EE/Ma 126b Information Theory - Homework Set #5

Ling Li, ling@cs.caltech.edu

February 20, 2001

5.1 *A channel with two independent looks at Y.*

(a) Since Y_1 and Y_2 are conditionally independent and conditionally identically distributed given X, we have

$$H(Y_1, Y_2|X) = H(Y_1|X) + H(Y_2|X) = 2H(Y_1|X),$$

and $p_{Y_1|X}(y|x) = p_{Y_2|X}(y|x)$. Thus

$$p_{Y_1}(y) = \sum_x p_{Y_1|X}(y|x)p(x) = \sum_x p_{Y_2|X}(y|x)p(x) = p_{Y_2}(y).$$

Therefore $H(Y_1) = H(Y_2)$. Then we have

$$\begin{aligned} I(X; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\ &= (H(Y_1) + H(Y_2) - I(Y_1; Y_2)) - 2H(Y_1|X) \\ &= 2\,(H(Y_1) - H(Y_1|X)) - I(Y_1; Y_2) \\ &= \boxed{2I(X; Y_1) - I(Y_1; Y_2)}. \end{aligned} \tag{1}$$

(1) also holds for continuous channels. (All changes needed for the proof is replacing symbols H, p, $\sum_x$ with their continuous versions h, f, and $\int_S dx$.)

(b) For the first channel, the capacity is $C_1 = \max_{p(x)} I(X; Y_1, Y_2)$. For the second channel, the capacity is $C_2 = \max_{p(x)} I(X; Y_1)$. Thus from (1),

$$\begin{aligned} C_1 &= \max_{p(x)} I(X; Y_1, Y_2) \\ &= \max_{p(x)} (2I(X; Y_1) - I(Y_1; Y_2)) \\ &\le 2 \max_{p(x)} I(X; Y_1) - \min_{p(x)} I(Y_1; Y_2) \\ &\le 2C_2. \end{aligned}$$

5.2 *The two-look Gaussian channel.* I am not sure whether the power constraint implies $EX = 0$. So let $\hat{X} = X - EX$ and $\hat{Y}_i = Y_i - EX$. If we assume $EX = 0$, the $\hat{X}$ and X, $\hat{Y}_i$ and Y_i are just the same in this solution.

Since X and Z_i are independent, $\hat{X}$ and Z_i are also independent. Then for $1 \le i, j \le 2$,

$$\begin{aligned} E(\hat{Y}_i \hat{Y}_j) &= E\left[(\hat{X} + Z_i)(\hat{X} + Z_j)\right] \\ &= E(\hat{X}^2) + E(\hat{X})E(Z_i) + E(\hat{X})E(Z_j) + E(Z_i Z_j) \\ &= E(\hat{X}^2) + E(Z_i Z_j). \end{aligned}$$

Thus the covariance of $(\hat{Y}_1, \hat{Y}_2)$ is

$$\hat{K} = E\begin{bmatrix} \hat{Y}_1^2 & \hat{Y}_1\hat{Y}_2 \\ \hat{Y}_1\hat{Y}_2 & \hat{Y}_2^2 \end{bmatrix} = \begin{bmatrix} E\hat{X}^2 + N & E\hat{X}^2 + N\rho \\ E\hat{X}^2 + N\rho & E\hat{X}^2 + N \end{bmatrix}.$$

Since $E(\hat{Y}_1, \hat{Y}_2) = \mathbf{0}$ and translation doesn't change the entropy, we have

$$h(Y_1, Y_2) = h(\hat{Y}_1, \hat{Y}_2) \le \frac{1}{2}\log(2\pi e)^2\left[N^2(1-\rho^2) + 2N(1-\rho)E\hat{X}^2\right],$$

with equality iff $(\hat{Y}_1, \hat{Y}_2) \sim \mathcal{N}_2(\mathbf{0}, \hat{K})$. Here $\left[N^2(1-\rho^2) + 2N(1-\rho)E\hat{X}^2\right]$ is the determinant of $\hat{K}$.

From $E\hat{X}^2 = EX^2 - (EX)^2 \le P - (EX)^2$ and $h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2 N^2(1-\rho^2)$, we get

$$\begin{aligned} I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= h(Y_1, Y_2) - h(X + Z_1, X + Z_2|X) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2) \\ &\le \frac{1}{2}\log\frac{\left[N^2(1-\rho^2) + 2N(1-\rho)E\hat{X}^2\right]}{N^2(1-\rho^2)} \\ &= \frac{1}{2}\log\left[1 + \frac{2EX^2 - 2(EX)^2}{N(1+\rho)}\right] \\ &\le \frac{1}{2}\log\left[1 + \frac{2P}{N(1+\rho)}\right], \end{aligned}$$

with equality iff $EX = 0$ and $E\hat{X}^2 = P$ and $(\hat{Y}_1, \hat{Y}_2) \sim \mathcal{N}_2(\mathbf{0}, \hat{K})$, i.e., $X \sim \mathcal{N}(0, P)$. Thus the channel capacity is

$$\boxed{C = \max_{p(x): EX^2 \le P} I(X; Y_1, Y_2) = \frac{1}{2}\log\left[1 + \frac{2P}{N(1+\rho)}\right]}.$$

(a) $\rho = 1$, $\boxed{C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)}$. Just as a normal Gaussian channel with only 1 look.

(b) $\rho = 0$, $\boxed{C = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right)}$.

(c) $\rho = -1$, $\boxed{C = \infty}$. Since now $Z_2 = -Z_1$, we can get $X = \frac{1}{2}(Y_1 + Y_2)$. Thus all the bits of X are transfered through the channel and therefore the capacity is infinity.

5.3 *Parallel channels and waterfilling.* The capacity of such parallel channels is

$$C = \frac{1}{2}\log\left[1 + \frac{(\nu - \sigma_1^2)^+}{\sigma_1^2}\right] + \frac{1}{2}\log\left[1 + \frac{(\nu - \sigma_2^2)^+}{\sigma_2^2}\right],$$

where $P_i = (\nu - \sigma_i^2)^+$ is the power assigned to source X_i, and ν is chosen so that $P_1 + P_2 = 2P$. When $\nu \le \sigma_1^2$, the two channels behave just like a single channel, since now $P_1 = (\nu - \sigma_1^2)^+ = 0$ — no power is used for source X_1. When $\nu > \sigma_1^2$, since we assume $\sigma_1^2 > \sigma_2^2$, we have $P_2 > P_1 > 0$ and the channels behave like a pair of channels. Thus, when $\nu = \sigma_1^2$ and $\boxed{2P = \sigma_1^2 - \sigma_2^2}$, the channels stop behaving like a single channel.

5.4 *Slepian-Wolf.* Let q denote $p_Y(1)$. Since X and Z are independent and $Y = X \oplus Z$, we get

$$q = p_X(0)p_Z(1) + p_X(1)p_Z(0) = (1-p)r + p(1-r). \tag{2}$$

So $H(Y) = H(q)$. From

$$\begin{aligned} H(Y|X) &= H(X \oplus Z|X) = H(Z|X) = H(Z) = H(r), \\ H(X,Y) &= H(X) + H(Y|X) = H(p) + H(r), \end{aligned}$$

we get

$$H(X|Y) = H(X,Y) - H(Y) = H(p) + H(r) - H(q).$$

Thus the region of rates allowing recovery of $\mathbf{X}, \mathbf{Y}$ is defined by

$$\begin{aligned} R_1 &> H(X|Y) = H(p) + H(r) - H(q), \\ R_2 &> H(Y|X) = H(r), \\ R_1 + R_2 &> H(X,Y) = H(p) + H(r). \end{aligned}$$

Note that from (2), q is between p and $(1-p)$ since it is a weighted sum of p and $(1-p)$ and weights are non-negative. Therefore $H(q) \geq H(p)$ with equality iff $r = 0, 1$. We also have $H(q) \geq H(r)$. So $H(X|Y) \leq H(p)$ and $H(X|Y) \leq H(r)$.

5.5 *A mutual information game.* We have known that for a Gaussian channel with noise Z^*,

$$I(X; X + Z^*) \leq I(X^*; X^* + Z^*) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right).$$

The entropy power inequality says (Theorem 16.6.3 in Cover's book): if $\mathbf{X}$ and $\mathbf{Y}$ are independent random n-vectors with densities, then

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

Let $n = 1$, $\mathbf{X} = X^*$ and $\mathbf{Y} = Z$. We get $2^{2h(X^*+Z)} \geq 2^{2h(X^*)} + 2^{2h(Z)}$, or

$$2^{2[h(X^*+Z)-h(Z)]} \geq 2^{2[h(X^*)-h(Z)]} + 1.$$

Since 2^x is a monotonic increasing function of x, and $h(Z) \leq h(Z^*)$, we get

$$2^{2[h(X^*+Z)-h(Z)]} \geq 2^{2[h(X^*)-h(Z^*)]} + 1 = \frac{P}{N} + 1.$$

Taking the logarithm,

$$h(X^* + Z) - h(Z) \geq \frac{1}{2}\log\left(1 + \frac{P}{N}\right).$$

Since X^* and Z are independent,

$$I(X^*; X^* + Z) = h(X^* + Z) - h(X^* + Z|X^*) = h(X^* + Z) - h(Z),$$

we finally have

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

Thus

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

Generally, we have

$$\min_x \max_y f(x,y) \ge \max_y \min_x f(x,y),$$

since for any x', we have $\max_y f(x',y) \ge \max_y \min_x f(x,y)$. Thus

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

From (3) and (4), we have

$$\min_Z \max_X I(X;X+Z) = \max_X \min_Z I(X;X+Z) = \frac{1}{2}\log\left(1+\frac{P}{N}\right). \tag{5}$$

For a game, any player (say, X) wants to maximize his[*] 'award' ($I(X;X+Z)$) in any situation. However, since a game is played simultaneously by several players, (in this game, 2 players,) he must take into consideration other players' actions. Thus, he would suppose other players are also very smart and thus what he eventually could get is the minimum maximum award ($\min_Z \max_X I(X;X+Z)$), where maximum is over all his actions and minimum is over all possible situations. If 'award' is replaced by 'punishment', then the player expects to get a maximum minimum 'punishment'. Usually, in a 2-player game, the award of one player is the punishment of the other player. Thus we will have balance like (5).

For this game, either player has an optimal strategy: using normal distribution. By using such strategies, they achieve the balance between them. If either of them deviates from the optimal strategy, the 'award' decreases, which means the mutual information decreases, from that player's standpoint.

[*] For convenience, I will use he instead of she/he.