EE4-40
EE9-CS7-26
EE9-SO20

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2018

MSc and EEE/EIE PART IV: MEng and ACGI

Corrected copy

INFORMATION THEORY

Thursday, 10 May 10:00 am

Time allowed: 3;00 hours

There are FOUR questions on this paper.

Answer ALL questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : C. Ling
Second Marker(s) : D. Gunduz

Information for students

Notation:

(a) Random variables are shown in Tahoma font. x, $\mathbf{x}$, $\mathbf{X}$ denote a random scalar, vector and matrix respectively.

(b) The size of a set A is denoted by $|A|$.

(c) By default, the logarithm is to the base 2.

(d) $\oplus$ denotes the exclusive-or operation, or modulo-2 addition.

(e) "i.i.d." means "independent identically distributed".

(f) $H(\cdot)$ is the entropy function.

(g) $C(x)=\frac{1}{2}\log_2(1+x)$ is the capacity function for the Gaussian channel in bits/channel use.

1. a) i) $H(X) = -\sum P(x)\log P(x) = 1$. $H(Y) = H(X) = 1$

ii) $H(X|Y) = -\sum P(x,y)\log P(x|y) = -\frac{1}{4}\left[\log\frac{1}{2}\times 4\right] = 1$

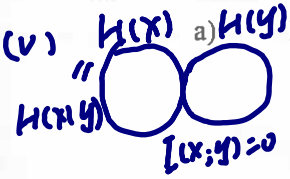
$H(Y|X) = H(X|Y) = 1$

The Questions

iii) $H(X,Y) = -\sum P(x,y)\log P(x,y) = -\frac{1}{4}\left[\log\frac{1}{4}\times 4\right] = 2$

iv) $I(X;Y) = H(X) - H(X|Y) = 0$

1. Basics of information theory.

v) 

a) Let the joint distribution of two random variables X and Y be given by

$p(X,Y)$	$Y=0$	$Y=1$	
$X=0$	1/4	1/4	$P(X=0)=\frac{1}{2}$
$X=1$	1/4	1/4	$P(X=1)=\frac{1}{2}$
	$P(Y=0)=\frac{1}{2}$	$P(Y=1)=\frac{1}{2}$	

Compute the following quantities:

i) $H(X)$, $H(Y)$

ii) $H(X|Y)$, $H(Y|X)$

iii) $H(X,Y)$

iv) $I(X;Y)$

v) Draw a Venn diagram for the above quantities.

[10]

b) Let X_i ($i = 1, 2$) be i.i.d. Bernoulli ($p = 1/4$) and $Y_1 = X_2$ and $Y_2 = X_1$. Calculate $I(X_i; Y_i)$ and $I(X_{1,2}; Y_{1,2})$.

[7]

c) Given two binary distributions $\mathbf{p} = \{\overset{p_1}{p}, \overset{p_2}{1-p}\}$ and $\mathbf{q} = \{\overset{q_1}{r}, \overset{q_2}{1-r}\}$. Calculate relative entropies $D(\mathbf{p}||\mathbf{q})$ and $D(\mathbf{q}||\mathbf{p})$ for $p = \frac{1}{4}$ and $r = \frac{1}{2}$.

[8]

b) $I(X_1;Y_1) = 0$. $I(X_2;Y_2) = 0$

$I(X_1,X_2;Y_1,Y_2) = \underbrace{I(X_1;Y_2)}_{H(X_1)-\underset{\overset{\shortparallel}{0}}{H(X_1|Y_2)}} + \underbrace{I(X_2;Y_1)}_{H(X_2)-\underset{\overset{\shortparallel}{0}}{H(X_2|Y_1)}} = H(X_1)+H(X_2) = 2H(p)$

c) $D(p\|q) = \sum_i p_i \log\frac{p_i}{q_i}$

$= p\log\frac{p}{r} + (1-p)\log\frac{1-p}{1-q}$

$D(q\|p) = r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p}$

2. a)

(i) 1 → decomposition of mutual info.

2 → when var known. Gaussian dist. max entropy [$\frac{1}{2}\log 2\pi e\sigma^2$]

3 → reducing condition increase entropy

2. Gaussian sources and channels.

4 → known var. Gaussian max entropy

a) Rate-distortion of Gaussian sources. Assume $X \sim N(0,\sigma^2)$ and $E(X-\hat{X})^2 \le D$.

5 → def

Justify each step of the following derivations.

i) Lower bound on mutual information.

6 → mutual info. (1) is nonnegative

$$I(X;\hat{X}) = h(X) - h(X|\hat{X}) = \tfrac{1}{2}\log 2\pi e\sigma^2 - h(X-\hat{X}|\hat{X})$$

$$\overset{(3)}{\ge} \tfrac{1}{2}\log 2\pi e\sigma^2 - h(X-\hat{X}) \overset{(4)}{\ge} \tfrac{1}{2}\log 2\pi e\sigma^2 - \tfrac{1}{2}\log\left(2\pi e\,\mathrm{Var}(X-\hat{X})\right)$$

$$\overset{(5)}{\ge} \tfrac{1}{2}\log 2\pi e\sigma^2 - \tfrac{1}{2}\log 2\pi e D$$

$$\overset{(6)}{\Rightarrow} I(X;\hat{X}) \ge \max\left(\tfrac{1}{2}\log\frac{\sigma^2}{D}, 0\right)$$

(ii)

$\hat{x} \sim N(0, \sigma^2 - D)$ → reconstruction

$z \sim (0, D)$ → noise

$x \sim (0, \sigma^2)$ → original signal

ii) Achievability. To show that we can find a distribution $p(\hat{x}, x)$ that achieves the lower bound, we construct a test channel that introduces distortion $D<\sigma^2$ shown in Fig. 2.1.

7 → $z = x - \hat{x}$

8 → noise z is indep. of $\hat{x}$

$z \sim (0, D)$. Gaussian max entropy given var.

9 → mutual info is nonnegative

10 → $R = I(X;\hat{X})$

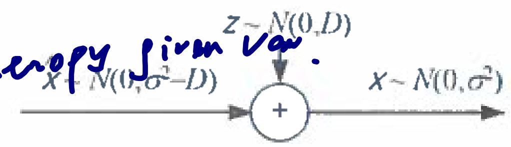


Fig. 2.1. Test channel.

$$I(X;\hat{X}) = h(X) - h(X|\hat{X}) = \tfrac{1}{2}\log 2\pi e\sigma^2 - h(X-\hat{X}|\hat{X})$$

$$\overset{(7)}{=} \tfrac{1}{2}\log 2\pi e\sigma^2 - h(Z|\hat{X}) \overset{(8)}{=} \tfrac{1}{2}\log\frac{\sigma^2}{D}$$

$$\overset{(9)}{\Rightarrow} I(X;\hat{X}) = \max\left(\tfrac{1}{2}\log\frac{\sigma^2}{D}, 0\right)$$

$$\overset{(10)}{\Rightarrow} D(R) = \frac{\sigma^2}{2^{2R}}$$

[10]

b)

$z_1 \sim (0, \sigma_1^2)$: $x_1 \to \oplus \to y_1$

$z_2 \sim (0, \sigma_2^2)$: $x_2 \to \oplus \to y_2$

$z_n \sim (0, \sigma_n^2)$: $x_n \to \oplus \to y_n$

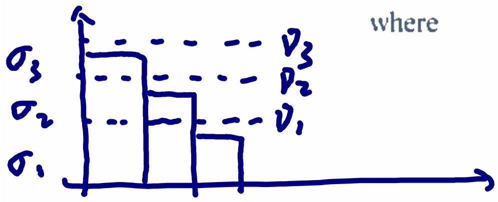
$\max I(x_1, \dots x_n; y_1, \dots y_n)$

s.t. $\sum_i P_i \le P$ (P_i → avg. power of channel i)

b) Waterfilling. Consider three parallel Gaussian channels, i.e.,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$$

$P_i = (\nu - \sigma_i^2)^+$



where

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}\right)$$

power: $\nu_1 - \sigma_1^2 > 0$: all to ch. 1

$\nu_1 - \sigma_1^2 = 3P \Rightarrow \nu_1 = 3P + \sigma_1^2$

$I(X_1, X_2, X_3; Y_1, Y_2, Y_3) = I(X_1; Y_1) = \frac{1}{2}\log\left(1 + \frac{\nu_1 - \sigma_1^2}{\sigma_1^2}\right)$ ($3P$)

iii) $\begin{cases} P_1 = \nu_2 - \sigma_1^2 \\ P_2 = \nu_2 - \sigma_2^2 \\ P_3 = 0 \end{cases}$ $2\nu_2 - \sigma_1^2 - \sigma_2^2 \le 3P \Rightarrow \nu_2 = \frac{3P + \sigma_1^2 + \sigma_2^2}{2}$

$C = I(X_1; Y_1) + I(X_2; Y_2)$

$= \frac{1}{2}\log(1 + \frac{\nu_2 - \sigma_1^2}{\sigma_1^2}) + \frac{1}{2}\log(1 + \frac{\nu_2 - \sigma_2^2}{\sigma_2^2})$

and there is a power constraint $E(X_1^2 + X_2^2 + X_3^2) \le 3P$. Assume that $\sigma_1^2 \ge \sigma_2^2 \ge \sigma_3^2$. At what power does the channel behaves like

i) a single channel with noise variance σ_3^2? Find the channel capacity in this case.

ii) a pair of channels with noise variances σ_3^2 and σ_2^2? Find the channel capacity in this case.

iii) three channels with noise variances σ_3^2, σ_2^2, and σ_1^2? Find the channel capacity in this case.

[10]

iii $\begin{cases} P_1 = \nu_3 - \sigma_1^2 \\ P_2 = \nu_3 - \sigma_2^2 \\ P_3 = \nu_3 - \sigma_3^2 \end{cases}$ $3\nu_3 - \sigma_1^2 - \sigma_2^2 - \sigma_3^2 \le 3P \Rightarrow \nu_3 = \frac{3P + \sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3}$

$C = I(X_1; Y_1) + I(X_2; Y_2) + I(X_3; Y_3)$

$= \frac{1}{2}\log(1 + \frac{\nu_3 - \sigma_1^2}{\sigma_1^2}) + \frac{1}{2}\log(1 + \frac{\nu_3 - \sigma_2^2}{\sigma_2^2}) + \frac{1}{2}\log(1 + \frac{\nu_3 - \sigma_3^2}{\sigma_3^2})$

c) Denote by $\phi(\mathbf{x})$ a multivariate Gaussian distribution with zero mean and covariance matrix $\mathbf{K}$. Given any continuous probability density $f(\mathbf{x})$ with the same mean and covariance matrix, prove that their relative entropy is given by

$$D(f||\phi) = h_\phi(\mathbf{x}) - h_f(\mathbf{x})$$

where $h_\phi(\mathbf{x})$ and $h_f(\mathbf{x})$ are their differential entropies, respectively.

[5]

C) multivariant Gaussian $X^n \sim N(\mu, k)$

$\frac{1}{\sqrt{2\pi|k|}} e^{-\frac{1}{2}(x-\mu)^T k^{-1}(x-\mu)}$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = E\begin{bmatrix} x_1x_1 & x_1x_2 \\ x_2x_1 & x_2x_2 \end{bmatrix}$

$\mu = E[X^n] \to (n \times 1)$ vector

$k = E[X^n X^{nT}] \to (n \times n)$ matrix

$\Phi(x) \to$ MVG with $\mu = 0$

$f(x) \to$ any distribution with same mean and variance matrix.

$D(f||\Phi) = \int f(x)\log\frac{f(x)}{\Phi(x)}dx$

$= \underbrace{\int f(x)\log f(x)dx}_{-H(f)} - \int f(x)\log\Phi(x)dx$

$\int f(x)\log(\frac{1}{\sqrt{2\pi|k|}}e^{-\frac{1}{2}x^Tk^{-1}x})dx$

$= \int f(x)[-\frac{1}{2}\log|2\pi k| - |\frac{1}{2}x^Tk^{-1}x|]dx$

$= -\frac{1}{2}\log|2\pi k| - \frac{1}{2}E[|x^Tk^{-1}x|]$

$= -\frac{1}{2}\log|2\pi k| - \underbrace{\frac{1}{2}E[|xx^Tk^{-1}|]}_{\frac{1}{2}\underbrace{E[|xx^T|]}_{k}k^{-1}} = -\frac{1}{2}\log|2\pi k| - \frac{1}{2}|I|$

diff. entropy of MVG

$= -H(f) - (\frac{1}{2}\log|2\pi k| + \frac{1}{2}|I|)$

$= -H(f) + \Phi(x)$

3. a) weakly symmetric ch.
1 → def. of entropy
3 → because ch. transition matrix is symmetric
4 → def
5 → def.
6 → max entropy
7 → def of capacity
2 → the first row of ch. transition matrix

3. Channel capacity.

a) The transition matrix **Q** of a *weakly symmetric channel* has the following properties: all columns of **Q** have the same sum $= |X||Y|^{-1}$, and all rows are permutations of each other. Justify each step of the following derivations.

$$H(y\,|\,x) \overset{(1)}{=} \sum_{x\in X} p(x)H(y\,|\,X=x) \overset{(2)}{=} H(\mathbf{Q}_{1,:})\sum_{x\in X} p(x) \overset{(3)}{=} H(\mathbf{Q}_{1,:})$$

where $\mathbf{Q}_{1,:}$ is the entropy of the first row of the **Q** matrix. Thus

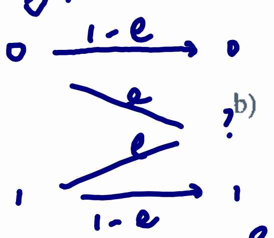
$$I(x;y) \overset{(4)}{=} H(y) - H(y\,|\,x) \overset{(5)}{=} H(y) - H(Q_{1,:}) \overset{(6)}{\leq} \log|Y| - H(Q_{1,:})$$

$$\Rightarrow C \overset{(7)}{=} \log|Y| - H(Q_{1,:})$$

[7]

symm. channel,
1. rows are permutations
2. sum of columns are the same.

b)



$$Q: \begin{array}{c|ccc} & 0 & 1 & e \\ \hline 0 & 1-e & 0 & e \\ 1 & 0 & 1-e & e \\ \hline & \downarrow & \downarrow & \downarrow \\ & 1-e & 1-e & 2e \end{array}$$

not a symmetric channel unless $e = \frac{1}{3}$

?b) In fact, the property that the uniform input distribution achieves the capacity of weakly symmetric channels can be extended to *generally symmetric channels* (to be proved in part c). In these channels, the set of outputs can be partitioned into subsets in a way, for each subset of the transition matrix: all rows are permutations of each other, and all columns are permutations of each other (similar to symmetric channels).

i) Show that the binary erasure channel is a generally symmetric channel, but is not a weakly symmetric channel.

ii) Which one of the following two channels is generally symmetric? Compute the capacity of the generally symmetric channel.

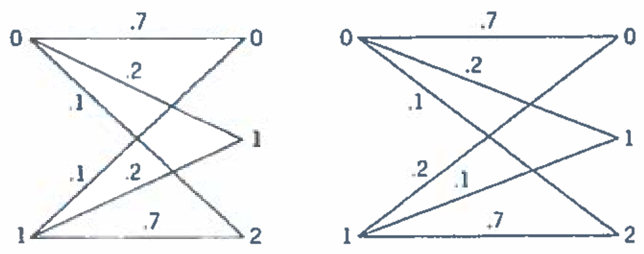


Fig. 3.1. Transition probabilities of two discrete memoryless channels (DMC).

ch. 1

$$Q: \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0.1 & 0.2 & 0.7 \\ \hline & \downarrow & \downarrow & \downarrow \\ & 0.8 & 0.4 & 0.8 \end{array}$$

[8]

ch. 2

$$Q: \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0.2 & 0.1 & 0.7 \\ \hline & \downarrow & \downarrow & \downarrow \\ & 0.9 & 0.3 & 0.8 \end{array}$$

c) Now, let us prove the uniform input distribution achieves the capacity of *generally symmetric channels*, in two steps.

i) Firstly, using the method of Lagrange multipliers, show that for any DMC, if there is an input distribution $Q(x) > 0$ for all x, such that

$$I(x;Y) = C \qquad \text{for all } x \tag{3.1}$$

where

$$I(x;Y) = \sum_y P(y|x)\log\frac{P(y|x)}{\sum_x Q(x)P(y|x)},$$

then C is the channel capacity.

ii) Show that Equation (3.1) holds for generally symmetric channels if the input distribution is uniform.

[10]

4. Network information theory.

a) Multi-access channel.

i) Describe the capacity region of a two-user multiple access Gaussian channel. Interpret the corner points (i.e., why can one of the users achieve the full capacity of a single-user channel as if the other user were absent?)

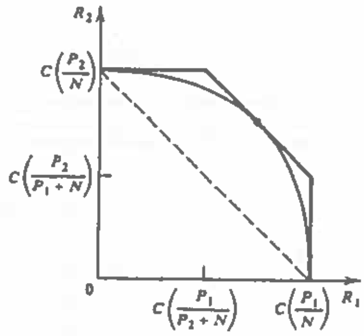


Fig. 4.1. Capacity region of multiple access channel.

ii) Verify the following equality for the corner point:

$$C\left(\frac{P_1}{N}\right)+C\left(\frac{P_2}{P_1+N}\right)=C\left(\frac{P_1+P_2}{N}\right)$$

where $C(x) = (\log(1+x))/2$ is the capacity function for the Gaussian channel.

iii) Now consider a multi-access channel of m users, each user with the same power P. Define the degrees of freedom (DoF) as

$$d = \lim_{P\to\infty} \frac{C\left(\frac{mP}{N}\right)}{C\left(\frac{P}{N}\right)}$$

Calculate the DoF for fixed m. Discuss the DoF per user as m increases.

[15]

b) Slepian-Wolf coding. Let (X, Y) have the joint probability mass function

$p(x,y)$	0	1
0	½	¼
1	0	¼

Calculate and sketch the Slepian-Wolf rate region for this source pair.

[10]