ECE 231A HW7 Shengze Ye

1. 4-ary Hamming distortion

A random variable X uniformly takes on values $\{0,1,2,3\}$. The distortion function is :

$$d(x,\hat{x}) = \begin{cases} 0 & x=\hat{x} \\ 1 & x\neq\hat{x} \end{cases}$$

Compute the rate distortion function $R(D)$ by finding a lower bound $I(x.\hat{x})$

$$R(D) = \min_{p(\hat{x}|x),\, E[d]\le D} I(x;\hat{x})$$

$$E[d] = \sum_{x,\hat{x}} p(x,\hat{x})\, d(x,\hat{y}) = Pr(X\neq X) \le D$$

$$\boxed{I(x;\hat{x}) = H(x) - H(x|\hat{x})}$$

we know $d = \begin{cases} 1 & \text{if } x\neq\hat{x} \\ 0 & \text{if } x=0 \end{cases}$

$$H(d,x|\hat{x}) = H(x|\hat{x}) + H(d|x,\hat{x})$$
$$= H(d|\hat{x}) + H(x|d,\hat{x})$$

$H(d|x,\hat{x})=0$

$\therefore\ H(x|\hat{x}) = H(d|\hat{x}) + H(x|d,\hat{x})$

From Fano's inequality we know that

$$H(x|d,\hat{x}) = P(d=0)\,H(x|\hat{x},d=0)$$
$$+ P(d=1)\,H(x|\hat{x},d=1)$$
$$= P(\hat{x}\neq x)\,H(x|\hat{x},d=1)$$

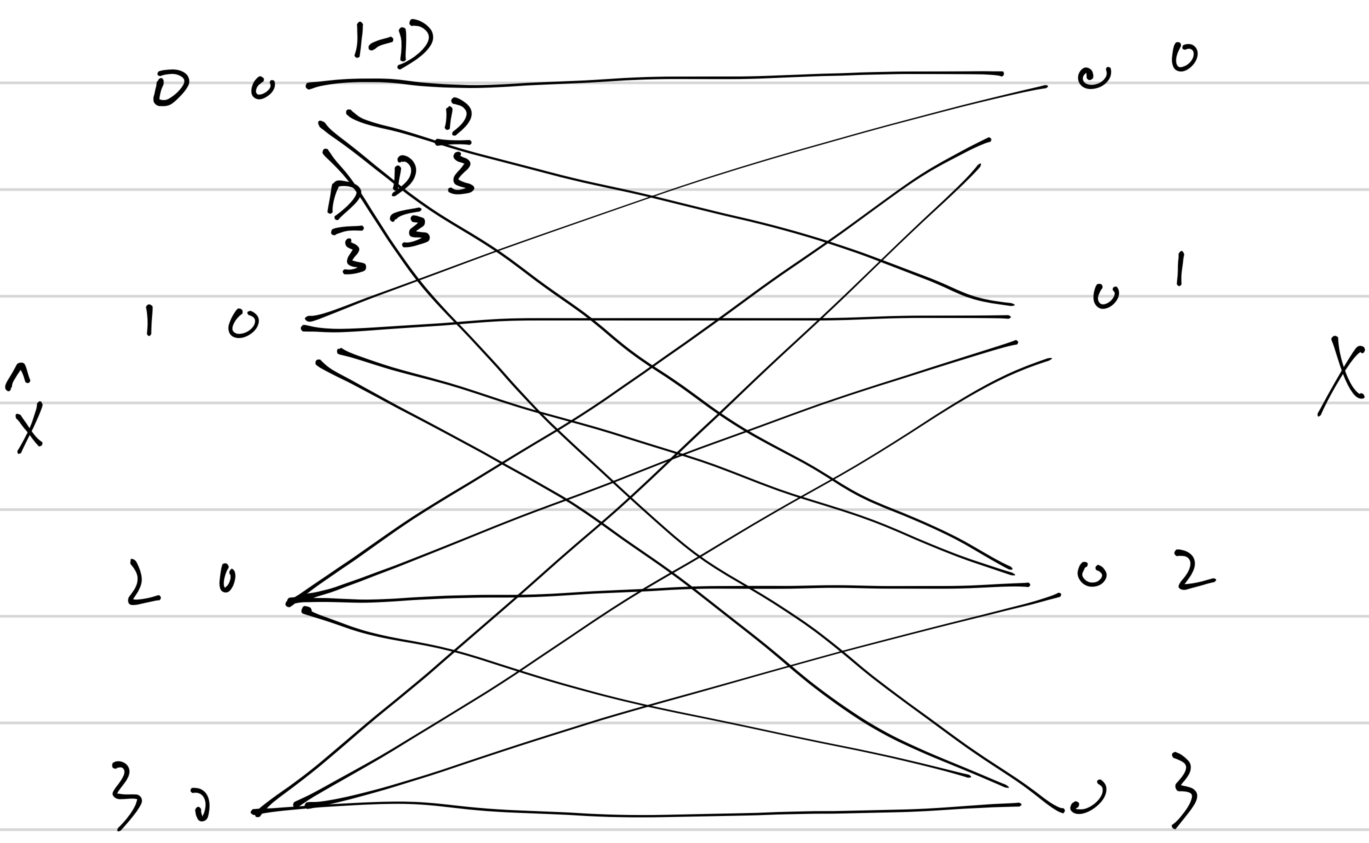
$$\le p(\hat{x} \ne x) \log |w-1| = p(\hat{x} \ne x) \log 3$$

$$\begin{aligned} H(x|\hat{x}) &= H(d|\hat{x}) + H(x|d,\hat{x}) \\ &\le H(Pr(\hat{x} \ne x)) + p(\hat{x} \ne x) \log 3 \\ &= H(D) + D \cdot \log 3 \end{aligned}$$

$$\begin{aligned} \therefore\ I(x;\hat{x}) &= H(x) - H(x|\hat{x}) \\ &\ge H(x) - H(D) - D \cdot \log 3 \\ &= \log 4 - H(D) - D \cdot \log 3 \\ &= 2 - H(D) - D \cdot \log 3 \quad \text{for } 0 \le D \le \tfrac{3}{4} \end{aligned}$$

$$\therefore R(D) \ge 2 - H(D) - D \log 3$$

consider the test channel



$2 - H(D) - D \cdot \log 3$ is the $I(X;\hat{X})$ for the test channel.

$p(x)$ is a uniform distribution

we can achieve this lower bound since $p(x)$ is uniform distribution

$\Rightarrow$ $p(\hat{x})$ is also uniform distribution

$$p(\hat{x}) = \begin{cases} \frac{1}{4} & \hat{x} = 0 \\ \frac{1}{4} & \hat{x} = 1 \\ \frac{1}{4} & \hat{x} = 2 \\ \frac{1}{4} & \hat{x} = 3 \end{cases}$$

And the conditional distribution $p(x|\hat{x})$ is

$$p(x|\hat{x}=0) \begin{cases} 1-D & x = 0 \\ \frac{D}{3} & x = 1 \\ \frac{D}{3} & x = 2 \\ \frac{D}{3} & x = 3 \end{cases} \qquad p(x|\hat{x}=1) = \begin{cases} \frac{D}{3} & x = 0 \\ 1-D & x = 1 \\ \frac{D}{3} & x = 2 \\ \frac{D}{3} & x = 3 \end{cases}$$

$$p(x|\hat{x}=2) = \begin{cases} \frac{D}{3} & x = 0 \\ \frac{D}{3} & x = 1 \\ 1-D & x = 2 \\ \frac{D}{3} & x = 3 \end{cases} \qquad p(x|\hat{x}=3) = \begin{cases} \frac{D}{3} & x = 0 \\ \frac{D}{3} & x = 1 \\ \frac{D}{3} & x = 2 \\ 1-D & x = 3 \end{cases}$$

$\therefore$ we can achieve the lower bound with $p(x,\hat{x}) = p(\hat{x}) \cdot p(x|\hat{x})$ described above, so:

$$R(D) = \begin{cases} 2 - H(D) - D \cdot \log 3 & 0 \le D \le \frac{3}{4} \\ 0 & D > \frac{3}{4} \end{cases}$$

2. Rate distortion for uniform source with Hamming distortion

Consider a source X uniformly distributed on the set $\{1, 2, \cdots m\}$. Find the rate distortion function for this source

$$d(x, \hat{x}) = \begin{cases} 0 & x = \hat{x} \\ 1 & x \neq \hat{x} \end{cases}$$

This is a generalization of problem 1. in the previous problem, $m=4$

$$E[d] = \sum_{x,\hat{x}} p(x,\hat{x})\, d(x,\hat{x}) = \Pr(x \neq \hat{x}) \leq D$$

$$R(D) = \min_{E[d] \leq D,\ p(x,\hat{x})} I(x;\hat{x})$$

$$I(x;\hat{x}) = H(x) - H(x|\hat{x})$$

From the previous problem and Fano's inequality,

$$\begin{aligned} H(x|\hat{x}) &= H(d|\hat{x}) + H(x|d,\hat{x}) \\ &\leq H(\Pr(x \neq \hat{x})) + \Pr(\hat{x} \neq x)\log(m-1) \\ &= H(D) + D \cdot \log(m-1) \end{aligned}$$

$$\begin{aligned} \therefore I(x;\hat{x}) &= H(x) - H(x|\hat{x}) \\ &\geq H(x) - H(D) - D\log(m-1) \end{aligned}$$

As x is uniform distribution $H(x) = \log m$

$$\therefore I(x;\hat{x}) \geq \log m - H(D) - D\log(m-1)$$

Similar to the previous problem, we choose

such that if $\hat{x} = x$ $= 1-D$

if $\hat{x} \neq x$ $= D/(m-1)$

we can achieve the lower boundary with $p(x,\hat{x}) = p(x|\hat{x}) \cdot p(\hat{x})$

$$\therefore p(x|\hat{x}) = \begin{cases} 1-D & \text{if } \hat{x} = x \\ D/(m-1) & \text{if } \hat{x} \neq x \end{cases}$$

($p(\hat{x})$ is uniform distribution)

To make sure the boundary

if $D > 1 - \frac{1}{m}$, it is always satisfied that

$E[d] = Pr(\hat{x} \neq x) < D$

$\therefore$ if $D > 1 - \frac{1}{m}$ $R(D) = 0$

$$\therefore R(D) = \begin{cases} \log m - H(D) - D\log(m-1) & 0 \le D \le 1 - \frac{1}{m} \\ 0 & D > 1 - \frac{1}{m} \end{cases}$$

3. Scaled Hamming distortion

A random variable X uniformly takes on $\{1, 2, 3\}$.

$$d(x,\hat{x}) = \begin{cases} 0 & x = \hat{x} \\ 2 & x \neq \hat{x} \end{cases}$$

Compute the rate distortion function $R(D)$

$$R(D) = \min_{p(x,\hat{x})\ E[d] \le D} I(X;\hat{X})$$

$$E[d] = \sum_{p(x,\hat{x})} p(x,\hat{x})\, d(x,\hat{x}) = Pr(x \neq \hat{x}) \cdot 2 \le D$$

$\therefore \ pr(x \neq \hat{x}) \leq \frac{D}{2}$

$$I(x,\hat{x}) = H(x) - H(x|\hat{x})$$

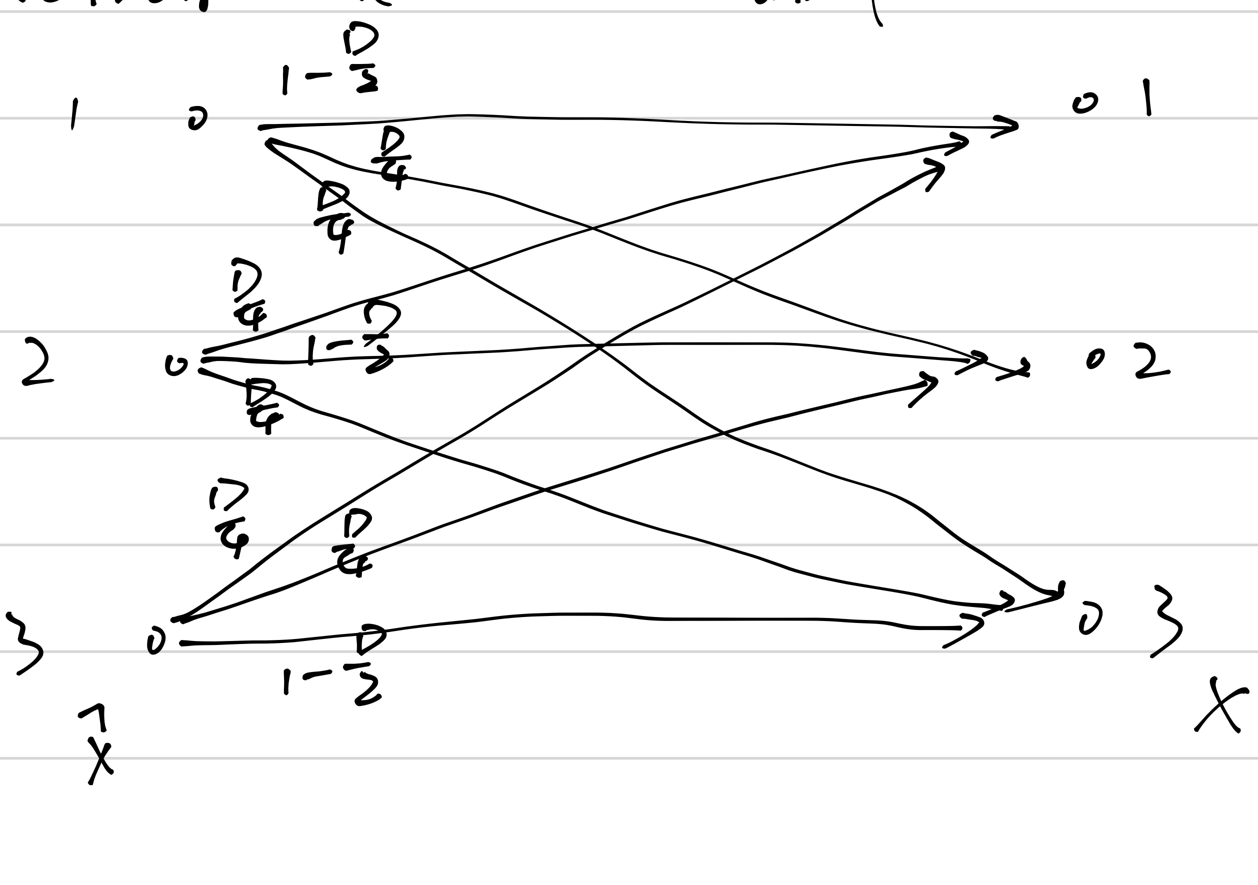
From the previous problems and the Fano's inequality we know that

$$\begin{aligned} H(x|\hat{x}) &= H(d\,|\,\hat{x}) + H(x|d,\hat{x}) \\ &\leq H(Pr(x \neq \hat{x})) + Pr(\hat{x} \neq x)\cdot \log(3-1) \\ &= H\left(\tfrac{D}{2}\right) + \tfrac{D}{2}\cdot 1 \end{aligned}$$

$\therefore \ H(x|\hat{x}) \leq H\left(\frac{D}{2}\right) + \frac{D}{2}$

$$\begin{aligned} I(x;\hat{x}) &= H(x) - H(x|\hat{x}) \\ &\geq H(x) - H\left(\tfrac{D}{2}\right) + \tfrac{D}{2} \\ &= \log 3 - H\left(\tfrac{D}{2}\right) + \tfrac{D}{2} \end{aligned}$$

consider the ~~test~~ channel



$\log 3 - H(\frac{D}{2}) + \frac{D}{2}$ is the $I(X;\hat{X})$ for this channel.

$\therefore$ The lower boundary can be achieved when $p(\hat{X})$ is uniform distribution. And the conditional distribution $p(x|\hat{x})$ is:

$$p(x|\hat{x}=1) = \begin{cases} 1-\frac{D}{2} & x=1 \\ \frac{D}{4} & x=2 \\ \frac{D}{4} & x=3 \end{cases}$$

$$p(x|\hat{x}=2) = \begin{cases} \frac{D}{4} & x=1 \\ 1-\frac{D}{2} & x=2 \\ \frac{D}{4} & x=3 \end{cases} \qquad p(x|\hat{x}=3) = \begin{cases} \frac{D}{4} & x=1 \\ \frac{D}{4} & x=2 \\ 1-\frac{D}{2} & x=3 \end{cases}$$

For the boundary

$0 \le \frac{D}{2} \le 1 - \frac{1}{3}$ $\quad\therefore 0 \le D \le \frac{4}{3}$

$$R(D) = \begin{cases} \log 3 - H(\frac{D}{2}) - \frac{D}{2} & 0 \le D \le \frac{4}{3} \\ 0 & D > \frac{4}{3} \end{cases}$$

4. Rate distortion function with finite distortion.

Find the rate distortion function $R(D) = \min I(X;\hat{X})$ for $X \sim B(\frac{1}{2})$ and

$$d(\hat{x},x) = \begin{cases} 0 & x=\hat{x} \\ 1 & x=1 \quad \hat{x}=0 \\ \infty & x=0 \quad \hat{x}=1 \end{cases}$$

$$R(D) = \min_{p(x,\hat{x})\, E[d] \le D} I(X;\hat{X})$$

$$E[d] = \sum p(x,\hat{x})\, d(x,\hat{x}) \le D$$

$\because d(0,1) = \infty \quad \therefore p(0,1)$ must be 0, otherwise $E[d]$ is infinite

$\therefore p(1,0) = D$

$X \sim B(\frac{1}{2}) \quad \therefore p(X=0) = \frac{1}{2} = p(X=1)$

Here, we have the joint distribution $p(x,\hat{x})$

$p(x,\hat{x})$	$\hat{x}=0$	$\hat{x}=1$
$x=0$	$\frac{1}{2}$	0
$x=1$	D	$\frac{1}{2}-D$

assume $D \le \frac{1}{2}$

$$I(X;\hat{X}) = H(X) - H(X|\hat{X})$$
$$= H(\tfrac{1}{2}) - \sum_i p(\hat{x}=x_i)\, H(X|\hat{X}=x_i)$$

$$= 1 - [p(\hat{X}=1)H(X|\hat{X}=1) + p(\hat{X}=0)H(X|\hat{X}=0)]$$

$$= 1 - [0 + (\tfrac{1}{2}+D)\, H(\frac{\frac{1}{2}}{\frac{1}{2}+D}, \frac{D}{\frac{1}{2}+D})]$$

$$= 1 + (\tfrac{1}{2}+D)\left[\frac{\frac{1}{2}}{\frac{1}{2}+D}\log\frac{\frac{1}{2}}{\frac{1}{2}+D} + \frac{D}{\frac{1}{2}+D}\log\frac{D}{\frac{1}{2}+D}\right]$$

$$= 1 + \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{2}+D} + D\log\frac{D}{\frac{1}{2}+D}$$

$$\therefore R(D) = \begin{cases} 1 + \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{2}+D} + D\log\frac{D}{\frac{1}{2}+D} & 0 \le D \le \frac{1}{2} \\ 0 & D > \frac{1}{2} \end{cases}$$

5. Erasure distortion.

consider $X \sim B(\frac{1}{2})$

$$d(x,\hat{x}) = \begin{bmatrix} 0 & 1 & \infty \\ \infty & 1 & 0 \end{bmatrix}$$

(a) $E[d] = \sum p(x,\hat{x})\, d(x,\hat{x}) \le D$

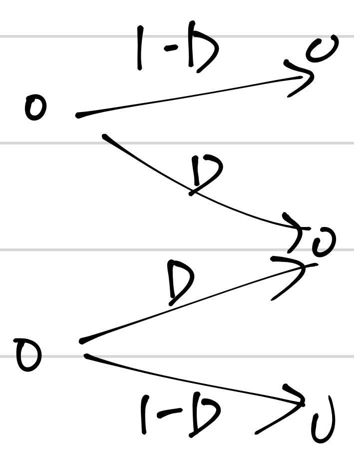
$\because$ $d(0,1) = d(1,0) = \infty$

D is a finite number

$\therefore p(0,1) = p(1,0) = 0$

The conditional distribution can be described as a erasure channel

$\therefore I(X;\hat{X}) = H(X) - H(X|\hat{X}) = 1 - D$



$$\therefore R(D) = \begin{cases} 1 - D & 0 \le D \le 1 \\ 0 & D > 1 \end{cases}$$

(b) Suggest a simple scheme to achieve any value of the rate distortion function of this source

if D can be described as a/b, then we can send the first b-a bits of a n bits block, then do not send the last b bits.
we recover the first b-a bits
$\Rightarrow$ we can send the information at rate 1-D, and achieve the distortion of D.
if D is not rational, we can use a very large b and a to make a/b as close to D as possible.

6. Bounds on the rate distortion function.
For the case of a continuous random variable X with zero mean and variance σ^2,
show that :

$$h(x) - \frac{1}{2}\log(2\pi e)D \le R(D) < \frac{1}{2}\log\frac{\sigma^2}{D}$$

① For lower bound
Since we use the Gaussian distortion function
$E[(X-\hat{X})^2] \le D$

$$I(X;\hat{X}) = h(X) - h(X|\hat{X})$$
$$= h(X) - h(X-\hat{X}|\hat{X})$$
$$\geq h(X) - h(X-\hat{X})$$
$$\geq h(X) - h(\mathcal{N}(0, E[(X-\hat{X})^2]))$$
$$= h(X) - \frac{1}{2}\log(2\pi e D)$$
$$\therefore \quad R(D) \geq h(X) - \frac{1}{2}\log(2\pi e D)$$

② For upper bound

consider a Gaussian channel

$$\begin{tikzcd}
 & Z \sim N\left(0, \frac{D\sigma^2}{\sigma^2-D}\right) \arrow[d] & & \frac{\sigma^2-D}{\sigma^2} \arrow[dl] & \\
X \arrow[r] & \oplus \arrow[r] & \odot \arrow[rr] & & \hat{X}
\end{tikzcd}$$

$$\hat{X} = \frac{\sigma^2-D}{\sigma^2}(X+Z)$$
$$E(X-\hat{X})^2 = E\left(\frac{D}{\sigma^2}X - \frac{\sigma^2-D}{\sigma^2}Z\right)^2$$
$$= \left(\frac{D}{\sigma^2}\right)^2 EX^2 + \left(\frac{\sigma^2-D}{\sigma^2}\right)^2 EZ^2$$
$$= \left(\frac{D}{\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2-D}{\sigma^2}\right)^2 \frac{D\sigma^2}{\sigma^2-D}$$
$$= \frac{D^2}{\sigma^2} + \frac{(\sigma^2-D)\cdot D}{\sigma^2} = D$$

$$E[\hat{X}]^2 = \left(\frac{\sigma^2-D}{\sigma^2}\right)^2 E(X+Z)^2$$

$$= \left(\frac{\sigma^2-D}{\sigma^2}\right)^2 (EX^2+EZ^2)$$

$$= \left(\frac{\sigma^2-D}{\sigma^2}\right)^2 \left(\sigma^2+\frac{D\sigma^2}{\sigma^2-D}\right)$$

$$= \sigma^2-D$$

$$I(X;\hat{X}) = h(\hat{X}) - h(\hat{X}|X)$$

$$= h(\hat{X}) - h\left(\frac{\sigma^2-D}{\sigma^2}Z\right)$$

$$= h(\hat{X}) - h(Z) - \log\frac{\sigma^2-D}{\sigma^2}$$

$$\le h(N(0,\sigma^2-D)) - \frac{1}{2}\log 2\pi e\frac{D\sigma^2}{\sigma^2-D} - \log\frac{\sigma^2-D}{\sigma^2}$$

$$= \frac{1}{2}\log(2\pi e)(\sigma^2-D) - \frac{1}{2}\log(2\pi e)\frac{D\sigma^2}{\sigma^2-D} - \log\frac{\sigma^2-D}{\sigma^2}$$

$$= \frac{1}{2}\log\frac{2\pi e(\sigma^2-D)}{2\pi e}\cdot\frac{\sigma^2-D}{D\sigma^2}\,\frac{\sigma^4}{(\sigma^2-D)^2}$$

$$= \frac{1}{2}\log\frac{\sigma^2}{D}$$

$$\therefore\ I(X;\hat{X}) \le \frac{1}{2}\log\frac{\sigma^2}{D} \qquad R(D) \le \frac{1}{2}\log\frac{\sigma^2}{D}$$

If X is Gaussian, the lower bound is equal to the upper bound, if X is not Gaussian, lower bound is less than upper bound.
Therefore, Gaussian random variables are harder to describe than others, since the Gaussian random variables have the maximum entropy.

7. Simplicity is best
Source W with zero mean and variance P

$$W \sim N(0, P)$$

$Y = X + Z$ with power constraint P and $Z \sim N(0, N)$

(a) Derive the smallest distortion possible in this scenario.

$$R(D) = \min I(W; \hat{W})$$
$$= h(W) - h(W|\hat{W}) = \frac{1}{2}\log(2\pi e P) - h(W - \hat{W}|\hat{W})$$
$$\geq \frac{1}{2}\log(2\pi e P) - h(W - \hat{W})$$
$$\geq \frac{1}{2}\log(2\pi e P) - h(N(0, E(X-\hat{X})^2))$$
$$= \frac{1}{2}\log(P/D)$$

$$C = \max I(X;Y)$$
$$= H(Y) - H(Y|X) = H(Y) - H(Z)$$
$$\leq \frac{1}{2}\log(2\pi e)(P+N) - \frac{1}{2}\log(2\pi e N)$$

$= \frac{1}{2}\log\left(1+\frac{P}{N}\right) = C$

The Gaussian channel coding theorem tells us that if $R \le C$, we can recover the information such that the average error probability close to zero

when R is high, distortion D is low
when $R = C$, we can achieve the smallest D
in this scenario

$\therefore \frac{1}{2}\log\left(\frac{P}{D}\right) \le \frac{1}{2}\log\left(1+\frac{P}{N}\right)$

$\therefore \frac{P}{D} \le 1+\frac{P}{N} = \frac{N+P}{N}$

$\Rightarrow D \ge \frac{PN}{N+P}$ $\quad\therefore$ the smallest D is $\frac{PN}{N+P}$

(b) $X = W$ $\quad W \sim \mathcal{N}(0,P)$ $\therefore X \sim \mathcal{N}(0,P)$

$Y = X + Z$ $\quad Z \sim \mathcal{N}(0,N)$ $\quad\therefore Y \sim \mathcal{N}(0,P+N)$

consider the channel $\hat{W} = Y$ $\quad W = X$

$R(D) = I(\hat{W};W) = I(X;Y)$

$= \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ as $\frac{P}{N} \to \infty$ $\quad R(D) \to \infty$

The R, D pair as $R \to \infty$ $D \to 0$
if we have large SNR, distortion is close to zero

Therefore, we can approach the theoretical performance Limit

(c) Now show that the proper choice of a produces a biased $\hat{w} = aY$ that achieves the performance limit of part a for every value of P/N assuming $X=W$

Considering the channel

$$W = X \rightarrow \oplus \longrightarrow \odot \longrightarrow \hat{w}$$
$$\uparrow_{Z} \qquad \uparrow_{a}$$

$Z \sim \mathcal{N}(0, N)$ if we could achieve the $D = \frac{PN}{P+N}$

$N = PD/(P-D)$

consider the channel is problem 6.

$$w = x \rightarrow \oplus \longrightarrow \odot \longrightarrow \hat{w}$$

$Z \sim (0, \frac{PD}{P-D})$ $\quad \frac{P-D}{P}$ $\qquad \hat{w} = \frac{P-D}{P}(x+z)$

$$E(w-\hat{w})^2 = E\left(\frac{D}{P}w - \frac{P-D}{P}z\right)^2$$
$$= D^2/P + (P-D)\cdot D/P = D = \frac{PN}{P+N}$$

$\therefore$ For every value of P/N, we can choose $a = \frac{P-D}{P}$ to achieves the performance Limit of part a

As $D = \frac{PN}{P+N}$, a is $a = \frac{(P/N)^2 + 2P/N}{P/N + 1}$ (proper a)

8. Properties of optimal rate distortion code.

A good (R, D) rate distortion code $R \approx R(D)$ puts severe constraints on the relationship of the source X^n and the representations $\hat{X}^n$.

Examine the chain of inequalities considering the conditions for equality and interpret as properties of a good code.

The converse of the rate distortion theorem relies on the following chain of inequalities

$$
\begin{aligned}
nR &\overset{(a)}{\geq} H(\hat{X}^n) \\
&\overset{(b)}{\geq} H(\hat{X}^n) - H(\hat{X}^n \mid X^n) \\
&= I(X^n ; \hat{X}^n) \\
&= H(X^n) - H(X^n \mid \hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - H(X^n \mid \hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid \hat{X}^n, X_{i-1}, \cdots, X_1) \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid \hat{X}_i) \\
&= \sum_{i=1}^{n} I(X_i ; \hat{X}_i)
\end{aligned}
$$

(d)

$$\geq \sum_{i=1}^{n} R(Ed(X_i, \hat{X}_i)) = n \sum_{i=1}^{n} \frac{1}{n} R(Ed(X_i, \hat{X}_i))$$

(e)

$$\geq nR\left(\frac{1}{n} \sum_{i=1}^{n} Ed(X_i ; \hat{X}_i)\right)$$

$$= nR\left(Ed(X^n, \hat{X}^n)\right) = n R(D)$$

(a) is equal if all the codewords are equally likely
(b) is equal if $\hat{X}^n$ is a deterministic function of X^n
(c) is equal if each X_i depends only on the corresponding $\hat{X}_i$

(d) is equal if the joint distribution X_i and $\hat{X}_i$ is the one achieving the minimum in the definition of the rate distortion function
(e) is equal if all the distortions are equal.

The optimal rate distortion code is deterministic, and the joint distribution between the source symbol and the codeword are indeppendent, equal to the joint distribution that achieves the minimum $R(D)$. For each instant, the $R(D)$ are the same.