Problem 10.2:

(a) Since Y_1 and Y_2 are conditionally independent given X, we have $\forall y_1, y_2, x$:

$$\Pr\{Y_1 = y_1, Y_2 = y_2 | X = x\} = \Pr\{Y_1 = y_1 | X = x\} \Pr\{Y_2 = y_2 | X = x\}. \tag{1}$$

Furthermore, since Y_1 and Y_2 are conditionally identically distributed given X, we have $\forall y, x$:

$$\Pr\{Y_1 = y | X = x\} = \Pr\{Y_2 = y | X = x\}. \tag{2}$$

Multiplying the above by $\Pr\{X = x\}$ and summing over all x, we get

$$\Pr\{Y_1 = y\} = \Pr\{Y_2 = y\} \qquad \forall y. \tag{3}$$

Thus Y_1 and Y_2 are (unconditionally) identically distributed. Now

$$\begin{aligned} I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2 | X) \\ \textit{from } (1) &= h(Y_1, Y_2) - h(Y_1 | X) - h(Y_2 | X) \\ \textit{from } (2) &= h(Y_1, Y_2) - 2h(Y_1 | X) \\ &= h(Y_1) + h(Y_2) - I(Y_1; Y_2) - 2h(Y_1 | X) \\ \textit{from } (3) &= 2h(Y_1) - I(Y_1; Y_2) - 2h(Y_1 | X) \end{aligned}$$

Thus,

$$I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2).$$

(b) Let C_2 be the capacity of the channel with output (Y_1, Y_2) and C_1 be the capacity of the channel with output Y_1.

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

with equality iff the distribution which achieves C_2 makes Y_1 and Y_2 independent. [Note that this result also holds for discrete channels.]

Problem 10.3:

Let $\mathbf{Y} = (Y_1, Y_2)$ and $\mathbf{Z} = (Z_1, Z_2)$. Consider the general case:

$$I(X; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y}|X) = h(\mathbf{Y}) - h(\mathbf{Z}),$$

where $h(\mathbf{Z})$ is fixed and $h(\mathbf{Y})$ is to be maximized subject to $E[X^2] \le S$. We have

$$\begin{aligned} E[Y_1^2] &= E[(X + Z_1)^2] = E[X^2] + N, \\ E[Y_2^2] &= E[(X + Z_2)^2] = E[X^2] + N, \\ E[Y_1 Y_2] &= E[(X + Z_1)(X + Z_2)] = E[X^2] + N\rho, \end{aligned}$$

and $Var(Y_1) = Var(Y_2) = Var(X) + N$, $Cov(Y_1, Y_2) = Var(X) + N\rho$. The determinant of the covariance matrix of $\mathbf{Y}$ is given by

$$\begin{aligned} |K_{\mathbf{Y}}| &= (Var(X) + N)^2 - (Var(X) + N\rho)^2 \\ &= N(1 - \rho)[2Var(X) + N(1 + \rho)] \end{aligned}$$

which is maximized when $Var(X) = S$. Note that, from Theorem 9.6.5, the differential entropy of a vector with given covariance matrix is maximized when that vector is Gaussian. For $\mathbf{Y}$ to be Gaussian, X must be Gaussian. Thus the capacity achieving distribution is $X \sim \mathcal{N}(0, S)$ and

$$\begin{aligned} C(S) &= \frac{1}{2}\log_2 \frac{|K_{\mathbf{Y}}|}{|K_{\mathbf{Z}}|} = \frac{1}{2}\log_2 \frac{N(1-\rho)[2S + N(1+\rho)]}{N^2(1-\rho)(1+\rho)} \\ &= \frac{1}{2}\log_2 \frac{2S + N(1+\rho)}{N(1+\rho)} \quad \frac{\text{bits}}{\text{channel use}} \end{aligned}$$

(a) $\rho = 1 \Longrightarrow C(S) = \frac{1}{2}\log_2(1 + S/N)$. This is just the capacity of the one-look channel. Because $\rho = 1$ implies that $Z_1 = Z_2$ w.p.1 which implies that $Y_1 = Y_2$ w.p.1. This is really just one channel output.

(b) $\rho = 0 \Longrightarrow C(S) = \frac{1}{2}\log_2(1 + 2S/N)$. This is just the capacity of a one-look channel with two times the power.

(c) $\rho = -1 \Longrightarrow C(S) = +\infty$. This case implies that $Z_1 = -Z_2$ w.p.1. This channel can be converted into a noiseless channel with the operation

$$\frac{Y_1 + Y_2}{2} = X \quad \text{w.p.1.}$$

Thus, any input can be perfectly reconstructed.

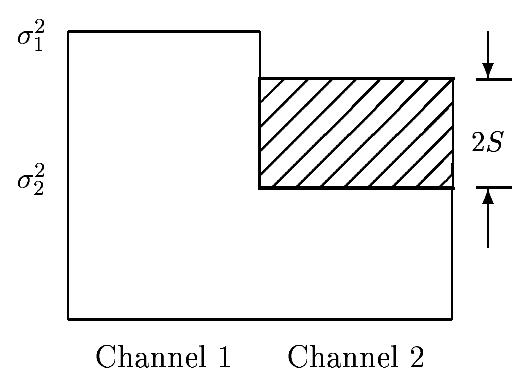
Problem 10.4: Consider the parallel channel

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

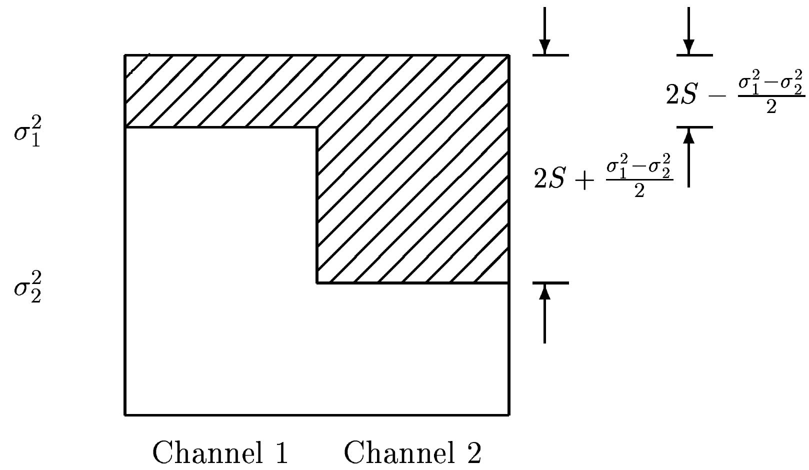
$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right).$$

$$\sigma_1^2 > \sigma_2^2, \qquad E[X_1^2 + X_2^2] \leq 2S$$

<u>**Case 1:**</u> $S \leq \frac{\sigma_1^2 - \sigma_2^2}{2}$. By waterfilling, we use only the second channel with input power $2S$.



Case 2: $S > \frac{\sigma_1^2-\sigma_2^2}{2}$. By waterfilling, we use $S - \frac{\sigma_1^2-\sigma_2^2}{2}$ of the power on the first channel and $S + \frac{\sigma_1^2-\sigma_2^2}{2}$ of the power on the second channel.



Problem 13.1: The one-bit scalar quantizer operates as follows:

$$Q(x) = \beta(\alpha(x)) = \begin{cases} -a & \text{if } x < 0 \\ +a & \text{if } x \geq 0 \end{cases}.$$

Here, the average mean-squared distortion is given by

$$\begin{aligned} D_{ave} = E[(X - Q(X))^2] &= \int_{-\infty}^{0} (x-(-a))^2 f(x)dx + \int_0^\infty (x-a)^2 f(x)dx \\ &= 2\int_0^\infty (x-a)^2 f(x)dx, \end{aligned} \quad (4)$$

where $f(x) \sim \mathcal{N}(0, \sigma^2)$. To find the optimal reproduction level a, we take the derivative and set it to zero:

$$\frac{dD_{ave}}{da} = -4\int_0^\infty (x-a)f(x)dx = 0.$$

This implies that

$$\begin{aligned} a &= \frac{\int_0^\infty xf(x)dx}{\int_0^\infty f(x)dx} \\ &= 2\int_0^\infty \frac{x}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}dx \\ &= \sqrt{\frac{2\sigma^2}{\pi}}\int_0^\infty e^{-u}du = \sqrt{\frac{2\sigma^2}{\pi}} \end{aligned}$$

where $u = x^2/(2\sigma^2)$ and $du = x/\sigma^2 dx$. To find the optimal distortion, we plug $a = \sqrt{\frac{2\sigma^2}{\pi}}$ into equation (1) above. This yields

$$\begin{aligned} D_{ave} &= 2\left[\int_0^\infty x^2 f(x)dx - 2a\int_0^\infty xf(x)dx + a^2\int_0^\infty f(x)dx\right] \\ &= 2\left[\sigma^2/2 - 2a\sqrt{\frac{\sigma^2}{2\pi}} + a^2/2\right] \\ &= 2\left[\sigma^2/2 - \frac{2\sigma^2}{\pi} + \frac{\sigma^2}{\pi}\right] \\ &= \frac{\pi-2}{\pi}\sigma^2 = 0.363\sigma^2 \end{aligned}$$

This should be compared with the rate-distortion limit (for infinite-dimensional vector quantizer) of $D = 0.25\sigma^2$.

Problem 13.2: The minimum distortion for a rate-zero code is $D = d_{max} = 1/2$ which is achieved by setting $\hat{x}_0 = 0$. Now assume that $0 \le D \le d_{max}$ and $E[d(X,\hat{X})] \le D$. This implies that $\Pr\{\hat{X} = 1|X = 0\} = 0$ otherwise we would end up with infinite distortion. In this situation, we have $E[d(X,\hat{X})] = \Pr\{\hat{X} = 0|X = 1\}(1/2) = \gamma/2$. Now,

$$\begin{aligned} I(X;\hat{X}) &= H(\hat{X}) - H(\hat{X}|X) \\ &= h_b((1-\gamma)/2) - H(\hat{X}|X=1)/2 \qquad [H(\hat{X}|X=0)=0] \\ &= h_b((1-\gamma)/2) - h_b(\gamma)/2. \end{aligned}$$

The function $h_b((1-\gamma)/2) - h_b(\gamma)/2$ is monotonically decreasing with γ for $\gamma \in [0,1]$. Since $\gamma/2 \le D$, we have $I(X;\hat{X}) \ge h_b((1-2D)/2) - h_b(2D)/2$.
Thus $R(D) = h_b((1-2D)/2) - h_b(2D)/2$ when $D \le 1/2$.

Problem 13.5: The minimum distortion for a rate-zero code is $D = d_{max} = (m-1)/m$ which is achieved by setting $\hat{x}_0 = 1$. Now assume that $0 \le D \le d_{max}$ and $E[d(X,\hat{X})] \le D$. From Problem 13.6 (see below),

$$R(D) = H(X) - \phi(D) = \log_2 m - \phi(D),$$

where

$$\phi(D) = \max_{\mathbf{p}:\sum p_i d_i \le D} H(\mathbf{p}),$$

$d_1 = 0, d_2 = d_3 = \ldots = d_m = 1$. We want to maximize

$$-\sum_{i=1}^{m} p_i \log_2 p_i$$

subject to the constraints

$$\sum_{i=1}^{m} p_i = 1 \tag{5}$$

and

$$\sum_{i=1}^{m} p_i d_i \le D. \tag{6}$$

Using the Lagrange multipliers λ, γ:

$$\frac{\partial}{\partial p_i}\left[\sum_{i=1}^{m} -p_i \log_2 p_i + \lambda \sum_{i=1}^{m} p_i d_i + \gamma \sum_{i=1}^{m} p_i\right] = \frac{-p_i}{(\ln 2) p_i} - \log_2 p_i + \lambda d_i + \gamma = 0.$$

This implies that

$$p_i = 2^{\gamma - \log_2 e + \lambda d_i}.$$

Since d_i is either 0 or 1,

$$p_i = \begin{cases} \alpha & i = 1 \\ \beta & i \ne 1 \end{cases},$$

where $\alpha = 2^{\gamma - \log_2 e}$ and $\beta = 2^{\gamma - \log_2 e + \lambda}$. To find α and β, we use (2) and (3).

$$\begin{aligned} (2) &\implies \alpha + (m-1)\beta = 1 \implies \beta = \frac{1-\alpha}{m-1} \\ (3) &\implies (m-1)\beta = D \implies \beta = \frac{D}{m-1} \end{aligned}$$

Thus $\alpha = 1 - D$ and $\beta = D/(m-1)$. Finally,

$$\begin{aligned}\phi(D) &= -(1-D)\log_2(1-D) - (m-1)\frac{D}{m-1}\log_2\frac{D}{m-1} \\ &= -(1-D)\log_2(1-D) - D\log_2 D + D\log_2(m-1) \\ &= h_b(D) + D\log_2(m-1)\end{aligned}$$

Thus $R(D) = \log_2 m - h_b(D) - D\log_2(m-1)$. Note that for $D = 0$, $R(0) = \log_2 m$ which is the entropy rate of the i.i.d. uniformly distributed source.

Problem 13.6:

(a) Concavity of $\phi(D)$: Let $\mathbf{p}_1$ and $\mathbf{p}_2$ achieve $\phi(D_1)$ and $\phi(D_2)$, respectively. Then $\mathbf{p} = \lambda\mathbf{p}_1 + (1-\lambda)\mathbf{p}_2$ is an admissible distribution for $\phi(\lambda D_1 + (1-\lambda)D_2)$ since

$$\sum_{i=1}^{m}(\lambda p_{1,i} + (1-\lambda)p_{2,i}) \le \lambda D_1 + (1-\lambda)D_2.$$

By the concavity of $H(\cdot)$,

$$H(\lambda\mathbf{p}_1 + (1-\lambda)\mathbf{p}_2) \ge \lambda H(\mathbf{p}_1) + (1-\lambda)H(\mathbf{p}_2) = \lambda\phi(D_1) + (1-\lambda)\phi(D_2).$$

Thus

$$\phi(\lambda D_1 + (1-\lambda)D_2) \ge \lambda\phi(D_1) + (1-\lambda)\phi(D_2).$$

(b) Equations (13.152) and (13.153) are by definitions. Inequality (13.154) is by the definition of $\phi(\cdot)$ since $p(\cdot|\hat{x})$ is admissible. Inequality (13.155) is by the concavity of $\phi(\cdot)$ (part (a)). Inequality (13.156) is by $\sum p(\hat{x})D_{\hat{x}} = E[d(X,\hat{X})] \le D$ and $\phi(D)$ is a non-increasing function of D.

(c) From part (b), $I(X;\hat{X}) \ge H(X) - \phi(D)$ whenever $E[d(X,\hat{X})] \le D$. Thus,

$$R(D) \ge H(X) - \phi(D).$$

(d) The lower bound is tight iff inequalities (13.154), (13.155) and (13.156) can be achieved with equality. This happens if

$$H(X|\hat{X} = \hat{x}) = \phi(D) \qquad \forall\hat{x}$$

and

$$p(\hat{x}) = \frac{1}{m} \qquad \forall\hat{x}.$$

Let $\mathbf{p}$ be the distribution which achieves $\phi(D)$. Define a backward test channel

$$p(x|\hat{x}) = \text{the permutation of } \mathbf{p} \text{ induced by the ordering } d_1, d_2, \ldots, d_m \text{ on the } \hat{x}\text{-column of the distortion matrix}$$

For example, for the previous problem:

$$p(x|\hat{x}) = \begin{bmatrix} \alpha & \beta & \cdots & \beta \\ \beta & \alpha & \cdots & \beta \\ \vdots & \vdots & \vdots & \vdots \\ \beta & \beta & \cdots & \alpha \end{bmatrix}.$$

This channel is strongly symmetric. If $p(\hat{x})$ is uniform, so is $p(x)$, which is consistent with the hypothesis. Clearly, $H(X|\hat{X} = \hat{x}) = \phi(D)$ and

$$R(D) = H(X) - \phi(D).$$

Problem 13.7: The minimum distortion for a rate-zero code is $D = d_{max} = 1$ which is achieved by setting $\hat{x}_0 = E$. Now assume that $0 \leq D \leq d_{max}$ and $E[d(X, \hat{X})] \leq D$. To have finite distortion, we must not allow transition from 0 to 1 or from 1 to 0. Consider the forward test channel:

$$\begin{aligned}
\Pr\{\hat{X} = 0 | X = 0\} &= 1 - \alpha \\
\Pr\{\hat{X} = E | X = 0\} &= \alpha \\
\Pr\{\hat{X} = 1 | X = 0\} &= 0 \\
\Pr\{\hat{X} = 0 | X = 1\} &= 0 \\
\Pr\{\hat{X} = E | X = 1\} &= \alpha \\
\Pr\{\hat{X} = 1 | X = 1\} &= 1 - \alpha
\end{aligned}$$

Then $E[d(X, \hat{X})] = \alpha = D$ and $I(X; \hat{X}) = 1 - \alpha = 1 - D$.
Thus $R(D) = 1 - D$.
A simple encoding procedure: To achieve distortion D at rate $R(D)$, take a block of n samples.

- Encoding: Ignore the first nD samples. Encode the remaining $n(1-D)$ samples by themselves.
- Decoding: Decode the first nD samples as "E". Decode the rest as themselves.

The rate is obviously $R = \frac{n(1-D)}{n} = 1 - D$ and the distortion is D.

Please let me know if you see a mistake in this solution.