EEE 551 Information Theory (Spring 2022)

Homework 5

---

1. *Cost-constrained binary channels.* Consider a cost-constrained binary input channel with cost function $b(x) = x$ for $x \in \{0, 1\}$. Find the capacity-cost function $C(B)$ (a) for the binary symmetric channel and (b) for the binary erasure channel.

   Solution: Recall that the capacity-cost function is given by

   $$C(B) = \max_{p(x):\mathbb{E}b(X)\leq B} I(X;Y).$$

   (a) For the binary symmetric channel with crossover probability p, let $X \sim \text{Bern}(r)$. To satisfy the cost constraint $\mathbb{E}b(X) \leq B$, we need $r \leq B$. Since $r = 1/2$ is the optimal choice of input distribution without a cost constraint, if $B \geq 1/2$ then $r = 1/2$ satisfies the cost constraint, so the capacity-cost function is simply the capacity of the channel with no constraint, i.e. $C(B) = 1 - H(p)$.

   Now consider the case $B < 1/2$. Note that $\Pr\{Y = 1\} = r(1-p) + (1-r)p = p + r(1-2p)$. If $p < 1/2$ and $r < 1/2$ then this quantity is increasing in r and less than $1/2$, meaning $H(Y)$ is increasing in r. If $p > 1/2$ and $r < 1/2$, then this quantity is decreasing in r but greater than $1/2$, meaning again that $H(Y)$ is increasing in r. Thus

   $$C(B) = \max_{r\leq B} H(Y) - H(Y|X) = H(B(1-p) + (1-B)p) - H(p)$$

   where equality is achieved by $r = B$.

   In summary,

   $$C(B) = \begin{cases} H\big(B(1-p) + (1-B)p\big) - H(p) & B < 1/2 \\ 1 - H(p) & B \geq 1/2. \end{cases}$$

   (b) For the binary erasure channel with erasure probability α, again let $X \sim \text{Bern}(r)$. We may write the mutual information as

   $$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(r(1-\alpha), (1-r)(1-\alpha), \alpha) - H(\alpha) \\ &= H(\alpha) + (1-\alpha)H(r) - H(\alpha) \\ &= (1-\alpha)H(r). \end{aligned}$$

   This is maximized over $r \leq B$ at $r = \min\{B, 1/2\}$, so

   $$C(B) = \begin{cases} (1-\alpha)H(B) & B < 1/2 \\ 1-\alpha & B \geq 1/2. \end{cases}$$

2. *Parallel Gaussian channels.* Consider 3 parallel Gaussian channels with noise variances given by

   $$N_1 = 1, \qquad N_2 = 4, \qquad N_3 = 10.$$

   Find the capacity as a function of total power P. Make sure your answer is in completely closed form (i.e., solve for $\alpha^\star$ as a function of P).

   Solution: The water filling solution is

   $$C = \sum_i \frac{1}{2}\log\left(1 + \frac{P_i}{N_i}\right)$$

where $P_i = (\alpha^\star - N_i)^+$, and $\alpha^\star$ is chosen such that $\sum_i P_i = P$. Let us consider different ranges for $\alpha^\star$. For $1 \le \alpha^\star \le 4$, we have

$$P_1 = \alpha^\star - 1, \qquad P_2 = P_3 = 0.$$

Thus this range works for $P = \alpha^\star - 1$, so we must have $P \in [0, 3]$. Thus, in this range we have

$$C = \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 1}{1}\right) = \frac{1}{2}\log(1 + P).$$

The next range is $4 \le \alpha^\star \le 10$, where

$$P_1 = \alpha^\star - 1, \qquad P_2 = \alpha^\star - 4, \qquad P_3 = 0.$$

Thus the total power is

$$P = 2\alpha^\star - 5$$

which means $3 \le P \le 15$, and $\alpha^\star = \frac{P+5}{2}$, so

$$C = \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 1}{1}\right) + \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 4}{4}\right) = \frac{1}{2}\log\left(1 + \frac{P+3}{2}\right) + \frac{1}{2}\left(1 + \frac{P-3}{8}\right).$$

The final range is $\alpha^\star \ge 10$, where

$$P_1 = \alpha^\star - 1, \qquad P_2 = \alpha^\star - 4, \qquad P_3 = \alpha^\star - 10.$$

Thus the total power is

$$P = 3\alpha^\star - 15$$

which means $P \ge 15$, and $\alpha^\star = \frac{P+15}{3}$, so

$$\begin{aligned} C &= \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 1}{1}\right) + \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 4}{4}\right) + \frac{1}{2}\log\left(1 + \frac{\alpha^\star - 10}{10}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{P+12}{3}\right) + \frac{1}{2}\log\left(1 + \frac{P+3}{12}\right) + \frac{1}{2}\log\left(1 + \frac{P-15}{30}\right). \end{aligned}$$

To summarize,

$$C = \begin{cases} \frac{1}{2}\log(1+P), & 0 \le P \le 3, \\ \frac{1}{2}\log\left(1 + \frac{P+3}{2}\right) + \frac{1}{2}\left(1 + \frac{P-3}{8}\right), & 3 \le P \le 15, \\ \frac{1}{2}\log\left(1 + \frac{P+12}{3}\right) + \frac{1}{2}\log\left(1 + \frac{P+3}{12}\right) + \frac{1}{2}\log\left(1 + \frac{P-15}{30}\right), & P \ge 15. \end{cases}$$

3. *Parallel binary erasure channels.* Consider k parallel binary erasure channels with an overall cost constraint. That is, the input is $X = (X_1, X_2, \ldots, X_k)$ where $X_j \in \{0, 1\}$. The output is $Y = (Y_1, Y_2, \ldots, Y_k)$, where Y_j is the output of a binary erasure channel with X_j as the input and erasure probability p_j. There is a joint cost constraint on the input codeword $x^n = (x_1^n, x_2^n, \ldots, x_k^n)$ given by

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ji} \le B.$$

Find the capacity of this channel. You may give your answer in parametric form (similar to the water filling solution, in which the capacity is written in terms of a Lagrange variable).

Solution: We know that the capacity-cost function will be

$$C(B) = \max_{p(x_1,\ldots,x_k):\mathbb{E}[\sum_{j=1}^{k} X_j] \le B} I(X_1, \ldots, X_k; Y_1, \ldots, Y_k).$$

We may upper bound the mutual information by

$$\begin{aligned}
I(X_1,\ldots,X_k;Y_1,\ldots,Y_k) &= H(Y_1,\ldots,Y_k) - H(Y_1,\ldots,Y_k|X_1,\ldots,X_k) \\
&\le \sum_{j=1}^k H(Y_j) - \sum_{j=1}^k H(Y_j|X_j) \\
&= \sum_{j=1}^k I(X_j;Y_j).
\end{aligned}$$

This upper bound is achievable by taking $X_1,\ldots,X_k$ to be independent. If we assume that $X_j \sim \mathrm{Bern}(r_j)$, then

$$\begin{aligned}
I(X_j;Y_j) &= H(Y_j) - H(Y_j|X_j) \\
&= H(r_j(1-p_j),(1-r_j)(1-p_j),p_j) - H(p_j) \\
&= H(p_j) + (1-p_j)H(r_j) - H(p_j) = (1-p_j)H(r_j).
\end{aligned}$$

Moreover

$$\mathbb{E}[\sum_{j=1}^k X_j] = \sum_{j=1}^k r_j.$$

Thus we may write the capacity-cost function as

$$C(B) = \max_{r_1,\ldots,r_k:\sum_{j=1}^k r_j \le B} \sum_{j=1}^k (1-p_j)H(r_j).$$

Without considering the cost constraint, the optimal choice of r_j is $r_j = 1/2$ for all j; thus if $B \ge k/2$, then $C(B) \ge \sum_{j=1}^k (1-p_j)$. Otherwise, we have the constrained optimization problem

$$\begin{array}{ll}
\text{maximize} & \sum_{j=1}^k (1-p_j)\left(-r_j \log r_j - (1-r_j)\log(1-r_j)\right) \\
\text{subject to} & \sum_{j=1}^k r_j \le B, \\
& r_j \ge 0.
\end{array}$$

The Lagrangian for this optimization problem is

$$\sum_{j=1}^k (1-p_j)\left(r_j \ln r_j + (1-r_j)\ln(1-r_j)\right) + \nu\left(\sum_{j=1}^k r_j - B\right) - \sum_{j=1}^k \lambda_j r_j.$$

Differentiating with respect to r_j, the optimal point satisfies the condition

$$0 = (1-p_j)\left(\ln r_j - \ln(1-r_j)\right) + \nu - \lambda_j$$

Solving for r_j, we find

$$r_j = \frac{e^{\frac{\lambda_j-\nu}{1-p_j}}}{1+e^{\frac{\lambda_j-\nu}{1-p_j}}}.$$

Note that for any finite λ_j,ν, this is always positive. That is, $r_j > 0$, so by complementary slackness, we actually have $\lambda_j = 0$, so

$$r_j = \frac{e^{\frac{-\nu}{1-p_j}}}{1+e^{\frac{-\nu}{1-p_j}}}. \tag{1}$$

Therefore, for $B \le k/2$, we have

$$C(B) = \sum_{j=1}^k (1-p_j)H(r_j)$$

where r_j is given by (1), and ν is chosen so that $\sum_{j=1}^k r_j = B$.

4. Problem 10.5 from Cover-Thomas: *Rate distortion for uniform source with Hamming distortion.* Consider a source X uniformly distributed on the set $\{1, 2, \ldots, m\}$. Find the rate distortion function for this source with Hamming distortion; that is,

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

Solution: Consider any $p(\hat{x}|x)$ such that $\mathbb{E}d(X, \hat{X}) \leq D$. Let $D' = \mathbb{E}d(X, \hat{X})$. Note that $d(X, \hat{X})$ is a binary random variable with entropy $H(D')$. We may write

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &\overset{\text{(a)}}{=} \log m - H(X, d(X, \hat{X})|\hat{X}) \\ &= \log m - H(d(X, \hat{X})|\hat{X}) - H(X|\hat{X}, d(X, \hat{X})) \\ &\overset{\text{(b)}}{\geq} \log m - H(d(X, \hat{X})) - \Pr\{d(X, \hat{X}) = 1\} H(X|\hat{X}, d(X, \hat{X}) = 1) \\ &\qquad\qquad - \Pr\{d(X, \hat{X}) = 0\} H(X|\hat{X}, d(X, \hat{X}) = 0) \\ &\overset{\text{(c)}}{=} \log m - H(D') - D' H(X|\hat{X}, d(X, \hat{X}) = 1) \\ &\overset{\text{(d)}}{\geq} \log m - H(D') - D' \log(m - 1) \end{aligned}$$

where:

- in (a) we have used the fact that $d(X, \hat{X})$ is a function of X and $\hat{X}$,
- in (b) we have used the fact that conditioning reduces entropy, and expanded the third term around $d(X, \hat{X})$
- in (c) we have used the fact that if $d(X, \hat{X}) = 0$, then $X = \hat{X}$ so X is determined by $\hat{X}$
- in (d) we have used the fact that if $d(X, \hat{X}) = 1$, then $X \neq \hat{X}$ so it may only take on $m - 1$ possible values, so its entropy is at most $\log(m - 1)$.

We may achieve equality in the above analysis with the following test channel. (Recall that a test channel is a channel from $\hat{X}$ back to X.) Given $\hat{X} = \hat{x}$, $X = \hat{x}$ with probability $1 - D'$, and $X = x$ with probability $\frac{D'}{m-1}$ for all $x \neq \hat{x}$. This test channel is valid because a uniform $\hat{X}$ leads to a uniform X. Moreover, $\mathbb{E}d(X, \hat{X}) = \Pr(X \neq \hat{X}) = D'$, and $H(X|\hat{X}) = H(D) + (1 - D) \log(m - 1)$, so equality is achieved in the above analysis. Therefore we may write the rate-distortion function as

$$R(D) = \min_{D' \leq D} \log m - H(D') - D' \log(m - 1).$$

Note that $H(D') + D' \log(m - 1)$ is a concave function of D'. We may find the value of D' that maximizes $H(D') + D' \log(m - 1)$ as follows:

$$\frac{d}{dD'}[-D' \log D' - (1 - D') \log(1 - D') + D' \log(m - 1)] = \log \frac{1 - D'}{D'} + \log(m - 1).$$

The above equals 0 (i.e. is a critical point) if $D' = \frac{m-1}{m}$. Thus $H(D') + D' \log(m - 1)$ is increasing for $D' \leq \frac{m-1}{m}$ and decreasing for larger values. Hence if $D \leq \frac{m-1}{m}$, the best choice is $D' = D$, and if $D > \frac{m-1}{m}$, the best choice is $D' = \frac{m-1}{m}$. Therefore the rate distortion function is

$$R(D) = \begin{cases} \log m - H(D) - D \log(m - 1) & D \leq \frac{m-1}{m} \\ \log m - H\left(\frac{m-1}{m}\right) - \frac{m-1}{m} \log(m - 1) & D > \frac{m-1}{m}. \end{cases}$$

5. Problem 10.7 from Cover-Thomas: *Erasure distortion.* Consider $X \sim$ Bernoulli($\frac{1}{2}$), and let the distortion measure be given by a matrix (each row is a letter of $\mathcal{X}$, each columns is a letter of $\hat{\mathcal{X}}$)

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

Calculate the rate distortion function for this source. Can you suggest a simple scheme to achieve any value of the rate distortion function for this source?

Solution: Denote the reconstruction alphabet by $\{0, e, 1\}$. To give a finite distortion, we cannot have $(X, \hat{X}) = (0, 1)$ or $(X, \hat{X}) = (1, 0)$ occur with positive probability. Assuming these do not occur, $\mathbb{E}d(X, \hat{X}) = \Pr\{\hat{X} = e\}$. If $D \geq 1$, then we may simply set $\hat{X} = e$ and automatically satisfy the distortion constraint, giving $R(D) = 0$.

Now consider the case $D < 1$. For a distribution that satisfies the distortion constraint $\Pr\{\hat{X} = e\} \leq D$, we may write

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= 1 - \Pr\{\hat{X} = 0\}H(X|\hat{X} = 0) - \Pr\{\hat{X} = 1\}H(X|\hat{X} = 1) - \Pr\{\hat{X} = e\}H(X|\hat{X} = e) \\ &\overset{(a)}{=} 1 - \Pr\{\hat{X} = e\}H(X|\hat{X} = e) \\ &\overset{(b)}{\geq} 1 - \Pr\{\hat{X} = e\} \\ &\overset{(c)}{\geq} 1 - D \end{aligned}$$

where (a) follows because when $\hat{X} = 0$ or $\hat{X} = 1$, we must have $X = \hat{X}$, (b) follows by the uniform bound, and (c) follows since by the distortion constraint.

The following test channel achieves equality in the above analysis. If $\hat{X} = 0$ or $\hat{X} = 1$, then $X = \hat{X}$, and if $\hat{X} = e$, then X is uniformly distributed between 0 and 1. This gives the proper distribution on X as long as $\hat{X} = 0$ and $\hat{X} = 1$ are equally likely. Moreover, by choosing $\Pr\{\hat{X} = e\} = D$, we satisfy the distortion constraint and achieve equality in the above. Therefore

$$R(D) = \begin{cases} 1 - D & D < 1 \\ 0 & D \geq 1. \end{cases}$$

The following simple schemes achieves the above rate-distortion function. Assuming $D < 1$, let $k = \lfloor n(1 - D) \rfloor$. Given x^n, we encode the first k bits. This requires at most $n(1 - D)$ bits, so we achieve rate $R = 1 - D$. At the decoder, we are able to determine $x_1, \ldots, x_k$ with no error, so we decode to $\hat{x}^n = (x_1, \ldots, x_k, e, e, \ldots, e)$. Thus $d(x^n, \hat{x}^n) = n - k \leq n - n(1 - D) + 1 = nD + 1$, so

$$\lim_{n \to \infty} \frac{1}{n} d(x^n, \hat{x}^n) \leq \lim_{n \to \infty} \frac{1}{n}(nD + 1) = D.$$

6. *Rate-distortion for exponential random variable.* Let X be an exponential random variable with expectation μ; i.e., $f_X(x) = \frac{1}{\mu}e^{-x/\mu}$ for $x \geq 0$. Define a distortion function for $x, \hat{x} \in \mathbb{R}$ as

$$d(x, \hat{x}) = \begin{cases} x - \hat{x}, & \hat{x} \leq x \\ \infty, & \hat{x} > x. \end{cases}$$

Find the rate-distortion function. *Hint:* Remember the results from problems 7 and 8 from Homework 4.

Solution: Recall from Homework 4 that for an exponential random variable with mean μ, the differential entropy is $\log(e\mu)$. Moreover, for any non-negative variable Z, if $\mathbb{E}[Z] \leq \mu$, then $h(Z) \leq \log(e\mu)$. Consider first $D \geq \mu$. We can choose $\hat{X} = 0$, so

$$\mathbb{E}[d(X, \hat{X})] = \mathbb{E}[X - \hat{X}] = \mathbb{E}[X] = \mu \leq D$$

where we have used the fact that since $\hat{X} = 0$, then X is always greater than $\hat{X}$. Moreover, $I(X; \hat{X}) = 0$, so $R(D) = 0$.

Now consider any $D < \mu$. Consider any $f(\hat{x}|x)$ such that $\mathbb{E}d(X, \hat{X}) \leq D$. Since $d(x, \hat{x}) = \infty$ when $\hat{x} > x$, we may assume that $\hat{X} \leq X$, so $\mathbb{E}d(X, \hat{X}) = \mathbb{E}[X - \hat{X}] \leq D$. Thus we may lower bound the mutual information by

$$\begin{aligned} I(X; \hat{X}) &= h(X) - h(X|\hat{X}) \\ &= \log(e\mu) - h(X - \hat{X}|\hat{X}) \\ &\geq \log(e\mu) - h(X - \hat{X}) \\ &\overset{\text{(a)}}{\geq} \log(e\mu) - \log(eD) \\ &= \log\left(\frac{\mu}{D}\right). \end{aligned}$$

where in (a) we have used the fact that $X - \hat{X}$ is non-negative with expectation less than D, so by the earlier result $h(X - \hat{X}) \leq \log(eD)$. To achieve equality in the above bound, we must have $Z = X - \hat{X}$ be independent of $\hat{X}$, and Z be an exponential random variable with $\mathbb{E}[Z] = D$. That is, $X = \hat{X} + Z$, where $\hat{X}$ and Z are independent, and both X and Z are exponential random variables. In order to show that this is achievable, we need to show that there exists a distribution for $\hat{X}$ that gives the correct distribution for X. We use moment generation functions. In particular,

$$\begin{aligned} \mathbb{E}[e^{tX}] &= \frac{1}{1 - \mu t}, \\ \mathbb{E}[e^{tZ}] &= \frac{1}{1 - Dt}. \end{aligned}$$

Since $\hat{X}$ and Z should be independent, we have

$$\mathbb{E}[e^{tX}] = \mathbb{E}[e^{t(\hat{X}+Z)}] = \mathbb{E}[e^{t\hat{X}}]\,\mathbb{E}[e^{tZ}].$$

Thus we should have

$$\mathbb{E}[e^{t\hat{X}}] = \frac{\mathbb{E}[e^{tX}]}{\mathbb{E}[e^{tZ}]} = \frac{1 - Dt}{1 - \mu t} = \frac{\frac{D}{\mu}(1 - \mu t) + 1 - \frac{D}{\mu}}{1 - \mu t} = \frac{D}{\mu} + \frac{1 - \frac{D}{\mu}}{1 - \mu t}.$$

The last expression is the moment generation function of the following distribution. Letting $\hat{X}'$ be an exponential distribution with mean μ,

$$\hat{X} = \begin{cases} 0 & \text{w.p. } \frac{D}{\mu} \\ \hat{X}' & \text{w.p. } 1 - \frac{D}{\mu}. \end{cases}$$

This distribution exists because $D < \mu$. Therefore, the lower bound is achievable, so

$$R(D) = \begin{cases} \log\left(\frac{\mu}{D}\right), & D < \mu, \\ 0, & D \geq \mu. \end{cases}$$

7. Problem 10.17 from Cover-Thomas: *Source-channel separation theorem with distortion.* Let $V_1, V_2, \ldots, V_n$ be a finite alphabet i.i.d. source with is encoded as a sequence of n input symbols X^n of a discrete memoryless channel. The output of the channel Y^n is mapped onto the reconstruction alphabet $\hat{V} = g(Y^n)$. Let $D = \mathbb{E}d(V^n, \hat{V}^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}d(V_i, \hat{V}_i)$ be the average distortion achieved by this combined source and channel coding scheme.

$$V^n \longrightarrow X^n(V^n) \longrightarrow \boxed{\text{Channel Capacity } C} \longrightarrow Y^n \longrightarrow \hat{V}^n$$

(a) Show that if $C > R(D)$, where $R(D)$ is the rate distortion function for V, it is possible to find encoders and decoders that achieve an average distortion arbitrarily close to D.

(b) (Converse) Show that if the average distortion is equal to D, the capacity of the channel C must be greater than $R(D)$.

Solution:

(a) Choose rate R where $C > R > R(D)$. Thus, since $R < C$, there exists a channel code over the noisy channel with rate R and arbitrarily small probability of error. Moreover, since $R > R(D)$, there exists a rate-distortion code for source V and distortion function d that achieves expected distortion arbitrarily close to D. Plugging these two codes together (i.e. the output of the rate-distortion encoder is used as the message for the channel code), the overall distortion is arbitrarily close to D.

(b) Using the data processing inequality, we have

$$I(V^n; \hat{V}^n) \leq I(X^n; Y^n).$$

Reproducing part of the converse argument for rate-distortion, we have (assuming the expected distortion is D)

$$\begin{aligned} I(V^n; \hat{V}^n) &\geq \sum_{i=1}^{n} I(V_i; \hat{V}_i) \\ &\geq \sum_{i=1}^{n} R(\mathbb{E}d(V_i, \hat{V}_i)) \\ &\geq nR(\mathbb{E}d(V^n, \hat{V}^n)) \\ &= nR(D). \end{aligned}$$

Now reproducing part of the converse argument for channel coding, we have

$$\begin{aligned} I(X^n; Y^n) &\leq \sum_{i=1}^{n} I(X_i; Y_i) \\ &\leq nC. \end{aligned}$$

Chaining together all of the above inequalities and dividing by n gives $R(D) \geq C$.

8. *Source-channel separation for a Gaussian source with a Gaussian channel.* Consider a special case of the setup from problem 7 where $V \sim \mathcal{N}(0, \sigma^2)$, the channel is a Gaussian noise channel with power constraint P and noise variance N, and the distortion function is the squared error distortion; i.e., $d(v, \hat{v}) = (v - \hat{v})^2$.

(a) Using the result from problem 7, find the smallest possible expected distortion D given the other parameters.

(b) Now consider the following simple approach without coding. For each $i = 1, \ldots, n$, the encoder sends $X_i = \alpha V_i$, and the decoder estimates the source using $\hat{V}_i = \beta Y_i$. Here, α and β are constants that must be determined. What is the minimum expected distortion D achievable using this scheme? Be sure to choose α to satisfy the power constraint of the channel, and β to minimize the expected distortion. Compare your to answer to that from part (a).

Solution:

(a) From problem 7, we need $C > R(D)$. The capacity is $C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$, and the rate-distortion function is $R(D) = \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right)$. Thus we can achieve any distortion D where

$$\frac{1}{2}\log\left(1 + \frac{P}{N}\right) > \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right).$$

We can rewrite this condition as

$$1 + \frac{P}{N} < \frac{\sigma^2}{D}.$$

Rearranging this, we need

$$D > \frac{\sigma^2}{1 + \frac{P}{N}} = \frac{N\sigma^2}{N + P}.$$

Thus, the smallest achievable distortion is $\frac{N\sigma^2}{N+P}$.

(b) In order to satisfy the power constraint, we need $\mathbb{E}[X_i^2] \leq P$. Since $X_i = \alpha V_i$, we have

$$\mathbb{E}[X_i^2] = \alpha^2\,\mathbb{E}[V_i^2] = \alpha^2\sigma^2.$$

Thus we need $\alpha^2 \leq \frac{P}{\sigma^2}$. We now calculate the expected distortion (in the following $Z_i \sim \mathcal{N}(0, N)$ is the channel noise):

$$\begin{aligned}
D &= \mathbb{E}(V_i - \hat{V}_i)^2 \\
&= \mathbb{E}(V_i - \beta Y_i)^2 \\
&= \mathbb{E}(V_i - \beta(X_i + Z_i))^2 \\
&= \mathbb{E}(V_i - \beta(\alpha V_i + Z_i))^2 \\
&= \mathbb{E}((1 - \alpha\beta)V_i - \beta Z_i)^2 \\
&= (1 - \alpha\beta)^2\mathbb{E}[V_i^2] + \beta^2\mathbb{E}[Z_i^2] \\
&= (1 - \alpha\beta)^2\sigma^2 + \beta^2 N.
\end{aligned}$$

We can minimize this distortion over β by setting with respect to β to 0:

$$0 = \frac{\partial D}{\partial \beta} = -2\alpha(1 - \alpha\beta)\sigma^2 + 2\beta N = 2(\alpha^2\sigma^2 + N)\beta - 2\alpha\sigma^2$$

so the best choice of beta is

$$\beta = \frac{\alpha\sigma^2}{\alpha^2\sigma^2 + N}$$

and the minimal distortion is

$$\begin{aligned}
D &= \left(1 - \alpha\frac{\alpha\sigma^2}{\alpha^2\sigma^2 + N}\right)^2\sigma^2 + \left(\frac{\alpha\sigma^2}{\alpha^2\sigma^2 + N}\right)^2 N \\
&= \frac{N^2\sigma^2}{(\alpha^2\sigma^2 + N)^2} + \frac{\alpha^2\sigma^4 N}{(\alpha^2\sigma^2 + N)^2} \\
&= \frac{\sigma^2 N(N + \alpha^2\sigma^2)}{(\alpha^2\sigma^2 + N)^2} \\
&= \frac{\sigma^2 N}{\alpha^2\sigma^2 + N}.
\end{aligned}$$

Recalling the power constraint that $\alpha^2 \leq \frac{P}{\sigma^2}$, the smallest we can make the distortion is

$$D = \frac{\sigma^2 N}{P + N}.$$

This is exactly the same as we found in part (a)!