An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 29, 2016

Lecture #12C: Problem solving session-III

Channel Capacity

- **Problem #1:** Given in Figure is a two-input, eight-output DMC

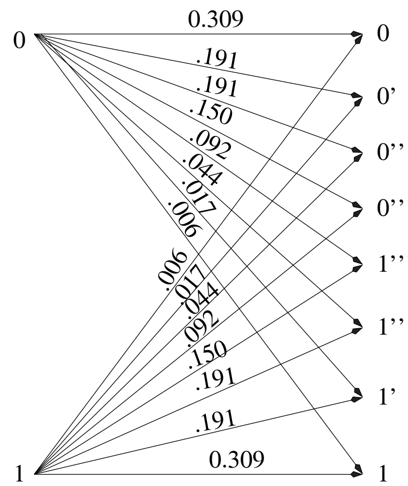


Figure: Figure for Problem 1

Channel Capacity

- **Problem #1:** Given in Figure is a two-input, eight-output DMC

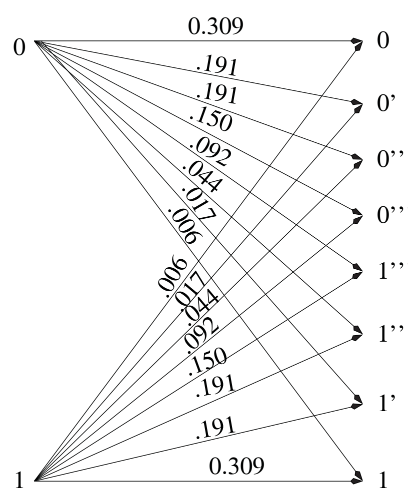


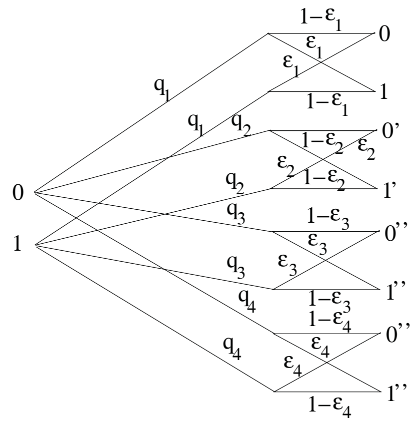
Figure: Figure for Problem 1

- Find the capacity C_8 of this DMC.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #1 (contd.)

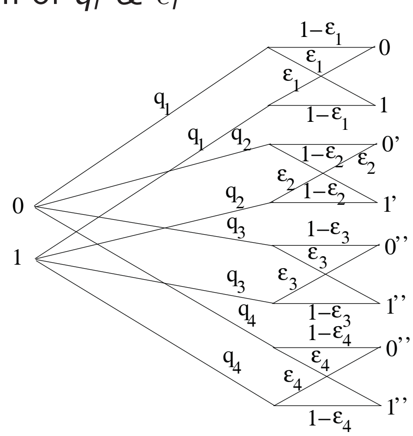
- **Solution:** Calculation of q_i & ϵ_i



$$q_1(1-\epsilon_1) = 0.309$$
$$q_1\epsilon_1 = 0.006$$

Problem #1 (contd.)

- **Solution:** Calculation of q_i & ϵ_i



$$q_1(1-\epsilon_1) = 0.309$$
$$q_1\epsilon_1 = 0.006$$

- This implies that $\epsilon_1 = 0.019$ and $q_1 = 0.315$

Problem #1 (contd.)

- Also

$$q_2(1-\epsilon_2) = 0.191$$
$$q_2\epsilon_2 = 0.017$$

Problem #1 (contd.)

- Also

$$q_2(1-\epsilon_2) = 0.191$$
$$q_2\epsilon_2 = 0.017$$

- This implies that $\epsilon_2 = 0.082$ and $q_2 = 0.208$

Problem #1 (contd.)

- Also

$$q_2(1-\epsilon_2) = 0.191$$
$$q_2\epsilon_2 = 0.017$$

- This implies that $\epsilon_2 = 0.082$ and $q_2 = 0.208$
- Similarly,

$$q_3(1-\epsilon_3) = 0.191$$
$$q_3\epsilon_3 = 0.044$$

Problem #1 (contd.)

- Also

$$q_2(1-\epsilon_2) = 0.191$$
$$q_2\epsilon_2 = 0.017$$

- This implies that $\epsilon_2 = 0.082$ and $q_2 = 0.208$
- Similarly,

$$q_3(1-\epsilon_3) = 0.191$$
$$q_3\epsilon_3 = 0.044$$

- This implies that $\epsilon_3 = 0.187$ and $q_3 = 0.235$

Problem #1 (contd.)

- Also,

$$q_4(1-\epsilon_4) = 0.150$$
$$q_4\epsilon_4 = 0.092$$

Problem #1 (contd.)

- Also,

$$q_4(1-\epsilon_4) = 0.150$$
$$q_4\epsilon_4 = 0.092$$

- This implies that $\epsilon_4 = 0.242$ and $q_4 = 0.380$

Problem #1 (contd.)

- Also,

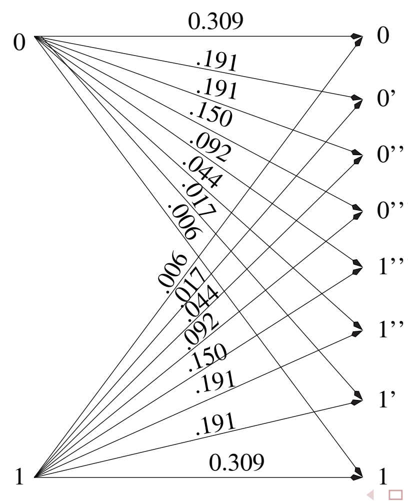
$$\begin{aligned} q_4(1-\epsilon_4) &= 0.150 \\ q_4\epsilon_4 &= 0.092 \end{aligned}$$

- This implies that $\epsilon_4 = 0.242$ and $q_4 = 0.380$
- Hence the capacity, C_8 is given by

$$\begin{aligned} C_8 &= \sum_{i=1}^{4} q_i C_i = \sum_{i=1}^{4} q_i(1-h(\epsilon_i)) \\ &= 0.315 \times 0.864 + 0.208 \times 0.592 + 0.235 \times 0.304 + 0.242 \times 0.042 \\ &= 0.477 \ \text{bits/use} \end{aligned}$$

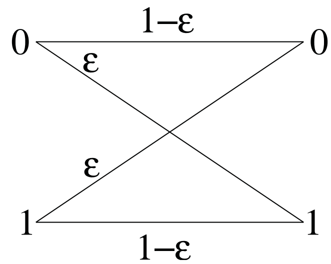
Problem #1 (contd.)

- **Problem 1(b):** If "hard-decision" demodulation had been used, then 0, 0', 0'', and 0''' would all be converted to 0 while 1, 1', 1'' and 1''' would all be converted to 1. Find the capacity C_2 of the resulting BSC and compute the decibel loss in capacity (i.e., compute $10\log_{10}(C_8/C_2)$.)



Problem #1 (contd.)

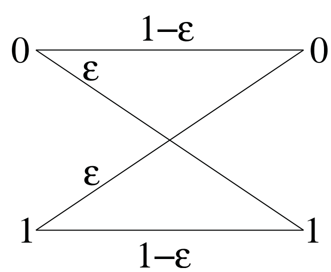
- **Solution:** ϵ is given by



$$
\begin{aligned}
\epsilon &= P_{Y/X}(0/1) + P_{Y/X}(0'/1) + P_{Y/X}(0''/1) + P_{Y/X}(0'''/1) \\
&= 0.006 + 0.017 + 0.044 + 0.092 \\
&= 0.159
\end{aligned}
$$

Problem #1 (contd.)

- **Solution:** ϵ is given by



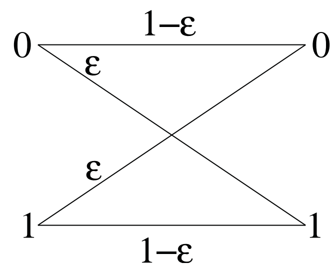
$$
\begin{aligned}
\epsilon &= P_{Y/X}(0/1) + P_{Y/X}(0'/1) + P_{Y/X}(0''/1) + P_{Y/X}(0'''/1) \\
&= 0.006 + 0.017 + 0.044 + 0.092 \\
&= 0.159
\end{aligned}
$$

- Thus $C_2 = 1 - h(\epsilon) = 0.368$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #1 (contd.)

- **Solution:** ϵ is given by



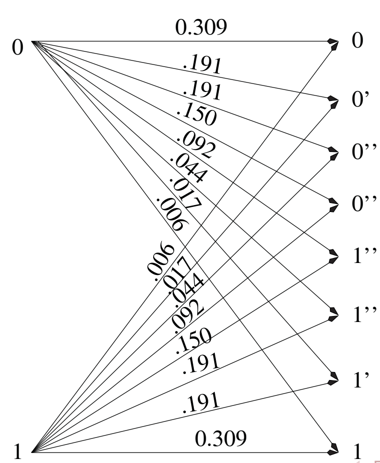
$$\begin{aligned}\epsilon &= P_{Y/X}(0/1) + P_{Y/X}(0'/1) + P_{Y/X}(0''/1) + P_{Y/X}(0'''/1) \\ &= 0.006 + 0.017 + 0.044 + 0.092 \\ &= 0.159\end{aligned}$$

- Thus $C_2 = 1 - h(\epsilon) = 0.368$.
- dB loss in capacity compared to C_8

$$D_{2/8} = 10\log_{10}\frac{C_8}{C_2} = 1.13dB$$

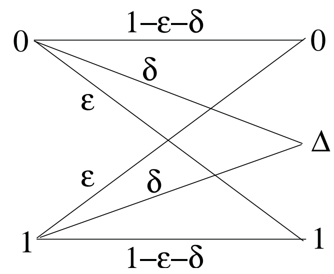
Problem #1 (contd.)

- **Problem 1(c):** There are three different sensible ways that the above channel could be converted to the Binary Symmetric Erasure Channel (BSEC) (one way is to convert 0''' and 1''' to Δ). Find the capacity C_3 for the way that gives the greatest capacity. Find also, the decibel loss compared to eight-level demodulation, and the decibel gain over hard-decision demodulation.



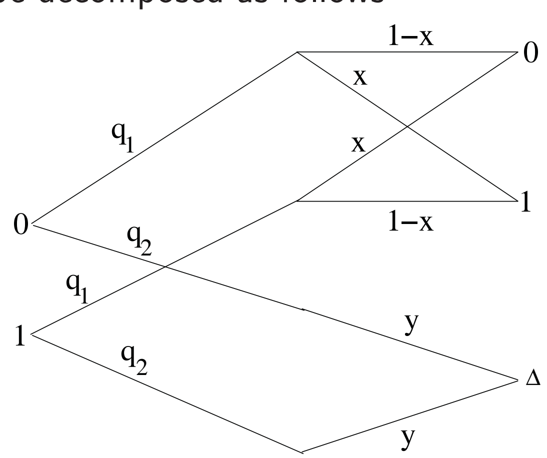
Problem #1 (contd.)

- **Solution:** Binary symmetric erasure channel



Problem #1 (contd.)

- This DMC can be decomposed as follows



Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #1 (contd.)

- Thus,

$$\begin{aligned} q_1(1-x) &= 1-\epsilon-\delta \\ q_1 x &= \epsilon \\ \Longrightarrow \frac{1-x}{x} &= \frac{1-\epsilon-\delta}{\epsilon} \end{aligned}$$

Problem #1 (contd.)

- Thus,

$$\begin{aligned} q_1(1-x) &= 1-\epsilon-\delta \\ q_1 x &= \epsilon \\ \Longrightarrow \frac{1-x}{x} &= \frac{1-\epsilon-\delta}{\epsilon} \end{aligned}$$

- This $\Longrightarrow x = \frac{\epsilon}{1-\delta}$ and $q_1 = 1-\delta$

Problem #1 (contd.)

- Thus,

$$\begin{aligned} q_1(1-x) &= 1-\epsilon-\delta \\ q_1 x &= \epsilon \\ \Longrightarrow \frac{1-x}{x} &= \frac{1-\epsilon-\delta}{\epsilon} \end{aligned}$$

- This $\Longrightarrow x = \frac{\epsilon}{1-\delta}$ and $q_1 = 1-\delta$
- Also, $q_2 = 1 - q_1 = \delta$ and $y = 1$.

$$\begin{aligned} C_1 &= 1 - h\left(\frac{\epsilon}{1-\delta}\right) \\ C_2 &= 0 \end{aligned}$$

Problem #1 (contd.)

- Capacity of BSEC is given by

$$C_3 = \sum_{i=1}^{2} q_i C_i = (1-\delta)\left[1 - h\left(\frac{\epsilon}{1-\delta}\right)\right]$$

	ϵ	δ	$1-\epsilon-\delta$	C
$0''', 1''' \Longrightarrow \Delta$	0.067	0.242	0.691	0.431
$0'', 0''', 1'', 1''' \Longrightarrow \Delta$	0.023	0.477	0.500	0.387
$0', 0'', 0''', 1', 1'', 1''' \Longrightarrow \Delta$	0.006	0.685	0.309	0.272

$0''', 1''' \Longrightarrow \Delta$ is the best.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #1 (contd.)

- dB loss is given by

$$D_{3/8} = 10\log_{10}\frac{0.477}{0.431}$$
$$= 0.44 \;\; dB$$

Problem #1 (contd.)

- dB loss is given by

$$D_{3/8} = 10\log_{10}\frac{0.477}{0.431}$$
$$= 0.44 \;\; dB$$

- dB gain is given by

$$D_{3/2} = 10\log_{10}\frac{0.431}{0.368}$$
$$= 0.6863 \;\; dB$$

Channel Capacity

- **Problem #2:** Consider the diagram shown in Figure

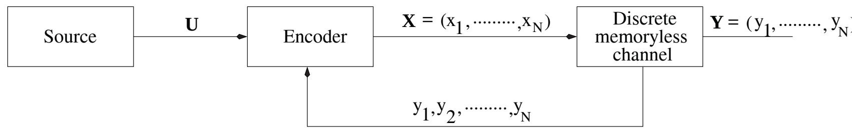


Figure: Channel with feedback

Each channel output as received is fed back into the encoder and can affect the selection of subsequent channel inputs. Show that

$$I(\mathbf{U}; Y^N) \leq \sum_{n=1}^{N} I(X_n, Y_n) \leq NC$$

The capacity of a DMC is not increased by the presence of feedback from the receiver to transmitter, then why use feedback?

Problem #2 (contd.)

- **Solution:**

$$\begin{aligned}
& I(U_1, \cdots, U_k; Y_1, \cdots, Y_N) \\
= \; & H(Y_1 \cdots Y_N) - \sum_{n=1}^{N} H(Y_n | Y_1 \cdots Y_{n-1} U_1 \cdots U_k) \\
= \; & H(Y_1 \cdots Y_N) - \sum_{n=1}^{N} H(Y_n | Y_1 \cdots Y_{n-1} U_1 \cdots U_k X_1 \cdots X_n) \\
= \; & H(Y_1 \cdots Y_N) - \sum_{n=1}^{N} H(Y_n | Y_1 \cdots Y_{n-1} X_1 \cdots X_n) \\
= \; & \sum_{n=1}^{N} (H(Y_n | Y_1 \cdots Y_{n-1}) - H(Y_n | Y_1 \cdots Y_{n-1} X_1 \cdots X_n)) \\
= \; & I(X_1 \cdots X_N; Y_1 \cdots Y_N)
\end{aligned}$$

Problem #2 (contd.)

- **Solution:**

$$
\begin{aligned}
& I(U_1, \cdots, U_k; Y_1, \cdots, Y_N) \\
= \; & I(X_1 \cdots X_N; Y_1 \cdots Y_N) \\
= \; & \sum_{n=1}^{N} I(X_1 \cdots X_n; Y_n | Y_1 \cdots Y_{n-1}) \\
= \; & \sum_{n=1}^{N} H(Y_n | Y_1 \cdots Y_{n-1}) - H(Y_n | X_1 \cdots X_n Y_1 \cdots Y_{n-1}) \\
= \; & \sum_{n=1}^{N} (H(Y_n | Y_1 \cdots Y_{n-1}) - H(Y_n | X_n)) \\
\leq \; & \sum_{n=1}^{N} (H(Y_n) - H(Y_n | X_n)) \\
= \; & \sum_{n=1}^{N} I(X_n; Y_n) \leq \sum_{n=1}^{N} C = NC
\end{aligned}
$$

Feedback reduces the complexity of the encoder and decoder.

Channel Capacity

- **Problem #3:** Arjun is a meteorologist with KTV station. His record in Kanpur city is given in the table below, the numbers indicating the relative frequency of the indicated event.

	Actual	
Prediction	Rain	No Rain
Rain	1/8	3/16
No Rain	1/16	10/16

Table 1: The weatherman Arjun's predictions.

Amrita notices that the weatherman, Arjun is right only 12/16 of the time, but could be right 13/16 of the time by always predicting no rain. She explains this situation and applies for the weatherman's job, but the weatherman's supervisor Rakesh who is an information theorist, turns her down. Why?

Problem #3 (contd.)

- **Solution:** Let's define a random variable X representing the actual weather such that X = 0 if there is no rain and X = 1 if it rains.

Problem #3 (contd.)

- **Solution:** Let's define a random variable X representing the actual weather such that X = 0 if there is no rain and X = 1 if it rains.
- Similarly, define a random variable Y to be the weatherman's (or the student's) prediction, with Y = 0 denoting predicting no rain and Y = 1 when rain is predicted.

Problem #3 (contd.)

- **Solution:** Let's define a random variable X representing the actual weather such that X = 0 if there is no rain and X = 1 if it rains.
- Similarly, define a random variable Y to be the weatherman's (or the student's) prediction, with Y = 0 denoting predicting no rain and Y = 1 when rain is predicted.
- If we view the weatherman as a channel that conveys information about some underlying event (rain or no rain in this case), then we can describe it by a set of four transition probabilities: $p(Y=0|X=0)$, $p(Y=1|X=0)$, $p(Y=0|X=1)$, $p(Y=1|X=1)$ and the two prior probabilities $p(X=0)$ and $p(X=1)$.

Problem #3 (contd.)

- **Solution:** Let's define a random variable X representing the actual weather such that X = 0 if there is no rain and X = 1 if it rains.
- Similarly, define a random variable Y to be the weatherman's (or the student's) prediction, with Y = 0 denoting predicting no rain and Y = 1 when rain is predicted.
- If we view the weatherman as a channel that conveys information about some underlying event (rain or no rain in this case), then we can describe it by a set of four transition probabilities: $p(Y=0|X=0)$, $p(Y=1|X=0)$, $p(Y=0|X=1)$, $p(Y=1|X=1)$ and the two prior probabilities $p(X=0)$ and $p(X=1)$.
- We have

$$p(X=0) = \sum_{y=0}^{1} p(X=0,y) = \frac{3}{16} + \frac{10}{16} = \frac{13}{16}$$

$$p(X=1) = 1 - p(X=0) = 1 - \frac{13}{16} = \frac{3}{16}$$

Problem #3 (contd.)

- The transition probabilities for the Arjun's channel is given by

$$
\begin{aligned}
p(Y=0/X=0) &= \frac{p(X=0,Y=0)}{p(X=0)} = \frac{10/16}{13/16} = \frac{10}{13} \\
p(Y=1/X=0) &= \frac{p(X=0,Y=1)}{p(X=0)} = \frac{3/16}{13/16} = \frac{3}{13} \\
p(Y=0/X=1) &= \frac{p(X=1,Y=0)}{p(X=1)} = \frac{1/16}{3/16} = \frac{1}{3} \\
p(Y=1/X=1) &= \frac{p(X=1,Y=1)}{p(X=1)} = \frac{1/8}{3/16} = \frac{2}{3}
\end{aligned}
$$

Problem #3 (contd.)

- For Arjun's channel,

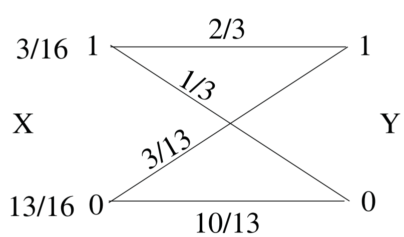


Figure: Arjun's Channel

$$
\begin{aligned}
p(Y=0) &= \frac{3}{16}\cdot\frac{1}{3} + \frac{13}{16}\cdot\frac{10}{13} = \frac{11}{16} \\
p(Y=1) &= 1 - P(Y=0) = \frac{5}{16}
\end{aligned}
$$

$$
I(X;Y) = \sum_{x=0}^{1}\sum_{y=0}^{1} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} = 0.090636022
$$

Problem #3 (contd.)

- The transition probabilities for the Amrita's channel is given by

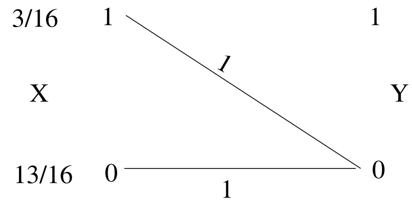


Figure: Amrita's Channel

$$
\begin{aligned}
p(Y = 0/X = 0) &= 1 \\
p(Y = 1/X = 0) &= 0 \\
p(Y = 0/X = 1) &= 1 \\
p(Y = 1/X = 0) &= 0
\end{aligned}
$$

Problem #3 (contd.)

- The transition probabilities for the Amrita's channel is given by

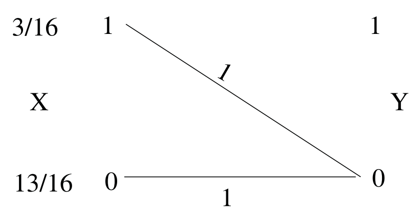


Figure: Amrita's Channel

$$
\begin{aligned}
p(Y = 0/X = 0) &= 1 \\
p(Y = 1/X = 0) &= 0 \\
p(Y = 0/X = 1) &= 1 \\
p(Y = 1/X = 0) &= 0
\end{aligned}
$$

- For Amrita's channel, $p(Y = 0) = 1$ and $p(Y = 1) = 0$.

Problem #3 (contd.)

- Mutual information is given by

$$I(X;Y) = \sum_{x=0}^{1}\sum_{y=0}^{1} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} = 0$$

Problem #3 (contd.)

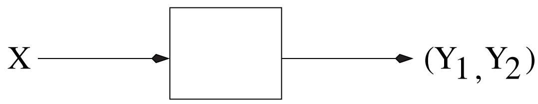
- Mutual information is given by

$$I(X;Y) = \sum_{x=0}^{1}\sum_{y=0}^{1} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} = 0$$

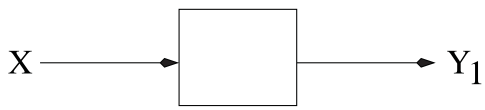
- Thus the Arjun's channel is better than the Amrita's channel.

Channel Capacity

- **Problem #4:** Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X, *i.e.* $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$. Show that $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1, Y_2)$. Conclude that the capacity of the channel



is less than twice the capacity of the channel



Problem #4 (contd.)

- **Solutions:**

$$\begin{aligned} I(X; Y_1 Y_2) &= H(Y_1 Y_2) - H(Y_1 Y_2/X) \\ &= H(Y_1) + H(Y_2) - I(Y_1; Y_2) - H(Y_1/X) - H(Y_2/X) \\ &= I(X; Y_1) + I(X; Y_2) - I(Y_1; Y_2) \\ &= 2I(X; Y_1) - I(Y_1; Y_2) \end{aligned}$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #4 (contd.)

- **Solutions:**

$$\begin{aligned} I(X;Y_1Y_2) &= H(Y_1Y_2) - H(Y_1Y_2/X) \\ &= H(Y_1) + H(Y_2) - I(Y_1;Y_2) - H(Y_1/X) - H(Y_2/X) \\ &= I(X;Y_1) + I(X;Y_2) - I(Y_1;Y_2) \\ &= 2I(X;Y_1) - I(Y_1;Y_2) \end{aligned}$$

- Capacity of channel with single look at Y:

$$C_1 = \max_{p(x)} I(X;Y_1)$$

Problem #4 (contd.)

- Capacity of channel with two independent looks at Y:

$$\begin{aligned} C_2 &= \max_{p(x)} I(X;Y_1Y_2) \\ &= \max_{p(x)} 2I(X;Y_1) - I(Y_1;Y_2) \\ &\leq \max_{p(x)} 2I(X;Y_1) \\ &= 2C_1 \end{aligned}$$

Channel Capacity

- **Problem #5:** Consider a pair of parallel Gaussian channels, i.e.,

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

  where

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

  and there is a power constraint $E(X_1^2 + X_2^2) \leq 2P$. Assume that $\sigma_1^2 > \sigma_2^2$. At what power does the channel stop behaving like a single channel with noise variance σ_2^2 and begin behaving like a pair of channels?

Problem #5 (contd.)

- **Solutions:** We will put all the signal power into the channel with less noise until the total power of noise and signal in that channel equals the noise power in the other channel.

Problem #5 (contd.)

- **Solutions:** We will put all the signal power into the channel with less noise until the total power of noise and signal in that channel equals the noise power in the other channel.
- After that, we will split any additional power evenly between the two channels.

Problem #5 (contd.)

- **Solutions:** We will put all the signal power into the channel with less noise until the total power of noise and signal in that channel equals the noise power in the other channel.
- After that, we will split any additional power evenly between the two channels.
- Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e. when $2P = \sigma_1^2 - \sigma_2^2$.

Mutual Information

- **Problem #6:** Find the mutual information I(X;Y), where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

Evaluate I(X;Y) for $\rho = 1, \rho = 0,$ and $\rho = -1$ and comment.

Problem #6

- **Solution:**

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

Problem #6

- **Solution:**

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2 \left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right)$$

- Using the expression for the entropy of a multivariate normal derived in class, we get

$$h(X, Y) = \frac{1}{2} \log (2\pi e)^2 |K| = \frac{1}{2} \log (2\pi e)^2 \sigma^4 (1 - \rho^2).$$

Problem #6

- **Solution:**

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2 \left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right)$$

- Using the expression for the entropy of a multivariate normal derived in class, we get

$$h(X, Y) = \frac{1}{2} \log (2\pi e)^2 |K| = \frac{1}{2} \log (2\pi e)^2 \sigma^4 (1 - \rho^2).$$

- Since X and Y are individually normal with variance σ^2,

$$h(X) = h(Y) = \frac{1}{2} \log 2\pi e \sigma^2.$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #6

- **Solution:**

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(0, \quad \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

- Using the expression for the entropy of a multivariate normal derived in class, we get

$$h(X, Y) = \frac{1}{2} \log (2\pi e)^2 |K| = \frac{1}{2} \log (2\pi e)^2 \sigma^4 (1 - \rho^2).$$

- Since X and Y are individually normal with variance σ^2,

$$h(X) = h(Y) = \frac{1}{2} \log 2\pi e \sigma^2.$$

- Hence

$$I(X; Y) = h(X) + h(Y) - h(X, Y) = -\frac{1}{2} \log(1 - \rho^2).$$

Problem #6 (contd.)

- Mutual information for different values of ρ:

Problem #6 (contd.)

- Mutual information for different values of ρ:
  a) $\rho = 1$. In this case, X = Y, and knowing X implies perfect knowledge about Y. Hence the mutual information is infinite, which agrees with the formula.

Problem #6 (contd.)

- Mutual information for different values of ρ:
  a) $\rho = 1$. In this case, X = Y, and knowing X implies perfect knowledge about Y. Hence the mutual information is infinite, which agrees with the formula.
  b) $\rho = 0$. In this case, X and Y are independent, and hence I(X;Y)=0, which agrees with the formula.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Problem #6 (contd.)

- Mutual information for different values of ρ:
  a) $\rho = 1$. In this case, X = Y, and knowing X implies perfect knowledge about Y. Hence the mutual information is infinite, which agrees with the formula.
  b) $\rho = 0$. In this case, X and Y are independent, and hence I(X;Y)=0, which agrees with the formula.
  c) $\rho =$ -1. In this case, X = -Y, and again the mutual information is infinite as in the case when $\rho = 1$.