**ECE 776 - Information theory**
**Final (Fall 2008)**

**Q.1**. (1 point) Consider the following bursty transmission scheme for a Gaussian channel with noise power $N$ and average power constraint $P$ (i.e., $1/n\sum_{i=1}^{n} x_i^2(w) \leq P$ for all messages $w$). The transmitter sends "0" except at times 0, $k$, $2k$, $3k$,... for a given integer $k > 0$. Notice that in these time instants, the power can be increased to $kP$ without violating the average power constraint. Find the maximum achievable rate with this type of schemes. Then, consider $P$ to be very small. Verify that in this regime the considered bursty scheme is optimal (i.e., it is equal to the capacity of the Gaussian channel).

*Sol.*: The channel at hand is equivalent to a Gaussian channel with $n/k$ channel uses and average power constraint $kP$ so that the maximum achievable rate (bits per total channel use) is

$$\begin{aligned} R &= \frac{1}{n}\cdot\frac{n}{k}\frac{1}{2}\log_2(1+kP) \\ &= \frac{1}{2k}\log_2(1+kP). \end{aligned}$$

Now, for small $P$, $\log_2(1+kP) = \log_2 e \log(1+kP) \simeq (\log_2 e)\cdot kP$, so that

$$R \simeq \frac{1}{2}(\log_2 e)\cdot P.$$

Comparing this with the capacity $C = 1/2\log_2(1+P) \simeq 1/2(\log_2 e)\cdot P$, we see that the bursty coding scheme suffers no loss of optimality in the low-SNR regime.

**Q.2.** (1 point) The histogram (or type) of a given sequence $x^n \in \mathcal{X}^n$ is defined as $N(a|x^n) = 1/n \cdot \sum_{i=1}^{n} 1\{x_i = a\}$, where $1\{\cdot\}$ is an indicator function (equal to 1 if the argument is true and zero otherwise).
(a) Show that the number of all possible types for a sequence of $n$ letters is less or equal than $(n+1)^{|\mathcal{X}|}$ (*Hint*: How many values can the function $N(a|x^n)$ take for every possible value $a \in \mathcal{X}$?).
(b) Based on the previous result, show that transmitting information about the type of a sequence requires a rate (bit/ symbol) that vanishes ($\rightarrow 0$) for $n \rightarrow \infty$.
(c) (*optional*) Consider now the case of two sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ observed by two independent encoders. Assume that the receiver wants to recover the joint type $N(a, b|x^n, y^n)$, do you think (*intuitively*) that a similar conclusion to point (b) would apply?

*Sol.*: (a) For every possible value $a \in \mathcal{X}$, the function $N(a|x^n)$ can take only $n+1$ values, namely $0, 1/n, 2/n, ..., 1$. Therefore, an upper bound to the number of functions $N(a|x^n)$ is obtained by assuming that every value $N(a|x^n)$ is selected from this set with replacement. It follows that $(n+1)^{|\mathcal{X}|}$ is an upper bound.
(b) Since the number of histograms (types) is less or equal than $(n+1)^{|\mathcal{X}|}$, the required rate per symbol is

$$R \leq \frac{\log_2(n+1)^{|\mathcal{X}|}}{n} \rightarrow 0 \text{ for } n \rightarrow \infty.$$

(c) (optional) The number of joint types, following point (a), is clearly less or equal than $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$. However, since $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ are observed by two independent encoders, one cannot directly index the joint type with a vanishing rate as at point (b). In fact, encoder 1 only knows the type $N(a|x^n)$ and similarly encoder 2 only knows $N(b|y^n)$. Therefore, it is expected that a vanishing rate would not be enough here. It turns out that in many cases one has to choose rates so that the entire sequences $x^n$ and $y^n$ (not only their type, which is what is actually requested) can be recovered at the destination. This corresponds to the Slepian-Wolf rate region. Further details on this can be found in the paper: R. Ahlswede and I. Csiszar, "To get a bit of information may be as hard as to get full information," IEEE Trans. Inf. Theory, vol. IT-27, no. 4, July 1981.

**Q.3**. (1 point) Consider a Gaussian multiple access channel in which the received signal is $Y = [Y_1\ Y_2]^T$ with

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix},$$

with independent Gaussian noises $N_1$ and $N_2$ with unit power. The average power constraints on $X_1$ and $X_2$ are $P_1$ and $P_2$, respectively.
(a) Find the capacity region.
(b) Consider the capacity region for the case where the two transmitters can perfectly cooperate, that is, $R_1 + R_2 \leq \max_{f(x_1,x_2):\ E[X_1^2]\leq P_1,\ E[X_2^2]\leq P_2} I(X_1, X_2; Y)$. Compare with point (a) and comment.

*Sol*: (a) We have

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) = h(Y|X_2) - h(Y|X_1, X_2) \\ &= h(Y_1, N_2|X_2) - h(N_1, N_2) \\ &= h(Y_1) + h(N_2) - h(N_1) - h(N_2) \\ &\leq \frac{1}{2}\log_2(1 + P_1), \end{aligned}$$

where in the third line we have used the fact that $N_1$, $N_2$, $X_1$ and $X_2$ are independent, and the last inequality is an equality if $X_1$ is Gaussian. We similarly have that the second bound $R_2 \leq I(X_2; Y|X_1)$ is maximized by a Gaussian distribution for $X_2$. Finally, for the sum-rate

$$\begin{aligned} R_1 + R_2 &\leq I(X_1, X_2; Y) = \\ &= h(Y_1) + h(Y_2) - h(N_1) - h(N_2) \\ &\leq \frac{1}{2}\log_2(1 + P_1) + \frac{1}{2}\log_2(1 + P_2), \end{aligned}$$

which is again maximized by Gaussian distributions. We have concluded that the capacity region is given by the rectangle $R_1 \leq \frac{1}{2}\log_2(1+P_1)$, $R_2 \leq \frac{1}{2}\log_2(1+P_2)$. This is not surprising since the two users have orthogonal channels to the destination (and, in fact, we could have concluded the above by this simple observation alone).
(b) For the perfect cooperation case

$$\begin{aligned} R_1 + R_2 &\leq I(X_1, X_2; Y) \\ &\leq h(Y_1) + h(Y_2) - h(N_1) - h(N_2) \\ &\leq \frac{1}{2}\log_2(1 + P_1) + \frac{1}{2}\log_2(1 + P_2), \end{aligned}$$

where the second inequality is tight if $X_1$ and $X_2$ are independent, i.e., $f(x_1,x_2) = f(x_1)f(x_2)$. The full cooperation region then contains rate points not contained in the multiple-access region derived at the previous point.

**Q.4.** (1 point) Assume that you have two parallel Gaussian channels with inputs $X_1$ and $X_2$ and noises $Z_1$ and $Z_2$, respectively. Assume that the noise powers are $E[Z_1^2] = 0.5$ and $E[Z_2^2] = 0.7$, while the total available power is $P = E[X_1^2] + E[X_2^2] = 0.4$. Find the optimal power allocation and corresponding capacity.

*Sol.*: The waterfilling conditions are:

$$\begin{aligned} P_1 &= E[X_1^2] = (\mu - 0.5)^+ \\ P_2 &= E[X_2^2] = (\mu - 0.7)^+ \end{aligned}$$

with

$$P_1 + P_2 = 0.4.$$

It follows that

$$(\mu - 0.5)^+ + (\mu - 0.7)^+ = 0.4,$$

which is satisfied by $\mu = 0.8$, yielding $P_1 = 0.3$ and $P_2 = 0.1$
The corresponding capacity is then

$$C = \frac{1}{2}\log_2\left(1 + \frac{0.3}{0.5}\right) + \frac{1}{2}\log_2\left(1 + \frac{0.1}{0.7}\right) = 0.435 \text{ bits/ channel use.}$$

**P.1.** (2 points) Consider the network in Fig. 1-(a), where two nodes communicate with each other, with rates $R_1$ ($W_1 \in [1, 2^{nR_1}]$) from node 1 to 2, and $R_2$ ($W_2 \in [1, 2^{nR_2}]$) from 2 to 1. Available coding strategies are of the type $X_{1,k} = f_{1,k}(W_1, Y_1^{k-1})$ and $X_{2,k} = f_{2,k}(W_2, Y_2^{k-1})$, while the decoding functions are $\hat{W}_1 = g_1(Y_1^n)$ and $\hat{W}_2 = g_2(Y_2^n)$. This is the two-way channel studied by Shannon.
($a$) Evaluate the cut-set outer bound to the capacity region of the rates $(R_1, R_2)$.
($b$) Consider now the special case of the network of Fig. 1-(a) shown in Fig. 1-(b). Here we have $Y_{1,k} = Y_{2,k} = X_{1,k} \oplus X_{2,k}$ with $X_{1,k}, X_{2,k} \in \{0,1\}$ (binary inputs) and $\oplus$ represents the mod-2 (binary) sum. Evaluate the capacity region for this network (*Hint*: first evaluate the cut-set bound for this network and then propose a simple scheme that achieves the cut-set bound).

*Sol.*:
(a) Only two cuts are possible. Accordingly, we get that if $(R_1, R_2)$ are achievable, then

$$\begin{aligned} R_1 &\leq I(X_1; Y_2|X_1) \\ R_2 &\leq I(X_2; Y_1|X_2) \end{aligned}$$

for some joint distribution $p(x_1, x_2)$.
(b) Let us evaluate the bound above for the network in Fig. 1-(b)

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) = I(X_1; X_1|X_2) = H(X_1|X_2) \leq H(X_1) \leq 1 \\ R_2 &\leq I(X_2; Y|X_1) = I(X_2; X_2|X_1) = H(X_2|X_1) \leq H(X_2) \leq 1 \end{aligned}$$

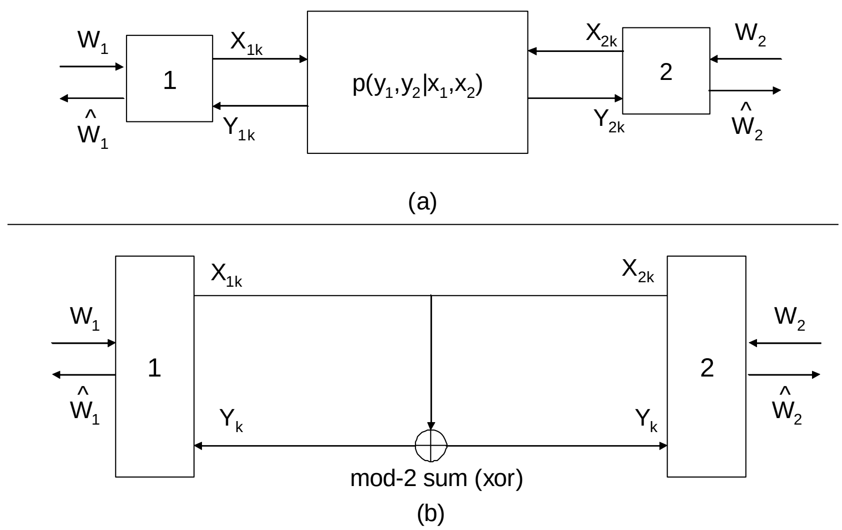


Figure 1:

where equality in both bounds is obtained for $p(x_1, x_2) = p(x_1)p(x_2)$ with $p(x_1)$ and $p(x_2)$ uniform.
A simple scheme that achieves this bound is the following. Every user $j$ transmits $x_{j,k} = 0$ or 1 with equal probability. Having received, say for user 1, $y_{1,k} = x_{1,k} \oplus x_{2,k}$, the transmitted bit $x_{1,k}$ can be clearly subtracted off and user 1 obtains $y_{1,k} \oplus x_{1,k} = x_{2,k}$. As a result, $R_1 \leq 1$ and $R_2 \leq 1$ are achievable.

**P.2.** (2 points) Consider a source $V_i$, $i = 1, 2, ..., n$, i.i.d. and uniformly distributed in the set $\mathcal{V}$ =$\{1, 2, 3, 4, 5\}$. We are interested in conveying this source to a destination over a Gaussian channel with signal to noise ratio $P/N = 10$. We use as measure for the quality of reconstruction the Hamming distortion (symbol-wise probability of error)

$$d(v, \hat{v}) = \begin{cases} 0 & \text{if } v = \hat{v} \\ 1 & \text{if } v \neq \hat{v} \end{cases}.$$

(a) What is the minimum average distortion $D = 1/n \cdot \sum_{i=1}^{n} E[d(v_i, \hat{v}_i)]$ that can be obtained? Provide a rough estimate of the solution of the final equation by plotting the two sides (*Hint*: What is an optimal scheme to convey the source to the destination?)
(b) Fix distortion $D = 0$ and $P/N = 2$. Assume now that the channel has $k$ channel uses per source symbol (rather than one, as assumed at the previous point). In other words, for every source symbol, the channel can send $k$ channel symbols to the destination. What is the minimum value of $k$ that is necessary to achieve the desired distortion over this channel?

*Sol.*: (a) Using the source-channel separation theorem, we have that the minimum average distortion $D$ is given by the equation

$$R(D) = C,$$

where $C = 1/2 \log_2(1+10)$ and $R(D)$ is the rate-distortion function for the source at hand. It can be derived (see problem 10.5 and assignments) that

$$\begin{aligned} R(D) &= \log_2 5 - H(D) - D \log_2(5-1) \\ &= \log_2 5 - H(D) - 2D \end{aligned}$$

if $0 \leq D \leq 4/5$, and

$$R(D) = 0,$$

if $D > 4/5$. We then need to solve (assuming there is a solution in the interval $0 \leq D \leq 4/5 = 0.8$):

$$\begin{aligned} \log_2 5 - H(D) - 2D &= 1/2 \log_2(1+10) \\ H(D) + 2D &= \log_2(5/\sqrt{11}), \end{aligned}$$

which turns out to lead to $D \simeq 0.085 < 0.8$.
(b) Since $D = 0$, we need to convey $R(0) = H(V) = \log_2 5 \simeq 2.32$ bits/ source symbol to the destination. Via a capacity-achieving channel code, we can send $1/2 \log_2(1+2) \simeq 0.79$ bits/ channel use. But, since we have $k$ channel uses per source symbol, the number of bits we can send to the destination per source symbol is $\simeq k \cdot 0.79$. Setting $k \cdot 0.79 \geq 2.32$ we get a minimum value of $k$, $k = \lceil 2.32/0.79 \rceil = 3$.

**P.3.** (2 points) You are given two independent i.i.d. sources $X_{1,i}$ and $X_{2,i}$, $i = 1, 2, ..., n$, such that $X_{1,i} \sim Ber(0.5)$ and $X_{2,i} \sim Ber(0.2)$.
(a) Suppose that the sequences $X_1^n$ and $X_2^n$ are observed by two independent encoders and that a common destination is interested in recovering losslessly $X_1^n$ and $X_2^n$. What is the set of rates $R_1$ (from the first encoder to the destination) and $R_2$ (from the second encoder to the destination) that guarantees vanishing probability of error at the destination (Slepian-Wolf problem)?
(b) Repeat the previous point by considering the sources $Z_{1,i} = X_{1,i} + X_{2,i}$ and $Z_{2,i} = X_{1,i} - X_{2,i}$. In other words, here the two encoders observe $Z_1^n$ and $Z_2^n$. Compare the rate region with that found at the previous point.
(c) Comment on the comparison of points (a) and (b).

*Sol.*: (a) Since $X_1$ and $X_2$ are independent, we have

$$\begin{aligned} R_1 &\geq H(X_1|X_2) = H(0.5) = 1 \\ R_2 &\geq H(X_2|X_1) = H(0.2) \simeq 0.72 \\ R_1 + R_2 &\geq H(X_1, X_2) = H(0.2) + H(0.5) \simeq 1.72, \end{aligned}$$

so the region is the intersection of the half spaces $R_1 \geq 1$ and $R_2 \geq 0.72$.
(b) We need to find the joint distribution $p(x_1, x_2)$ of $Z_1$ and $Z_2$, which is given by

| $Z_1 \backslash Z_2$ | $0$ | $1$ | $-1$ |
|---|---|---|---|
| $0$ | $0.5 \cdot 0.8$ | $0$ | $0$ |
| $1$ | $0$ | $0.5 \cdot 0.8$ | $0.5 \cdot 0.2$ |
| $2$ | $0.5 \cdot 0.2$ | $0$ | $0$ |

,

since, e.g., $\Pr[Z_1 = 0, Z_2 = 0] = \Pr[X_1 = 0, X_2 = 0] = 0.5 \cdot 0.8$. We can now calculate the Slepian-Wolf region:

$$\begin{aligned} R_1 &\geq H(Z_1|Z_2) = \Pr[Z_2 = 0]H(Z_1|Z_2 = 0) = 0.5 \cdot H(0.2) \simeq 0.36 \\ R_2 &\geq H(Z_2|Z_1) = \Pr[Z_2 = 1]H(Z_2|Z_1 = 1) = 0.5 \cdot H(0.2) \simeq 0.36 \\ R_1 + R_2 &\geq H(Z_1, Z_2) = -2 \cdot 0.1 \log_2 0.1 - 2 \cdot 0.4 \log_2 0.4 \simeq 1.72. \end{aligned}$$

It is easy to see that this region is larger than the one found at point (a) in terms of the individual rates $R_1$ and $R_2$ (the sum-rate constraint is the same). This is because here the sources are correlated and thus one can exploit the gains arising from distributed coding to reduce rates $R_1$ and $R_2$.