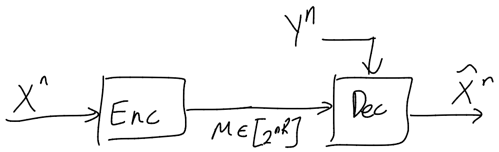
Semih Yagli

12/15/2016

Thursday

ELE538-Information Theoretic Security

Homework-4

1. 

Assume that the blocklength n, the encoder f and the decoder g satisfy the average distortion constraint $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(X_i, \hat{X}_i)] \leq D$ when operated at the rate R. Let M denote the number of the recovery points, then

$$
\begin{aligned}
nR &\geq H(M) \\
&\geq H(M|Y^n) && (\text{Conditioning Reduces Entropy}) \\
&= I(X^n; M|Y^n) && (H(M|X^n, Y^n)=0,\ M \text{ is determined from } (X^n, Y^n)) \\
&= \sum_{i=1}^{n} I(X_i; M|Y^n, X^{i-1}) && (\text{Chain rule}) \\
&= \sum_{i=1}^{n} \left(I(X_i; X^{i-1}, Y^{i-1}, Y_{i+1}^n | Y_i) + I(X_i; M|Y^n, X^{i-1}) \right) && (X_i - Y_i - (X^{i-1}, Y^{i-1}, Y_{i+1}^n)) \\
&= \sum_{i=1}^{n} I(X_i; M, Y^{i-1}, Y_{i+1}^n, X^{i-1}|Y_i) && \text{Let } U_i = (M, Y^{i-1}, Y_{i+1}^n) \\
&= \sum_{i=1}^{n} I(X_i; U_i, X^{i-1}|Y_i) \\
&\geq \sum_{i=1}^{n} I(X_i; U_i|Y_i) && (I(X_i; X^{i-1}|Y_i, U_i) \geq 0) \\
&= \sum_{i=1}^{n} I(X_i; U_i) - I(Y_i; U_i) && (U_i - X_i - Y_i) \\
&= n\left(I(X_T; U_T|T) - I(Y_T; U_T|T)\right) && \text{Let } T \perp\!\!\!\perp X^n, Y^n, U^n \\
&\geq n\left(I(X_T; U_T) - I(Y_T; U_T)\right) && I(X_T; U_T) \leq I(X_T; U_T|T) \\
& && I(Y_T; U_T) = I(Y_T; U_T|T) \\
&\geq n R\left(\mathbb{E}[d(X_T, \hat{X}_T)]\right) \\
&= n R\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(X_i; \hat{X}_i)]\right) \\
&\geq n R(D).
\end{aligned}
$$

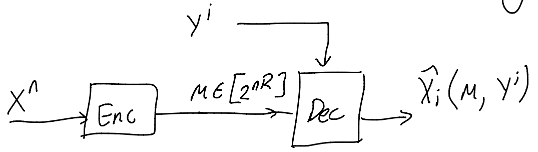
It follows that any achievable rate R must satisfy $R \geq R(D)$.

Note that $U_i = (M, Y^{i-1}, Y_{i+1}^n)$ depends on the side information Y^n in a non-causal fashion. Moreover, since $\hat{X}_i$ is a function of (M, Y^n), it follows that restricting $\hat{X}_i$ to be a function of $U_i = (M, Y^{i-1}, Y_{i+1}^n)$ and Y_i does not hurt us.

(Note: I called $R(D)$ as the rate distortion function and showed any achievable rate R must be greater than $R(D)$, the question is stated in a way that $R(D) \geq$ ~~ must be the end result. I hope this does not cause any discrepancy.)

→ In the solution of this problem, the book "Network Information Theory" by El Gamal and Kim is used as a reference.

2. Let us first describe our setting in a block-diagram. For each $i \in \{1, 2, \ldots, n\}$,



Rate distortion function with <u>causal</u> side information available at the decoder (denoted by $R(D)$ throughout this solution) is the infimum of rates R s.t. there exists a sequence of $(2^{nR}, n)$ codes with $\mathbb{E}[d(X^n, \hat{X}^n)] \leq D$.

We shall now prove the following theorem:

Theorem: Let (X, Y) denote two correlated discrete memoryless sources and let $d(x, \hat{x})$ be a distortion measure. The rate distortion function for X with side information Y causally available at the decoder satisfies the following equality

$$R(D) = \min_{\substack{P_{U|X} \\ P_{\hat{X}|UY} : \, \mathbb{E}[d(X,\hat{X})] \leq D}} I(X; U)$$

Proof of Theorem:

(Achievability)

We use strongly joint typicality encoding to prove the achievability.

<u>Codebook:</u> Fix the conditional pmf $(u|x)$ and function $\hat{x}(u, y)$ that attain $R\left(\frac{D}{1+\epsilon}\right)$ where D denotes the allowed distortion. For each $m \in \{1, 2, \ldots, 2^{nR}\}$, generate independent and random codewords $u^n(m)$ according to $\prod_{i=1}^{n} P_U(u_i)$

Encoding: Given a source sequence x^n, find an index m such that $u^n(m)$ such that x^n and $u^n(m)$ are strongly jointly typical, i.e., $(u^n(n), x^n) \in \mathcal{T}_{\varepsilon'}^{(n)}$. If there are more than one $m \in [2^{nR}]$ s.t. $(u^n(m), x^n) \in \mathcal{T}_{\varepsilon'}^{(n)}$, choose the smallest index. If there is no index, set $m=1$.

Encoder outputs m.

Decoding: Decoder outputs the reconstruction sequence $\hat{x}^n(m, y^n)$ by setting $\hat{x}_i = \hat{x}(u_i(m), y_i)$ for each $i \in \{1, 2, \dots, n\}$.

(Recall: ε' is used in joint typical encoding.)

Expected Distortion: Denote the chosen index by M and let $\varepsilon > \varepsilon'$. Note that error occurs when $(U^n(M), X^n, Y^n)$ are not jointly typical.

Let $\mathcal{E} = \{(U^n(M), X^n, Y^n) \notin \mathcal{T}_{\varepsilon}^{(n)}\}$ denote the error event. Note that $\mathcal{E}$ denotes the decoding error, we may also consider the encoding error $\mathcal{E}_0 = \{(U^n(M), X^n) \notin \mathcal{T}_{\varepsilon'}^{(n)} \ \forall m \in [2^{nR}]\}$

In that case, we have

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}[\mathcal{E}_0 \cap \mathcal{E}] + \mathbb{P}[\mathcal{E}_0^c \cap \mathcal{E}] \leq \mathbb{P}[\mathcal{E}_0] + \mathbb{P}[\mathcal{E}_0^c \cap \mathcal{E}]$$

We know that $\mathbb{P}[\mathcal{E}_0] = \mathbb{P}[(U^n(M), X^n) \notin \mathcal{T}_{\varepsilon'}^{(n)}] \to 0$ as long as $R > I(X; U) + 3\varepsilon$ (This is done when we proved the achievability of plain lossy source coding theorem. There, we had $\hat{X}(M)$ instead of $U^n(M)$). Now consider $\mathbb{P}[\mathcal{E}_0^c \cap \mathcal{E}]$ term.

Since $\varepsilon > \varepsilon'$, $(U^n(M), X^n) \in \mathcal{T}_{\varepsilon'}^{(n)}$ and conditioned on $U^n(M) = u^n$, $X^n = x^n$

Y^n is distributed as $\prod_{i=1}^{n} P_{Y|UX}(y_i | u_i, x_i) = \prod_{i=1}^{n} P_{Y|X}(y_i | x_i)$

That is, we have $U - X - Y$. And by a property of conditional joint typicality (in particular, item 7c* in class notes) we have $\mathbb{P}[\mathcal{E}_0^c \cap \mathcal{E}] \longrightarrow 0$ as well. Thus, the asymptotic expected distortion when the expectation is taken over codebooks can be upper as follows (We assume distortion function satisfies $\max_{x,\hat{x}} d(x,\hat{x}) < \infty$.)

$$\mathbb{E}[d(X^n; \hat{X}^n)] = \mathbb{P}[\mathcal{E}].\ \mathbb{E}[d(X^n, \hat{X}^n) | \text{error}] + \mathbb{P}[\mathcal{E}^c]\ \mathbb{E}[d(X^n, \hat{X}^n) | \text{no error}]$$

$$\leq \mathbb{P}[\mathcal{E}].\ \max_{x,\hat{x}} d(x,\hat{x}) + \mathbb{P}[\mathcal{E}^c]\ (1+\varepsilon)\ \mathbb{E}[d(X,\hat{X})]$$

Due to convexity of $d(x^n; \hat{x}^n)$ and property 4 of Joint typicality

$$\leq \mathbb{P}[\mathcal{E}].\ \max_{x,\hat{x}} d(x,\hat{x}) + \mathbb{P}[\mathcal{E}^c].\ D$$

Taking $n \to \infty$, we see that $\mathbb{P}[\mathcal{E}] \to 0$ and $\mathbb{P}[\mathcal{E}^c] \to 1$ if $R > I(X;U) + 3\varepsilon' = R(\frac{D}{1+\varepsilon}) + 3\varepsilon'$ taking $\varepsilon \to 0$ (and $\varepsilon' \to 0$ as $\varepsilon' < \varepsilon$ is needed) we see that

$$\min_{\substack{P_{U|X} \\ P_{\hat{X}|UY} : \mathbb{E}[d(X,\hat{X})] \leq D}} I(X;U)$$

is an achievable rate.

Converse proof is on the next page.

(Converse)

Assume that the blocklength n, the encoder f and the decoder g satisfy the average distortion constraint $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(X_i, \hat{X}_i)] \leq D$ when operated at the rate R. Let M denote the number of the recovery points. Motivated by the previous problem, since $\hat{X}_i$ is a function of M and Y^i, we set $U_i = (M, Y^{i-1})$. Note that $U_i - X_i - Y_i$ forms a Markov Chain and just like in previous problem $\hat{X}_i$ is a function of U_i and Y_i. We have,

$$
\begin{aligned}
nR \geq H(M) &= I(X^n; M) && (M \text{ is a function of } X^n) \\
&= \sum_{i=1}^{n} I(X_i; M \mid X^{i-1}) && (\text{chain rule}) \\
&= \sum_{i=1}^{n} I(X_i; M, X^{i-1}) && (X_i \perp\!\!\!\perp X^{i-1}) \\
&= \sum_{i=1}^{n} I(X_i; M, X^{i-1}) + I(X_i; Y^{i-1} \mid M, X^{i-1}) && (X_i - (M, X^{i-1}) - Y^{i-1}) \\
&= \sum_{i=1}^{n} I(X_i; M, Y^{i-1}, X^{i-1}) \\
&= \sum_{i=1}^{n} I(X_i; U_i, X^{i-1}) \\
&\geq \sum_{i=1}^{n} I(X_i; U_i) && (I(X_i; X^{i-1} \mid U_i) \geq 0) \\
&= n\, I(X_T; U_T \mid T) && (\text{where } T \perp\!\!\!\perp U^n, X^n \text{ and } Y^n) \\
&\geq n\, I(X_T; U_T) && (I(X_T; T) = 0 \text{ and } I(X_T; T \mid U_T) \geq 0) \\
&\geq n R(\mathbb{E}[d(X_T; \hat{X}_T)]) && (\text{by definition of rate distortion function}) \\
&= n R\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(X_i; \hat{X}_i)]\right) && \left(R(D) \text{ is decreasing in } D \text{ and } \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[d(X_i; \hat{X}_i)]\right) \\
&\geq n R(D)
\end{aligned}
$$

It follows that any achievable rate R must satisfy $R \geq R(D)$.

<u>Discussion:</u> Why do we lose efficiency provided by $-I(U;Y)$?

Recall,

Plain Lossy-source coding rate distortion function:

$$R(D) = \min_{P_{\hat{X}|X}:\, \mathbb{E}[d(X,\hat{X})] \leq D} I(X;\hat{X})$$

Lossy-source coding with non-causal side information at the decoder:

$$R(D) = \min_{\substack{P_{U|X} \\ P_{\hat{X}|Y,U}:\, \mathbb{E}[d(X;\hat{X})] \leq D}} I(X;U) - I(Y;U)$$

In this case there are two efficiencies thanks to the side information at the decoder. First one is that we now have an auxiliary random variable U and we minimize over $P_{U|X}$. Note that

$$\min_{\substack{P_{\hat{X}|X}: \\ \mathbb{E}[d(X;\hat{X})] \leq D}} I(X;\hat{X}) \overset{(*)}{\geq} \min_{\substack{P_{U|X} \\ P_{\hat{X}|U,Y}:\, \mathbb{E}[d(X;\hat{X})] \leq D}} I(X;U)$$

because we may choose $Y \neq \emptyset$ and $U = \hat{X}$ to recover the left side in $(*)$. The second efficiency is that we now have a $-I(U;Y)$ term in the rate distortion function. This is because decoder has some information about X_{j+1}^{n} and can adjust itself to recover some more points. Hence, given the same distortion constraint, in the non-causal version we can compress with lower rates.

Now, lets go back to the setting of problem 2

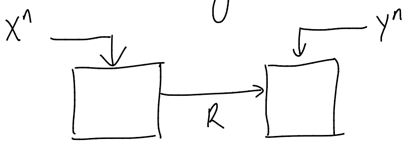
Lossy source coding with causal side information:

$$R(D) = \min_{\substack{P_{X|U} \\ P_{\hat{X}|Y,U}:\ \mathbb{E}[d(X,\hat{X})] \leq D}} I(X;U)$$

In this case, we are losing the extra $-I(U;Y)$ term that is provided by the non-causality. This is because all we have is some past information uncompressed string X^i. This helps us better understand what is compressed into M but does not help us anticipate what message is being compressed next, therefore we loose the efficiency provided by $-I(U;Y)$.

→ Hope this was enough of an argument to make a case.

3. (a) Block - Diagram of the setting:



We are trying to find the key capacity C_k in the one-way rate limited communication setting. In this setting
$$C_k = \max_{\substack{P_{U|X}: \\ I(X;U|Y) \leq R}} I(U;Y)$$

Note that $I(X;U|Y) = P_e \, I(X;U|Y=e)$ because

$I(X;U|Y=1) = 0$ since $Y=1 \Rightarrow X=1$ a.s.

$I(X;U|Y=0) = 0$ since $Y=0 \Rightarrow X=1$ a.s.

(Note that $I(X;U) \leq H(X) = 1$ and, when $P_e < R$, if $I(X;U) \leq 1$ $I(X;U|Y) \leq P_e < R$ and the constraint still satisfied)

Note also that $I(X;U|Y=e) = I(X;U)$

Hence, our constraint $I(X;U|Y) \leq R$ reduces to $I(X;U) \leq \min\left\{\frac{R}{P_e}, 1\right\}$

Now, we wish to maximize $I(U;Y)$ under the constraint $I(X;U) \leq \frac{R}{P_e}$

but note that since $U - X - Y$, we have

$$I(U;Y) + I(U;X|Y) = I(U;X)$$

$$\Rightarrow \quad I(U;Y) + P_e \, I(U;X) = I(U;X) \quad \Rightarrow \quad I(U;Y) = (1-P_e)\, I(U;X)$$
$$\leq \min\left\{\left(\frac{1-P_e}{P_e}\right)R,\ 1-P_e\right\}$$

Claim: $$\max_{\substack{P_{U|X}: \\ I(X;U|Y) \leq R}} I(U;Y) = \min\left\{\left(\frac{1-P_e}{P_e}\right)R,\ 1-P_e\right\}$$

Assuming $H(X) = 1$ bit

If $X \sim Ber(a)$

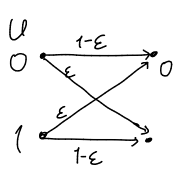
$$\max_{\substack{P_{U|X}: \\ I(X;U|Y) \leq R}} I(U;Y) = \min\left\{\left(\frac{1-P_e}{P_e}\right)R,\ (1-P_e)h(a)\right\}$$

Proof of claim: To simplify the algebra, let $X \sim Ber(1/2)$ o.w. the construction below remains identical. First, consider the case when $\frac{R}{P_e} \geq 1$. Then in that case, we pick $U = X$ and we get $I(U;X) = H(X) = 1$. In this case $I(U;Y) = 1 - P_e$ is achieved.

Therefore, we may assume $\frac{R}{P_e} < 1$. In this case, it suffices to show that $\exists U$ s.t. $I(U;X) = \frac{R}{P_e}$ can be achieved. Below is a construction method.

Given $P_X \sim Ber(1/2)$, we construct $P_{U|X}$ to achieve $I(U;X) = \frac{R}{P_e}$. To do so, we first construct reverse channel $P_{X|U}$ then we find $P_{U|X}$ using Bayes' Rule.

$$\underbrace{H(X)}_{\overset{1}{=}} - H(X|U) = \frac{R}{P_e} \Rightarrow H(X|U) = 1 - \frac{R}{P_e}. \text{ Consider the channel } P_{X|U} \overset{(\text{Let})}{=} BEC(\varepsilon)$$

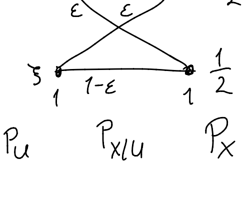


$$\rightarrow H(X|U) = h(\varepsilon) = 1 - \frac{R}{P_e} \rightsquigarrow \text{Thus pick } \varepsilon = h^{-1}\left(1 - \frac{R}{P_e}\right)$$

To proceed, let

$$P_{X|U}(x|u) = \begin{cases} 1-\varepsilon & (x,u) = (0,0) \text{ or } (x,u) = (1,1) \\ \varepsilon & (x,u) = (0,1) \text{ or } (x,u) = (1,0) \end{cases}$$

Now, Given $\varepsilon = h^{-1}(1 - \frac{R}{P_e})$ find $P_U(1) = \xi = 1 - P_U(0)$



$$\frac{1}{2} = (1-\varepsilon)\xi + \varepsilon(1-\xi) \Rightarrow \xi = 1/2$$

So $P_U(1) = \frac{1}{2} = P_U(0)$, $P_{X|U} \sim BSC(\varepsilon)$ with $\varepsilon = h^{-1}(1 - \frac{R}{P_e}) \Rightarrow I(U;X) = \frac{R}{P_e}$

By Bayes' rule $P_{U|X} = \dfrac{P_{X|U} \cdot P_U}{P_X}$. Hence, we need $P_{U|X} = P_{X|U}$, i.e

$P_{U|X}(0|0) = 1-\varepsilon$ $\quad$ $P_{U|X}(1|1) = 1-\varepsilon$

$P_{U|X}(0|1) = \varepsilon$ $\quad$ $P_{U|X}(1|0) = \varepsilon$

where $\varepsilon = h^{-1}\left(1 - \frac{R}{P_e}\right)$

In that case, $I(X;U) = \frac{R}{P_e}$.

So we can indeed achieve $I(U;Y) = \left(\frac{1-P_e}{P_e}\right) R$ by picking $P_{U|X}$ as above.

Hence, the key capacity is $C_k = \min\left\{\left(\frac{1-P_e}{P_e}\right) R,\ 1-P_e\right\}$

Note, if $X \sim Ber(a)$ similar construction yields the following result for key capacity:

$$C_k = \min\left\{\left(\frac{1-P_e}{P_e}\right) R,\ (1-P_e)\, h(a)\right\}$$

where $h(\cdot)$ denotes binary entropy function.

4. (a) Let X_1 be distributed according to Laplace law with mean μ and variance $2b^2$.

Assume that $X_2 = X_1 + c$. Then

$$f_{X_1}(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \quad \text{and} \quad f_{X_2}(x) = \frac{1}{2b} e^{-\frac{|x-\mu-c|}{b}}$$

Hence
$$\log \frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{|x-\mu-c|}{b} - \frac{|x-\mu|}{b}$$

since $|a|-|b| \leqslant |a-b| \leqslant |a|+|b|$ we have

$$|x-\mu| - |c| \leqslant |x-\mu-c| \leqslant |x-\mu| + |c|$$

$$\Rightarrow \quad -\frac{|c|}{b} \leqslant \log \frac{f_{X_1}(x)}{f_{X_2}(x)} \leqslant \frac{|c|}{b}$$

$\sqrt{2}\,b^{\cancel{2}} = \frac{\sqrt{\mathrm{Var}}}{\sqrt{2}}$

$$\Rightarrow \quad \left| \log \frac{f_{X_1}(x)}{f_{X_2}(x)} \right| \leqslant \frac{|c|}{b} = \frac{\sqrt{2}\,|c|}{\mathrm{Var}(X_1)}$$

4 (b) Let X_1 be distributed according to Gaussian law with mean μ and variance σ^2.

Assume that $X_2 = X_1 + c$. Then

$$f_{X_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{and} \quad f_{X_2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu-c)^2}{2\sigma^2}}$$

Hence,

$$\log \frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{(x-\mu-c)^2}{2\sigma^2} - \frac{(x-\mu)^2}{2\sigma^2} = \frac{c^2 - 2(x-\mu)c}{2\sigma^2}$$

and this quantity cannot be bounded uniformly for all x.