ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 21 Information Theory and Coding
Solutions to homework 10 December 4, 2012

PROBLEM 1. 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, $C = \frac{1}{2}\log(1 + P/N)$, 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,

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

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, and so is equivanent to having a noiseless channel.

Note that the capacity of the above channel in all cases is the same as the capacity of the channel $X \to Y_1 + Y_2$. This is not true in general.

Problem 2. (a) By the water-filling solution discussed in 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 $2P = \sigma_1^2 - \sigma_2^2$.

(b) Since we are interested in the asymptotics $P/\sigma_1^2 \longrightarrow \infty$ without loss of generality we assume the waterpourring level to be greater than σ_1^2. Hence $P_i = \lambda - \sigma_i^2$, $i = 1, 2$. It follows that

$$\begin{aligned} C_1(P) - C_2(P) &= \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right) \\ &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}\right) + \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right) \end{aligned}$$

Now

$$\begin{aligned} \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2}\right) &= \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2} + \frac{P - P_1}{\sigma_1^2}\right) \\ &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2} + \frac{P - P_1}{\sigma_1^2}\right) \end{aligned}$$

and similarly

$$\frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right) = \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2} + \frac{P - P_2}{\sigma_2^2}\right) = \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2} - \frac{P - P_1}{\sigma_2^2}\right) .$$

We then deduce that

$$\begin{aligned} C_1(P) - C_2(P) &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}\right) + \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}\right) - \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2} + \frac{P - P_1}{\sigma_1^2}\right) - \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2} - \frac{P - P_1}{\sigma_2^2}\right) \\ &= -\frac{1}{2}\log\left(1 + \frac{P - P_1}{\lambda}\right) - \frac{1}{2}\log\left(1 - \frac{P - P_1}{\lambda}\right) . \end{aligned}$$

We conclude by showing that $\frac{P-P_1}{\lambda}$ tends to zero as P/σ_1^2 tends to infinity. Since $P_i = \lambda - \sigma_i^2$, $i = 1, 2$, and since $2P = 2\lambda - \sigma_1^2 - \sigma_2^2$, we have that

$$\begin{aligned} \frac{P - P_1}{\lambda} &= \frac{P - (\lambda - \sigma_1^2)}{\lambda} \\ &= \frac{2(\sigma_2^2 - \sigma_1^2)}{2P + \sigma_1^2 + \sigma_2^2} . \end{aligned}$$

By assumption P/σ_1^2 tends to infinity, and because $\sigma_2^2 < \sigma_1^2$, we have that $\frac{2(\sigma_2^2 - \sigma_1^2)}{2P + \sigma_1^2 + \sigma_2^2}$ tends to zero as P/σ_1^2 tends to infinity. This gives the desired result.

Problem 3. (a) All rates less than $\frac{1}{2}\log_2(1 + \frac{P}{\sigma^2})$ are achievable.

(b) The new noise $Z_1 - Z_2$ has zero mean and variance $\mathrm{E}((Z_1 - Z_2)^2) = 2\sigma^2 - 2\rho\sigma^2$. Therefore, all rates less than $\frac{1}{2}\log_2(1 + \frac{P}{2(1-\rho)\sigma^2})$ are achievable.

(c) The capacity is $C = \max I(X; Y_1, Y_2) = \max(h(Y_1, Y_2) - h(Z_1, Z_2)) = \frac{1}{2}\log_2(1 + \frac{P}{2(1-\rho)\sigma^2})$. This shows that the scheme used in (b) is a way to achieve capacity.

Problem 4.

a) For all $x \in \mathcal{X}$, since $P(x^*) \geq P(x)$, then $\log(\frac{1}{P(x)}) \geq \log(\frac{1}{P(x^*)})$. Hence,

$$H(X) = \sum_{x \in \mathcal{X}} P(x) \log(\frac{1}{P(x)}) \geq (\log(\frac{1}{P(x^*)})) \sum_{x \in \mathcal{X}} P(x) = \log(\frac{1}{P(x^*)}).$$

b) As we have seen in class, we define

$$Z = \begin{cases} 0, & \hat{X} = X \\ 1, & \hat{X} \neq X \end{cases}$$

Then, $H(X, Z|Y) = H(X|Y) + H(Z|X, Y) = H(Z|Y) + H(X|Z, Y)$.
Moreover, $H(Z|X, Y) \leq H(Z|X, g(Y) = \hat{X}) = 0$ and $H(Z|Y) \leq H(P_e)$. Therefore,

$$H(X|Y) \leq H(P_e) + H(X|Z, Y) = H(P_e) + P_e H(X|Z = 1, Y) \leq H(P_e) + P_e \log(|\mathcal{X}| - 1).$$

c) Assume that $\hat{x} = g(y)$ for some observation y. This means that $P(\hat{x}|y) \geq P(x|y)$ for all $x \in \mathcal{X}$. According to part (a), $H(X|Y = y) \geq \log(\frac{1}{P(\hat{x}|y)})$. Combining these, we obtain

$$P(\hat{x}|y) \geq e^{-H(X|Y=y)}.$$

On the other hand, $P_e = P\{\hat{X} \neq X\} = 1 - P\{\hat{X} = X\}$. So,

$$\begin{aligned} P_e &= 1 - \sum_{y \in \mathbf{Y}} P(Y = y) P(\hat{x}|y) \\ &\leq 1 - \sum_{y \in \mathcal{Y}} P(Y = y) e^{-H(X|Y=y)} \\ &\leq 1 - e^{-\sum_{y \in \mathcal{Y}} P(Y=y) H(X|Y=y)} \\ &= 1 - e^{H(X|Y)}. \end{aligned}$$

where we used the hint in the last inequality.

Problem 5. This property is also known as “martingale processes have orthogonal increments”. You can show the result by using:

$$E[E[X_i|X^{i-1}]] = E[X_i].$$

Problem 6. The $-$ channel is also a binary symmetric channel with cross-over probability given by $\epsilon^- = 2\epsilon - 2\epsilon^2$. For the $+$ channel, using the formula

$$P(y_1 y_2 u_1 | u_2) = \frac{1}{2} P(y_1 | u_1 \oplus u_2) P(y_2 | u_2),$$

we can compute the transition probabilities. Assume first $u_2 = 0$, then

$$
\begin{aligned}
P(000|0) &= \frac{1}{2}(1-\epsilon)^2 \\
P(001|0) &= \frac{1}{2}\epsilon(1-\epsilon) \\
P(010|0) &= \frac{1}{2}(1-\epsilon)\epsilon \\
P(011|0) &= \frac{1}{2}\epsilon^2 \\
P(100|0) &= \frac{1}{2}\epsilon(1-\epsilon) \\
P(101|0) &= \frac{1}{2}(1-\epsilon)^2 \\
P(111|0) &= \frac{1}{2}(1-\epsilon)\epsilon
\end{aligned}
$$

You can compute the result similarly for $u_2 = 1$.