final

E4.40
SO20

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2010

MSc and EEE/ISE PART IV: MEng and ACGI

INFORMATION THEORY

Monday, 10 May 10:00 am

Time allowed: 3:00 hours

There are SIX questions on this paper.

Answer FOUR 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) : A. Manikas

Information for students

Notation:

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

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

(c) The normal distribution is denoted by $N(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

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

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

(f) Entropy function for a binary source $H(p)=-p\log_2 p-(1-p)\log_2(1-p)$; its derivative $H'(p)=\log_2(1-p)-\log_2 p$.

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

The Questions

1.

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

$p(X,Y)$	$Y=0$	$Y=1$	$Y=2$
$X=0$	0	1/8	1/8
$X=1$	1/8	1/8	0
$X=2$	1/4	0	1/4

Compute:

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

ii) The conditional entropy $H(X|Y)$, $H(Y|X)$

iii) The joint entropy $H(X,Y)$

iv) The mutual information $I(X;Y)$

v) Draw a Venn diagram for the above quantities.

[10]

b) Let X be a discrete random variable taking integer values. What is the general inequality relationship between $H(X)$ and $H(Y)$ if

i) $Y = e^X$

ii) $Y = \sin(X^2)$.

[8]

c) Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X, i.e., $H(f(X)) \le H(X)$ for any function f.

[7]

2.

a)

i) Explain what is meant by a uniquely decodable code.

ii) Explain what is meant by an instantaneous code.

iii) Suppose an instantaneous code for a source X has codeword lengths l_1, l_2, …, $l_{|X|}$. Use the code tree to justify the Kraft inequality $\sum_{i=1}^{|X|} 2^{-l_i} \le 1$ for any discrete sources. The following example shows such a tree for a four-symbol source and is only meant to be a hint.

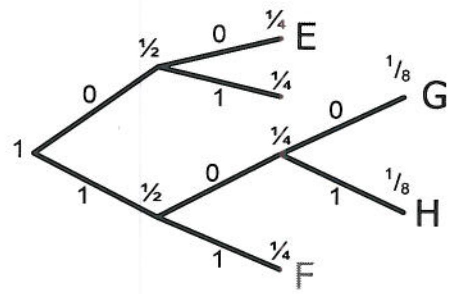


Fig. 2.1. The coding tree.

[10]

b) Lempel-Ziv coding. Give the LZ78 parsing and encoding of the following sequence:

$$1100101101010011010101$$

[Note: For this question, you will see less than 15 phrases; so ALWAYS use four bits to represent the location of a phrase. Do not worry about how to save such bits.]

[5]

c) If $X_i \in \{A, B\}$ are i.i.d. with probability mass vector $\{0.8, 0.2\}$. Using binary Huffman codes, determine the coding redundancy (i.e. the difference between the entropy and the number of bits used per symbol) when

i) Each X_i is encoded individually,

ii) Pairs of X_i are encoded together (i.e. X_1X_2 followed by X_3X_4 etc) and

iii) Triplets of X_i are encoded together.

[10]

3.

a) Justify each step of the following proof that mutual information is concave in the input distribution.

Proof: Consider two random variables U and V with probability mass vectors $\mathbf{p}_u$ and $\mathbf{p}_v$. Define a Bernoulli random variable Z with $p(Z=1) = \lambda$. Let $X = U$ if $Z=1$ and $X=V$ if $Z=0$. Thus,

$$\mathbf{p}_x \overset{(1)}{=} \lambda \mathbf{p}_u + (1-\lambda)\mathbf{p}_v .$$

Since $I(X,Z;Y) \overset{(2)}{=} I(X;Y) + I(Z;Y \mid X) \overset{(3)}{=} I(Z;Y) + I(X;Y \mid Z)$

and $I(Z;Y \mid X) \overset{(4)}{=} H(Y \mid X) - H(Y \mid X,Z) \overset{(5)}{=} 0,$

We have $I(X;Y) \overset{(6)}{\geq} I(X;Y \mid Z)$

$$\overset{(7)}{=} \lambda I(X;Y \mid Z=1) + (1-\lambda) I(X;Y \mid Z=0)$$

$$\overset{(8)}{=} \lambda I(U;Y) + (1-\lambda) I(V;Y)$$

$\overset{(9)}{\Rightarrow} I(X;Y)$ is concave in $\mathbf{p}_x$.

[9]

b) Calculate the capacity of the following channels with probability transition matrix

i)
$$\mathbf{Q} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix} \quad X, Y \in \{0,1,2\}$$

ii)
$$\mathbf{Q} = \begin{bmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-q & q \end{bmatrix} \quad X, Y \in \{0,1,2,3\}$$

[Hint: This channel is a union of two binary symmetric channels (BSC), whose capacity C satisfies $2^C = 2^{C_1} + 2^{C_2}$ where C_1 and C_2 are the capacity of the BSC's.]

[6]

c) Consider a cascade of n identical independent binary symmetric channels (BSC), each with raw error probability p. No encoding or decoding takes place at the intermediate terminals $X_0, X_1, \ldots, X_n$. Show that it is equivalent to a single BSC with error probability $\frac{1}{2}(1-(1-2p)^n)$, compute the channel capacity and comment on it limit as n tends to infinity.

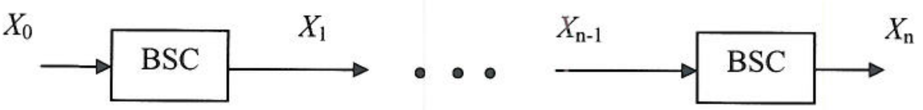


Fig. 3.1. The cascade channel.

[10]

4.

a) Given a correlation matrix $\mathbf{K}$ of zero-mean continuous-valued random process, it is well known that the Gaussian distribution

$$\varphi(\mathbf{x}) = |2\pi\mathbf{K}|^{-1/2} \exp\left(-1/2\, \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}\right)$$

has the maximum entropy. Justify the steps in the following proof.

$$\begin{aligned}
& 0 \overset{(1)}{\leq} D(f \,\|\, \varphi) \overset{(2)}{=} -h_f(\mathbf{X}) - E_f \log \varphi(\mathbf{X}) \\
\Rightarrow\ & h_f(\mathbf{X}) \overset{(3)}{\leq} -(\log e) E_f\left(-1/2 \ln\left(|2\pi\mathbf{K}|\right) - 1/2\, \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}\right) \\
& \overset{(4)}{=} 1/2 (\log e)\left(\ln\left(|2\pi\mathbf{K}|\right) + \operatorname{tr}\left(E_f \mathbf{x}\mathbf{x}^T \mathbf{K}^{-1}\right)\right) \\
& \overset{(5)}{=} 1/2 (\log e)\left(\ln\left(|2\pi\mathbf{K}|\right) + \operatorname{tr}(\mathbf{I})\right) \\
& \overset{(6)}{=} 1/2 \log\left(|2\pi e\mathbf{K}|\right) \overset{(7)}{=} h_\varphi(\mathbf{X})
\end{aligned}$$

[7]

b) Parallel channels and waterfilling. 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 $\sigma_1^2 \geq \sigma_2^2$. Compute the channel capacity.

[8]

c) Let X be a non-negative continuous random variable with mean constraint, i.e., $E[X] = m$. Show that the exponential distribution $f(x) = \lambda e^{-\lambda x}, x \geq 0, (m = 1/\lambda)$ has the maximum differential entropy $h(X) = -\int f(x) \log_2 f(x) dx = \log_2(e/\lambda)$.

[10]

5.

a) Consider the rate-distortion function $R(D) = \min I(X;\hat{X}),\ E_{X,\hat{X}} d(X,\hat{X}) \le D$. Justify each step in the following derivation of the rate-distortion function for a Gaussian source $X \sim N(0,\sigma^2)$ and $d(x,\hat{x}) = (x-\hat{x})^2$.

$$I(X;\hat{X}) \overset{(1)}{=} h(X) - h(X \mid \hat{X}) \overset{(2)}{=} ½\log 2\pi e\sigma^2 - h(X-\hat{X} \mid \hat{X})$$
$$\overset{(3)}{\ge} ½\log 2\pi e\sigma^2 - h(X-\hat{X}) \overset{(4)}{\ge} ½\log 2\pi e\sigma^2 - ½\log\left(2\pi e \operatorname{Var}\left(X-\hat{X}\right)\right)$$
$$\overset{(5)}{\ge} ½\log 2\pi e\sigma^2 - ½\log 2\pi e D$$
$$\Rightarrow R(D) \overset{(6)}{\ge} \max\left(½\log\frac{\sigma^2}{D}, 0\right)$$

[6]

b) Consider the inference channel in Fig. 5.1. There are two senders with equal power P, two receivers, with crosstalk coefficient a. The noise is Gaussian with zero mean and variance N. Show that the capacity under very strong interference (i.e., $a^2 \ge 1 + P/N$) is equal to the capacity under no interference at all.

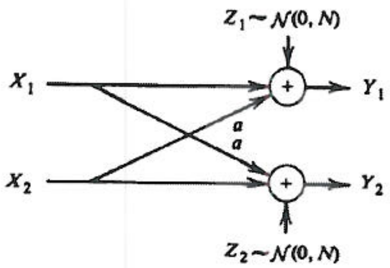


Fig. 5.1. Interference channel.

[9]

c) Consider a two-user multiple access Gaussian channel with reference to Fig. 5.2, where $C(x)$ is the capacity function, N is the noise power, and P_1 and P_2 are the powers of the two users.

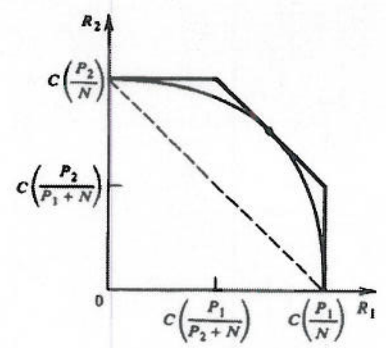


Fig. 5.2. Capacity region of multi-access channel.

The "quarter circle" in Fig. 5.2 is the capacity region of FDMA

$$R_1 \le W_1 C\left(\frac{P_1}{N_0 W_1}\right), \quad R_2 \le W_2 C\left(\frac{P_2}{N_0 W_2}\right)$$

where N_0 is the single-sided power spectral density of the noise. Show that the FDMA region is not larger than the capacity region of the multiple access channel (i.e., CDMA region) with the same bandwidth $W = W_1 + W_2$. (Hint: the function $x\log(1+1/x)$ is increasing and concave for $x > 0$.)

[10]

6. Consider discrete-valued random vectors $\mathbf{x}$ and $\mathbf{y}$ of length n where each pair (X_i, Y_i) is drawn i.i.d. from the joint probability distribution function $p_{XY}(x,y)$. The jointly typical set $J_\varepsilon^{(n)}$ is the set of vector pairs satisfying the following conditions:

$$J_\varepsilon^{(n)} = \Big\{\mathbf{x},\mathbf{y} : \left|-n^{-1}\log_2 p_X(\mathbf{x}) - H(\mathbf{x})\right| < \varepsilon, \quad \left|-n^{-1}\log_2 p_Y(\mathbf{y}) - H(\mathbf{y})\right| < \varepsilon, \quad \left|-n^{-1}\log_2 p_{XY}(\mathbf{x},\mathbf{y}) - H(\mathbf{x},\mathbf{y})\right| < \varepsilon\Big\}$$

where $p_X(x)$ and $p_Y(y)$ are the probability distribution functions of X_i and Y_i respectively. Since the sequences are i.i.d., the probability $p_X(\mathbf{x}) = \prod_{i=1}^{n} p_X(x_i)$ and $p_X(\mathbf{x})$ and $p_{XY}(\mathbf{x},\mathbf{y})$ can be written in a similar fashion.

a) Let X' and Y' be mutually independent random variables with probability distribution functions $p_{X'}(x') = p_X(x)$ and $p_Y(y) = p_{Y'}(y')$ respectively. Generate vectors $\mathbf{x}'$ and $\mathbf{y}'$ where each pair is also i.i.d. Show that the probability that $\mathbf{x}'$ and $\mathbf{y}'$ are jointly typical is small by justifying each step (1) to (7) in the following derivation:

$$p\left(\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}\right) \overset{(1)}{=} \sum_{\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}} p(\mathbf{x}',\mathbf{y}') \overset{(2)}{=} \sum_{\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}} p(\mathbf{x}')p(\mathbf{y}')$$

$$p\left(\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}\right) \overset{(3)}{\le} \left|J_\varepsilon^{(n)}\right| \max_{\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}} p(\mathbf{x}')p(\mathbf{y}') \overset{(4)}{\le} 2^{n(H(X,Y)+\varepsilon)} 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)} \overset{(5)}{=} 2^{-n(I(X;Y)-3\varepsilon)}$$

$$p\left(\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}\right) \overset{(6)}{\ge} \left|J_\varepsilon^{(n)}\right| \min_{\mathbf{x}',\mathbf{y}' \in J_\varepsilon^{(n)}} p(\mathbf{x}')p(\mathbf{y}') \overset{(7)}{\ge} (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)} 2^{-n(H(X)+\varepsilon)} 2^{-n(H(Y)+\varepsilon)} \overset{(8)}{\ge} (1-\varepsilon) 2^{-n(I(X;Y)+3\varepsilon)}$$

[10]

b) Suppose $n = 5$, $\varepsilon = 0$, and the joint distribution $p_{XY}(x,y)$ is given by

$p_{XY}(x,y)$	$y = 0$	$y = 1$
$x = 0$	1/5	1/5
$x = 1$	2/5	1/5

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are drawn i.i.d. from the above distribution.

i) Show that sequence $\mathbf{x}$ is in the typical set $A_\varepsilon^{(n)}(\mathbf{X}) = \{\mathbf{x} : \left|-n^{-1}\log_2 p_X(\mathbf{x}) - H(X)\right| < \varepsilon\}$ if and only if $\mathbf{x}$ contains three 1's.

Compute the probability $P(\mathbf{x} \in A_\varepsilon^{(n)}(\mathbf{X}))$ of the typical set.

ii) Of the 32 possible sequences $\mathbf{y}$, how many of them are in the typical set $A_\varepsilon^{(n)}(\mathbf{Y}) = \{\mathbf{y} : \left|-n^{-1}\log_2 p_Y(\mathbf{y}) - H(Y)\right| < \varepsilon\}$? Compute the probability $P(\mathbf{y} \in A_\varepsilon^{(n)}(\mathbf{Y}))$ of the typical set.

iii) Compute the probability $P(\mathbf{X},\mathbf{Y} \in J_\varepsilon^{(n)})$ of the jointly typical set.

[15]

Information Theory 2010 SOLUTIONS

B - bookwork
E - new example
A - new application

E40
So20

1.

a) Marginal distribution of X.

i) $P(X=0) = \frac{1}{4}$ $\quad P(X=1) = \frac{1}{4}$ $\quad P(X=2) = \frac{1}{2}$

$$H(X) = \frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{1}{2}\log 2 = \frac{3}{2}$$

Marginal distribution of Y:

$p(y=0) = \frac{3}{8}$ $\quad p(y=1) = \frac{1}{4}$ $\quad p(y=2) = \frac{3}{8}$ [2E]

$$H(Y) = \frac{3}{8}\log\frac{8}{3} + \frac{1}{4}\log 4 + \frac{3}{8}\log\frac{8}{3} = 1.56$$

ii) Conditional entropy $H(Y|X)$ = average row entropy

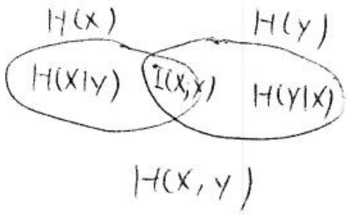
$$H(Y|X) = p(X=0)\,H(Y|X=0) + p(X=1)\,H(Y|X=1) + p(X=2)\,H(Y|X=2)$$
$$= \frac{1}{4}H(\tfrac{1}{2}) + \frac{1}{4}H(\tfrac{1}{2}) + \frac{1}{2}H(\tfrac{1}{2})$$
$$= 1$$ [2E]

$$H(X|Y) = p(y=0)\,H(\tfrac{1}{3}) + p(y=1)\,H(\tfrac{1}{2}) + p(y=2)\,H(\tfrac{1}{3})$$
$$= \frac{3}{8}\cdot 0.918 + \frac{1}{4}\cdot 1 + \frac{3}{8}\cdot 0.918$$
$$= 0.94$$

iii) $H(X,Y) = H(X) + H(Y|X) = \frac{3}{2} + 1 = \frac{5}{2} = 2.5$ [2E]

iv) $I(X;Y) = H(X) - H(X|Y) = \frac{3}{2} - 0.94 = 0.56$ [2E]

v)



[2B]

b) i) $H(X) = H(Y)$ [4E]

ii) $H(Y) \leq H(X)$ [4E]

c) $H(f(X), X) = H(X) + H(f(X)|X) = H(X) + 0 = H(X)$

$= H(f(X)) + H(X|f(X)) \geq H(f(X))$

[7A]

2. a) i) "uniquely decodable" means than the mapping between symbols and codewords is one-to-one, and there is no ambiguity in decoding [3B]

ii) "instantaneous" means that the codewords are instantaneously decodable, i.e., no codewords are a prefix of another codeword [3B]

iii) We plot a coding tree where each node has at most two branches. Codewords are the leaves, and the number of branches starting from the root is the length of the codeword. We assign probability 1 to the root, and a node on level l has probability 2^{-l}. Obviously, the probabilities of all leaves is never greater than 1. $\Rightarrow \sum_{i=1}^{|X|} 2^{-l_i} \le 1$ [4B]

b) Parsing: 1, 10, 0, 101, 1010, 100, 11, 01, 10101 [5E]

location: 1 2 3 4 5 6 7 8 9

encoding: $(\overset{0}{0000}, 1)$, $(\overset{1}{0001}, 0)$, $(\overset{0}{0000}, 0)$, $(\overset{2}{0010}, 1)$, $(\overset{4}{0100}, 0)$

$(\overset{2}{0010}, 0)$, $(\overset{1}{0001}, 1)$, $(\overset{3}{0011}, 1)$, $(\overset{5}{0101}, 1)$

c) $H(X) = -0.2\log 0.2 - 0.8\log 0.8 = 0.72$ [2B]

i) Redundancy $= 1 - 0.72 = 0.28$ [2E]

ii)

0	AA	$\cancel{x_1}$	0.64	0.64	0.64
11	AB	$\cancel{x_2}$	0.16	0.2	0.36
100	BA	$\cancel{x_3}$	0.16	0.16	
101	BB	$\cancel{x_4}$	0.04		

average length
$= 1\times 0.64 + 2\times 0.16 + 3\times 0.16 + 3\times 0.04$
$= 0.64 + 0.32 + 0.6$ [3E]
$= 1.56$

Redundancy $= \frac{1.56}{2} - 0.72 = 0.78 - 0.72 = 0.06$

iii)

0	AAA	0.512	0.512	0.512	0.512	0.512	0.512	0.512
100	AAB	0.128	0.128	0.128	0.128	0.232	0.256	0.488
101	ABA	0.128	0.128	0.128	0.128	0.128	0.232	
110	BAA	0.128	0.128	0.128	0.128	0.128		
11100	ABB	0.032	0.04	0.064	0.104			
11101	BAB	0.032	0.032	0.04				
11110	BBA	0.032	0.032					
11111	BBB	0.008						

[3E]

average length
$= 1\times 0.512 + 3\times 0.384 + 5\times 0.104$
$= 2.184$

2 Redundancy $= \frac{2.18}{3} - 0.72 = 0.728 - 0.72 = 0.008$

3. a) (1) total probability theorem [1 B] each step

(2) Chain rule (3) Chain rule

(4) definition (5) $H(Y|X,Z) = H(Y|X)$ if X is known, Z is known.

(6) $I(Z;Y|X) = H(Y|X) - H(Y|X,Z) = 0$

$\Rightarrow I(X;Y) = I(Z;Y) + I(X;Y|Z) \geq I(X;Y|Z)$

(7) total prob. theorem

(8) $Z=1 \Rightarrow X = U$, $Z=0 \Rightarrow X = V$

(9) definition of concavity

b) i) symmetric channel [3E]

$$C = \log|Y| - H(Q_{1,:}) = \log 3 - H(\tfrac{1}{2}) = 0.58$$

ii) $C_1 = 1 - H(p)$ $C_2 = 1 - H(q)$ [3 E]

$$2^C = 2^{C_1} + 2^{C_2}$$

$$C = \log_2 \left(2^{1-H(p)} + 2^{1-H(q)} \right)$$

c) There are many proofs. One is by induction: [2 A]

When $n=1$, $P_e = \frac{1}{2}(1-(1-2p)) = p$

[4 A]

Suppose it's correct when $n = k-1$.

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \begin{pmatrix} 1-\frac{1}{2}(1-(1-2p)^{k-1}) & \frac{1}{2}(1-(1-2p)^{k-1}) \\ \frac{1}{2}(1-(1-2p)^{k-1}) & 1-\frac{1}{2}(1-(1-2p)^{k-1}) \end{pmatrix} = \begin{pmatrix} 1 - \cdots & \frac{1}{2}(1-(1-2p)^{k}) \\ \frac{1}{2}(1-(1-2p)^{k}) & 1 - \cdots \end{pmatrix}$$

Done.

$$C = 1 - H\left(\tfrac{1}{2}(1-(1-2p)^n)\right)$$ [2 A]

As $n \to \infty$, $(1-2p)^n \to 0$ unless $p = 0$.

Therefore $C \to 1 - H(\frac{1}{2}) = 0$. [2 A]

4. a) (1) $D(f\|g)$ is nonnegative [1B] each step

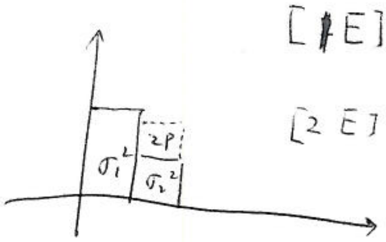
(2) definition (3) algebra

(4) $tr(AB) = tr(BA)$ (5) $E_f XX^T = K$

(6) $tr(I) = n$ (7) differential entropy of Gaussian

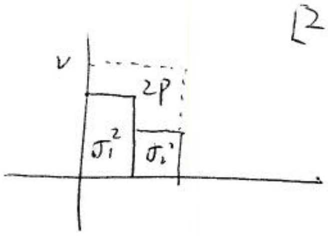
b) It is when $2P = \sigma_1^2 - \sigma_2^2$ that the channel starts to behave like two channels [1E]

When $2P < \sigma_1^2 - \sigma_2^2$, all water is filled into the second channel. [2E]



Thus [2E]

$$C = \frac{1}{2}\log\left(1 + \frac{2P}{\sigma_2^2}\right)$$



When $2P \geq \sigma_1^2 - \sigma_2^2$, two channels are filled with water [3E]

$$2P = v - \sigma_1^2 + v - \sigma_2^2 \Rightarrow v = \frac{2P + \sigma_1^2 + \sigma_2^2}{2} = P + \frac{\sigma_1^2 + \sigma_2^2}{2}$$

$$P_1 = v - \sigma_1^2 = P - \frac{\sigma_1^2 - \sigma_2^2}{2}, \quad P_2 = P + \frac{\sigma_1^2 \ \sigma_2^2}{2}$$

$$C = \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right)$$

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

c) If X has exponetial distribution, then

$$m = E[X] = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda}$$ [2A]

Differential entropy $h(X) = -\int_0^\infty f(x)\log_2 f(x)\, dx = -\log_2 e \int_0^\infty f(x) \ln f(x)\, dx$

$= -\log e \int_0^\infty \lambda e^{-\lambda x}(\ln\lambda - \lambda x)\, dx = -\log e\left(\ln\lambda - \int_0^\infty \lambda x e^{-\lambda x} d\lambda x\right) = \log(e/\lambda)$ bits

Apply relative entropy, for any distribution g

$$0 \leq D(g\|f) = -h_g(X) - E_g \log f(X)$$ [2A]

$$\Rightarrow h_g(X) \leq -\log e\ E_g(\ln\lambda - \lambda X)$$ [3A]

$$= -\log e\,(\ln\lambda - \lambda E_g(X)) = -\log e\,(\ln\lambda - 1)$$

$$= \log(e/\lambda) = h_f(X)$$ [3A]

5. a) (1) definition [1B] each

(2) differential entropy of Gaussian, translation invariance

(3) conditioning reduces entropy

(4) Gaussian has maximum entropy

(5) entropy of Gaussian

(6) R is not negative

b) Without interference: $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ [2B]

With very strong interference: Y_1 firstly decodes X_2, (treating X_1 as noise) subtracts it out, then decodes X_1. [3B]

Y_1 can decode X_2 up to rate $\frac{1}{2}\log\left(1+\frac{a^2P}{P+N}\right)$

Capacity is not changed as long as [4B]

$$\frac{1}{2}\log\left(1+\frac{a^2P}{P+N}\right) \geq \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

$$\Rightarrow \quad \frac{a^2P}{P+N} \geq \frac{P}{N} \quad \Rightarrow a^2 \geq \frac{P+N}{N}$$

c) We need to show that for FDMA

$$\left.\begin{aligned} R_1 &\leq C\left(\frac{P_1}{N}\right) \\ R_2 &\leq C\left(\frac{P_2}{N}\right) \\ R_1+R_2 &\leq C\left(\frac{P_1+P_2}{N}\right) \end{aligned}\right\} \text{CDMA region} \quad \text{noise power } N = N_0W$$

[2B]

$$\frac{W_1}{2}\log\left(1+\frac{P_1}{N_0W_1}\right) \leq \frac{W}{2}\log\left(1+\frac{P_1}{N_0W}\right) \quad (1)$$

$$\frac{W_2}{2}\log\left(1+\frac{P_2}{N_0W_2}\right) \leq \frac{W}{2}\log\left(1+\frac{P_2}{N_0W}\right) \quad (2)$$

[2A]

$$\frac{W_1}{2}\log\left(1+\frac{P_1}{N_0W_1}\right)+\frac{W_2}{2}\log\left(1+\frac{P_2}{N_0W_2}\right) \leq \frac{W}{2}\log\left(1+\frac{P_1+P_2}{N_0W}\right) \quad (3)$$

(1), (2) are true because $x\log(1+\frac{1}{x})$ is increasing and $W \geq W_1$, $W \geq W_2$ $(W = W_1+W_2)$ [1A]

(3) follows from concavity.

Since

$$\frac{W}{P_1+P_2} = \frac{P_1}{P_1+P_2}\,\frac{W_1}{P_1} + \frac{P_2}{P_1+P_2}\,\frac{W_2}{P_2}, \qquad [1A]$$

One has

$$\frac{W}{2}\log\left(1+\frac{P_1+P_2}{N_0 W}\right)$$

$$= (P_1+P_2)\,\frac{W}{2(P_1+P_2)}\log\left(1+\frac{P_1+P_2}{N_0 W}\right) \qquad [2A]$$

$$\geq (P_1+P_2)\left[\frac{P_1}{P_1+P_2}\cdot\frac{W_1}{2P_1}\log\left(1+\frac{P_1}{N_0 W_1}\right) + \frac{P_2}{P_1+P_2}\cdot\frac{W_2}{2P_2}\log\left(1+\frac{P_2}{N_0 W_2}\right)\right]$$

$$= \frac{W_1}{2}\log\left(1+\frac{P_1}{N_0 W_1}\right) + \frac{W_2}{2}\log\left(1+\frac{P_2}{N_0 W_2}\right). \qquad [2A]$$

6. a) (1) definition [1B]

(2) independence [1B]

(3) $p(x')p(y') \le \max p(x')p(y')$ [1B]

(4) $|J_\varepsilon^{(n)}| \le 2^{n(H(X,Y)+\varepsilon)}$, typicality [2B] (5) algebra [1B]

(6) $p(x',y') \ge \min p(x',y')$ [1B] (7) typicality [2B] (8) algebra [1B]

b) $\varepsilon = 0 \Rightarrow$ typical sequences are those containing correct portions of 0's and 1's

i) $p(x=0) = \frac{2}{5}$ $p(x=1) = \frac{3}{5}$ $P(x \in A_\varepsilon^{(n)}(X)) = C_5^3 \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^3 = 0.3456$

$\Rightarrow$ x is typical if it contains three 1's. [5A]

ii) $p(y=0) = \frac{3}{5}$ $p(y=1) = \frac{2}{5}$ $P(y \in A_\varepsilon^{(n)}(X)) = C_5^2 \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^2 = 0.3456$

$\Rightarrow$ y is typical if it contains two 1's

There are $C_5^2 = 10$ such sequences. [5A]

iii) $P(X, Y \in J_\varepsilon^{(n)}) = P(X, Y \in J_\varepsilon^{(n)} \mid X \in A_\varepsilon^{(n)}(X))\, P(X \in A_\varepsilon^{(n)}(X))$

$$= C_2^1 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot C_3^1 \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} \cdot C_5^3 \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^3 = 0.0768$$

1 $y_i = 1$ out of 2 1's when $x_i = 0$

1 $y_i = 1$ out of 3 1's when $X_i = 1$

[5A]