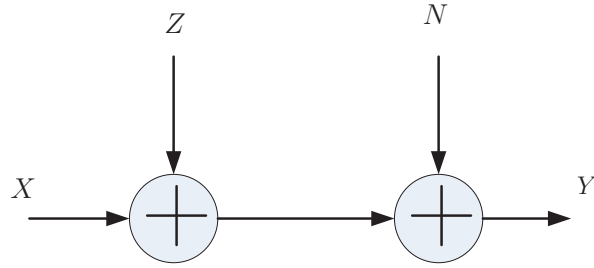


Figure 3: Additive Gaussian noise channel.

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

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.

Solution: Additive Gaussian channel where the noise might be a relay

See attached file.

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

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

Solution This claim is **false** since the mutual information term represents a Gaussian channel with no noise. For this channel the capacity is infinite. Where the entropy of X is bounded by the the entropy of a gaussian RV which is $\frac{1}{2} \log(2\pi e)\sigma_x^2$.

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 $P_{Y_1 Y_2 | X_1 X_2}$. A message M is randomly and uniformly chosen from the message set $\mathcal{M} = \{1, 2, \dots, 2^{nR}\}$ and is to be transmitted from the encoder to the decoder in a lossless manner (as defined in class).

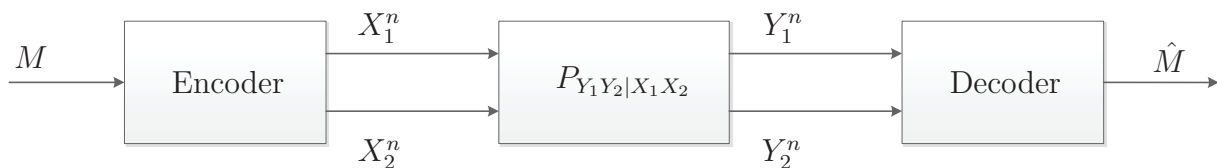


Figure 4: Two antenna point-to-point DMC.

(a) What is the capacity of the channel?

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

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

$$Y_{1,i} = X_{1,i} + Z_1, \tag{12}$$

$$Y_{2,i} = X_{1,i} + X_{2,i} + Z_1 + Z_2, \tag{13}$$

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 $Z_2 \sim \mathcal{N}(0, N_2)$. The input signals are bound to an average power constraints,

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_{1,i}^2 \right] \leq P_1 \quad ; \quad \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_{2,i}^2 \right] \leq P_2. \tag{14}$$

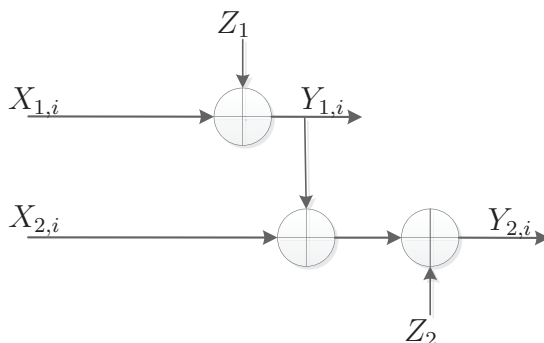


Figure 5: A Gaussian two antenna point-to-point DMC.

- (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.

Solution:

- (a) Let us denote the input pair (X_1^n, X_2^n) by \tilde{X}^n and the output pair (Y_1^n, Y_2^n) by \tilde{Y}^n . An equivalent channel to the one considered in this question is the point-to-point DMC for which $(\tilde{X}^n, \tilde{Y}^n)$ serve as the channel's input and output sequences, respectively, and the channel transition matrix is $P_{\tilde{Y}|\tilde{X}}$. Recalling that the point-to-point channel capacity is given by $\max_{P_{\tilde{X}}} I(\tilde{X}; \tilde{Y})$, and substituting $\tilde{X}^n = (X_1^n, X_2^n)$ and $\tilde{Y}^n = (Y_1^n, Y_2^n)$ we obtain:

$$C = \max_{P_{X_1 X_2}} I(X_1, X_2; Y_1, Y_2). \quad (15)$$

- (b) First note that Y_2 can be rewritten as $Y_2 = Y_1 + X_2 + Z_2$. Now, we upper

bound the capacity as:

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&= h(Y_1) - h(Y_1|X_1, X_2) + h(Y_2|Y_1) - h(Y_2|X_1, X_2, Y_1) \\
&\stackrel{(a)}{=} h(Y_1) - h(Z_1) + h(Y_1 + X_2 + Z_2|Y_1) - h(Y_1 + X_2 + Z_2|X_1, X_2, Y_1) \\
&\stackrel{(b)}{=} h(Y_1) - h(Z_1) + h(X_2 + Z_2|Y_1) - h(Z_2) \\
&\stackrel{(c)}{\leq} h(Y_1) - h(Z_1) + h(X_2 + Z_2) - h(Z_2) \\
&= h(Y_1) - \frac{1}{2} \log(2\pi e N_1) + h(Y_2) - \frac{1}{2} \log(2\pi e N_2) \\
&\stackrel{(d)}{\leq} \frac{1}{2} \log(2\pi e(P_1 + N_1)) - \frac{1}{2} \log(2\pi e N_1) + \frac{1}{2} \log(2\pi e(P_2 + N_2)) - \frac{1}{2} \log(2\pi e N_2) \\
&= \frac{1}{2} \log \left(\prod_{i=1}^2 (P_i + N_i) \right)
\end{aligned}$$

where:

- (a) follows from the definitions of Y_1 and the fact that Z_1 is independent of X_1 ;
- (b) follows from the fact that Z_2 is independent of (X_1, X_2, Z_1) and therefore it is independent of (X_1, X_2, Y_1) ;
- (c) follows from the fact that conditioning reduces entropy;
- (d) follows by the maximum of differential entropy property.

This upper bound is achieved by choosing (X_1, X_2) to be jointly Gaussian RVs with the following distribution,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \right) \quad (16)$$

This distribution achieves (c) with an equality since by this choice we get that $Y_1 = X_1 + Z_1$ and $X_2 + Z_2$ are independent. Whereas (d) is achieved with equality since by this choice Y_1 and Y_2 are Gaussian RVs with variances $P_1 + N_1$ and $P_2 + N_2$, respectively (which achieves the maximum of entropy)

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,

$$0 = \mathbb{E}[Z] \quad ; \quad \gamma = \mathbb{E}[|Z|^2], \quad (17)$$

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.

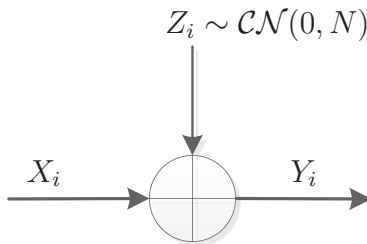


Figure 6: A complex Gaussian point-to-point channel.

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

$$Y_i = X_i + Z_i, \quad (18)$$

where X_i , for $i \in \{1, \dots, n\}$, 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,

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n |X_i|^2 \right] \leq P. \quad (19)$$

The capacity of the complex Gaussian channel is given by,

$$C = \max_{f_X: \mathbb{E}[|X|^2] \leq P} I(X; Y). \quad (20)$$

- (e) Express the capacity in (20) 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.

Solution: Complex Gaussian Channel.

See attached file.

16. True or False

Copy each relation to your notebook and write **true** or **false**. Then, if it's true, prove it. If it is false give a counterexample or prove that the opposite is true.

- (a) Let X, Y be a pair of random variables jointly distributed according to $P_{X,Y}$. For every $y \in \mathcal{Y}$ we have

$$H(X|Y = y) \leq H(X).$$

- (b) Consider the channel in Fig. 7 where Z is a Gaussian noise with power N and a is a deterministic constant. There is a power constraint on X such that $E[X^2] \leq P$. The capacity between X and Y is denoted by C . Is $C = \frac{1}{2} \log(1 + \frac{P}{N})$?

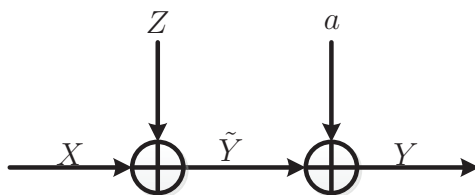


Figure 7: Cascaded AWGN channel.

Solutions

a) False. For example:

$Y \setminus X$	0	1
0	0.25	0.25
1	0.5	0

For this distribution we can calculate $H_b(0.25) = H(X) < H(X|Y = 0) = 1$.

b) True. For the converse, if multiplying by a would increase the capacity then we would use it in the point to point channel. To achieve, we divide by a and apply the decoding procedure as in the point to point channel.

c) True. This can be done as in the previous question. Another approach is by noting that the function $(\cdot)^3$ is a bijective function and therefore by having Y we indeed can recover \tilde{Y} and:

$$\begin{aligned}
 I(X; \tilde{Y}) &= H(\tilde{Y}) - H(\tilde{Y}|X) \\
 &= H(\tilde{Y}, Y) - H(Y, \tilde{Y}|X) \\
 &= H(Y) - H(Y|X) \\
 &= I(X; Y)
 \end{aligned}$$

and thus $\max_{p(x)} I(X; Y) = \max_{p(x)} I(X; \tilde{Y}) = \frac{1}{2} \log(1 + SNR)$.

17. **Parallel Gaussian channels** Consider a channel consisting of 2 parallel Gaussian channels, with inputs X_1 and X_2 and outputs given by

$$\begin{aligned}
 Y_1 &= X_1 + Z_1, \\
 Y_2 &= X_2 + Z_2.
 \end{aligned}$$

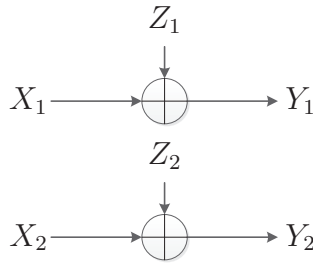


Figure 8: Parallel Gaussian channels.

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 $E[X^2] \leq P$. 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 $E[X_1^2] + E[X_2^2] \leq 2P$. 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 $E[X_1^2] + E[X_2^2] \leq 2P$. The outputs of the channels are given by

$$\begin{aligned} Y_1 &= aX + Z_1, \\ Y_2 &= bX + 2Z_1. \end{aligned}$$

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: Parallel Gaussian channels.

- (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

$$C = \frac{1}{2} \log(1 + \text{SNR}). \quad (21)$$

In our case,

$$\begin{aligned} \text{SNR} &= \frac{E[(X_1 + X_2)^2]}{E[(Z_1 + Z_2)^2]} \\ &\leq \frac{4P}{\sigma_1^2 + \sigma_2^2}. \end{aligned} \quad (22)$$

So, the capacity of this channel is given by

$$C = \frac{1}{2} \log \left(1 + \frac{4P}{\sigma_1^2 + \sigma_2^2} \right). \quad (23)$$

(b) Let

$$\begin{aligned} Y_1 &= aX + Z_1 \\ Y_2 &= bX + Z_2, \end{aligned} \tag{24}$$

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

$$\begin{aligned} I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2) \\ &= h(Y_1, Y_2) - \frac{1}{2} \log 2\pi e \sigma_1^2 \sigma_2^2, \end{aligned} \tag{25}$$

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

$$\Sigma_{Y_1, Y_2} = \begin{pmatrix} a^2P + \sigma_1^2 & abP \\ abP & b^2P + \sigma_2^2 \end{pmatrix}. \tag{26}$$

Then,

$$\begin{aligned} |\Sigma_{Y_1, Y_2}| &= (a^2P + \sigma_1^2)(b^2P + \sigma_2^2) - a^2b^2P^2 \\ &= P(a^2\sigma_2^2 + b^2\sigma_1^2) + \sigma_1^2\sigma_2^2 \\ &\leq P(a^2\sigma_2^2 + (2 - a^2)\sigma_1^2) + \sigma_1^2\sigma_2^2 \\ &= a^2P(\sigma_2^2 - \sigma_1^2) + (2P + \sigma_2^2)\sigma_1^2, \end{aligned} \tag{27}$$

and

$$h(Y_1, Y_2) \leq \log 2\pi e + \frac{1}{2} \log[a^2P(\sigma_2^2 - \sigma_1^2) + (2P + \sigma_2^2)\sigma_1^2]. \tag{28}$$

We can now see that, since $\sigma_1^2 < \sigma_2^2$, the expression in (28) 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

$$C_b = \frac{1}{2} \log \left(1 + \frac{2P}{\sigma_1^2} \right) \tag{29}$$

- (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

$$Y = X. \quad (30)$$

Thus, the capacity is infinite.

18. Fast fading Gaussian channel:

Consider a Gaussian channel given by $Y_i = G_i X_i + Z_i$, where $Z_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, N)$

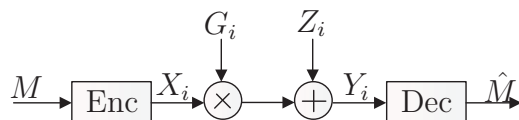


Figure 9: Fast fading Gaussian channel

and $G_i \stackrel{i.i.d.}{\sim} P_G(g)$.

The gains and noise are independent, i.e., $\{Z_i\} \perp \{G_i\}$, 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 .
- What is the capacity formula?
 - Find the optimal inputs distribution in the formula you gave.
 - 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 .
- What is the capacity formula?
 - 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 18b.

Solution: Fast fading Gaussian channel

- (a) i. As we saw in the lectures, the capacity is given by

$$C_1 = \sup_{P_X} I(X; Y|G) \quad (31)$$

where the maximum is taken over all X -distributions such that $E(X^2) \leq P$.

- ii. We show in the next item that the optimal input is $X \sim \mathcal{N}(0, P)$.
 iii. The states are known only to the decoder, and thus we may assume that $X \perp\!\!\!\perp G$. We have

$$\begin{aligned} I(X; Y|G) &= h(Y|G) - h(Y|X, G) \\ &= h(Y|G) - h(Y - GX|X, G) \\ &= h(Y|G) - h(Z|X, G) \\ &\stackrel{(a)}{=} h(Y|G) - h(Z) \\ &= P(G = 1) \cdot h(Y|G = 1) + P(G = 2) \cdot h(Y|G = 2) - h(Z) \\ &= \frac{1}{2} \cdot h(X + Z|G = 1) + \frac{1}{2} \cdot h(2X + Z|G = 2) - h(Z) \\ &\stackrel{(a)}{=} \frac{1}{2} \cdot h(X + Z) + \frac{1}{2} \cdot h(2X + Z) - \frac{1}{2} \log(2\pi eN) \end{aligned} \quad (32)$$

where (a) follows from the fact that $Z \perp\!\!\!\perp (X, G)$. Now, the maximal entropy lemma implies

$$h(X + Z) \leq \frac{1}{2} \log(2\pi e(P + N)), \quad (33)$$

$$h(2X + Z) \leq \frac{1}{2} \log(2\pi e(4P + N)), \quad (34)$$

with *equality* if and only if $X \sim \mathcal{N}(0, P)$. Thus,

$$I(X; Y|G) \leq \frac{1}{2} \log \left(\frac{\sqrt{(P + N)(4P + N)}}{N} \right), \quad (35)$$

again, with *equality* if and only if $X \sim \mathcal{N}(0, P)$. Therefore

$$\begin{aligned} C_1 &= \max_{P_X} I(X; Y|G) \\ &= \frac{1}{4} \log \left(1 + \frac{P}{N} \right) + \frac{1}{4} \log \left(1 + \frac{4P}{N} \right). \end{aligned} \quad (36)$$

- (b) i. Again, as we saw in the lectures, the capacity in this case is given by

$$C_2 = \sup_{P_{X|G}} I(X; Y|G) \quad (37)$$

where the maximum is over all distributions $P_{X|G}$ which satisfy the power constraint.

- ii. As before, we get:

$$I(X; Y|G) = \frac{1}{2} \cdot h(X + Z|G = 1) + \frac{1}{2} \cdot h(2X + Z|G = 2) - \frac{1}{2} \log(2\pi eN). \quad (38)$$

Define the functional:

$$\text{var}(X|W = w) \triangleq \mathbb{E}(X^2|W = w). \quad (39)$$

Then, let $\text{var}(X|G = i) \triangleq P_i$ for $i = 1, 2$. By the power constraint, we have that $P_1 + P_2 \leq 2P$. Accordingly, due to the maximal entropy lemma, we get

$$h(X + Z|G = 1) \leq \frac{1}{2} \log(2\pi e(P_1 + N)) \quad (40)$$

$$h(2X + Z|G = 2) \leq \frac{1}{2} \log(2\pi e(4P_2 + N)), \quad (41)$$

where both inequalities are achieved if $X|G = i$ is Gaussian with variance P_i . Hence,

$$I(X; Y|G) \leq \frac{1}{2} \log \left(\frac{\sqrt{(P_1 + N)(P_2 + N)}}{N} \right). \quad (42)$$

Therefore, the capacity is given by

$$C_2 = \sup_{(P_1, P_2): P_1 + P_2 \leq 2P} \frac{1}{2} \log \left(\frac{\sqrt{(P_1 + N)(4P_2 + N)}}{N} \right) \quad (43)$$

$$= \sup_{(P_1, P_2): P_1 + P_2 \leq 2P} \left\{ \frac{1}{4} \log \left(1 + \frac{P_1}{N} \right) + \frac{1}{4} \log \left(1 + \frac{4P_2}{N} \right) \right\}. \quad (44)$$

Remark: This optimization problem can be solved by Water-filling.

(c) As before we will get

$$\begin{aligned} I(X; Y|G) &= \frac{1}{2} \cdot h(Z|G=0) + \frac{1}{2} \cdot h(X+Z|G=1) - \frac{1}{2} \log(2\pi eN) \\ &= \frac{1}{2} \cdot h(X+Z|G=1) - \frac{1}{4} \log(2\pi eN). \end{aligned} \quad (45)$$

Also,

$$h(X+Z|G=1) \leq \frac{1}{2} \log(2\pi e(P_1+N)). \quad (46)$$

Thus,

$$\begin{aligned} C_3 &= \sup_{(P_1): P_1 \leq 2P} \frac{1}{4} \log \left(\frac{P_1+N}{N} \right) \\ &= \frac{1}{4} \log \left(1 + \frac{2P}{N} \right). \end{aligned} \quad (47)$$

Intuition: To achieve (47) the transmitter will not send any data when $G=0$ (because in this case our information will be lost), but rather all the data will be transmitted when $G=1$ (and with power $2P$ to satisfy the power constraint). Now, when $G=1$, the channel reduces to a simple Gaussian channel, $Y=X+Z$, with signal to noise ratio of $2P/N$, namely, we can achieve a rate of

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

However, since with high probability, $G=1$ half of the time, we need to multiply the last result by half, and we get the quarter factor in (47).