ECE563 Fall 2016 **Homework 5**

1. *Rate Splitting Multiple Access.* Consider the M-user Gaussian multiple access channel capacity region which is the set of rate vectors $(R_1, R_2, \ldots, R_M)$ satisfying

$$R_i \leq C(\sum_{i \in S} P_i, \sigma^2), \quad \forall S \subseteq \{1, 2, \ldots, M\}, \tag{1}$$

where $C(P, \sigma^2)$ denotes $0.5 \log(1 + \frac{P}{\sigma^2})$ and σ^2 is the noise variance and $P_1, \ldots P_M$ are the power constraints on the individual users. A rate vector $(R_1, R_2, \ldots, R_M)$ is said to be achievable by successive decoding if, after a possible reindexing of the users, we have

$$R_j \leq C(P_j, \sigma^2 + \sum_{i=1}^{j} P_i), \quad \forall j = 1 \ldots M. \tag{2}$$

In this case one can first decode user M treating the others as noise, then user $M-1$, and so on, if the codes for the individual users are chosen appropriately. In a creative paper in 1996, Rimoldi and Urbanke introduced an idea by which *every* point in the capacity region can be decoded using successive decoding – to achieve any given rate vector some of the M users have to act as if they are two different users, and one can always ensure that at most $M-1$ of the original M users will have to act this way. This problem illustrates this result for the case of $M = 2$.

The dominating face of the 2-user Gaussian multiple access channel capacity region is the portion of it where $R_1 + R_2 = C(P_1 + P_2, \sigma^2)$. Clearly, it suffices to consider rate pairs on the dominating face.

(a) Consider an arbitrary rate pair (R_1, R_2) on the dominating face of the 2-user Gaussian multiple access channel capacity region. Show that if, in addition, either of the equalities $R_1 = C(P_1, \sigma^2)$ or $R_2 = C(P_2, \sigma^2)$ holds (i.e., if we are one of the corners on the dominating face), then the point is already successively decodable, without either user having to split its rate.

(b) Suppose we are at an interior point on the dominating face, i.e., $R_1 < C(P_1, \sigma^2)$, $R_2 < C(P_2, \sigma^2)$ and $R_1 + R_2 = C(P_1 + P_2, \sigma^2)$. Define $\delta > 0$ by $R_2 = C(P_2, \sigma^2 + \delta)$. Define the power triple (p_1, p_2, p_3) by $p_1 = \delta, p_2 = P_2, p_3 = P_1 - \delta$ and the rate triple (r_1, r_2, r_3) by $r_1 = C(p_1, \sigma^2), r_2 = C(p_2, \sigma^2 + p_1), r3 = C(p_3, \sigma^2 + p_1 + p_2)$. Argue that, if the first user splits himself into two users of power constraint p_1 and p_3 respectively, with the former aiming to communicate at rate r_1 and the latter at rate r_3, then (r_1, r_2, r_3) is a successively decodable triple of the resulting 3-user problem with power constraints (p_1, p_2, p_3) which effectively provides communication at the rate pair (R_1, R_2) of the original 2-user problem with power constraints (P_1, P_2).

(c) *Bonus*: Construct an analogous scheme that works for a general $M > 2$.

2. Consider independent random variables X, Z with zero mean each and variances P, N respectively. Consider the following game involving payoff of the mutual information $I(X; X+Z)$. The "signal player" chooses a distribution on X to maximize $I(X; X+Z)$ while the "noise player" chooses a distribution on Z to minimize $I(X; X+Z)$.

   (a) Letting X^* and Z^* denote Gaussian random variables (with the appropriate variances of P and N respectively) show the saddle point conditions:

   $$I(X; X+Z^*) \leq I(X^*; X^*+Z^*) \leq I(X^*; X^*+Z). \tag{3}$$

   Hint: You might find it useful to use the entropy power inequality:

   $$e^{2h(X+Y)} \geq e^{2h(X)} + e^{2h(Y)}. \tag{4}$$

   (b) Conclude that

   $$\min_Z \max_X I(X; X+Z) = \max_X \min_Z I(X; X+Z). \tag{5}$$

   (c) What implications can you draw from this result for reliable communication over the additive noise channel?

   (d) *Bonus*: The entropy power inequality is one of the deep technical results in information theory. The main reason it is deep is because it can be viewed as a solution of a minimization of a concave function with Gaussian random variables as the resulting solution. It was "conjectured" by Shannon in his seminal 1948 paper but formally proved only in 1965 by Nelson Blachman. The connections between the entropy power inequality and the classical Brunn-Minkowski inequality from convex geometry was made in a 1991 paper by Dembo, Cover and Thomas by relating both these inequalities to the (reverse) Young inequality in functional analysis. Since Gaussians feature centrally in probability theory, the entropy power inequality also shows up as a certain strengthening of the central limit theorem (Barron, 1986, Annals of Probability). A reading exercise involves catching up on some beautiful mathematics, as uncovered by these cited papers.

3. Consider the following multiple access channel: $Y[m] = X_1[m] + \text{sign}(X_2[m])$, where X_1, X_2 are both real. Here $\text{sign}(x)$ is the function that is equal to $+1$ if $x > 0$ and -1 otherwise. There is the usual power constraint on the transmissions: $\sum_{m=1}^{n} (X_i[m])^2 \leq nP_i$ for $i = 1, 2$. Note that there is interference between the two users, but no noise in this channel.

   (a) Find the capacity region.

   (b) Describe a coding scheme that achieves the capacity region.

4. The multiple access channel problems below are separate ones.

   (a) Consider the multiple access channel: $Y = X_1^2 + X_2$ with $X_i \in \{-1, 0, 1\}$ for $i = 1, 2$. Find the capacity region and describe the optimizing distributions $p^*(x_1), p^*(x_2)$ for any point on the boundary of the capacity region.

   (b) Find the capacity region of the multiple access channel $Y = X_1^{X_2}$, where $X \in \{2, 4\}$ and $X_2 \in \{1, 2\}$. Would your answer change if the range of X_1 were $\{1, 2\}$ instead?

   (c) Find and sketch the capacity region of the multiplicative multiple access channel: $Y = X_1 X_2$ where $X_1 \in \{0, 1\}$ and $X_2 \in \{1, 2, 3\}$.

5. A speaker of Dutch, Spanish and French wishes to communicate simultaneously to three people: D, S and F. Person D knows only Dutch but can distinguish between a spanish and french word. Similarly, person S knows only Spanish but can distinguish between the other two languages. Person F knows only French but can distinguish between the other two languages. Suppose each of the three languages has M words. Assuming three words per second are spoken for the questions below.

   (a) What is the largest rate at which the speaker can speak to person D?

   (b) If he speaks to person D at the maximum rate, what is the maximum rate at which he can speak simultaneously to S?

   (c) If he is speaking to D and S at the joint rate above, what is the maximum rate at he can also speak to F?

6. Consider a point-to-point DMC. In the course we have met the "joint typicality" decoder, which needs to know the statistics of the channel. Consider the following decoder, known as the maximal mutual information (MMI) decoder, which doesn't need to know the channel statistics. The (standard random coding) assumptions hold: (a) 2^{nR} codewords are drawn i.i.d. according to distribution P_X. (b) Message J is chosen uniformly from the set $\{1, 2, ..., 2^{nR}\}$, and the corresponding codeword $X_n(J)$ is sent through the channel. (c) The channel is discrete memoryless, characterized by the conditional PMF $P_{Y|X}$, where both X and Y take values in the respective finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. Now suppose sequence $Y_n \in \mathcal{Y}^n$ is received by the decoder. Let $P_{X^n(j),Y^n} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ denote the joint *empirical distribution* of the jth codeword $X^n(j)$ and Y^n. The MMI decoder produces as its estimate the codeword with maximum empirical mutual information with Y^n:

$$\hat{J} = \arg \max_{j \in 1,2,\ldots,2^{nR}} I(P_{X^n(j),Y^n}) \tag{6}$$

   where, for $Q_{X,Y} \in P(\mathcal{X}, \mathcal{Y})$ and $I(Q_{X,Y})$ denotes the mutual information between X and Y when distributed according to $Q_{X,Y}$ (and ties in the maximization are broken arbitrarily).

(a) Does the MMI decoder have to know the channel statistics $P_{Y|X}$ in order to implement this decoding scheme?

(b) Prove that, for any $Q_{X,Y} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$:

$$D(Q_{X,Y} \| P_X \times P_Y) \geq I(Q_{X,Y}). \tag{7}$$

Hint: Use the fact that $I(Q_{X,Y}) = D(Q_{X,Y} \| Q_X \times Q_Y)$.

(c) Using the previous part, prove that $f(\theta) \geq \theta$, where

$$f(\theta) := \min_{Q_{X,Y} \in \mathcal{P}(\mathcal{X},\mathcal{Y}): I(Q_{X,Y}) \geq \theta} D(Q_{X,Y} \| P_X \times P_Y). \tag{8}$$

(d) Using the definition of typical sequences and the AEP (together called the "method of types" in an influential book by Cśiszár and Körner), show that

$$\lim_{n \to \infty} \frac{1}{n} \log P(I(P_{X^n(j),Y^n}) \geq \theta | J = 1) = -f(\theta). \tag{9}$$

(e) What is the supremum of rates R for which the error probability $P(\hat{J} \neq J) \to 0$ as $n \to \infty$ under the MMI decoding scheme? How does it compare to the supremum of rates for which joint typicality decoding would have achieved reliable communication?

Hint: First derive the following inequalities:

$$\begin{aligned}
P(\hat{J} \neq J) &= P(\hat{J} \neq J | J = 1) \\
&\leq P(I(P_{X^n(1),Y^n}) \leq \theta | J = 1) + P(I(P_{X^n(j),Y^n}) > \theta, \text{for some } j \neq 1 | J = 1) \\
&\leq P(I(P_{X^n(1),Y^n}) \leq \theta | J = 1) + 2^{nR} P(I(P_{X^n(2),Y^n}) > \theta | J = 1).
\end{aligned}$$

7. Consider communicating over the memoryless "additive Gaussian noise channel":

$$y = \sqrt{E}x + z \tag{10}$$

where the input x is restricted to ± 1 and the additive noise z is Gaussian (with zero mean and variance σ^2) and independent of the input x. We communicate one of M messages over n channel uses – the rate of communication is $R = \log M$ or equivalently $M = 2^{nR}$. A code is a mapping from the set of messages to a n-length string of ± 1 entries. For a code C we denote by $\mathbf{c}_i$ the N-length vector of inputs corresponding to the message i. Consider *maximum likelihood* (ML) decoding at the destination. In this exercise we evaluate the performance of the ML decoder and arrive at a formula for a rate of reliable communication over this channel. Suppose the received vector (output of the channel over the n uses) is $\mathbf{y}$.

(a) Derive a formula for the likelihood of the message being i and show that it depends only on the Euclidean distance $\|\mathbf{y} - \sqrt{E}\mathbf{c}_i\|$.

(b) Consider the binary hypothesis testing problem of distinguishing among two equally likely messages i and j from the received output vector $\mathbf{y}$. Calculate the average probability of error with the ML decoder (denoted by P_{2e}) in terms of the $Q(\cdot)$ function. Show that this formula only depends on the Euclidean distance $\|\mathbf{c}_i - \mathbf{c}_j\|$.

(c) Consider the ensemble of all possible codes (2^{nM} of them), with uniform probability on each. Show that (via the union bound) the average probability of the error, averaged over all messages and all codes is at most MP_{2e}.

(d) Using the well-known inequality (worth knowing how to prove it)

$$Q(x) < \frac{1}{2}e^{-\frac{x^2}{2}}, \tag{11}$$

show that as long as the rate of communication is less than $R_0 := 1 - \log_2(1 + e^{-\frac{2E}{\sigma^2}})$, then communication is arbitrarily reliable.

(e) Find the capacity C of this channel by simplifying the Shannon formula for capacity as much as possible (although there is no closed form solution). Numerically plot C and R against the "signal-to-noise ratio" $\frac{E}{\sigma^2}$ in the same graph with the horizontal axis in log-scale. Comment on the gap between the two quantities C and R_0 for different values of signal to noise ratio.

Notes: This problem is adapted from Chapter 5 of a classical book on communication theory by Jack Wozencraft and Irwin Jacobs, both teaching at MIT when the book was written in 1966. Irwin Jacobs went on to (co)found Qualcomm. Inc, a giant in the telecommunication industry.