Fundamentals of Information Theory

Homework 6

Name: Zhou Shouchen
Student ID: 2021533042

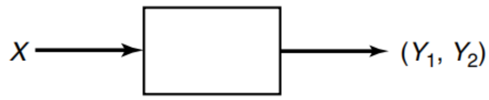
Due 23:59 (CST), Dec. 15, 2024

Problem 1

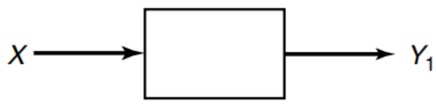
9.1 Channel with two independent looks at Y. Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X.

(a) Show that $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2)$.

(b) Conclude that the capacity of the channel



is less than twice the capacity of the channel



Solution

(a)

$$
\begin{aligned}
I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\
&= h(Y_1, Y_2) - h(Y_1|X) - h(Y_2|X) && \text{(Since } Y_1 \perp Y_2|X) \\
&= h(Y_1, Y_2) + I(X, Y_1) - h(Y_1) + I(X; Y_2) - h(Y_2) && \text{(Since } I(X; Y_i) = h(Y_i) - h(Y_i|X)) \\
&= I(X; Y_1) + I(X; Y_2) - (h(Y_1) + h(Y_2) - h(Y_1, Y_2)) \\
&= 2I(X; Y_1) - I(Y_1; Y_2) && \text{(Since } Y_1, Y_2 \text{ are identical distributed when given } X)
\end{aligned}
$$

(b) Suppose the channel capacity of the first channel is C_1 and the second channel is C_2. Then, we have

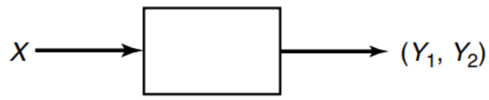
$$
\begin{aligned}
C_1 &= \max_{p(x)} \ I(X; Y_1, Y_2) \\
&= \max_{p(x)} \ 2I(X; Y_1) - I(Y_1; Y_2) \\
&\leq \max_{p(x)} \ 2I(X; Y_1) \qquad \text{(Since } I(Y_1, Y_2) \geq 0) \\
&= 2 \max_{p(x)} \ I(X; Y_1) \\
&= 2C_2
\end{aligned}
$$

If and only if $I(Y_1; Y_2) = 0$, i.e. $Y_1 \perp Y_2$, the equality holds.

So above all, we have proved that the capacity of the channel is less than twice the capacity of the channel.

Problem 2

9.2 Two-look Gaussian channel



Consider the ordinary Gaussian channel with two correlated looks at X, that is, $Y = (Y_1, Y_2)$, where

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2 \end{aligned}$$

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

$$C = \max_{p(x)} I(X; Y_1, Y_2)$$

$$\begin{aligned} I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2|X) \qquad (\text{since } Y_i = X + Z_i) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2) \qquad (\text{since } Z_i \perp X) \end{aligned}$$

Since $(Z_1, Z_2) \sim \mathcal{N}_2(0, K)$, and $|K| = N^2(1-\rho^2)$, we have

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

And since $Y_i = X + Z_i$, and since $X \perp Z_i$, so

$$\begin{aligned} \mathrm{Var}(Y_i) &= \mathrm{Var}(X + Z_i) = \mathrm{Var}(X) + \mathrm{Var}(Z_i) = \mathrm{Var}(X) + N \\ \mathrm{Cov}(Y_1, Y_2) &= \mathrm{Cov}(X + Z_1, X + Z_2) = \mathrm{Var}(X) + \rho N \end{aligned}$$

So the covariance matrix of (Y_1, Y_2) is

$$K_Y = \begin{bmatrix} \mathrm{Var}(Y_1) & \mathrm{Cov}(Y_1, Y_2) \\ \mathrm{Cov}(Y_1, Y_2) & \mathrm{Var}(Y_2) \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(X) + N & \mathrm{Var}(X) + \rho N \\ \mathrm{Var}(X) + \rho N & \mathrm{Var}(X) + N \end{bmatrix}$$

$$\Rightarrow |K_Y| = 2\mathrm{Var}(X) N (1-\rho) + N^2(1-\rho^2)$$

Since $\rho \in [-1, 1] \Rightarrow 1 - \rho \geq 0$, and since $\mathrm{Var}(X) \leq P$, so we can get that

$$|K_Y|_{\max} = 2PN(1-\rho) + N^2(1-\rho^2)$$

So

$$h(Y_1, Y_2) \leq \frac{1}{2} \log \left[(2\pi e)^2 |K_Y|_{\max}\right] = \frac{1}{2} \log \left[(2\pi e)^2 \left(2PN(1-\rho) + N^2(1-\rho^2)\right)\right]$$

So

$$\begin{aligned} C &= \max_{p(x), \mathrm{Var}(X) \leq P} I(X; Y_1, Y_2) \\ &= \frac{1}{2} \log \left(\frac{2PN(1-\rho) + N^2(1-\rho^2)}{N^2(1-\rho^2)} \right) \\ &= \frac{1}{2} \log \left(1 + \frac{2P}{N(1+\rho)} \right) \end{aligned}$$

If and only if $X \sim \mathcal{N}(0, P)$, and $(Y_1, Y_2) \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} P+N & P+\rho N \\ P+\rho N & P+N \end{bmatrix} \right)$,then C takes the maximum value among $I(X; Y_1, Y_2)$.

(a) $\rho = 1$:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right) \text{ bits per transmission}$$

(b) $\rho = 0$:

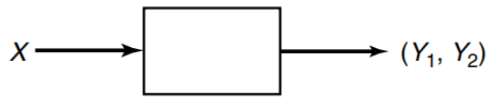
$$C = \frac{1}{2} \log \left(1 + \frac{P}{2N} \right) \text{ bits per transmission}$$

(c) $\rho = -1$:

$$C = +\infty \text{ bits per transmission}$$

Problem 3

9.5 Fading channel. Consider an additive noise fading channel



$$Y = XV + Z$$

where Z is additive noise, V is a random variable representing fading, and Z and V are independent of each other and of X. Argue that knowledge of the fading factor V improves capacity by showing that

$$I(X; Y \mid V) \geq I(X; Y)$$

Solution

From the chain rule of mutual information, we have

$$\begin{aligned} I(X;Y,V) &= I(X;Y) + I(X;V|Y) \\ &= I(X;V) + I(X;Y|V) \end{aligned}$$

Since $X \perp V$, so $I(X;V) = 0$. Therefore, we have

$$I(X;Y|V) = I(X;Y) + I(X;V|Y)$$

Since conditional mutual information is also a mutual information, which means that it has the same properties as mutual information, we have

$$I(X;V|Y) \geq 0$$

Therefore, we have

$$I(X;Y|V) \geq I(X;Y) \quad \square$$

Problem 4

9.6 Parallel channels and water-filling. Consider a pair of parallel Gaussian channels:

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

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 $\mathbb{E}\left(X_1^2 + X_2^2\right) \leq 2P$. 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?

Solution

From what we have learned about the parallel channels, we know that the capacity of the parallel channels is

$$C = \sum_{i=1}^{2} \frac{1}{2} \log\left(1 + \frac{P_i}{\sigma_i^2}\right)$$

Where $\sum\limits_{i=1}^{2} P_i \leq 2P$.

And with water-filling method, we have $P_i = \left(\nu - \sigma_i^2\right)_+$, $P_1 + P_2 \leq 2P$ and ν is the water level, $(\cdot)_+ = \max(\cdot, 0)$.

Since $\sigma_1^2 > \sigma_2^2$, so

1. When $\nu \in (\sigma_2^2, \sigma_1^2] \Rightarrow P_1 = 0, P_2 = \nu - \sigma_2^2$

$\Rightarrow 2P = P_1 + P_2 \in (0, \sigma_1^2 - \sigma_2^2)$

So when $0 \leq P \leq \dfrac{\sigma_1^2 - \sigma_2^2}{2}$, the channel behaves like a single channel with noise variance σ_2^2.

2. When $\nu > \sigma_1^2$, we have $P_1 = \nu - \sigma_1^2 > 0, P_2 = \nu - \sigma_2^2 > 0$

$\Rightarrow 2P = P_1 + P_2 > 2\nu - \sigma_1^2 - \sigma_2^2 > \sigma_1^2 - \sigma_2^2$

So when $P > \dfrac{\sigma_1^2 - \sigma_2^2}{2}$, the channel behaves like a pair of channels.

So when $P = \dfrac{\sigma_1^2 - \sigma_2^2}{2}$, the channel stops behaving like a single channel with noise variance σ_2^2, and begins behaving like a pair of channels.

Problem 5

9.9 Vector Gaussian channel. Consider the vector Gaussian noise channel

$$Y = X + Z$$

where $X = (X_1, X_2, X_3)\,, Z = (Z_1, Z_2, Z_3)\,, Y = (Y_1, Y_2, Y_3)$, $\mathbb{E}\|X\|^2 \le P$, and

$$Z \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}\right)$$

Find the capacity. The answer may be surprising.

Solution

$$I(X_1, X_2, X_3; Y_1, Y_2, Y_3) = h(Y_1, Y_2, Y_3) - h(Z_1, Z_2, Z_3)$$

Since this is a colored noise channel, we apply the eigenvalue decomposition to the noise covariance matrix K_Z. And we could find that its eigenvalues are $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 3$.

So

$$h(Z_1, Z_2, Z_3) = \frac{1}{2} \log\left[(2\pi e)^3 \cdot 0\right] = -\infty$$

And the conclusion is that the

$$|K_Y| = |K_X + K_Z| = \sum_{i=1}^{3} \left[(\nu - \lambda_i)_+ + \lambda_i\right] \ge 0 + 1 + 3 = 4$$

So we have

$$h(Y_1, Y_2, Y_3)_{\max} = \frac{1}{2} \log\left[(2\pi e)^3 \cdot |K_Y|\right] \ge \frac{1}{2} \log\left[(2\pi e)^3 \cdot 4\right] > 0$$

Thus, combine all above, we have

$$\begin{aligned} C &= \max_{p(x)} I(X_1, X_2, X_3; Y_1, Y_2, Y_3) \\ &= h(Y_1, Y_2, Y_3)_{\max} - h(Z_1, Z_2, Z_3) \\ &> 0 - (-\infty) \\ &= +\infty \end{aligned}$$

So above all, the capacity of this channel is $+\infty$ bits per transmission.

Problem 6

9.15 Discrete input, continuous output channel. Let $\Pr\{X = 1\} = p$, $\Pr\{X = 0\} = 1 - p$, and let $Y = X + Z$, where Z is uniform over the interval $[0, a], a > 1$, and Z is independent of X.

(a) Calculate

$$I(X;Y) = H(X) - H(X \mid Y)$$

(b) Now calculate $I(X;Y)$ the other way by

$$I(X;Y) = h(Y) - h(Y \mid X)$$

(c) Calculate the capacity of this channel by maximizing over p.

Solution

Since $Z \sim \text{Unif}(0, a)$, so $f_Z(z) = \frac{1}{a}, 0 \leq z \leq a$.

Then we can calculate $f_Y(y)$ by

$$f_Y(y) = \sum_{x=0,1} f_{Y|X}(y|X = x)p(X = x) = \sum_{x=0,1} f_Z(y - x)p(X = x)$$

1. When $x = 0$:

<1> when $y - x \in [0, a] \Rightarrow y \in [0, a] \Rightarrow f_Z(y - x) = \frac{1}{a}$

<2> otherwise, $f_Z(y - x) = 0$

So we have

$$f_{Y|X=0}(y) = \begin{cases} \frac{1}{a}, & 0 \leq y \leq a \\ 0, & \text{otherwise} \end{cases}$$

2. When $x = 1$:

<1> when $y - x \in [0, a] \Rightarrow y \in [1, a + 1] \Rightarrow f_Z(y - x) = \frac{1}{a}$

<2> otherwise, $f_Z(y - x) = 0$

So we have

$$f_{Y|X=1}(y) = \begin{cases} \frac{1}{a}, & 1 \leq y \leq a + 1 \\ 0, & \text{otherwise} \end{cases}$$

And since $a > 1$, and combined with what we have, we can get that:

$$f_Y(y) = \begin{cases} \frac{1-p}{a}, & 0 \leq y < 1 \\ \frac{1}{a}, & 1 \leq y \leq a \\ \frac{p}{a}, & a < y \leq a + 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Since $X \sim \text{Bern}(p)$, so $H(X) = H(p)$.

1. When $y \in [0, a]$, X is deterministic, i.e. $X = 0$.

2. When $y \in [a, a + 1]$, X is deterministic, i.e. $X = 1$.

So in these two cases, $H(X|Y = y) = 0$.

3. When $y \in [1, a]$:

Since $P(X = 0|Y = y) = \dfrac{f_{Y|X=0}(y)p(X = 0)}{f_Y(y)} = \dfrac{\frac{1}{a}(1 - p)}{\frac{1}{a}} = 1 - p$, i.e. $H(X|Y = y) = H(p)$.

So we can get that

$$
\begin{aligned}
H(X|Y) &= \int_0^{a+1} H(X|Y=y) f_Y(y)\ \mathrm{dy} \\
&= \int_0^{1} H(X|Y=y) f_Y(y)\ \mathrm{dy} + \int_1^{a} H(X|Y=y) f_Y(y)\ \mathrm{dy} + \int_a^{a+1} H(X|Y=y) f_Y(y)\ \mathrm{dy} \\
&= \int_0^{1} 0 \cdot f_Y(y)\ \mathrm{dy} + \int_1^{a} H(p) \cdot \frac{1}{a}\ \mathrm{dy} + \int_a^{a+1} 0 \cdot f_Y(y)\ \mathrm{dy} \\
&= \frac{a-1}{a} H(p)
\end{aligned}
$$

So above all, we can get that

$$
I(X;Y) = H(X) - H(X|Y) = H(p) - \frac{a-1}{a} H(p) = \frac{1}{a} H(p)
$$

(b) $I(X;Y) = h(Y) - h(Y|X)$

$$
\begin{aligned}
h(Y) &= -\int_0^{a+1} f_Y(y) \log f_Y(y)\ \mathrm{dy} \\
&= -\int_0^{1} \frac{1-p}{a} \log \frac{1-p}{a}\ \mathrm{dy} - \int_1^{a} \frac{1}{a} \log \frac{1}{a}\ \mathrm{dy} - \int_a^{a+1} \frac{p}{a} \log \frac{p}{a}\ \mathrm{dy} \\
&= -\frac{1-p}{a} \log \frac{1-p}{a} - (a-1) \frac{1}{a} \log \frac{1}{a} - \frac{p}{a} \log \frac{p}{a} \\
&= \frac{1}{a} H(p) + \log a
\end{aligned}
$$

Since $Y = X + Z$, and $X \perp Z$, $Z \sim \text{Unif}(0, a)$, so

$$
h(Y|X) = h(Z|X) = h(Z) = \log a
$$

So above all, we can get that

$$
I(X;Y) = h(Y) - h(Y|X) = \frac{1}{a} H(p)
$$

(c) Let C be the capacity of this channel, then

$$
\begin{aligned}
C &= \max_{p(x)}\ I(X;Y) \\
&= \max_{p(x)}\ \frac{1}{a} H(p) \\
&= \frac{1}{a} H\left(\frac{1}{2}\right) \\
&= \frac{1}{a}
\end{aligned}
$$

So above all, the channel capacity is $\frac{1}{a}$, when $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$ it achieves.