EE376A - Information Theory
Final, Monday March 16th

Instructions:

- You have **three hours**, 3.30PM - 6.30PM

- The exam has 4 questions, totaling 120 points.

- Please start answering each question on a new page of the answer booklet.

- You are allowed to carry the textbook, your own notes and other course related material with you. Electronic reading devices [including kindles, laptops, ipads, etc.] are allowed, provided they are used solely for reading pdf files already stored on them and not for any other form of communication or information retrieval.

- You are required to provide detailed explanations of how you arrived at your answers.

- You can use previous parts of a problem even if you did not solve them.

- As throughout the course, entropy (H) and Mutual Information (I) are specified in bits.

- log is taken in base 2.

- Throughout the exam 'prefix code' refers to a variable length code satisfying the prefix condition.

- Good Luck!

1. **Three Shannon Codes** *(25 points)*

Let $\{U_i\}_{i \geq 1}$ be a stationary finite-alphabet source whose alphabet size is r. Note that the stationarity property implies that $P(u_i), P(u_i | u_{i-1})$ do not depend on i. Throughout this problem, assume that $-\log P(u_i)$ and $-\log P(u_i | u_{i-1})$ are integers for all (u_i, u_{i-1}). Recall the definition of a Shannon Code given in the lecture. Your TA's decided to compress this source in a lossless fashion using Shannon coding. However, each of them had a different idea:

- Idoia suggested to code symbol-by-symbol, i.e., concatenate Shannon codes on the respective source symbols $U_1, U_2, \ldots$.

- Kartik suggested to code in pairs. In other words, first code (U_1, U_2) with a Shannon code designed for the pair, then code (U_3, U_4), and so on.

- Jiantao suggested to code each symbol given the previous symbol by using the Shannon code for the conditional pmf $\{P(u_i | u_{i-1})\}$. In other words, first code U_1, then code U_2 given U_1, then code U_3 given U_2, and so on.

In this problem, you will investigate which amongst the three schemes is best for a general stationary source.

(a) (10 points) If the source is memoryless (i.e. i.i.d.), compare the expected codeword length per symbol, i.e., $\frac{1}{n} E[l(U^n)]$, of each scheme, assuming $n > 2$ is even.

(b) (15 points) Compare the schemes again, for the case where the source is no longer memoryless and, in particular, is such that U_{i-1} and U_i are not independent.

2. Channel coding with side information *(35 points)*

Consider the binary channel given by

$$Y_i = X_i \oplus Z_i, \tag{1}$$

where X_i, Y_i, Z_i all take values in $\{0, 1\}$, and $\oplus$ denotes addition modulo-2. There are channel states S_i which determine the noise level of Z_i as follows.

- S_i is binary valued, taking values in the set $\{G, B\}$, distributed as

$$S_i = \begin{cases} G, & \text{with probability } \frac{2}{3} \\ B, & \text{with probability } \frac{1}{3} \end{cases}$$

- The conditional distribution of Z_i given S_i is characterized by

$$P(Z_i = 1 | S_i = s) = \begin{cases} \frac{1}{4}, & \text{if } s = G \\ \frac{1}{3}, & \text{if } s = B \end{cases}$$

  In other words, $Z_i | \{S_i = s\} \sim Bernoulli(p_s)$, where

$$p_s = \begin{cases} \frac{1}{4}, & \text{if } s = G \\ \frac{1}{3}, & \text{if } s = B \end{cases}$$

  $\{(S_i, Z_i)\}$ are i.i.d. (in pairs), independent of the channel input sequence $\{X_i\}$.

(a) (10 points) What is the capacity of this channel when *both* the encoder *and* the decoder have access to the state sequence $\{S_i\}_{i \geq 1}$?

(b) (10 points) What is the capacity of this channel when *neither* the encoder *nor* the decoder have access to the state sequence $\{S_i\}_{i \geq 1}$?

(c) (10 points) What is the capacity of this channel when *only the decoder* knows the state sequence $\{S_i\}_{i \geq 1}$?

(d) (5 points) Which is largest and which is smallest among your answers to parts (a), (b) and (c)? Explain.

3. Modulo-3 additive noise channel *(25 points)*

(a) (5 points) Consider the modulo-3 additive white noise channel given by

$$Y_i = X_i \oplus Z_i, \tag{2}$$

where X_i, Z_i, Y_i all take values in the alphabet $\{0, 1, 2\}$, $\oplus$ denotes addition modulo-3, and $\{Z_i\}$ are i.i.d. $\sim Z$ and independent of the channel input sequence $\{X_i\}$.
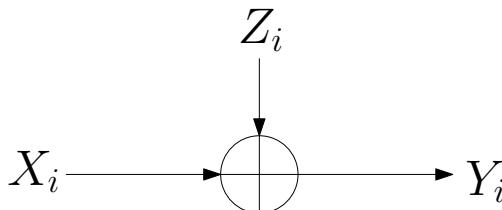

Figure 1: Ternary additive channel.

Show that the capacity of this channel is given by

$$C = \log 3 - H(Z). \tag{3}$$

(b) (7 points) For $\epsilon \geq 0$ define

$$\phi(\epsilon) = \max_{Z:Pr(Z\neq 0)\leq \epsilon} H(Z), \tag{4}$$

where the maximization is over ternary random variables Z that take values in $\{0, 1, 2\}$ (and that satisfy the indicated constraint). Obtain $\phi(\epsilon)$ explicity, as well as the distribution of the random variable, Z_ϵ, that achieves the associated maximum.

[Distinguish between the ranges $0 \leq \epsilon < 2/3$ and $\epsilon \geq 2/3$.]

(c) (5 points) Consider the problem of rate distortion coding of a memoryless source $U_i \sim U$, where the source and the reconstruction alphabets are both equal and ternary, i.e., $\mathcal{U} = \mathcal{V} = \{0, 1, 2\}$. Let the distortion measure be Hamming loss

$$d(u, v) = \begin{cases} 0, & \text{if } u = v \\ 1, & \text{otherwise.} \end{cases}$$

For U, V that are jointly distributed such that $E[d(U, V)] \leq D$, justify the following chain of equalities and inequalities

$$I(U; V) \overset{(i)}{=} H(U) - H(U|V)$$

$$\overset{(ii)}{=} H(U) - H(U \ominus V | V)$$

$$\overset{(iii)}{\geq} H(U) - H(U \ominus V)$$

$$\overset{(iv)}{\geq} H(U) - \phi(D),$$

where $\ominus$ denotes subtraction modulo-3 and $\phi(D)$ was defined in Equation (4). Argue why this implies that the rate distortion function of the source U is lower bounded as

$$R(D) \geq H(U) - \phi(D). \tag{5}$$

The above inequality is known as the 'Shannon lower bound' (specialized to our setting of ternary alphabets and Hamming loss).

(d) (8 points) Show that when U is uniform (on $\{0, 1, 2\}$), the Shannon lower bound holds with equality, i.e.,

$$R(D) = H(U) - \phi(D) = \log 3 - \phi(D), \quad 0 \leq D \leq 1. \tag{6}$$

[Hint: establish, by construction, existence of a joint distribution on U, V such that U is uniform and the inequalities in Part (c) hold with equalities]

4. **Gaussian source and channel** *(35 points)*

- **Gaussian Channel**

  Consider the parallel Gaussian channel which has two inputs $X = (X_1, X_2)$ and two outputs $Y = (Y_1, Y_2)$, where

  $$Y_1 = X_1 + Z_1$$
  $$Y_2 = X_2 + Z_2,$$

  and $Z_i \sim \mathcal{N}(0, \sigma_i^2), i = 1, 2$, are independent Gaussian random variables. We impose an average power constraint on the input X, which is

  $$\mathbb{E}[\|X\|^2] = \mathbb{E}[X_1^2 + X_2^2] \leq P$$

  (a) (10 points) Give an explicit formula for the capacity of this channel in terms of $P, \sigma_1^2, \sigma_2^2$.

  (b) (7 points) Suppose you had access to capacity-achieving schemes for the scalar AWGN channel whose capacity we derived in class. How would you use them to construct capacity-achieving schemes for this parallel Gaussian channel?

- **Gaussian Source**

  Consider a two-dimensional real valued source $U = (U_1, U_2)$ such that $U_1 \sim \mathcal{N}(0, \sigma_1^2)$, and $U_2 \sim \mathcal{N}(0, \sigma_2^2)$, and U_1 is independent of U_2. Let $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be the distortion measure

  $$d(u, v) = \|u - v\|^2 = |u_1 - v_1|^2 + |u_2 - v_2|^2$$

  We wish to compress i.i.d. copies of the source U, with average per-symbol distortion no greater than D, i.e. the usual lossy compression setup discussed in class.

  (a) (10 points) Evaluate the rate-distortion function $R(D)$ explicitly in terms of the problem parameters $D, \sigma_1^2, \sigma_2^2$.

  (b) (8 points) Suppose you had access to good lossy compressors for the scalar Gaussian source whose rate-distortion function we derived in class. How would you use them to construct good schemes for this two-dimenstional Gaussian source?