Homework 6 Solutions

Problem 1

5) **Using two channels at once.** Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1 \mid x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2 \mid x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1 \mid x_1) \times p(y_2 \mid x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of this channel.

5) *Using two channels at once.* Suppose we are given two channels, $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$, which we can use at the same time. We can define the product channel as the channel, $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$. To find the capacity of the product channel, we must find the distribution $p(x_1, x_2)$ on the input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ that maximizes $I(X_1, X_2; Y_1, Y_2)$. Since the joint distribution

$$p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2), \tag{601}$$

$Y_1 \to X_1 \to X_2 \to Y_2$ forms a Markov chain and therefore

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) && (602)\\
&= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) && (603)\\
&= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (604)\\
&\le H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (605)\\
&= I(X_1; Y_1) + I(X_2; Y_2), && (606)
\end{aligned}$$

where (603) and (604) follow from Markovity, and we have equality in (605) if Y_1 and Y_2 are independent. Equality occurs when X_1 and X_2 are independent. Hence

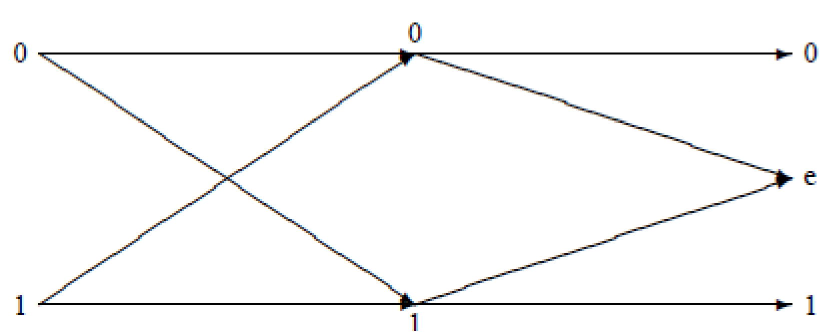
$$\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) && (607)\\
&\le \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) && (608)\\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) && (609)\\
&= C_1 + C_2. && (610)
\end{aligned}$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions that maximize C_1 and C_2 respectively.

Part 2 was done in a problem-solving session.

Problem 2

Suppose a binary symmetric channel of capacity C_1 is immediately followed by a binary erasure channel of capacity C_2. Find the capacity C of the resulting channel.



a) *Brute force method:* Let $C_1 = 1 - H(p)$ be the capacity of the BSC with parameter p, and $C_2 = 1 - \alpha$ be the capacity of the BEC with parameter α. Let $\tilde{Y}$ denote the output of the cascaded channel, and Y the output of the BSC. Then, the transition rule for the cascaded channel is simply

$$p(\tilde{y}|x) = \sum_{y=0,1} p(\tilde{y}|y)p(y|x)$$

for each $(x, \tilde{y})$ pair.
Let X ~Bern(π) denote the input to the channel. Then,

$$H(\tilde{Y}) = H((1-\alpha)(\pi(1-p) + p(1-\pi)),\, \alpha,\, (1-\alpha)(p\pi + (1-p)(1-\pi)))$$

and also

$$\begin{aligned} H(\tilde{Y}|X=0) &= H((1-\alpha)(1-p),\, \alpha,\, (1-\alpha)p) \\ H(\tilde{Y}|X=1) &= H((1-\alpha)p,\, \alpha,\, (1-\alpha)(1-p)) \;=\; H(\tilde{Y}|X=0). \end{aligned}$$

Therefore,

$$\begin{aligned} C &= \max_{p(x)} I(X;\tilde{Y}) \\ &= \max_{p(x)} [H(\tilde{Y}) - H(\tilde{Y}|X)] \\ &= \max_{p(x)} [H(\tilde{Y})] - H(\tilde{Y}|X) \\ &= \max_{\pi} [H((1-\alpha)(\pi(1-p) + p(1-\pi)),\, \alpha,\, (1-\alpha)(p\pi + (1-p)(1-\pi)))] \\ &\quad - H((1-\alpha)(1-p),\, \alpha,\, (1-\alpha)p). \end{aligned} \tag{1769}$$

Note that the maximum value of $H(\tilde{Y})$ occurs when $\pi = 1/2$ by the concavity and symmetry of $H(\cdot)$. (We can check this also by differentiating Eq. (1769) with respect to π.)
Substituting the value $\pi = 1/2$ in the expression for the capacity yields

$$\begin{aligned} C &= H((1-\alpha)/2,\, \alpha,\, (1-\alpha)/2) - H((1-p)(1-\alpha),\, \alpha,\, p(1-\alpha)) \\ &= (1-\alpha)(1 + (1-p)\log(1-p) + p\log p) \\ &= C_1 C_2. \end{aligned}$$

b) *Elegant method:*
For the cascade of an arbitrary discrete memoryless channel (with capacity C) with the erasure channel (with the erasure probability α), we will show that

$$I(X;\tilde{Y}) = (1-\alpha)I(X;Y). \tag{1770}$$

Then, by taking suprema of both sides over all input distributions $p(x)$, we can conclude the capacity of the cascaded channel is $(1-\alpha)C$.
Proof of Eq. (1770):
Let

$$E = \begin{cases} 1, & \tilde{Y} = e \\ 0, & \tilde{Y} = Y \end{cases}.$$

Then, since E is a function of Y,

$$\begin{aligned} H(\tilde{Y}) &= H(\tilde{Y}, E) \\ &= H(E) + H(\tilde{Y}|E) \\ &= H(\alpha) + \alpha H(\tilde{Y}|E=1) + (1-\alpha)H(\tilde{Y}|E=0) \\ &= H(\alpha) + (1-\alpha)H(Y), \end{aligned}$$

where the last equality comes directly from the construction of E. Similarly,

$$\begin{aligned} H(\tilde{Y}|X) &= H(\tilde{Y}, E|X) \\ &= H(E|X) + H(\tilde{Y}|X, E) \\ &= H(E) + \alpha H(\tilde{Y}|X, E=1) + (1-\alpha)H(\tilde{Y}|X, E=0) \\ &= H(\alpha) + (1-\alpha)H(Y|X), \end{aligned}$$

whence

$$I(X;\tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|X) = (1-\alpha)I(X;Y).$$

Problem 3 – Questions from Chapter 10

10.1

1) **One bit quantization of a single Gaussian random variable.** Let $X \sim \mathcal{N}(0, \sigma^2)$ and let the distortion measure be squared error. Here we do not allow block descriptions. Show that the optimum reproduction points for 1 bit quantization are $\pm\sqrt{\frac{2}{\pi}}\sigma$, and that the expected distortion for 1 bit quantization is $\frac{\pi-2}{\pi}\sigma^2$. Compare this with the distortion rate bound $D = \sigma^2 2^{-2R}$ for $R = 1$.

1) *One bit quantization of a Gaussian random variable.* Let $X \sim \mathcal{N}(0, \sigma^2)$ and let the distortion measure be squared error. With one bit quantization, the obvious reconstruction regions are the positive and negative real axes. The reconstruction point is the centroid of each region. For example, for the positive real line, the centroid a is

$$\begin{aligned} a &= \int_0^\infty x \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx && (941) \\ &= \int_0^\infty \sigma\sqrt{\frac{2}{\pi}} e^{-y}\, dy && (942) \\ &= \sigma\sqrt{\frac{2}{\pi}}, && (943) \end{aligned}$$

using the substitution $y = x^2/2\sigma^2$. The expected distortion for one bit quantization is

$$\begin{aligned}
D &= \int_{-\infty}^{0}\left(x+\sigma\sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx && (944)\\
&\quad + \int_{0}^{\infty}\left(x-\sigma\sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx && (945)\\
&= 2\int_{-\infty}^{\infty}\left(x^2+\sigma^2\frac{2}{\pi}\right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx && (946)\\
&\quad -2\int_{0}^{\infty}\left(-2x\sigma\sqrt{\frac{2}{\pi}}\right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx && (947)\\
&= \sigma^2 + \frac{2}{\pi}\sigma^2 - 4\frac{1}{\sqrt{2\pi}}\sigma^2\sqrt{\frac{2}{\pi}} && (948)\\
&= \sigma^2\frac{\pi-2}{\pi}. && (949)
\end{aligned}$$

10.3

3) **Rate distortion for binary source with asymmetric distortion.** Fix $p(\hat{x}|x)$ and evaluate $I(X;\hat{X})$ and D for

$$X \sim \text{Bern}(1/2),$$

$$d(x,\hat{x}) = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$$

3) *Binary source with asymmetric distortion.* $X \sim \text{Bern}(\frac{1}{2})$, and the distortion measure is

$$d(x,\hat{x}) = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}. \tag{955}$$

Proceeding with the minimization to calculate $R(D)$ as

$$R(D) = \min_{p(\hat{x}|x):\sum p(x)p(\hat{x}|x)d(x,\hat{x})\le D} I(X;\hat{X}), \tag{956}$$

we must choose the conditional distribution $p(\hat{x}|x)$. Setting $p(0|0) = \alpha$ and $p(1|1) = \beta$, we get the joint distribution

$$p(x,\hat{x}) = \begin{bmatrix} \frac{\alpha}{2} & \frac{1-\alpha}{2} \\ \frac{1-\beta}{2} & \frac{\beta}{2} \end{bmatrix}. \tag{957}$$

Hence the distortion constraint can be written as

$$\frac{1-\alpha}{2}a + \frac{1-\beta}{2}b \le D. \tag{958}$$

The function to be minimized, $I(X;\hat{X})$, can be written

$$I(X;\hat{X}) = H(\hat{X}) - H(\hat{X}|X) = H(\frac{\alpha+1-\beta}{2}) - \frac{1}{2}H(\alpha) - \frac{1}{2}H(\beta). \tag{959}$$

Using the method of Lagrange multipliers, we have

$$J(\alpha,\beta,\lambda) = H(\frac{\alpha+1-\beta}{2}) - \frac{1}{2}H(\alpha) - \frac{1}{2}H(\beta) + \lambda(\frac{1-\alpha}{2}a + \frac{1-\beta}{2}b) \tag{960}$$

and differentiating to find the maximum, we have the following equations:

$$\frac{1}{2}\left(\log\frac{\frac{1-\alpha+\beta}{2}}{\frac{\alpha+1-\beta}{2}}\right) - \frac{1}{2}\left(\log\frac{1-\alpha}{\alpha}\right) - \frac{\lambda a}{2} = 0 \tag{961}$$

$$-\frac{1}{2}\left(\log\frac{\frac{1-\alpha+\beta}{2}}{\frac{\alpha+1-\beta}{2}}\right) - \frac{1}{2}\left(\log\frac{1-\beta}{\beta}\right) - \frac{\lambda b}{2} = 0 \tag{962}$$

$$\frac{1-\alpha}{2}a + \frac{1-\beta}{2}b = D \tag{963}$$

In principle, these equations can be solved for α, β, and λ and substituted back in the definition to find the rate distortion function. This problem unfortunately does not have an explicit solution.

10.4

4) **Properties of** $R(D)$. Consider a discrete source $X \in \mathcal{X} = \{1,2,\ldots,m\}$ with distribution $p_1, p_2, \ldots, p_m$ and a distortion measure $d(i,j)$. Let $R(D)$ be the rate distortion function for this source and distortion measure. Let $d'(i,j) = d(i,j) - w_i$ be a new distortion measure and let $R'(D)$ be the corresponding rate distortion function. Show that $R'(D) = R(D+\bar{w})$, where $\bar{w} = \sum p_i w_i$, and use this to show that there is no essential loss of generality in assuming that $\min_{\hat{x}} d(i,\hat{x}) = 0$, i.e., for each $x \in \mathcal{X}$, there is one symbol $\hat{x}$ which reproduces the source with zero distortion.

4) *Properties of the rate distortion function.* By definition,

$$R'(D') = \min_{p(\hat{x}|x):\sum p(\hat{x}|x)p(x)d'(x,\hat{x}) \le D'} I(X;\hat{X}). \tag{964}$$

For any conditional distribution $p(\hat{x}|x)$, we have

$$\begin{aligned}
D' &= \sum_{x,\hat{x}} p(x)p(\hat{x}|x)d'(x,\hat{x}) && (965)\\
&= \sum_{x,\hat{x}} p(x)p(\hat{x}|x)(d(x,\hat{x}) - w_x) && (966)\\
&= \sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) - \sum_{x} p(x)w_x \sum_{\hat{x}} p(\hat{x}|x) && (967)\\
&= D - \sum_{x} p(x)w_x && (968)\\
&= D - \bar{w}, && (969)
\end{aligned}$$

or $D = D' + \bar{w}$. Hence

$$\begin{aligned}
R'(D') &= \min_{p(\hat{x}|x):\sum p(\hat{x}|x)p(x)d'(x,\hat{x}) \le D'} I(X;\hat{X}) && (970)\\
&= \min_{p(\hat{x}|x):\sum p(\hat{x}|x)p(x)d(x,\hat{x}) \le D'+\bar{w}} I(X;\hat{X}) && (971)\\
&= R(D'+\bar{w}). && (972)
\end{aligned}$$

For any distortion matrix, we can set $w_i = \min_{\hat{x}} d(i,\hat{x})$, hence ensuring that $\min_{\hat{x}} d'(x,\hat{x}) = 0$ for every x. This produces only a shift in the rate distortion function and does not change the essential theory. Hence, there is no essential loss of generality in assuming that for each $x \in \mathcal{X}$, there is one symbol $\hat{x}$ which reproduces it with zero distortion.

Problem 4

We are given a set of k parallel independent additive Gaussian noise channels with noise variances $N_1, \ldots, N_k$ respectively. A single transmitter is permitted to communicate to a single receiver over this set of channels. The transmitter is power constrained to P. Find the capacity of the system (in bits per use) in each of the following scenarios :

(a) The transmitter can distribute its available power among the k channels in any way it likes and can choose the inputs to each channel in any way it likes (as a function of the message it wants to send) subject to the power constraints determined by the way it distributes power over the channels. The receiver receives information from each of the k channels separately.

(b) The transmitter is constrained to use exactly the same input in each of the k channels (as a function of the message it wants to send). The receiver receives information from each of the k channels separately.

(c) The transmitter can distribute its available power among the k channels in any way it likes. The inputs to each channel have to be scaled versions of a single input (as a function of the message the transmitter wishes to send) and subject to the individual power constraints determined by the way the transmitter distributes power over the channels. The receiver receives information from each of the k channels separately.

(d) The transmitter is constrained to use exactly the same input in each of the k channels (as a function of the message it wants to send). The receiver, however, only sees the sum of the outputs of the k channels.

Source: https://people.eecs.berkeley.edu/~ananth/229ASpr07/soln_6_229Aspr07.pdf

(a) The capacity is given by water pouring the available power over the profile of the noise, as in Section 9.4 of the text. Thus, we choose a level $\nu = \nu(P)$ such that

$$\sum_{l=1}^{k} (\nu - N_l)^+ = P \ ,$$

and the capacity is then given by

$$C(P) = \sum_{l=1}^{k} \max(0, \frac{1}{2}\log(\frac{\nu}{N_l})) \ .$$

(b) With the given constraints on the communication scheme the channel is in effect a scalar input vector output additive Gaussian noise channel described at each channel use by an equation of the form

$$\mathbf{Y} = \mathbf{1}X + \mathbf{Z}$$

where $\mathbf{Z}$ is Gaussian with mean zero and covariance matrix $diag(N_1, \ldots, N_k)$, $\mathbf{1}$ denotes a column vector of ones, and X is power constrained to $\frac{P}{k}$. The capacity of this channel is

$$C(P) = \max_{p_X} I(X; \mathbf{Y})$$

where the maximization is over all distributions satisfying $E[X^2] \leq \frac{P}{k}$. We have

$$I(X; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y} \mid X) = h(\mathbf{Y}) - \sum_{l=1}^{k} \frac{1}{2}\log(2\pi e N_l) \ .$$

Now, the covariance matrix of $\mathbf{Y}$ is

$$E[\mathbf{Y}\mathbf{Y}^T] = E[X^2]\mathbf{1}\mathbf{1}^T + diag(N_1, \ldots, N_k)$$

Hence

$$I(X; \mathbf{Y}) \leq \frac{1}{2}\log det(I + E[X^2] \begin{bmatrix} N_1^{-1} & \ldots & N_k^{-1} \\ \vdots & \ddots & \vdots \\ N_1^{-1} & \ldots & N_k^{-1} \end{bmatrix}) \ .$$

Since the matrix in this expression is of rank 1 with its only nonzero eigenvalue being $\sum_{l=1}^{k} N_l^{-1}$ and since $E[X^2] \leq \frac{P}{k}$ we get

$$C(P) = \frac{1}{2}\log(1 + \frac{P}{k}\sum_{l=1}^{k} N_l^{-1})$$

where equality holds since this can be achieved by choosing $X \sim N(0, \frac{P}{k})$.

(c) Consider the allocation of powers P_l to channel l, with $\sum_{l=1}^{k} P_l \leq P$ Given this choice, the channel at each use looks like

$$\mathbf{Y} = X\mathbf{a} + \mathbf{Z}$$

where $\mathbf{Z}$ is Gaussian with mean zero and covariance matrix $diag(N_1, \ldots, N_k)$, $\mathbf{a}$ denotes the column vector $[\sqrt{P_1}, \ldots, \sqrt{P_k}]^T$, and X is power constrained to 1. Repeating the calculation of the preceding part of the problem we see the capacity, given this allocation of power between channels, is

$$\frac{1}{2}\log(1 + \sum_{l=1}^{k} \frac{P_l}{N_l})$$

which is achieved when $X \sim N(0, 1)$. This expression is to be maximized over the power allocation to arrive at the capacity of the channel, and this is seen to require that all the power be allocated to the best channel, so that

$$C(P) = \frac{1}{2}\log(1 + \frac{P}{N_{min}}) .$$

(d) The channel is now a scalar input scalar output channel that looks at each use like

$$Y = kX + Z$$

where $Z \sim N(0, \sum_{l=1}^{k} N_l)$ and $E[X^2] \leq \frac{P}{k}$. The capacity of the channel is therefore

$$C(P) = \frac{1}{2}\log(1 + \frac{kP}{\sum_{l=1}^{k} N_l}) .$$

Problem 5

A memoryless source U is uniformly distributed on $\{0, \ldots, r-1\}$. The following distortion function is given by

$$d(u,v) = \begin{cases} 0, & u = v, \\ 1, & u = v \pm 1 \mod r, \\ \infty, & \text{otherwise.} \end{cases}$$

Show that the rate distortion function is

$$R(D) = \begin{cases} \log r - D - h_2(D), & D \leq \frac{2}{3}, \\ \log r - \log 3, & D > \frac{2}{3}. \end{cases}$$

Source: http://www.ee.bgu.ac.il/~haimp/multi2/HW/hw1sol_sanov_rate_distortion.pdf

From the symmetry of the problem, we can assume the conditional distribution of $p(v|u)$ as

$$p(v|u) = \begin{cases} 1-p & u = v, \\ p/2 & u = v \pm 1 \quad \mod r, \\ 0 & \text{otherwise} \end{cases}$$

Then, $E(d(U,V)) = p$. Therefore, the rate distortion function is

$$R(D) = \min_{p \le D} I(U;V).$$

Now, we know that

$$\begin{aligned} I(U;V) =& H(V) - H(V|U) \\ =& \log r - H(1-p, p/2, p/2), \end{aligned}$$

since U is uniform, and due to symmetry, V is also uniform. We know that $H(1-p,p,p) \le \log 3$, and this is achieved when $p = 2/3$. Therefore,

$$R(D) = \log r - \log 3, \quad \text{if } D > 2/3.$$

Now, let's consider the case when $D \le 2/3$. Denote

$$\begin{aligned} f(p) = H(1-p, p/2, p/2) =& -(1-p)\log(1-p) - p/2 \log p/2 \times 2 \\ =& -(1-p)\log(1-p) - p\log p + p, \end{aligned}$$

We know that $f(p)$ is a concave function. By differentiating with respect to p,

$$\frac{df(p)}{dp} = \log(1-p) + 1 - \log p - 1 + 1 = \log \frac{1-p}{p} + 1$$

and setting $f(p) = 0$, $f(p)$ becomes maximum when $p = 2/3$. Therefore, if $D \le 2/3$, $f(p)$ is an increasing function of p. Thus,

$$R(D) = \log 3 - D - h_2(D), \quad \text{if } D \le 2/3.$$