EPFL Winter Semester 2009/2010
Handout # 23, Thursday, 17 December, 2009

Solutions: Homework Set # 6

Problem 1 (Cascade Network)

(a) We know that the capacity of the channel is equal to

$$C = \max_{P_X} I(X;V).$$

From the problem setup we observe that we have the following Markov chain

$$X \leftrightarrow Y \leftrightarrow U \leftrightarrow V.$$

By data processing inequality we have $I(X;V) \leq I(X;Y)$ and $I(X;V) \leq I(U;V)$. Then we can proceed as follows. We have $I(X;V) \leq I(X;Y)$ so we can find the $\max_{P_X}$ of both side so we have

$$C = \max_{P_X} I(X;V) \leq \max_{P_X} I(X;Y) = C(p).$$

To show $C \leq C(q)$ is a little bit more tricky and it should be done in two step of maximization. Again we have $I(X;V) \leq I(U;V)$ so we can write

$$\begin{aligned} C = \max_{P_X} I(X;V) &\leq I(U;V)|_{\text{for some } P'_U \text{ dictated by choosing } P^*_X \text{ to be the maximizer of } I(X;V)} \\ &\leq \max_{P_U} I(U;V) = C(q). \end{aligned}$$

So finally we have

$$C \leq \min[C(p), C(q)].$$

Note that the above argument works for every two cascade channel, not only the binary symmetric channel.

(b) In this case, when there is no processing at the relay, $U = Y$, the overall channel from X to V can be regarded as a new binary symmetric channel with some new transition probability. To find the transition probability of the overall channel we process as follows

$$\begin{aligned} \mathbb{P}[V=0|X=0] &= \sum_{i\in\{0,1\}} \mathbb{P}[V=0, Y=i|X=0] \\ &\overset{(1)}{=} \sum_{i\in\{0,1\}} \mathbb{P}[V=0|X=0, Y=i] \cdot \mathbb{P}[Y=i|X=0] \\ &\overset{(2)}{=} \sum_{i\in\{0,1\}} \mathbb{P}[V=0|Y=i] \cdot \mathbb{P}[Y=i|X=0] \\ &= (1-p)(1-q) + pq \end{aligned}$$

where (1) follows from the chain rule for probability, (2) follows from the Markov chain we have. Then we can write

$$\begin{aligned}
\mathbb{P}[V=0|X=1] &= \sum_{i\in\{0,1\}} \mathbb{P}[V=0, Y=i|X=1] \\
&= \sum_{i\in\{0,1\}} \mathbb{P}[V=0|X=1, Y=i]\cdot\mathbb{P}[Y=i|X=1] \\
&= \sum_{i\in\{0,1\}} \mathbb{P}[V=0|Y=i]\cdot\mathbb{P}[Y=i|X=1] \\
&= p(1-q)+(1-p)q.
\end{aligned}$$

Similarly we can find $\mathbb{P}[V=1|X=0]$ and $\mathbb{P}[V=1|X=1]$ using the same method but the answer will be the same.

So the for the capacity in this case we have $C' = 1 - h_2(p(1-q)+(1-p)q)$.

(c) In this part we assume that relay can do some processing. The scheme that we suggest is as follows. Let $r = \min[C(p), C(q)]$. First source S used some channel code with rate r to encode its data and send it over the first channel. Then relay wait until receive the whole block of data and decode it. Because the source sends information at rate below the capacity of first channel ($r \le C(p)$) we can make the decoding error as small as possible. After decoding, the relay re-encode the information using some channel code with rate r (this time for the second channel) and send it to the destination D. Again because we have $r \le C(q)$ the destination can decode the data with a very small probability of error. So our scheme achieves the rate $r = \min[C(p), C(q)]$.

(d) We have

$$C = \min[1-h_2(p), 1-h_2(q)],$$

and

$$C' = 1 - h_2(p(1-q)+(1-p)q).$$

Without loss of generality let us assume that $0 \le p \le q \le 1/2$. Note that for a binary symmetric channel that has cross probability larger than $1/2$ we can change the role of 1 and 0 in its output so we can always assume that the cross probability of a binary symmetric channel is some parameter in the interval $[0, 1/2]$.

The we observe that the binary entropy function $h_2(x)$ is an increasing function for $x \in [0, 1/2]$. So using the above assumptions, for C we have

$$C = 1 - h_2(q).$$

Then we will show that $C' \le C$. To this end we have to show that $h_2(p(1-q)+(1-p)q) \ge h_2(q)$. Because the $h_2(\cdot)$ is a concave function and has the symmetry $h_2(x) = h_2(1-x)$ for $x \in [0,1]$ then we should prove that

$$q \le p(1-q)+(1-p)q \le 1-q.$$

For the left hand side inequality we can write

$$q \le p(1-q)+(1-p)q \quad \Longleftrightarrow \quad q \le q + \underbrace{p(1-2q)}_{\ge 0},$$

which is always true (under the above assumptions on p and q).

For the right hand side inequality we can write

$$p(1-q) + (1-p)q \leq 1 - q \quad \Longleftrightarrow \quad 0 \leq \underbrace{(1-2q)}_{\geq 0}\underbrace{(1-p)}_{\geq 0},$$

which is again always true (under the above assumptions on p and q). So we are done.

We should expect the result $C' \leq C$ before doing any calculation. When we let the relay R to do some processing one of its option is to do not anything and just forwards the bits. So obviously we can deduce that by performing some extra processing we may do better, so we conclude $C' \leq C$.

Problem 2 (Binary Multiplier Channel)

(a) The receiver observes both Y and Z. Based on the value of random variable Z it can decide whether the value of Y provides any information about X or not. If $Z = 1$ then $Y = X$ and if $Z = 0$ then the value of Y is zero independently of X. So by observing both Y and Z the equivalent channel $\mathcal{C} : X \longrightarrow YZ$ is an erasure channel with capacity

$$1 - \mathbb{P}[\text{Erasure}] = 1 - \mathbb{P}[Z = 0] = \alpha.$$

(b) Now we assume that the receiver does not have access to the random variable Z. In this part, the goal is to find the capacity of the channel $\mathcal{C}' : X \longrightarrow Y$.

To this end let us expand the mutual information

$$I(X;Y) = H(Y) - H(Y|X).$$

Then we have to find the expressions of $H(Y)$ and $H(Y|X)$ with respect to α and p.

To calculate $H(Y)$ we need to find probability $\mathbb{P}[Y = 1]$. We can write

$$\mathbb{P}[Y = 1] = \mathbb{P}[X = 1, Z = 1] = \mathbb{P}[X = 1] \cdot \mathbb{P}[Z = 1] = \alpha p,$$

so we have

$$H(Y) = h_2(\alpha p),$$

where $h_2(x) \triangleq -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

Then we have to find an expression for $H(Y|X)$. We can expand it as follows

$$\begin{aligned} H(Y|X) &= H(Y|X=0)\mathbb{P}[X=0] + H(Y|X=1)\mathbb{P}[X=1] \\ &= 0 \times \mathbb{P}[X=0] + H(Z)\mathbb{P}[X=1] \\ &= p \cdot h_2(\alpha). \end{aligned}$$

Putting everything together we have

$$I(X;Y) = h_2(\alpha p) - p \cdot h_2(\alpha),$$

and we want to find the maximum value of $I(X;Y)$ with respect to the input distribution which here is only described by the parameter p. So to find the maximum of this mutual information we find its derivative with respect to p.

First, note that for the derivative of $h_2(x)$ we have

$$h_2'(x) = \log_2 \frac{1-x}{x},$$

so we can write

$$\begin{aligned}\frac{\partial I(X;Y)}{\partial p} &= \frac{\partial h_2(\alpha p)}{\partial p} - h_2(\alpha) \\ &= \alpha \cdot \log_2 \frac{1-\alpha p}{\alpha p} - h_2(\alpha).\end{aligned}$$

To find the value of p which maximize $I(X;Y)$ we have to set $\frac{\partial I(X;Y)}{\partial p} = 0$ so we have

$$\alpha \cdot \log_2 \frac{1-\alpha p}{\alpha p} - h_2(\alpha) = 0,$$

or if $\alpha \neq 0$

$$\frac{1-\alpha p}{\alpha p} = 2^{h_2(\alpha)/\alpha}.$$

Then we can rearrange the above expression to obtain

$$p = \frac{1}{\alpha \left[1 + 2^{h_2(\alpha)/\alpha}\right]}.$$

Finally, for the capacity of the channel we have

$$C = h_2\left(\frac{1}{1+2^{h_2(\alpha)/\alpha}}\right) - \frac{1}{\left[1+2^{h_2(\alpha)/\alpha}\right]} \cdot \frac{h_2(\alpha)}{\alpha}.$$

For sanity check let us consider two limiting cases $\alpha \to 0$ and $\alpha \to 1$. For the first case Z is always zero and the output of the channel is independent from its input so we expect the channel has zero capacity in this case. For the second case Z is always one so we have a perfect binary channel that its capacity should be 1 bit. Now let us consider these two cases

$$\alpha \longrightarrow 1 \quad \Rightarrow \quad \frac{h_2(\alpha)}{\alpha} \longrightarrow 0 \quad \Rightarrow \quad C \longrightarrow h_2(1/2) = 1,$$

and

$$\alpha \longrightarrow 0 \quad \Rightarrow \quad \frac{h_2(\alpha)}{\alpha} \longrightarrow +\infty \quad \Rightarrow \quad C \longrightarrow h_2(0) = 0.$$

Problem 3 (Jointly Typical Sequences)

(a) We need to prove that

$$(x^n, y^n) \in A_\epsilon^{(n)}(X,Y) \Leftrightarrow x^n \in A_\epsilon^{(n)}(X), z^n \in A_\epsilon^{(n)}(Z),$$

where $y^n = x^n z^n$. First we calculate some entropies:

$$\begin{aligned}H(X) =&1 \\ H(Z) =&H(Y|X) = H_2(\alpha) \\ H(X,Y) =&H(X) + H(Y|X) = 1 + H(Y|X) = 1 + H_2(\alpha) \\ H(Y) =&-\alpha\log(\alpha) - (1-\alpha)\log(\frac{1-\alpha}{2}).\end{aligned}$$

Notice that $p(x^n) = (\frac{1}{2})^n$, $\forall x^n$, hence $-\frac{1}{n}\log p(x^n) = 1$ which implies $x^n \in A_\epsilon^{(n)}(X)$ for any x^n. On the other hand, $(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)$ implies that $-\frac{1}{n}\log p(x^n, y^n) \in (H(X,Y)-\epsilon, H(X,Y)-\epsilon)$. Since $p(x^n,y^n) = (\frac{1}{2})^n(1-p)^{n-k}p^k = p(x^n)p(z^n)$, where k is the number of places x and y differ, it follows that

$$\begin{aligned}
-\frac{1}{n}\log p(x^n,y^n) =& -\frac{1}{n}\log(\frac{1}{2})^n(1-p)^{n-k}p^k \\
=& -\frac{1}{n}\log p(x^n)p(z^n) \\
=& -\frac{1}{n}\log p(x^n) - \frac{1}{n}\log p(z^n) \\
=& 1 - \frac{1}{n}\log p(z^n) \in (H(X,Y)-\epsilon, H(X,Y)-\epsilon).
\end{aligned}$$

So, $-\frac{1}{n}\log p(z^n) \in (H(Z)-\epsilon, H(Z)-\epsilon)$, i.e. z^n is typical. Using the same equations in reverse direction, we see that assuming typical z^n implies that $-\frac{1}{n}\log p(x^n,y^n) \in (H(X,Y)-\epsilon, H(X,Y)-\epsilon)$.

It remains to show that y^n is typical. Notice that $y_i = x_i z_i$. Then

$$\begin{aligned}
p(Y_i = 0) =& p(X_i = 1, Z_i = 0) + p(X_i = -1, Z_i = 0) \\
=& p(X_i = 1)p(Z_i = 0) + p(X_i = -1)p(Z_i = 0) \\
=& \alpha.
\end{aligned}$$

$$\begin{aligned}
p(Y_i = 1) =& p(X_i = 1, Z_i = 1) \\
=& p(X_i = 1)p(Z_i = 1) \\
=& \frac{1}{2}(1-\alpha) = \frac{1}{2}(1-\alpha).
\end{aligned}$$

Hence, we should verify that for large enough n, $-\frac{1}{n}\log p(y^n) = H(Y)$, $\forall y^n$ i.e. for large enough n, every y^n is typical: Assume that $\mathcal{E}$ is the set of indices in y^n where 0 is observed.

$$\begin{aligned}
p(y^n) =& \prod_{i=1}^{n} p(y_i) \\
=& (\alpha)^{|\mathcal{E}|}(\frac{1}{2}(1-\alpha))^{n-|\mathcal{E}|}
\end{aligned}$$

Thus,

$$-\frac{1}{n}\log p(y^n) = -\frac{|\mathcal{E}|}{n}\log(\alpha) - \frac{n-|\mathcal{E}|}{n}\log(\frac{1}{2}(1-\alpha))$$

which converges in probability to

$$-\alpha\log(\alpha) - (1-\alpha)\log(\frac{1}{2}(1-\alpha)) = H(Y).$$

(b) By joint AEP We know that the probability that y^n is jointly typical with an independent $x^n(i)$ is roughly $2^{-nI(X;Y)}$.

(c)

$$\begin{aligned} Pr(\text{error}|x^n(1) \text{ was sent}) \leq & Pr((x^n, y^n) \notin A_\epsilon^{(n)}(X,Y) + Pr(A) \\ < & \epsilon + Pr(A), \end{aligned}$$

where A denotes the event that there exists a codeword $x^n(i),\ i \neq 1$ which is jointly typical with Y^n. Using the union bound and the fact that $x^n(i)$'s are *i.i.d*, we see that

$$\begin{aligned} Pr(A) = & Pr\left(E_2 \cup \ldots \cup E_{2^{nR}} | \, x^n(1) \text{ was sent}\right) \\ \leq & \sum_{i=1}^{2^{nR}} Pr(E_i|x^n(1)) \\ = & \sum_{i=1}^{2^{nR}} Pr(E_2|x^n(1)) \\ \leq & 2^{nR}2 \end{aligned}$$

Thus,

$$\begin{aligned} Pr(\text{error}|x^n(1) \text{ was sent}) \leq & Pr((x^n, y^n) \notin A_\epsilon^{(n)}(X,Y) + Pr(A) \\ < & \epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)}., \end{aligned}$$

which is smaller than 2ϵ when $R < I(X;Y)$ and n is sufficiently large. The probability of error is then

$$Pr(\text{error}) = \sum_{i=1}^{2nR} Pr(\text{error}|x^n(i) \text{ was sent})Pr(x^n(i) \text{ was sent}) = Pr(\text{error}|x^n(1) \text{ was sent}) < 2\epsilon.$$

Problem 4 (Deterministic Channel)

(a) Let us write

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &\overset{(1)}{=} H(Y) \\ &\leq \log_2 m, \end{aligned}$$

where (1) follows because $f(\cdot)$ is a deterministic function so we have $H(Y|X) = 0$; if we know X then we know Y completely. Furthermore, if $f(\cdot)$ is a surjective function, the all values $\{1, \ldots, m\}$ can be taken by Y. So by assigning an input distribution $P_X(x)$ such that $P_Y(y)$ become uniform we would obtain the maximum of $I(X;Y)$ which is $\log_2 m$ so we have

$$C = \max_{P_X} I(X;Y) = \log_2 m.$$

However, when $f(\cdot)$ is not surjective then with the same argument as stated above we have

$$C = \max_{P_X} I(X;Y) = \log_2 |\text{Im}(f)|,$$

where $\text{Im}(f)$ denotes the image of f

$$\text{Im}(f) \triangleq \{y \mid y = f(x), \quad \forall x \in \{1, \ldots, n\}\}.$$

(b) Here we have $Y = AX$ and again we can write

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) \\ &\overset{(1)}{\leq} \log_2 2^{\mathrm{rank}(A)} \\ &= \mathrm{rank}(A), \end{aligned}$$

where (1) is true because of the following reason. For each vector X, the vector Y, $Y = AX$, would be one of the elements of the column space of A. This space has $2^{\mathrm{rank}(A)}$ elements and thus $H(Y) \leq \log_2 2^{\mathrm{rank}(A)}$. By choosing the input distribution to be uniform, $P_Y(y)$ would be uniform over all elements of this column space and we achieve the maximum of $H(Y)$. So we have

$$C = \max_{P_X} I(X;Y) = \mathrm{rank}(A).$$

(c) Consider two channels $Y = PX$ with capacity $C_1 = \mathrm{rank}(P)$ and $Y = TX$ with capacity $C_2 = \mathrm{rank}(T)$ with capacity $C_2 = \mathrm{rank}(T)$.

Then consider the cascade channel $Y_2 = PY_1$ and $Y_1 = TX$. Thus we have $Y_2 = PTX$ so the capacity of the cascade channel is $C_{\mathrm{cascade}} = \mathrm{rank}(PT)$.

From problem 1, we know that

$$C_{\mathrm{cascade}} \leq \min[C_1, C_2].$$

Thus we have

$$\mathrm{rank}(PT) \leq \min[\mathrm{rank}(P), \mathrm{rank}(T)].$$