FINAL EXAM

Friday, January 16, 2009, 14:15-18:00
This exam has 4 problems and 100 points in total.

Instructions

- You are allowed to use 2 sheet of paper for reference. No mobile phones or calculators are allowed in the exam.
- You can attempt the problems in any order as long as it is clear which problem is being attempted and which solution to the problem you want us to grade.
- If you are stuck in any part of a problem do not dwell on it, try to move on and attempt it later.
- Please solve every problem on **separate paper sheets**.
- It is your responsibility to **number the pages** of your solutions and write on the first sheet the **total number of pages** submitted.

Some Preliminaries

- The capacity of a Gaussian channel $Y = X + Z$ where $Z \sim \mathcal{N}(0, N)$ with an input power constraint is $C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$, which is achievable for a Gaussian input distribution $X \sim \mathcal{N}(0, P)$.
- The rate-distortion function for Gaussian source $X \sim \mathcal{N}(0, P)$ and Euclidean distortion measure $d(x, \hat{x}) = (x - \hat{x})^2$ is $R(D) = \frac{1}{2}\log\left(\frac{P}{D}\right)$.
- The entropy two jointly Gaussian random variables (X, Y) with covariance matrix

$$A = \operatorname{cov}(X, Y) = \mathbb{E}\left[\begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} X & Y \end{bmatrix}\right] = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

  is given by $h(X, Y) = \frac{1}{2}\log\left((2\pi e)^2 \det \mathbf{A}\right) = \frac{1}{2}\log\left((2\pi e)^2\sigma_x^2\sigma_y^2(1 - \rho^2)\right)$.
- If $X \sim \mathcal{N}(0, P)$ and $Y = aX + b$ where a and b are constants, then $Y \sim \mathcal{N}(b, \alpha^2 P)$.

GOOD LUCK!

Problem 1 (20 pts)

Consider the channel shown in Fig. 1 which is formed by two binary symmetric channels. The source S sends $X \in \{0,1\}$ through the first channel and the relay node R receives $Y \in \{0,1\}$ where $\Pr(Y \neq X) = p \leq \frac{1}{2}$. Then the relay R produces $U \in \{0,1\}$ and feeds it to the second channel and the destination D receives $V \in \{0,1\}$ where $\Pr(U \neq V) = q \leq \frac{1}{2}$. We denote by $C(p) = 1 - H_2(p)$ and $C(q) = 1 - H_2(q)$ the capacities of the first and the second channel, respectively, where $H_2(\cdot)$ is the binary entropy function. We use C to denote the capacity of the whole channel from X to V.

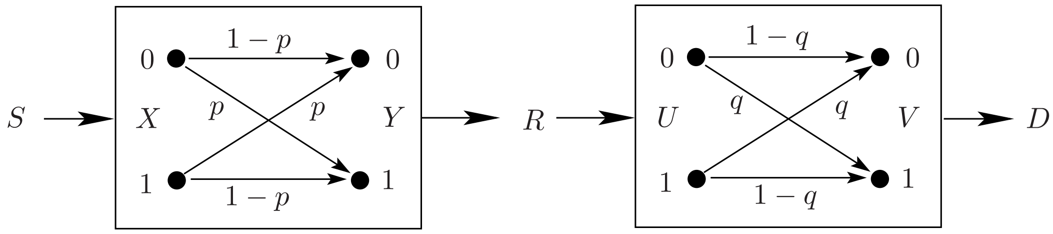

Figure 1: A cascade of system of two binary symmetric channels

(a) Show that $C \leq C(p)$ and $C \leq C(q)$. *[4pts]*

(b) Assume that there is no processing allowed at the relay, *i.e.,* the relay node just forwards its received bit to the destination. Find the capacity of the channel $C = \max_{p(x)} I(X;V)$. *[7pts]*
Hint: Do the two channels create a single binary symmetric channel in this scenario?

(c) Now assume that the relay node can wait to receive an arbitrary number of bits, process them, and produce a sequence of Y to transmit to the destination node. Show that $C' = \min\{C(p), C(q)\}$ is achievable. *[5pts]*

(d) Compare C and C'. Which one is larger? Explain why. *[4pts]*

Problem 2 (30 pts)

Consider a Gaussian channel as shown in Fig 2 with the input power constraint $\mathbb{E}[X_1^2] \leq P$. The receiver observes the noisy signal

$$Y_1 = X_1 + Z_1,$$

where Z_1 is Gaussian noise distributed as $Z_1 \sim \mathcal{N}(0, N)$ and independent of the channel inputs. The receiver has also access to a Gaussian random variables $U_1 \sim \mathcal{N}(0, N)$ which is independent of X_1, but correlated with Z_1, *i.e.,* $\mathbb{E}[Z_1 U_1] = \mu_1 N$.

(a) It can be shown that the capacity of the described channel is $C = \max_{p(x)} I(Y_1, U_1; X_1)$. Calculate C. *[8pts]*
Hint: Start by writing $I(Y_1, U_1; X_1) = h(Y_1, U_1) - h(Y_1, U_1|X_1) = h(Y_1, U_1) - h(Z_1, U_1)$.

(b) Let us apply a function on Y_1 and U_1 at the receiver to obtain $\tilde{Y}_1 = Y_1 + \gamma U_1$. Show that $I(\tilde{Y}_1; X_1) \leq I(Y_1, U_1; X_1)$. Assuming $X_1 \sim \mathcal{N}(0, P)$, find the value of γ, for which $\tilde{Y}_1$ would be a sufficient statistic for decoding X_1, *i.e.,* $I(\tilde{Y}_1; X_1) = I(Y_1, U_1; X_1)$. *[8pts]*

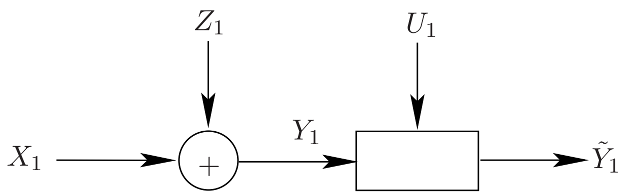

Figure 2: A Gaussian channel with side information at the receiver

(c) Define $\hat{X}_1 = \alpha_1 Y_1 + \beta_1 U_1$. Find the optimal values for α_1 and β_2 to minimize $\mathbb{E}[(\hat{X}_1 - X_1)^2]$. *[6pts]*
Hint:

$$\begin{aligned}
\mathbb{E}[(\hat{X}_1 - X_1)^2] &= \mathbb{E}[(\alpha_1 Y_1 + \beta_1 U_1 - X_1)^2] \\
&= \mathbb{E}[((\alpha_1 - 1)X_1 + \alpha_1 Z_1 + \beta_1 U_1)^2] \\
&= \mathbb{E}[((\alpha_1 - 1)X_1)^2 + (\alpha_1 Z_1 + \beta_1 U_1)^2] \\
&= (\alpha_1 - 1)^2 \mathbb{E}[X_1^2] + \alpha_1^2 \mathbb{E}[Z_1^2] + \beta_1^2 \mathbb{E}[U_1^2] + 2\alpha_1\beta_1 \mathbb{E}[Z_1 U_1]
\end{aligned}$$

Now Assume that we have two parallel channels each look like the channel we considered above, as shown in Fig 3. The correlation between the Gaussian noises is of the form

$$\mathbb{E}\left[\begin{bmatrix} Z_1 \\ U_1 \\ Z_2 \\ U_2 \end{bmatrix} \begin{bmatrix} Z_1 & U_1 & Z_2 & U_2 \end{bmatrix}\right] = \begin{bmatrix} N & \mu_1 N & 0 & 0 \\ \mu_1 N & N & 0 & 0 \\ 0 & 0 & N & \mu_2 N \\ 0 & 0 & \mu_2 N & N \end{bmatrix}.$$

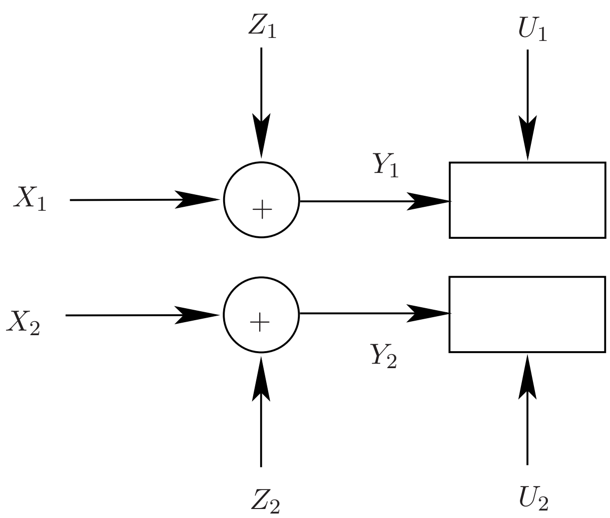

Figure 3: Two parallel Gaussian channel with side information at the receivers

It can be shown that

$$\max_{p(x_1,x_2):\mathbb{E}[X_1^2]+\mathbb{E}[X_2^2]\le P} I(X_1, X_2; U_1, Y_1, U_2, Y_2) = \max_{p(x_1,x_2)=p(x_1)p(x_2):\mathbb{E}[X_1^2]+\mathbb{E}[X_2^2]\le P} \{I(X_1; U_1, Y_1) + I(X_2; U_2, Y_2)\};$$

i.e., it is optimal to choose independent X_1 and X_2.

(d) Find the capacity of this parallel channel with a total power constraint $\mathbb{E}[X_1^2]+\mathbb{E}[X_2^2] \leq P$. What is the optimal power allocation for this channel, *i.e.*, the optimal values of P_1 and P_2 such that $\mathbb{E}[X_1^2] = P_1$ and $\mathbb{E}[X_2^2] = P_2$ and $P_1 + P_2 \leq P$. *[8pts]*

Hint: Use the result of part (d)

Problem 3 (30 pts)

Consider the Gaussian channel shown in Fig. 4 with

$$Y = X + Z,$$

where $Z \sim \mathcal{N}(0, N)$ is a Gaussian noise independent of X, and $X \in \mathbb{R}$ and $Y \in \mathbb{R}$ are the input and the output of the channel, respectively. Let $\rho : \mathcal{X} \to \mathbb{R}^+$ be a cost function for the channel input, and define the channel capacity for a given cost P as

$$C(P) = \max_{p(x):\mathbb{E}[\rho(X)]\leq P} I(X;Y).$$

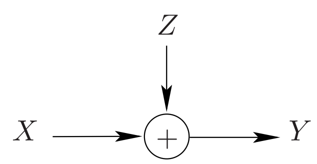

Figure 4: The Gaussian channel

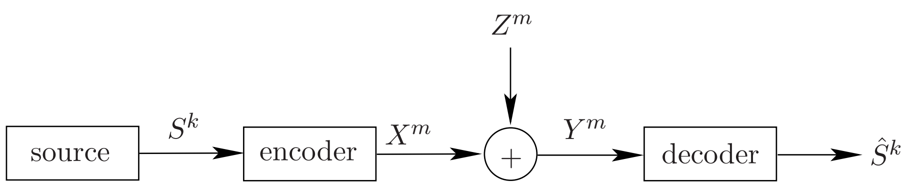

Figure 5: Transmission with distortion

The source in Fig 5 produces Gaussian symbols $S \sim \mathcal{N}(0, Q)$ with zero mean and variance Q. The encoder maps a sequence of length k of the source symbols to a sequence of length m of the channel input X^m, using the encoding function $X^m = f(S^k)$ and X^m is fed to the channel. The decoder uses the chanel output Y^m to estimate the source sequence $\hat{S}^k = g(Y^m)$. The quality of this reconstruction is measured by a distance function $d(\cdot,\cdot) : \mathcal{S} \times \hat{\mathcal{S}} \to \mathbb{R}^+$. The rate-distortion function for a given distortion D is defined as

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

(a) Prove that $R(D) \leq C(P)$. *[4pts]*

(b) Evaluate the functions $R(D)$ and $C(P)$ for $d(s, \hat{s}) = (s - \hat{s})^2$ and $\rho(x) = x^2$. Use the inequality $R(D) \leq C(P)$ to obtain a bound on D in terms of P, Q, and N. *[2pts]*

(c) Suppose we use $m = k = 1$ and take the source S_ℓ and scale to obtain the channel input which satisfies the power constraint, *i.e.,* $X_\ell = \alpha S_\ell$, such that $\mathbb{E}[X_\ell^2] = P$. Find the value of α. Given $Y_\ell = \alpha S_\ell + Z_\ell$, the decoder finds $\hat{S}_\ell = \beta Y_\ell$, such that $\mathbb{E}[(S_\ell - \hat{S}_\ell)^2]$ is minimized. Find the β, and show that $\mathbb{E}[(S_\ell - \hat{S}_\ell)^2] \leq D$. *[8pts]*
Hint: Compute $\mathbb{E}[(S_\ell - \hat{S}_\ell)^2]$ in terms of P, Q, N, and β. Take the derivative with respect to β.

The solution of part (b) shows that one achieve the optimal performance through uncoded transmission. In class, we have proved a source-channel separation which showed that it is optimal to separately encode the course and do channel coding on it. In parts (c), and (d) we show that even when separation holds, one can get very simple scheme if we combine source and channel coding, if the conditions of the *matching theorem* are satisfied.

The matching theorem states that using encoding and decoding functions of block length 1 ($m = k = 1$) is optimal if and only if the following conditions hold.

(i) $I(S, \hat{S}) = I(X; Y)$,

(ii) $\rho(x) = aD(p_{Y|X=x}(y|x) \parallel p_Y(y)) + b$, where a and b are constant, and $D(\cdot \parallel \cdot)$ denotes the Kullback-Leiber divergence,

(iii) $d(s, \hat{s}) = -c \log p_{S|\hat{S}}(s|\hat{s}) + d(s)$, where c is a constant and $d(s)$ is an arbitrary function of s (does not depend on $\hat{s}$).

Using this theorem, we seek conditions to have optimal code block length 1 for this problem.

(d) Let $\rho(x) = x^2$ be the cost function and $x = f(s) = \alpha' s$ be the encoding function. Show that condition (ii) is satisfied. *[8pts]*

(e) Let the distance function is $d(s, \hat{s}) = (s - \hat{s})^2$ and the decoding function is of the form $\hat{s} = g(y) = \beta' y$. For given $\alpha' = \sqrt{P/Q}$ find the value of β' such that condition (iii) is satisfied. Compare β' to β you have found in part (c). How can you explain it? *[8pts]*
Hint: You can use the facts that $\hat{S} \sim \mathcal{N}(0, \beta'^2(P + N))$, *i.e.,*

$$p_{\hat{S}}(\hat{s}) = \frac{1}{\sqrt{2\pi(P+N)}\beta'} \exp\left[-\frac{\hat{s}^2}{2\beta'^2(P+N)}\right],$$

and $\hat{S}|s \sim \mathcal{N}(\alpha'\beta' s, \beta'^2 N)$, *i.e,*

$$p_{\hat{S}|S}(\hat{s}|s) = \frac{1}{\sqrt{2\pi N}\beta'} \exp\left[-\frac{(\hat{s} - \alpha'\beta' s)^2}{2\beta'^2 N}\right].$$

Use the Bayes' rule

$$p_{S|\hat{S}}(s|\hat{s}) = \frac{p_{\hat{S}|S}(\hat{s}|s) p_S(s)}{p_{\hat{S}}(\hat{s})}.$$

to find $p_{S|\hat{S}}(s|\hat{s})$.

Problem 4 (20 pts)

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 $W_{Y|X=x}$ and P_Y as

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

(a) Given any distribution P_Y, show that for any choice of input distribution $\tilde{P}_X(x)$, *[7pts]*

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

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 *[8pts]*

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

and 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 *[5pts]*

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

and

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

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 iff $\tilde{P}_Y(y) = P_Y^*(y)$.