Final Exam Solution - Moed B

(1) True or False

(a.) True

$$H(X,Z) = H(X) + H(Z|X)$$
$$\geq H(X) + H(Z|X,Y) \quad \text{conditioning reduces entropy}$$
$$\geq H(X|Y) + H(Z|X,Y) \quad \text{conditioning reduces entropy}$$
$$= H(X|Y) + H(Z|Y) \quad X-Y-Z$$

(b.) False

Take $g(x)$ to be a bijection function.

Then $H(X-g(X)) = 0$ where $H(X) = H_b(\alpha)$.

(c.) False

(1) $I(X;Z|Y) = H(X|Y) - H(X|Y,Z) \overset{X-Y-Z}{=} H(X|Y) - H(X|Y) = 0$

(2) $H(X,Z|Y) - H(Z|Y) = H(X|Y) + H(Z|Y,X) - H(Z|Y)$
$$\overset{X-Y-Z}{=} H(X|Y) + H(Z|Y) - H(Z|Y)$$
$$= H(X|Y)$$

$$H(X|Y) \geq 0 \Rightarrow H(X,Z|Y) - H(Z|Y) \geq I(X;Z|Y)$$

(d.) True

Since g is non-increasing we have $g' \leq 0$.
Since g is convex we have $g'' \geq 0$.
Since f is concave we have $f'' \leq 0$.
So

$$[g(f)]' = g'(f)\, f'$$

$$[g(f)]'' = [g'(f) f']' = \underbrace{g''(f)(f')^2}_{\geq 0} + \underbrace{g'(f) f''}_{\geq 0} \geq 0$$

($g''(f)(f')^2 \geq 0$ since $g'' \geq 0$ and $(f')^2 \geq 0$; $g'(f) f'' \geq 0$ since $g' \leq 0$ and $f'' \leq 0$.)

(2) Additive Gaussian noise

(a.) We can look at this channel as a regular Gaussian point-to-point channel $Y = X + V$
where $V = Z + N$, $V \sim \mathcal{N}(0, G_1^2 + G_2^2)$
The capacity for this channel was given in class:

$$C = \frac{1}{2} \log\left(1 + \frac{P}{G_1^2 + G_2^2}\right)$$

(b.) Let us define

$$\Sigma_{XZ} = \begin{pmatrix} P_X & G_{ZX} \\ G_{ZX} & P_Z \end{pmatrix}$$

For this setting the capacity is given by

$$C = \max_{f(x,z)} I(X, Z; Y)$$

$$
\begin{aligned}
I(X,Z;Y) &= h(Y) - h(Y \mid X,Z) \\
&= h(X+Z+N) - h(X+Z+N \mid X,Z) \\
&= h(X+Z+N) - h(N) \\
&= h(X+Z+N) - \tfrac{1}{2}\log(2\pi e G_N^2) \\
&\le \tfrac{1}{2}\log(2\pi e G_Y^2) - \tfrac{1}{2}\log(2\pi e G_N^2) \\
&= \tfrac{1}{2}\log\left(2\pi e \left(\underbrace{G_X^2}_{P} + 2\underbrace{E[XZ]}_{G_{XZ}} + \underbrace{G_Z^2}_{P} + G_N^2\right)\right) - \tfrac{1}{2}\log(2\pi e G_N^2) \\
&= \tfrac{1}{2}\log\left(1 + \frac{P_X + P_Z + 2G_{XZ}}{G_N^2}\right)
\end{aligned}
$$

In order to maximize $I(X,Z;Y)$ we would like to maximize $\frac{1}{2}\log\left(1 + \frac{P_X+P_Z+2G_{XZ}}{G_N^2}\right)$. Therefore we need to maximize $G_{XZ} = E[XZ]$.

By Cauchy-Swartz inequality we have

$$E[XZ] \le \sqrt{E[X^2]\,E[Z^2]}$$

where equality is achieved when $Z = \alpha X$ so

$$E[XZ] = \sqrt{E[X^2]E[Z^2]} = \sqrt{P_X P_Z}$$

$$\alpha E[X^2] = \sqrt{P_Z P_X} \implies \alpha P_X = \sqrt{P_Z P_X} \implies \alpha = \sqrt{\frac{P_Z}{P_X}}$$

So we take $\boxed{Z = \sqrt{\frac{P_Z}{P_X}}\,X}$ and $\boxed{\Sigma_{XZ} = \begin{pmatrix} P_X & \sqrt{P_X P_Z} \\ \sqrt{P_X P_Z} & P_Z \end{pmatrix}}$

giving us that $I(X,Z;Y)$ is upper bounded by

$$\boxed{C = \frac{1}{2}\log\left(1 + \frac{P_X + P_Z + 2\sqrt{P_X P_Z}}{G_N^2}\right)}$$

hence it is indeed maximizable and achievable, and is the capacity for this setting.

(3) Differential entropy upper bound where the covariance matrix is fixed.

Lets denote the covariance matrix for X, Y by

$$\Sigma_{XY} = \begin{pmatrix} E[X^2] & E[XY] \\ E[XY] & E[Y^2] \end{pmatrix} = \begin{pmatrix} P_X & G_{XY} \\ G_{XY} & P_Y \end{pmatrix}$$

(a) Upper bound $h(X)$.

We saw that $h(X)$ is upper bounded by

$$h(X) \leq \frac{1}{2}\log\left(2\pi e\, E[X^2]\right)$$

$$= \frac{1}{2}\log\left(2\pi e\, (\Sigma_{XY})_{1,1}\right)$$

$$= \frac{1}{2}\log\left(2\pi e\, P_X\right)$$

(b) Upper bound $h(X,Y)$.

We saw in class that $h(X,Y)$ with $X,Y \sim \Sigma_{XY}$ is upper bounded by

$$h(X,Y) \leq \frac{1}{2}\log\left((2\pi e)^n \det(\Sigma_{XY})\right)$$

(c) Upper bound $h(X|Y)$

$$h(X|Y) \leq h(X^G | Y^G)$$

$$= h\left(X^G - \hat{X}^G_{l.in}(Y^G) \,|\, Y^G\right)$$

where $\hat{X}_{l.in}(Y)$ is the optimal linear estimator of X in the MMSE sense given the measurment Y

$$= h\left(X^G - \hat{X}^G_{l.in}(Y^G)\right)$$

for Gaussian r.v. the linear estimator is the optimal estimator there for it is independent of any function of

$$= h\left(X - \frac{6_{xy}}{P_y} Y\right)$$

$$= \frac{1}{2} \log\left[2\pi e \; E\left[\left(X - \frac{6_{xy}}{P_y} Y\right)^2\right]\right]$$

$$= \frac{1}{2} \log\left[2\pi e \left(P_x + \frac{6_{xy}^2}{P_y} - \frac{2 6_{xy}^2}{P_y}\right)\right]$$

$$= \frac{1}{2} \log\left[2\pi e \left(P_x - \frac{6_{xy}^2}{P_y}\right)\right]$$

(4.) Drawing a codebook

(a.) We have a sequence $x^n \in A_\varepsilon^{(n)}$ and we want to calculate the probability of the following event:

By the definition of the typical set, the probability of every $x^n \in A_\varepsilon^{(n)}$ to be drawn is bounded by:

$$2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$$

Note that the question regards to the probability of a specific sequence $x^n \in A_\varepsilon^{(n)}$ and not the probability of the whole set (which is almost 1)

(b) Our codebook has 2^{nR} codewords. Remember that every time a codeword is drawn, with probability 1 we get a sequence from the typical set. Moreover the probability of each of the sequences in $A_\varepsilon^{(n)}$ is uniform.
And finally: $(1-\varepsilon)\, 2^{n(H(X)-\varepsilon)} < |A_\varepsilon^{(n)}(X)| < 2^{n(H(X)+\varepsilon)}$

If we take $R < H(X) - \varepsilon$, i.e. the size of the codebook is less that $2^{nH(X)-\varepsilon}$ and we a get that the codebook exponentially smaller than the typical set since:

$$\frac{2^{nR}}{2^{n(H(X)-\varepsilon)}} \quad \frac{2^{n(H(X)-\varepsilon-\delta)}}{2^{n(H(X)-\varepsilon)}} = 2^{-n\delta} \xrightarrow{n\to\infty} 0$$

Therefore there exponentially more sequences in $A_\varepsilon^{(n)}(X)$ than we use for our codebook and the probability of repetition goes to 0 as $n \to \infty$.

So $\boxed{\bar{\mathcal{I}} = H(X) - \varepsilon}$

This is only a intuitive explanation, so consider:

We want to find an $\bar{\mathcal{I}}$ such that if $R < \bar{\mathcal{I}}$

$$P\left(\exists\, i \neq 1 \,:\, x^n(i) = x^n(1)\right)$$

$$= P\left(\bigcup_{i=2}^{2^{nR}} \left\{x^n(i) = x^n(1)\right\}\right)$$

$$\leq \sum_{i=2}^{2^{nR}} P\left(x^n(i) = x^n(1)\right)$$

Union bound ↗

$$\leq \sum_{i=2}^{2^{nR}} 2^{-n(H(X)-\varepsilon)}$$

all codeword are from $A_\varepsilon^{(n)}$ w.p 1 and this an upper bound on each prob. (from A)

$$\leq 2^{nR}\, 2^{-n(H(X)-\varepsilon)} = 2^{n(R-(H(X)-\varepsilon))}$$

If we want this probability to go to zero as $n \to \infty$ we need an $R < H(X) - \varepsilon$ so $\boxed{\bar{\mathcal{I}} = H(X) - \varepsilon}$

Another way to look at it is as a perfect channel where $Y_i = X_i$ for every $i \in \{1, \dots n\}$.
Since no error occur in transmission through this channel the only error possible is if two of the 2^{nR} codewords we draw are the same (since in this case the decoder won't be able to decode the message).
The capacity, as always, is given by:

$$I(X;Y) \underset{Y=X \text{ perfect channel}}{=} I(X;X) = H(X)$$

Transmitting in every rate which is less that the capacity assures as that $P_e^{(n)} \to 0$ as $n \to \infty$ so we need to take $\boxed{R < H(X) - \varepsilon = I}$

(Each of the three ways of solution were acceptable)

(c) This subsection was given as a bonus:

Intuitively, by drawing a codebook with rate $R > H(X) + \varepsilon$ we get that the codebook is exponentially larger that the typical set since: $\frac{2^{n(H(X)+\varepsilon+\delta)}}{2^{n(H(X)+\varepsilon)}} = 2^{n\delta} \xrightarrow{n\to\infty} \infty$

So the number of codewords in our codebook is so large that with probability which coverges to 1 as $n \to \infty$ we will have a repetition for the first codeword. (and any other as well) So $\boxed{I = H(X) + \varepsilon}$

Mathematically: we have $x^n(1)$ fixed.
First let as look at the prob. that all the codewords differ from $x^n(1)$

$$P(A) = \mathbb{P}\left[\forall i \neq 1 : x^n(i) \neq x^n(1)\right] = \prod_{i=2}^{2^{nR}} \mathbb{P}\left[x^n(i) \neq x^n(1)\right]$$

$$= \prod_{i=2}^{2^{nR}} \left(1 - \underbrace{\mathbb{P}\left[x^n(i) = x^n(1)\right]}_{\text{The prob. from (a)}}\right)$$

$$\leq \prod_{i=2}^{2^{nR}} \left(1 - 2^{-n(H(X)+\varepsilon)}\right)$$

$$\leq \left(1 - 2^{-n(H(X)+\varepsilon)}\right)^{2^{nR}}$$

adding $i=1$ to the product and did drawings

{ Now we use the inequality: $(1-x)^n \leq e^{-nx}$ }

$$\leq e^{-2^{nR} \cdot 2^{-n(H(X)+\varepsilon)}} = e^{-2^{n(R-H(X)-\varepsilon)}}$$

$\tilde{n} = 2^{nR}$

$\tilde{x} = 2^{-n(H(X)+\varepsilon)}$

Taking $R > H(X) + \varepsilon$ we have that $\mathbb{P}(A) \to 0$ as $n \to \infty$.

We want to calculate the prob. that there is another codeword similar to $x^n(1)$ which is in fact $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

So for $R > \underline{d} = H(X) + \varepsilon$ $\mathbb{P}(A) \to 0$ means that $\mathbb{P}(A^c) \xrightarrow[n \to \infty]{} 1$ as we wanted.

$$\boxed{\underline{d} = H(X) + \varepsilon}$$

(5) Network Coding With State

Assume a network $(\mathcal{N}, \mathcal{E})$ for which the capacity of the links changes over time i according to some binary state Z_i. For $Z=0$ the capacities are $C^{(0)}$ and for $Z=1$ the capacities are $C^{(1)}$.
The network is used once each time. $\mathcal{N}$ and $\mathcal{E}$ does not change over time.
$Z \overset{iid}{\sim} P(z)$ and all $v \in \mathcal{N}$ know Z^n non causally.
Z^n is independent of the source sequence.
The code may depend on the states up to time i, ie Z^i and the capacity is defined as in class.

(a) Provide a definition for the code and the capacity for this problem.

Code: A $(2^{nR}, n)$ code for noiseless state dependent relay network $G(\mathcal{N}, \mathcal{E}, \{C_z\}_{z \in \mathcal{Z}})$ consists of:
1) A source message set $\{1, .., 2^{nR}\}$ and uniform dist. of message M.
2) A source encoder that for each time i assigns $m_{ij}^{(i)} \in \{1, .., 2^{nC_{ij}^{(z_i)}}\}$ to each message $m \in \{1, .., 2^{nR}\}$ for each edge $(i,j) \in \mathcal{E}$.
3) A set of $(N-2)$ encoders:
Encoder k, where $k \in \{2, .., N-1\}$ assigns at each time i, an index $m_{kl}^{(i)} \in \{1, .., 2^{nC_{kl}^{(z_i)}}\}$ to each possible set of received messages $\{m_{jk}\}_{(j,k) \in \mathcal{E}}$ for each $(k,l) \in \mathcal{E}$.
4) Destination decoder that assign a message $\hat{m} \in \{1, .., 2^{nR}\}$ as a function of the received messages $\{m_{jN} : (j,N) \in \mathcal{E}\}$.

The definition of the prob. of error and an achievable rate stay as they were for the standard relay network.

Capacity: The capacity of $\mathcal{G}(\mathcal{N}, \mathcal{E}, \{C^{(z)}\}_{z\in\mathcal{Z}})$ is the supremum over the set of all achievable rates.

(b) Denote by $Cap(z)$ the capacity of the relay network $\mathcal{G}(\mathcal{N}, \mathcal{E}, C^{(z)})$, where $z \in \mathcal{Z} = \{0, 1\}$.

For the upper bound on the capacity of the network $\mathcal{G}(\mathcal{N}, \mathcal{E}, \{C^{(z)}\}_{z\in\mathcal{Z}})$ first consider:

$$Cap(0) = \min_{\{S\subset\mathcal{N}: 1\in S, N\in S^c\}} C^{(0)}(S)$$

$$Cap(1) = \min_{\{S\in\mathcal{N}: 1\in S, N\in S^c\}} C^{(1)}(S)$$

where $C^{(z)}(S) = \sum_{\{(j,k)\in\mathcal{E}: j\in S, k\in S^c\}} C_{jk}^{(z)}$ for $z \in \mathcal{Z}$.

The upper bound we want is:

$$\boxed{\sum_{z\in\mathcal{Z}} p(z)\, Cap(z) = p(z=0)\, Cap(0) + p(z=1)\, Cap(1)}$$

Proof: $nR = H(M)$ ($M \sim Unif\{1, \dots, 2^{nR}\}$)

$= H(M|Z)$ ($M \perp\!\!\!\perp Z$)

$= I(M; \hat{M}|Z) + H(M|\hat{M}, Z)$

$\leq H(\hat{M}|Z) - H(\hat{M}|M, Z) + n\varepsilon_n$

Fano: $H(M|\hat{M}) \leq n\varepsilon_n$, $\lim_{n\to\infty}\varepsilon_n = 0$ and $H(M|\hat{M}, Z) \leq H(M, \hat{M}) \leq n\varepsilon_n$

$= H(\hat{M}|Z) + n\varepsilon_n$

$\leq H(T_1^{(Z)}, T_2^{(Z)}, \dots, T_K^{(Z)}|Z) + n\varepsilon_n$

Markov chain as in class

$\leq \sum_{i=1}^{K} H(T_i^{(Z)}|Z) + n\varepsilon_n$

Conditioning reduces entropy

$$= \sum_{i=1}^{n} \sum_{z \in \mathcal{Z}} p(z) H(T_i^{(z)} | Z=z) + n\varepsilon_n$$

$$\leq \sum_{j \in S, k \in S^c} \sum_{z \in \mathcal{Z}} p(z) C_{jk}^{(z)} + n\varepsilon_n$$

Since it true for every possible cut $(S : S^c)$ it must hold for the minimal cut, i.e we get

$$nR \leq n \sum_{z \in \mathcal{Z}} p(z) Cap(z) + n\varepsilon_n$$

Taking n to infinity and using the fact that $\mathcal{Z} = \{0, 1\}$ we have

$$\boxed{R \leq p(z=0) Cap(0) + p(z=1) Cap(1)}$$ □

(c) Achievable scheme

In class we saw how to achieve $R^{(i)} = \min_{\{S \subset N : 1 \in S, N \in S^c\}} C^{(i)}(S)$, $i = 1, 2$

So here, we take the codes that achieve $R^{(0)}$ and $R^{(1)}$, and use them as follows:

All the nodes know the current state at time i, Z_i. For each time i, each node will use the code for $R^{(0)}$ if $Z_i = 0$, or the code $R^{(1)}$ if $Z_i = 1$.

Because no delay exists in the network, we can be sure that by using the code $R^{(z_i)}$ we transmit at the best rate posible for the channel use at time i.

Since the rates $R^{(0)}$ and $R^{(1)}$ achieve $Cap(0)$ and $Cap(1)$ respetivly, and we trasmit at $R^{(0)}$ $p(z=0)$ of the time, and at $R^{(1)}$ $p(z=1)$ of the time, we achieve the upper bound in (b)

(d.) for this case we define

$$\widetilde{Cap}(0) = \min_{i \in D} \min_{\{SCN, IES, iES^c\}} C^{(0)}(S)$$

$$\widetilde{Cap}(1) = \min_{i \in D} \min_{\{SCN, IES, iES^c\}} C^{(1)}(S)$$

where

$$C^{(z)}(S) = \sum_{\{(i,k) \in \mathcal{E},\ j \in S,\ k \in S^c\}} C_{jk}^{(z)} \quad \text{for} \quad z \in \mathcal{Z}$$

So we have

$$C = \sum_{z \in \mathcal{Z}} p(z)\, \widetilde{Cap}(z) = p(z=0)\, \widetilde{Cap}(0) + p(z=1)\, \widetilde{Cap}(1)$$

(e.) The answer won't change since we only need each node to know the current state Z_i, they don't use all the sequence Z^n.