EE 376B
Information Theory
Prof. T. Cover

Handout #12
Tuesday, April 28, 2009

Solutions to Homework Set #3

1. **Multiple layer waterfilling**
   Let $C(x) = \frac{1}{2}\log(1+x)$ denote the channel capacity of a Gaussian channel with signal to noise ratio x. Show
$$C\left(\frac{P_1}{N}\right) + C\left(\frac{P_2}{P_1+N}\right) = C\left(\frac{P_1+P_2}{N}\right).$$
   This suggests that 2 independent users can send information as well as if they had pooled their power.

   Solution: Multiple layer waterfilling
$$\begin{aligned} C\left(\frac{P_1+P_2}{N}\right) &= \frac{1}{2}\log\left(1+\frac{P_1+P_2}{N}\right) \\ &= \frac{1}{2}\log\left(\frac{N+P_1+P_2}{N}\right) \\ &= \frac{1}{2}\log\left(\frac{N+P_1+P_2}{N+P_1}\cdot\frac{N+P_1}{N}\right) \\ &= \frac{1}{2}\log\left(\frac{N+P_1+P_2}{N+P_1}\right)+\frac{1}{2}\log\left(\frac{N+P_1}{N}\right) \\ &= C\left(\frac{P_2}{P_1+N}\right)+C\left(\frac{P}{N_1}\right) \end{aligned}$$

2. **Parallel channels and waterfilling**
   Consider a pair of parallel Gaussian channels, i.e.,
$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$
   where
$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right),$$
   and there is a power constraint $E(X_1^2+X_2^2) \le P$. Assume that $\sigma_1^2 > \sigma_2^2$.

(a) At what power does the channel stop behaving like a single channel with noise variance σ_2^2, and begin behaving like a pair of channels, ie., at what power does the worst channel become useful?

(b) What is the capacity $C(P)$ for large P?

Solution: Parallel channels and waterfilling

(a) By the result of Section 9.5 of Cover and Thomas, it follows that we will put all the signal power into the channel with less noise until the total power of noise + signal in that channel equals the noise power in the other channel. After that, we will split any additional power evenly between the two channels.

Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e., when $P = \sigma_1^2 - \sigma_2^2$.

(b) Let $E(X_1^2) = P_1$ and $E(X_2^2) = P_2$. Therefore

$$P = P_1 + P_2. \tag{1}$$

From waterfilling we know

$$P_2 = P_1 + \sigma_1^2 - \sigma_2^2. \tag{2}$$

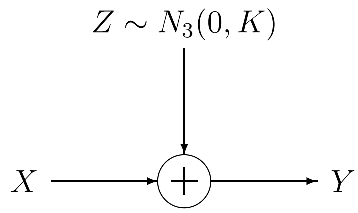
From equations (1) and (2) we get

$$\begin{aligned} P_1 &= \frac{P - (\sigma_1^2 - \sigma_2^2)}{2} \\ P_2 &= \frac{P + (\sigma_1^2 - \sigma_2^2)}{2}. \end{aligned}$$

Hence

$$C(P) = \frac{1}{2}\log\left(1 + \frac{P - (\sigma_1^2 - \sigma_2^2)}{2\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P + (\sigma_1^2 - \sigma_2^2)}{2\sigma_2^2}\right)$$

3. **Vector channel**
Consider the 3 input 3 output Gaussian channel



where $X, Y, Z \in \mathbb{R}^3$, $E||X||^2 = E(X_1^2 + X_2^2 + X_3^2) \leq P$, and $Z \sim N_3(0, K)$. Find the capacity for

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}.$$

Solution: Vector channel

We know that

$$C = \frac{1}{2} \log \left(\frac{|K_Y|}{|K|} \right)$$

where K_Y is the covariance matrix of the channel output. We can calculate the eigenvalues of the K matrix to be $\lambda_1 = 1$, $\lambda_2 = 1 - \rho$, and $\lambda_3 = 1 + \rho$. Hence $|K| = 1 - \rho$. We now need to maximize $|K_Y| = |K_X + K|$. From Section 9.5 of Cover and Thomas we see that

$$\sup_{K_X} |K_X + K| = \prod_{i=1}^{3} (A_i + \lambda_i),$$

where $A_i = (\nu - \lambda_i)^+$ and $A_1 + A_2 + A_3 = P$. We will first look at the case where $\rho > 0$. Hence we have

$$A_1 = \begin{cases} 0 & \text{if } P < \rho \\ \frac{P-\rho}{2} & \text{if } \rho \leq P < 3\rho \\ \frac{P}{3} & \text{if } 3\rho \leq P \end{cases}$$

$$A_2 = \begin{cases} P & \text{if } P < \rho \\ \frac{P+\rho}{2} & \text{if } \rho \leq P < 3\rho \\ \frac{P}{3} + \rho & \text{if } 3\rho \leq P \end{cases}$$

$$A_3 = \begin{cases} 0 & \text{if } P < \rho \\ 0 & \text{if } \rho \leq P < 3\rho \\ \frac{P}{3} - \rho & \text{if } 3\rho \leq P \end{cases}$$

Therefore,

$$\sup_{K_X} |K_X + K| = \begin{cases} (P+1-\rho)(1+\rho) & \text{if } P < \rho \\ \left(\frac{P-\rho}{2}+1\right)^2 (1+\rho) & \text{if } \rho \le P < 3\rho \\ \left(\frac{P}{3}+1\right)^3 & \text{if } P > 3\rho \end{cases}$$

and

$$C = \begin{cases} \frac{1}{2}\log\left(\frac{(P+1-\rho)(1+\rho)}{1-\rho^2}\right) & \text{if } P < \rho \\ \frac{1}{2}\log\left(\frac{\left(\frac{P-\rho}{2}+1\right)^2(1+\rho)}{1-\rho^2}\right) & \text{if } \rho \le P < 3\rho \\ \frac{1}{2}\log\left(\frac{\left(\frac{P}{3}+1\right)^3}{1-\rho^2}\right) & \text{if } P > 3\rho \end{cases}$$

Since if given a channel with parameter $\rho < 0$ we can negate Y_3 and get the same channel with a new parameter $\tilde{\rho} = -\rho > 0$, we know that the capacity for a channel with parameter ρ must be the same for the channel with parameter $-\rho$. Therefore

$$C = \begin{cases} \frac{1}{2}\log\left(\frac{(P+1-|\rho|)(1+|\rho|)}{1-\rho^2}\right) & \text{if } P < |\rho| \\ \frac{1}{2}\log\left(\frac{\left(\frac{P-|\rho|}{2}+1\right)^2(1+|\rho|)}{1-\rho^2}\right) & \text{if } |\rho| \le P < 3|\rho| \\ \frac{1}{2}\log\left(\frac{\left(\frac{P}{3}+1\right)^3}{1-\rho^2}\right) & \text{if } P > 3|\rho| \end{cases}$$

Note if the noise where white, i.e. $\rho = 0$ then

$$C = \frac{1}{2}\log((P/3+1)^3) = 3\frac{1}{2}\log(1+P/3)$$

as expected.

4. **Filtered noise**

We consider the previous vector channel where Z is filtered white noise

$$Z = BU,$$

where $U \sim N_3(0, I)$ is white Gaussian noise and

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{bmatrix}.$$

(a) Find the capacity.

(b) How would you signal over this channel?

Solution: Filtered noise

(a) First observe that $|B| = 0$ and therefore $|K| = 0$. From the capacity equation

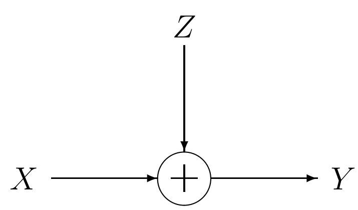
$$C = \frac{1}{2}\log\left(\frac{|K_Y|}{|K|}\right),$$

it is clear that $C = \infty$.

(b) Observe that $Z_3 = Z_1 + Z_2$. One way to exploit this structure in the noise is to send nothing at time 1 and 2, i.e. $X_1 = X_2 = 0$. Therefore the receiver knows Z_1 and Z_2 and hence also knows Z_3. The receiver can then decode Y_3 perfectly by noting that $X_3 = Y_3 + Z_2 + Z_1$.

5. **A mutual information game**

Consider the following channel:



Throughout this problem we shall constrain the signal power

$$EX = 0, \qquad EX^2 = P,$$

and the noise power

$$EZ = 0, \qquad EZ^2 = N,$$

and assume that X and Z are independent. The channel capacity is given by $I(X; X + Z)$.

Now for the game. The noise player chooses a distribution on Z to minimize $I(X; X + Z)$, while the signal player chooses a distribution on X to maximize $I(X; X + Z)$.

Letting $X^* \sim \mathcal{N}(0, P)$, $Z^* \sim \mathcal{N}(0, N)$, show that Gaussian X^* and Z^* satisfy the saddlepoint conditions

$$I(X; X + Z^*) \le I(X^*; X^* + Z^*) \le I(X^*; X^* + Z).$$

Thus

$$\min_Z \max_X I(X; X + Z) = \max_X \min_Z I(X; X + Z) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right),$$

and the game has a value. In particular, a deviation from normal for either player worsens the mutual information from that player's standpoint. Can you discuss the implications of this?

Note: Part of the proof hinges on the entropy power inequality from Chapter 16, which states that if $\mathbf{X}$ and $\mathbf{Y}$ are independent random n-vectors with densities, then

$$e^{\frac{2}{n}h(\mathbf{X}+\mathbf{Y})} \geq e^{\frac{2}{n}h(\mathbf{X})} + e^{\frac{2}{n}h(\mathbf{Y})}.$$

Solution: A mutual information game

Let X and Z be random variables with $EX = 0$, $EX^2 = P$, $EZ = 0$ and $EZ^2 = N$. Let $X^* \sim \mathcal{N}(0, P)$ and $Z^* \sim \mathcal{N}(0, N)$. Then as proved in class,

$$\begin{aligned}
I(X; X + Z^*) &= h(X + Z^*) - h(X + Z^*|X) \\
&= h(X + Z^*) - h(Z^*) \\
&\leq h(X^* + Z^*) - h(Z^*) \\
&= I(X^*; X^* + Z^*),
\end{aligned}$$

where the inequality follows from the fact that given the variance, the entropy is maximized by the normal.

To prove the other inequality, we use the entropy power inequality

$$2^{2h(X+Z)} \geq 2^{2h(X)} + 2^{2h(Z)}$$

with equality if X and Z are independent normal.

Now we have

$$\begin{aligned}
I(X^*; X^* + Z) &= h(X^* + Z) - h(X^* + Z|X^*) \\
&= h(X^* + Z) - h(Z) \\
&\geq \frac{1}{2}\log\left(2^{2h(X^*)} + 2^{2h(Z)}\right) - h(Z) \\
&= \frac{1}{2}\log\left(1 + \frac{2^{2h(X^*)}}{2^{2h(Z)}}\right) \\
&\geq \frac{1}{2}\log\left(1 + \frac{2^{2h(X^*)}}{2^{2h(Z^*)}}\right) \\
&= \frac{1}{2}\log\left(2^{2h(X^*)} + 2^{2h(Z^*)}\right) - h(Z^*) \\
&= h(X^* + Z^*) - h(Z^*) \\
&= I(X^*; X^* + Z^*),
\end{aligned}$$

where the first inequality follows from the entropy power inequality and the second inequality follows from the fact that $g(\xi) = \frac{1}{2}\log\left(1 + \frac{2^{2h(X^*)}}{2^{2\xi}}\right)$ is a nonincreasing function and $h(Z) \leq h(Z^*)$.

Combining the two inequalities, we have

$$I(X; X + Z^*) \le I(X^*; X^* + Z^*) \le I(X^*; X^* + Z).$$

Hence, using these inequalities, it follows directly that

$$\begin{aligned} \min_Z \max_X I(X; X + Z) &\le \max_X I(X; X + Z^*) \\ &= I(X^*; X^* + Z^*) \\ &= \min_Z I(X^*; X^* + Z) \\ &\le \max_X \min_Z I(X^*; X^* + Z). \end{aligned} \tag{3}$$

We have shown an inequality relationship in one direction between $\min_Z \max_X I(X; X + Z)$ and $\max_X \min_Z I(X; X+Z)$. We will now prove the inequality in the other direction is a general result for all functions of two variables.

For any function $f(a, b)$ of two variables, for all b, for any a_0,

$$f(a_0, b) \ge \min_a f(a, b).$$

Hence

$$\max_b f(a_0, b) \ge \max_b \min_a f(a, b).$$

Taking the minimum over a_0, we have

$$\min_{a_0} \max_b f(a_0, b) \ge \min_{a_0} \max_b \min_a f(a, b).$$

or

$$\min_a \max_b f(a, b) \ge \max_b \min_a f(a, b).$$

From this result,

$$\min_Z \max_X I(X; X + Z) \ge \max_X \min_Z I(X; X + Z). \tag{4}$$

From (3) and (4), we have

$$\begin{aligned} \min_Z \max_X I(X; X + Z) &= \max_X \min_Z I(X; X + Z) \\ &= \frac{1}{2} \log \left(1 + \frac{P}{N}\right). \end{aligned}$$

This inequality implies that we have a saddlepoint in the game, which is the value of the game. If signal player chooses X^*, the noise player cannot do any better than choosing Z^*. Similarly, any deviation by the signal player from X^* will make him do worse, if the noise player has chosen Z^*. Any deviation by either player will make him do worse.

Another implication of this result is that not only is the normal the best possible signal distribution, it is the worst possible noise distribution.

6. **Additive noise channel**

This problem has an instructive answer. Consider the channel $Y = X + Z$, where X is the transmitted signal with power constraint P, Z is independent additive noise, and Y is the received signal. Let

$$Z = \begin{cases} 0, & \text{with prob. } 1/10 \\ Z^*, & \text{with prob. } 9/10 \end{cases},$$

where $Z^* \sim N(0, N)$. Thus Z has a mixture distribution which is the mixture of a Gaussian distribution and a degenerate distribution with mass 1 at 0.

(a) What is the capacity of this channel?

(b) How would you signal in such a manner as to achieve capacity?

Solution: Additive noise channel

(a) The capacity of the channel is in fact infinite. Since Z has a discrete component, the differential entropy $h(Z) = -\infty$. Now choose any distribution for X such that $Y = X + Z$ has no atoms, e.g. $X \sim U[0, \sqrt{P}]$ will suffice. Since $h(Y) > -\infty$,

$$C \geq h(Y) - h(Z) \Rightarrow C = \infty$$

(a) Many different signalling schemes are possible. A rather simple one is for the transmitter to pick a rational number (between $-\sqrt{P}$ and $\sqrt{P}$) and transmit it. Since there are countably many rational numbers in $(-\sqrt{P}, \sqrt{P})$, $\mathbf{P}(\eta \in \mathcal{Q}) = 0$ where η is the standard normal random variable. Therefore if the receiver gets a rational number, she can correctly conclude that $Z = 0$ w.p.1. Now the transmitter transmits the same rational number repeatedly for n times so that with probability $1 - (9/10)^n$, the receiver can get a rational number at least once and hence decode the message correctly. This immediately implies that the achievable rate of this signalling scheme is infinite since there are infinitely many rational numbers that can be transmitted in this way.

Note: This scheme is impractical since there is no finite-time algorithm to determine whether a real number is rational or not. We can, however, modify the above scheme to one in which the signal set is an arbitrary finite subset of rational numbers. It is easy to see that the achievable rate is unbounded and we still have the infinite capacity.

7. **Estimation.**
Here is the estimation counterpart to Fano's inequality. Let X be a random variable with differential entropy $h(X)$. Let $\hat{X}$ be an estimate of X, and let $E(X-\hat{X})^2$ be the expected prediction error.

Given side information Y and estimator $\hat{X}(Y)$, show

$$E(X-\hat{X}(Y))^2 \geq \frac{1}{2\pi e} e^{2h(X|Y)}.$$

Solution: Estimation
Let $\hat{X}$ be any estimator of X (without involving Y yet). Then

$$\begin{aligned}
\mathbf{E}(X-\hat{X})^2 &\geq \min_{\hat{X}} \mathbf{E}(X-\hat{X})^2 \\
&= \mathbf{E}\left(X-\mathbf{E}(X)\right)^2 \\
&= \mathrm{Var}(X) \\
&\geq \frac{1}{2\pi e} e^{2h(X)},
\end{aligned}$$

where the last inequality follows from the fact that the Gaussian distribution has the maximum entropy for a given variance.

Conditioning on $Y = y$ one obtains from the previous part that

$$\mathbf{E}\left(X-\hat{X}(y)\right)^2 \geq \frac{1}{2\pi e} e^{2h(X|Y=y)}$$

We take expectation with respect to Y and use Jensen's inequality (e^x is convex) to conclude the proof.