Homework Set #5

Due 3rd December 2022, before 11:59pm.
Submit your solutions to Gradescope with Entry Code:**57DN5B**

Problem 1 (Capacity per unit cost)

Consider a DMC with cross-over probability $W_{Y|X}(\cdot|\cdot)$ with input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$. It costs $c(x)$ to send symbol $x \in \mathcal{X}$ over the channel. Assume that $c(x) > 0$, $\forall x \in \mathcal{X}$.

For some $x \in \mathcal{X}$, let us define the information divergence between $W_{Y|X=x}$ (which we write as $W_{Y|X}$ for shorthand) and P_Y as

$$D(W_{Y|X} \parallel P_Y) \triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|x) \log \frac{W_{Y|X}(y|x)}{P_Y(y)}.$$

Note that $D(W_{Y|X} \parallel P_Y)$ is still a function of $x \in \mathcal{X}$.

(a) Given any distribution P_Y, show that for any choice of input distribution $\tilde{P}_X(x)$,

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel P_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \max_{x \in \mathcal{X}} \frac{D(W_{Y|X} \parallel P_Y)}{c(x)}$$

Hint: You may use the following fact. For $a, b, c, d > 0$ with $\frac{a}{b} \leq \frac{c}{d}$, we always have $\frac{a}{b} \leq \frac{a+c}{b+d} \leq \frac{c}{d}$.

(b) Let $\tilde{P}_X$ and P_X be arbitrary input distributions and let $\tilde{P}_Y(y) = \sum_{x \in \mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x)$ and $P_Y(y) = \sum_{x \in \mathcal{X}} P_X(x) W_{Y|X}(y|x)$ be the resulting output distributions. Show that

$$\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel P_Y) - \sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y) \geq 0,$$

indicating when it is satisfied with equality, and then conclude that

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \max_{x \in \mathcal{X}} \frac{D(W_{Y|X} \parallel P_Y)}{c(x)}$$

Hint: Use properties of information divergence (Kullback-Leibler distance) or Jensen's inequality.

(c) Suppose we are given an input distribution $P_X^*(x)$ such that

$$\frac{D(W_{Y|X} \parallel P_Y^*)}{c(x)} \leq \lambda, \ \forall x \in \mathcal{X}$$

and

$$\frac{D(W_{Y|X} \parallel P_Y^*)}{c(x)} = \lambda, \ \forall x : P_X^*(x) > 0$$

where $P_Y^*(y) = \sum_x P_X^*(x) W_{Y|X}(y|x)$.

Using part (b) show that for any $\tilde{P}_X(x)$, and $\tilde{P}_Y(y) = \sum_{x\in\mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x)$,

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) c(x)} \leq \lambda$$

with equality if and only if $\tilde{P}_Y(y) = P_Y^*(y)$.

(d) Using the result found in (c), what can you conclude about the optimizing input distribution for finding the *capacity per unit cost* C_{cost}, defined as,

$$C_{\text{cost}} = \max_{P_X(x)} \frac{I(X;Y)}{\mathbb{E}[c(X)]}.$$

Problem 2 (Maximal differential entropy distributions)

(a) Suppose we have

$$Z = Z_1 + Z_2,$$
$$\mathbb{E}(Z_1^2) \leq \sigma^2, \mathbb{E}(Z_2^2) \leq \sigma^2,$$

and Z_1 and Z_2 are independent of each other. What is the maximal differential entropy of Z, and what joint distribution of Z_1 and Z_2 achieves this?

(b) Now suppose

$$Z = \sum_{i=1}^{L} Z_i, \tag{1}$$
$$\text{where } \mathbb{E}(Z_i^2) \leq \sigma^2, \forall i$$

Let $\{Z_i\}$ be independent of each other. What is the maximal differential entropy of Z, and what joint distribution of $\{Z_i\}$ achieves this?

(c) Now suppose we still have (1), and $\{Z_i\}$ can be dependent. What is the maximal differential entropy of Z, and what joint distribution of $\{Z_i\}$ achieves this?

Problem 3 (Gaussian channel with multiple looks)

A transmitter is communicating to a receiver over Gaussian channels.

(a) Consider two noisy versions of the transmitted signal X are received as shown in Fig. 1, where

$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2$$

where Z_1, Z_2 are independent Gaussian noises both of variance σ^2, *i.e.*, $Z_1, Z_2 \sim \mathcal{N}(0, \sigma^2)$. The transmit signal is constrained to have an average power of P. Find the capacity of this channel and describe how do you achieve it.

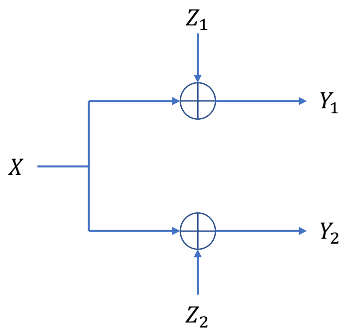

Figure 1: Two look Gaussian channels

(b) Now, suppose we have two independent transmitters available as shown in Fig. 2, from which we can transmit signals X_1 and X_2 respectively. Also, there is an overall power constraint, *i.e.,* the total power from X_1 and X_2 is constrained to be P. Suppose the transmitted signals are received through four channels with noise Z_1, Z_2, Z_3, and Z_4, *i.e.*,

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= X_1 + Z_2 \\ Y_3 &= X_2 + Z_3 \\ Y_4 &= X_2 + Z_4 \end{aligned}$$

As before Z_i are all independent of each other and Gaussian. We further assume that $Z_1, Z_2 \sim \mathcal{N}(0, \sigma_1^2)$, and $Z_3, Z_4 \sim \mathcal{N}(0, \sigma_2^2)$. Find the optimization problem for the channel capacity.

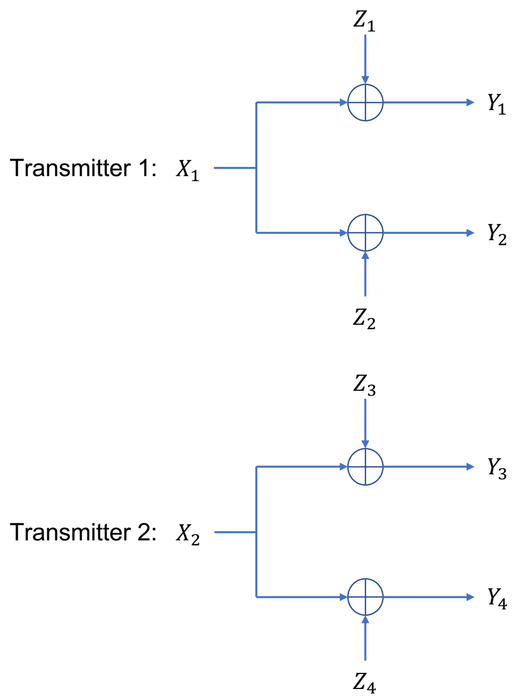

Figure 2: Two look Parallel Gaussian channels

	$\hat{x}=0$	$\hat{x}=E$	$\hat{x}=1$
$x=0$	0	1	∞
$x=1$	∞	1	0

Table 1: Values of $d(x,\hat{x})$ for problem $6(b)$

(c) Determine the optimal power allocation to each transmitter, the capacity of the channel, and describe how do we achieve the capacity.

(d) If $P=3$ and $\sigma_1^2=2$, then find σ_2^2 such that we do not transmit through Transmitter 2.

Problem 4 (Erasure side-information for rate-distortion)

Consider $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, and let Y be an "erased" version of the source X, *i.e.,* $Y \in \{0, E, 1\}$,

$$Y_i = \begin{cases} X_i & \text{with probability } 1-p \\ E & \text{with probability } p \end{cases}$$

Let Y be available to both the source encoder as well as the source decoder. It is easy to derive the *conditional* rate-distortion function, exactly like we did in class, by conditioning the encoder and decoder on the available side-information Y. This can be easily shown to be:

$$R_{X|Y}(D) = \min_{p(\hat{x}|x,y):\mathbb{E}[d(x,\hat{x})]\leq D} I(X;\hat{X}|Y)$$

(a) Consider the Hamming distortion measure, and find the conditional rate-distortion function for this source with the available side-information, *i.e.,* evaluate $R_{X|Y}(D)$.
Hint: Note that this does not need extensive calculations.

(b) Suppose the reconstruction itself can produce an erased version of the source, *i.e.,* $\hat{x} \in \hat{\mathcal{X}} = \{0, E, 1\}$, and the distortion measure be given by the matrix in Table 1.

Calculate the conditional rate distortion function for this source.

Problem 5 (Rate Distortion)

Consider two independent Gaussian sources X_1, X_2, where $X_1 \sim \mathcal{N}(0,2)$ and $X_2 \sim \mathcal{N}(0,1)$. We use the distortion measure $d(\underline{X},\hat{\underline{X}}) = ||\underline{X}-\hat{\underline{X}}||^2 = \sum_{i=1}^{2}\left|X_i-\hat{X}_i\right|^2$ for vectors $\underline{X} = [X_1, X_2], \hat{\underline{X}} = [\hat{X}_1, \hat{X}_2]$.

(a) For the distortion $D=\frac{5}{2}$, find the optimization problem for the rate distortion function $R(D)$.

(b) Determine how will you achieve the rate-distortion function $R(D)$ in (a).

(c) Now, consider the Gaussian sources $Y_1 \sim \mathcal{N}(0,\frac{3}{2})$ and $Y_2 \sim \mathcal{N}(0,\frac{3}{2})$ are correlated with the correlation coefficient $\rho=\frac{1}{3}$, *i.e.,* the covariance matrix $K_{\underline{Y}}$ of $\underline{Y}=[Y_1, Y_2]$ is given by

$$K_{\underline{Y}} = \begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

For the same distortion $D = \frac{5}{2}$, describe how will you achieve the rate-distortion function $R(D)$.

Can you transform the sources $\underline{Y} = [Y_1, Y_2]$ *in a manner to obtain a situation similar to independent parallel Gaussian sources?*