Random codes in communication 31 July 2011

Final Test (Moed Bet)

1) **True or False** (20 points)
Copy each relation to your notebook and write **true** or **false**. Then, if it's true, prove it. If it is false give a counterexample or prove that the opposite is true.
   a) For two random variables $X,\ Y$, $H(X) > H(Y)$, then Y can be compressed better than X [4 points].

   Solution: **True**. $H(X) > H(Y)$, hence $R_X > R_Y$ and Y can be compressed better.

   b) Which of the following sequence of code-lengths are a valid binary prefix codes(can be more than one answer) [4 points].?
   - 1,2,3,4 V
   - 1,2,2,4 X
   - 1,3,3,3 V
   - 2,2,2,3 V

   c) Let X be a continues alphabet random variable. For which of the following channels with power constraint $\frac{1}{n}\sum_{i=1}^{n} x_i^2(m) \leq P$, the capacity is not finite (there can be more than one answer) [12 points]:
   i) $Y = X$
   ii) $Y = X + Z$, where $Z \sim N(0,1)$
   iii) $Y = X + Z$, where $Z \sim N(0,1)$ with probability 0.9, and $Z = 0$ with probability 0.1.
   Provide a scheme that achieves the infinite capacity, if exists.

   Solution: Since X has continues alphabet, (a) and (c) - which has no noise - has infinite capacity. The scheme is for (c), but also apply for (a):
   - Take the interval $[0, \sqrt{P}]$ and divide it into 2^{nR} points- $j\frac{\sqrt{P}}{2^{nR}}$, $j = 1, 2, ..., 2^{nR}$.
   - For every message j, send $x_i(m) = j\frac{\sqrt{P}}{2^{nR}}$ for all $i = 1, 2, ..., n$.
   - The decoder declares the message at the first time that $y = j\frac{\sqrt{P}}{2^{nR}}$ for some j.

   An error accurse if $Z \neq 0$ for all i. Hence, $P(er) = 0.9^n$, which goes to zero with increasing n.

2) **Blahut-Arimoto's algorithm** (30 points) Consider the i.i.d. source coding model as in Fig. 1, where the input X^n is distributed i.i.d. $\sim p(x)$ which is given by nature, and the decoder is required to produce an output $\hat{X}^n$ that has a distortion constraint with the source, i.e.,

$$\frac{1}{n}\mathbb{E}\left[d(X^n, \hat{X}^n)\right] \leq D.$$

The solution to this problem, i.e., the minimum rate that satisfies the constraint above for a given

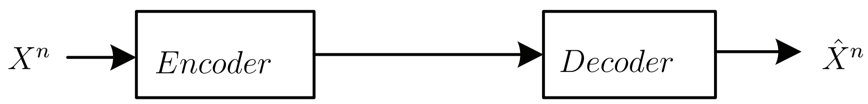

Fig. 1. Rate distortion. We require $\frac{1}{n}\mathbb{E}\left[d(X^n, \hat{X}^n)\right] \leq D$

distortion D, is

$$R(D) = \min_{p(\hat{x}|x):\mathbb{E}\left[d(X,\hat{X})\right]\leq D} I(X;\hat{X}).$$

Solving this optimization problem is a difficult task for the general source. In this question we develop an iterative algorithm for finding the solution for a fixed source $p(x)$.

a) First, prove that the mutual information as a function of $p(\hat{x}|x)$ and $p(\hat{x})$ as below. [3 points]

$$I(X;\hat{X}) = \sum_{x,y} p(x)p(\hat{x}|x)\log\frac{p(\hat{x}|x)}{p(\hat{x})}.$$

b) Show that $I(X;Y)$ as written above is concave in both $p(\hat{x}),\ p(\hat{x}|x)$ (Hint. You may use the Log-Sum-inequality). [6 points]
c) Find an expression for $p(\hat{x}|x)$ that minimizes $I(X;\hat{X})$ when $p(\hat{x})$ is fixed (Hint. You may use the Lagrange multipliers method with the constraints $\sum_{\hat{x}} p(\hat{x}|x) = 1$ for all x, and $\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) - D \leq 0$. See also the paragraph below!). [9 points]
d) Find an expression for $p(\hat{x})$ that minimizes $I(X;\hat{X})$ when $p(\hat{x}|x)$ is fixed (Hint. You may use the Lagrange multipliers method with constraints $\sum_{\hat{x}} p(\hat{x}) = 1$). [7 points]
e) Using (d), conclude that $R(D) = \min_{p(\hat{x}),p(\hat{x}|x)} I(X;\hat{X})$. [5 points]

Clarification: First, note that the solution to (c) is a function of $\lambda \geq 0$ - the parameter associated with the distortion constraint. This is ok, and is due to the fact that this parameter λ is correlated with the slope at the point (R,D).
Second, the algorithm is performed by minimizing in each iteration over another variable; first over $p(\hat{x})$ when $p\hat{x}|x)$ is fixed, then over $p(\hat{x}|x)$ when $p(\hat{x})$ is fixed, and so on. This iterative algorithm converges for every slope $-\lambda$, and hence one can find the rate distortion function $R(D)$ for every i.i.d. source X^n, with reasonable alphabet size.

Solution:

a) Since $I(X;\hat{X}) = H(\hat{X}) - H(\hat{X}|X)$, the answer is obvious.
b) Recall, that the Log-Sum inequality is

$$\sum_{i=1}^{n} a_i \log\frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right)\log\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

Hence

$$\begin{aligned}(\lambda p_1(\hat{x}|x)+&(1-\lambda)p_2(\hat{x}|x))\log\frac{\lambda p_1(\hat{x}|x)+(1-\lambda)p_2(\hat{x}|x)}{\lambda p_1(\hat{x})+(1-\lambda)p_2(\hat{x})}\\ &\leq \lambda p_1(\hat{x}|x)\log\frac{p_1(\hat{x}|x)}{p_1(\hat{x})} + (1-\lambda)p_2(\hat{x}|x)\log\frac{p_2(\hat{x}|x)}{p_2(\hat{x})}.\end{aligned}$$

Multiplying by $p(x)$ and summing over all $x,\ \hat{x}$, and letting $\mathcal{I}(p(\hat{x}),p(\hat{x}|x))$ be the mutual information as in (a), we obtain

$$\begin{aligned}\mathcal{I}\big(\lambda p_1(\hat{x})+&(1-\lambda)p_2(\hat{x}),\lambda p_1(\hat{x}|x)+(1-\lambda)p_2(\hat{x}|x)\big)\\ &\leq \lambda\mathcal{I}\big(p_1(\hat{x}),p_1(\hat{x}|x)\big)+(1-\lambda)\mathcal{I}\big(p_2(\hat{x}),p_2(\hat{x}|x)\big).\end{aligned}$$

c) Define the lagrangian

$$L = \sum_{x,\hat{x}} p(x)p(\hat{x}|x)\log\frac{p(\hat{x}|x)}{p(\hat{x})} + \mu(x)(\sum_{\hat{x}} p(\hat{x}|x) - 1)$$
$$+ \lambda(\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) - D),$$

and differentiate over $p(\hat{x}|x)$. Solving $\frac{\partial L}{\partial p(\hat{x}|x)} = 0$ provides us with

$$p(\hat{x}|x) = \frac{p(\hat{x})2^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} p(\hat{x})2^{-\lambda d(x,\hat{x})}}.$$

d) Define the lagrangian

$$J = \sum_{x,\hat{x}} p(x)p(\hat{x}|x)\log\frac{p(\hat{x}|x)}{p(\hat{x})} + \mu(\sum_{\hat{x}} p(\hat{x}) - 1),$$

and differentiate over $p(\hat{x})$. Solving $\frac{\partial J}{\partial p(\hat{x})} = 0$ provides us with

$$p(\hat{x}) = \sum_{x} p(x)p(\hat{x}|x) \tag{1}$$

e) The expression for $p(\hat{x})$ is the one that corresponds to $p(\hat{x}|x)$, and hence minimizing over $p(\hat{x}),\ p(\hat{x}|x)$ is the same as over $p(\hat{x}|x)$ alone.

3) **Bit loading algorithm** (50 points) Consider the parallel Gaussian channel as illustrated in Fig.2. As seen in class, Gaussian inputs distribution maximizes the mutual information for this channel. In order to determine the variance value in each subchannel, there are two types of loading algorithms for parallel Gaussian channel- those that try to maximize data rate (power loading) and those that try to minimize the energy at a given fixed data rate (bit-loading). We studied in class the power loading algorithm, i.e., loading procedure which maximizes the number of bits per symbol subject to a fixed energy constraint. Let us define the bit loading criterion.
Definition 1: (Bit loading criterion) A bit loading procedure minimizes the energy sum

$$\sum_{n=1}^{N} \epsilon_n, \tag{2}$$

subject to:

$$b = \frac{1}{2}\sum_{n=1}^{N} \log_2\left(1 + \epsilon_n g_n\right), \tag{3}$$

where g_n represents the subchannel signal-to-noise ratio, g_n is a fixed value for each of the subchannels. However, $\epsilon_n \geq 0$, which is the energy investment in the n-th subchannel, can be varied to minimizes the total energy, subject to a fixed data rate b.

a) Write the bit loading problem as a convex optimization problem as learned in class, and explain why this is a convex problem. [7 point]
b) Solve the bit loading optimization problem (hint: KKT conditions), and write the implicit equation from which we can derive the water lever out of the parameters $\{g_n\}_{n=1}^{N}$ and b. [15 points]
c) Find the energy distribution $\{\epsilon_n\}_{n=1}^{N}$, for the following channel:

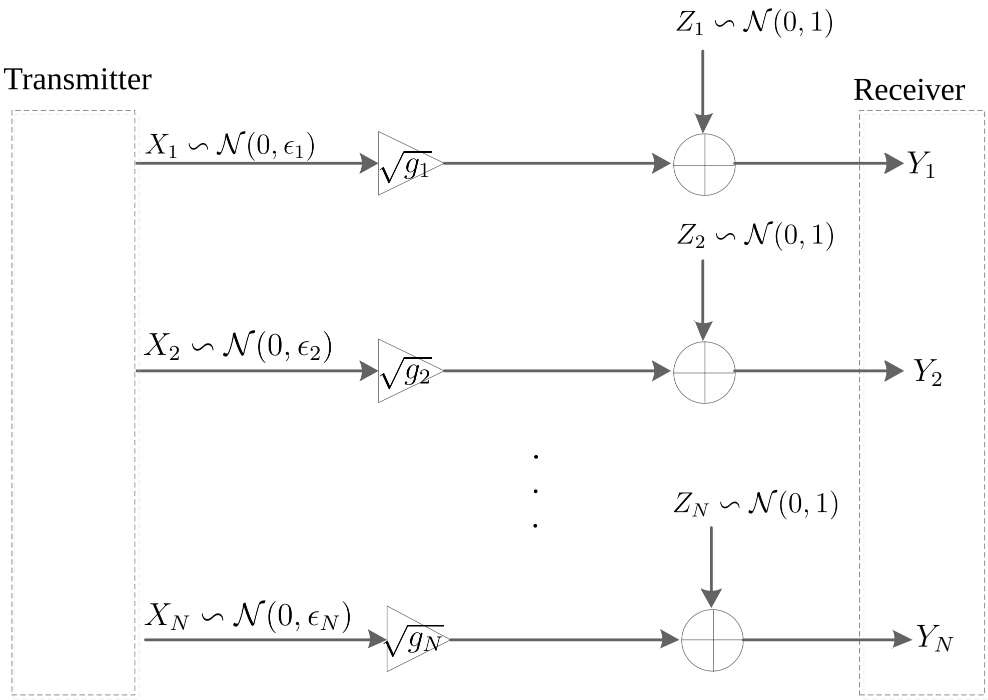

Fig. 2. Parallel Gaussian channel.

g_1	g_2	g_3	g_4	g_5	b
19	17	10	3	0.05	6

Hint: be careful, if you decide that $\epsilon_i = 0$ for some $i \in \{1, 2, .., N\}$, the water level constant changes. [10 points]

d) Find the new data rate b for sub-ex (c) in the case where constant energy is loaded in each subchannel, i.e., in each of the N' sub channels you decided to use in (c), the energy is $\frac{\min_{\epsilon^N} \sum_{n=1}^{N} \epsilon_n}{N'}$. [10 points]

e) Consider the case that the channel is $Y = G \cdot X + Z$, where $G \in R^{N \times N}$ (now-not diagonal), $X \in R^{N \times 1}$, $Z \in R^{N \times 1}$, and $Z \backsim \mathcal{N}(0, I)$. How Would you change your solution to (c)? [8 points]

Solution:

a) The convex optimization problem is, of course:

$$\min_{\epsilon^N} \sum_{n=1}^{n} \epsilon_n,$$

subject to

$$b - \frac{1}{2} \sum_{n=1}^{N} \log_2 (1 + \epsilon_n g_n) \leq 0.$$

This is a convex problem since the objective is affine, and the constraint is convex (since $\log$ is concave).

b) We write the Lagrangian as

$$J = \sum_{n=1}^{n} \epsilon_n + \lambda \left(b - \frac{1}{2} \sum_{n=1}^{N} \log_2(1 + \epsilon_n g_n) \right).$$

Solving $\frac{\partial J}{\partial \epsilon_n} = 0$ and we obtain

$$2(1 + \epsilon_n g_n) = \lambda g_n,$$

or

$$\epsilon_n = \frac{\lambda}{2} - \frac{1}{g_n}.$$

Since all ϵ_ns are positive, we have

$$\epsilon_n = \left(\nu - \frac{1}{g_n} \right)^+.$$

Now, if $N' = \{n : \ \nu > \frac{1}{g_n}\}$, i.e., all channels that should be in use, then:

$$\begin{aligned} b &= \frac{1}{2} \sum_{n=1}^{N} \log_2 (1 + \epsilon_n g_n) \\ &= \frac{1}{2} \sum_{n=1}^{N} \log_2 \left(1 + \left(\nu - \frac{1}{g_n} \right)^+ g_n \right) \\ &= \frac{1}{2} \sum_{n \in N'} \log_2 (\nu g_n) \\ &= \frac{1}{2} |N'| \log_2(\nu) + \frac{1}{2} \sum_{n \in N'} \log_2(g_n). \end{aligned}$$

and,

$$\log_2(\nu) = \frac{2b}{|N'|} - \frac{1}{|N'|} \sum_{n \in N'} \log_2(g_n).$$

c) First, we assume that all channels can be in use. Hence, $\nu = \frac{1}{\max g_n} = 20$. In that case we can see that the data rate is much larger than 6. On the other hand, if the worst channel is not in use, and $\nu = \frac{1}{g_4} = 0.667$, then

$$\begin{aligned} \frac{1}{2} |N'| \log_2(\nu) + \frac{1}{2} \sum_{n \in N'} \log_2(g_n) &= 2 \log_2(0.667) + \frac{1}{2} \log(3 \cdot 10 \cdot 17 \cdot 19) \\ &= 1.039. \end{aligned}$$

This data rate is lesser than $b = 6$, and hence we know that $0.667 \leq \nu \leq 20$ and $\epsilon_5 = 0$. Therefore, we can write

$$\begin{aligned} \log_2(\nu) &= \frac{2b}{|N'|} - \frac{1}{|N'|} \sum_{n \in N'} \log_2(g_n) \\ &= \frac{12}{4} - \frac{1}{4} \log_2(3 \cdot 10 \cdot 17 \cdot 19) \\ &= 2.0034, \end{aligned}$$

or $\nu = 4.009 \approx 4$, which is the water level. Now, the energy investment is $\epsilon^N = \{3.942, 3.941, 3.9, 3.667, 0\}$.

d) Now, that we know that the 5th channel is not in use, and the average power is

$$\epsilon = \frac{\min_{\epsilon^N} \sum_{n=1}^{N} \epsilon_n}{N'} = 3.863,$$

the data rate will be

$$b = \frac{1}{2} \sum_{n=1}^{4} \log_2 \left(1 + \epsilon g_n\right) = 3.196.$$

e) As discussed in class, any matrix has a SVD decomposition $G = UDV$, D is diagonal and U, V are unitary. Hence,

$$\begin{aligned} Y &= UDVX + Z, \\ U^{-1}Y &= D(VX) + U^{-1}Z, \\ Y' &= DX' + Z'. \end{aligned}$$

This problem can be solved, and thus solve the original problem.

Good Luck!