JAIST — School of Information Science — I232 Information Theory

Final Exam 2021

Instructors: Brian M. Kurkoski and Lei Liu

June 7, 2021 13:30–15:10

SOLUTIONS

Exam policy:

- You may not connect to the internet, except to use the LMS.
- You may not communicate with anyone during the exam, except the instructor.
- This is an open-book exam. You may use only
  - *Information Theory Lecture Notes* by the instructor,
  - *Elements of Information Theory* by Cover and Thomas,
  - anything written in your own hand.
  - You can also use any material on the LMS on your device. Your device screen should not display pages other than the LMS.
- **Not following the above rules could result in failing the course.**

In addition:

- Calculators are not allowed and not needed. Numerical answers such as 5^7 and $1 + 2\log 3$ are acceptable, but simplify as much as reasonably possible.
- You can write on the back of the paper.
- Students should wear a mask during the exam.

Question	Points	Score
1	25	
2	20	
3	25	
4	30	
Total:	100	

This exam has 6 pages.

1. *Information theory concepts* State the key ideas from information theory. (Do not write too much.)
   (a) (5 points) Does compressing over blocks of length n and letting $n \to \infty$ make the "performance" of a source code *strictly* better? If so, state why. If not, show why not.
   (b) (5 points) Your friend has a 16 pixel-by-16 pixel image, in which each pixel is either black or white with probability 0.5, and all pixels are independent. He wants to losslessly compress this to an 8 pixel-by-8 pixel image, again with black and white pixels. If this is possible, find a code or describe the procedure you would use to do so. If this is not possible, state the reason.
   (c) (5 points) If possible, find a prefix code over an alphabet of size $D = 3$ with codeword lengths $(1, 1, 2, 3)$. If not possible, explain the reason.
   (d) (5 points) Let X and Y be independent and discrete. Find $H(2\mathsf{X}, -2\mathsf{Y})$ in terms of the entropies $H(\mathsf{X})$, $H(\mathsf{Y})$ and $H(\mathsf{X}, \mathsf{Y})$.
   (e) (5 points) Consider a channel with capacity $C = 0.5$. Suppose a code with length $n = 10^6$ is used for this channel. What can you say about the number of message bits that can be reliably transmitted over this channel?

Solution:

(a) No, it does not *always* reduce the compression rate, i.e. the "performance." If the source is D-adic, for example $p_{\mathsf{X}}(x) \sim 2^{-i}$, then $n = 1$ compression is optimal.

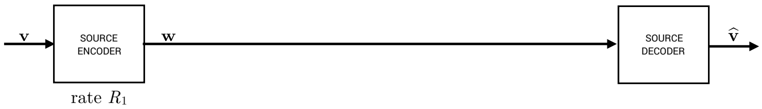
(b) This is not possible. Because the source consists of $(0.5, 0.5)$ uniform bits, $R > H(\mathsf{X}) = 1$ and lossless compression is not possible. (Lossy compression is possible.)

(c) This is possible because the Kraft inequality is satisfied. An example is $\mathcal{C} = (0, 1, 20, 210)$.

(d) Recall $H(a\mathsf{X}) = H(\mathsf{X})$ for any $a \neq 0$. In the same way, $H(2\mathsf{X}, -2\mathsf{Y}) = H(\mathsf{X}, \mathsf{Y}) = H(\mathsf{X}) + H(\mathsf{Y})$, where independence was used. Or: $H(2\mathsf{X}, -2\mathsf{Y}) = H(2\mathsf{X}) + H(-2\mathsf{Y}) = H(\mathsf{X}) + H(\mathsf{Y})$

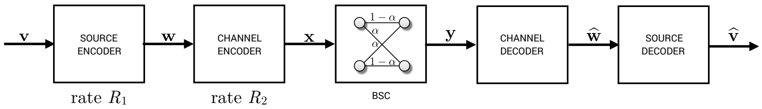
(e) By the channel coding theorem, if $R < C$ then reliable transmission is possible. So, no more than 5×10^5 message bits can be reliably transmitted.

2. *Source coding-channel coding* An information source consists of n_1 bits $\mathsf{V}_1, \ldots, \mathsf{V}_{n_1}$, i.i.d. distributed as $p_{\mathsf{V}}(v) = [1-p, p]$, $v \in \{0, 1\}$. A source encoder of rate $R_1 < 1$ compresses to a **binary** index $\mathbf{w} = (w_1, w_2, \ldots, w_{n_2})$.



(a) (5 points) How many bits n_2 are in the sequence $\mathbf{w}$? In terms of n_1 and p, what is a lower bound on n_2?

Now, to reliably transmit this index $\mathbf{w}$ over a binary symmetric channel with error probability α, a channel encoder of rate $R_2 < 1$ encodes $\mathbf{w}$ to a codeword $\mathbf{x}$.



(b) (5 points) How many bits n_3 are in the sequence $\mathbf{x}$? In terms of n_2 and α, what is a lower bound on n_3?

At the destination, a decoder estimates $\hat{\mathbf{w}}$, which is decompressed to $\hat{\mathbf{v}}$.

(c) (10 points) Let the total rate be $R = R_1/R_2$. In terms of p and α, find a condition on R such that the probability of error $\Pr[\mathsf{V} \neq \widehat{\mathsf{V}}]$ can go to 0 as $n \to \infty$.

Solution: (a) The rate of the code is R_1, so $n_2 = n_1 R_1$. Or, since the index set is $\{1, 2, \ldots, 2^{n_1 R_1}\}$, this is described using $\log\left(2^{n_1 R_1}\right) = n_1 R_1$ bits.

From optimal compression, an upper bound is $R_1 \geq H(\mathsf{V}) = h(p)$, so we have $n_2 \geq n_1 h(p)$.

(b) The rate of the code is R_2, so $n_3 = \frac{1}{R_2} n_2 = \frac{R_1}{R_2} n_1$. For the BSC, $C = 1 - h(\alpha)$. From $R_2 < C$, we have $n_3 > \dfrac{n_2}{1 - h(\alpha)}$

(c) As $n \to \infty$ we must have $R_1 > h(p)$ in order for source coding to succeed. And, we must have $R_2 < 1 - h(\alpha)$.

$$R = \frac{R_1}{R_2} > \frac{h(p)}{1 - h(\alpha)}$$

3. *Multiple access channel (MAC)* Consider the following multiple access channel (MAC):

$$Y = X_1 \oplus X_2 \oplus N,$$

where $\oplus$ denotes xor, $X_1, X_2 \in \{0,1\}$ are two binary inputs, and $N \in \{0,1\}$ is a binary noise with $p_N(0) = p$.

(a) (5 points) For fixed $X_1 = 0$, find the maximum achievable rate between X_2 and Y. Find the optimal distribution of X_2 which achieves this rate.

Solution: Since $X_1 = 0$, we have

$$Y = X_2 \oplus N.$$

Thus,

$$R = \max_{p_{X_2}} I(X_2; Y|X_1) = \max_{p_{X_2}} H(Y) - H(N) = 1 - h(p).$$

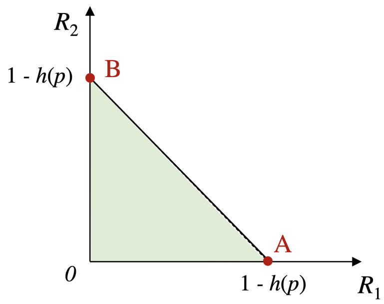
The optimal distribution of X_2 is uniform distribution, i.e., $p_{X_2}(0) = p_{X_2}(1) = 1/2$.

(b) (10 points) Let R_1 be the achievable rate of X_1 and R_2 be the achievable rate of X_2. Draw the capacity region (R_1, R_2) and label the rates on it.

Solution: The capacity region is given by

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2) = 1 - h(p), \\ R_1 &\le I(X_2; Y|X_1) = 1 - h(p), \\ R_1 + R_2 &\le I(X_1, X_2; Y) = 1 - h(p). \end{aligned}$$

Note: The mutual information above are respectively maximized by the uniform distribution of X_1, X_2 and $X_1 \oplus X_2$.



(c) (10 points) Give a scheme to achieve the whole capacity region in (b)?

Solution:

- Point A can be achieved by (a).
- Similarly, Point B can be achieved by $X_1 = 0$ and uniformly distributed X_2, i.e., $p_{X_2}(1) = p_{X_2}(0) = 1/2$.
- Any point on segment AB can be achieved by the time sharing between A and B:

  $$\alpha A + (1 - \alpha) B,$$

  where $\alpha \in [0,1]$. Since the capacity region is dominated by segment AB. Therefore, the whole capacity region is achieved.

4. *Multiple input and multiple output channel (MIMO) Channel* Consider a MIMO channel:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is a 3×1 noise vector, $\mathbf{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ is a message vector with power constraint $\sum_{i=1}^{3} \mathrm{E}\{x_i^2\} = 1$, and $\mathbf{A}$ is a channel matrix given by

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Assume that $\mathbf{A}$ is known at transmitter. The singular value decomposition (SVD) of $\mathbf{A}$ is given by

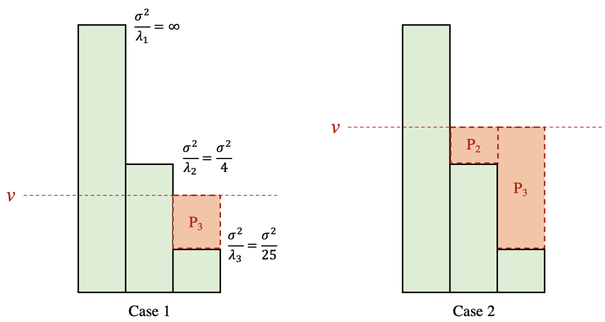
$$\mathbf{A} = \underbrace{\begin{bmatrix} 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}}_{\mathbf{V}^{\mathrm{T}}} \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}}_{\mathbf{V}},$$

where $\mathbf{V}$ is an orthogonal matrix.

(a) (5 points) Find the equivalent parallel AWGN channels for the MIMO channel.

> **Solution:** Let $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{x}$, $\tilde{\mathbf{y}} = \mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{y}$, and $\tilde{\mathbf{n}} = \mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{n}$. Then we obtain the following parallel Gaussian AWGN channels:
>
> $$\tilde{\mathbf{y}} = \mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{y} = \mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{V}\mathbf{x} + \mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{n} = \tilde{\mathbf{x}} + \tilde{\mathbf{n}},$$
>
> where $\tilde{\mathbf{n}} \sim \mathcal{N}(0, \sigma^2 \mathbf{\Lambda}^{-2})$. (Note: $1/0 = \infty$.)

(b) (10 points) Find the optimal power allocation for the parallel AWGN channels in (a).

> **Solution:** The optimal power allocation is given by water filling:
>
> $$p_i^* = (v - \sigma^2/\lambda_i^2)^{\dagger}, \quad i = 1, 2, 3.$$
>
> Since $\lambda_1 = 0$, we have $p_1^* = 0$. Furthermore, following the power constraint, we have $p_2^* + p_3^* = 1$.
>
> - Case 1: When $\sigma^2 > \frac{100}{21}$, we have
>
> $$p_1^* = p_2^* = 0,$$
> $$p_3^* = 1.$$
>
> - Case 2: When $\sigma^2 \leq \frac{100}{21}$, we have
>
> $$p_1^* = 0,$$
> $$p_2^* = \frac{1 - \frac{21}{100}\sigma^2}{2} = \frac{1}{2} - \frac{21}{200}\sigma^2,$$
> $$p_3^* = 1 - p_2^* = \frac{1}{2} + \frac{21}{200}\sigma^2.$$
>
> 
>

(c) (10 points) Based on the result in (b), find the capacity of the MIMO channel in (1).

Solution: The capacity of the MIMO channel is given by:

$$C = \frac{1}{2}\log\left(1+\frac{4p_2^*}{\sigma^2}\right)+\frac{1}{2}\log\left(1+\frac{25p_3^*}{\sigma^2}\right).$$

- Case 1: When $\sigma^2 > \frac{100}{21}$, we have

$$C = \frac{1}{2}\log\left(1+\frac{25}{\sigma^2}\right).$$

- Case 2: When $\sigma^2 \leq \frac{100}{21}$, we have

$$C = \frac{1}{2}\log\left(\frac{29}{50}+\frac{2}{\sigma^2}\right)+\frac{1}{2}\log\left(\frac{29}{8}+\frac{25}{2\sigma^2}\right).$$

(d) (2 points) Let $\sigma^2 = 6$. Calculate the optimal power allocation in (b) and the capacity in (c).

Solution: Since $\sigma^2 = 6 > \frac{100}{21}$, following Case 1 in (b), we have

$$\begin{aligned} p_1^* &= p_2^* = 0, \\ p_3^* &= 1. \end{aligned}$$

Following Case 1 in (c), we have

$$C = \frac{1}{2}\log\frac{31}{6}.$$

(e) (3 points) Let $\sigma^2 = 2$. Calculate the optimal power allocation in (b) and the capacity in (c).

Solution: Since $\sigma^2 = 2 < \frac{100}{21}$, following Case 2 in (b), we have

$$\begin{aligned} p_1^* &= 0, \\ p_2^* &= \frac{1}{2}-\frac{21}{200}\sigma^2 = \frac{29}{100}, \\ p_3^* &= 1-p_2^* = \frac{1}{2}+\frac{21}{200}\sigma^2 = \frac{71}{100}. \end{aligned}$$

Following Case 2 in (c), we have

$$C = \frac{1}{2}\log\frac{79}{50}+\frac{1}{2}\log\frac{79}{8} = \log\frac{79}{20}.$$