Model Answers to Exercise 10 of November 18, 2015

http://www.isi.ee.ethz.ch/teaching/courses/it1/

Problem 1  

Channel Coding for a “Double”-Channel

To prove the existence of such a codebook, we use techniques developed for the proof of the channel coding theorem [Cover & Thomas, pp. 201].

- \( Q(x), R < \min\{I(Q, W_1(y_1|x)), I(Q, W_2(y_2|x))\} \) and \( \epsilon > 0 \) are given.
- Fix \( n \).
- Generate a random codebook \( \mathcal{C} \) of block length \( n \) according to \( Q(x) \).
- Reveal \( \mathcal{C} \) to the encoder and both decoders.
- Design joint typicality decoders for each channel \( \phi_1 \) and \( \phi_2 \).
- Given a message \( m \) the encoder maps it to \( X(m) \) (according to \( \mathcal{C} \)).
- \( X(m) \) is transmitted over channels 1 and 2.
- Decode \( Y_1 \) according to \( \phi_1 \) and \( Y_2 \) according to \( \phi_2 \).
- If both decoders find a unique \( \hat{m} \) such that \( (X(\hat{m}), Y_1) \) and \( (X(\hat{m}), Y_2) \) are jointly typical, then declare \( \hat{m} \), otherwise declare an error.
- An error in decoder \( \nu, \nu = 1, 2 \), occurs if the decoder either finds no codeword that is jointly typical with \( Y_\nu, \nu = 1, 2 \), or it finds more than one jointly typical codeword. The probability of the first event approaches zero as \( n \) goes to infinity. The probability of the second event depends on the mutual information \( I(Q, W_\nu) \). Using the union of events bound, we know that the average probability of error \( P_e \) is smaller than or equal to \( P_{e,1} + P_{e,2} \) where \( P_{e,\nu} \) is the average probability of error for channel \( \nu, \nu = 1, 2 \). Taking the average over all codebooks, we get

\[
E[P_e] \leq 2\epsilon + \left(2^nR - 1\right) \left[2^{-n[I(Q, W_1) - 3\epsilon]} + 2^{-n[I(Q, W_2) - 3\epsilon]}\right]
\leq 2\epsilon + 2^{n(R - I(Q, W_1) + 3\epsilon)} + 2^{n(R - I(Q, W_2) + 3\epsilon)}.
\]

- If \( R < \min\{I(Q, W_1(y_1|x)), I(Q, W_2(y_2|x))\} \), then there must exist some \( \epsilon > 0 \) for which \( R < \min\{I(Q, W_1(y_1|x)), I(Q, W_2(y_2|x))\} - 3\epsilon \). For this \( \epsilon \) it thus holds that \( E[P_e] \) goes to zero as \( n \) goes to infinity. Therefore, for this \( \epsilon \) and any \( \delta > 0 \), we can find an \( n \) such that \( E[P_e] < \delta \).
- Since the average over all codebooks is less than \( \delta \), there must exist a (deterministic) codebook such that its \( P_e \leq \delta \).
By throwing away half the codewords we can get a maximal probability of error less than $2\delta$. Since for any message $m$, its probability of error is $\lambda_m = \lambda_{m,1} + \lambda_{m,2} \leq 2\delta \ \forall m$, then $\lambda_{m,\nu} \leq 2\delta \ \forall m$ and $\forall \nu = 1, 2$.

**Problem 2**

**Zero-Error Capacity**

The channel has the transition matrix $W(y|x)$ shown in Figure 1.

![Transition Matrix](image)

Figure 1: Transition matrix of a transmission channel.

a) Since the channel is strongly symmetric, its capacity is

$$C = \log |\mathcal{Y}| - H(\text{row of transition matrix}) = \log_2 5 - 1 = \log_2 2.5 \approx 1.322 \text{ bits}.$$ 

b) Let us construct a block code consisting of codewords of length 2. We want to achieve a capacity greater than 1 bit per channel use, i.e., greater than 2 bits per two channel uses. Thus, we need more than 4 codewords. Let’s pick 5 codewords with distinct first symbols: $\{0a, 1b, 2c, 3d, 4e\}$. We must choose $a, b, c, d, e$ so that the receiver will be able to determine which codeword was transmitted. A simple repetition code will not work, since if, say, 22 is transmitted, then 11 might be received, and the receiver could not tell whether the codeword was 00 or 22. Instead, we use the code $\{00, 13, 21, 34, 42\}$, i.e., the codewords can be described as $x_i = (i, 3i \mod 5)$ for all $i = 0, \ldots, 4$. Then each codeword will be received as one of 4 possible 2-tuples which are all distinct:

- $00 \rightarrow \{44, 14, 41, 11\}$
- $13 \rightarrow \{02, 22, 04, 24\}$
- $21 \rightarrow \{10, 30, 12, 32\}$
- $34 \rightarrow \{23, 43, 20, 40\}$
- $42 \rightarrow \{31, 01, 33, 03\}$.

Since there are 5 possible error-free messages with 2 channel uses, the zero-error capacity of this channel is at least $\frac{1}{2} \log_2 5 \approx 1.161$ bits.

© Amos Lapidoth, 2015/2016

Figure 2: Graphical interpretation.

Figure 3: Codebook of five codewords (the graph has to be understood cyclically).

Here is an alternative way of solving this problem graphically. We introduce a graph consisting of vertices and undirected edges. In the case of block length $n = 1$ we introduce for each input symbol one vertex and two vertices are adjacent (i.e., they are connected by an edge) if and only if there is at least one output symbol which can be the possible output of both input symbols corresponding to the two vertices. In our example, 0 and 2 are adjacent because both can lead to the output symbol 1; but 0 and 1 are not adjacent because they have no common output symbol (see Figure 2).

We see that two messages of length $n = 1$ cannot be confused if they are not adjacent, so to achieve error-freeness our codebook has to consist of non-adjacent vertices. Of course we try to take the maximum number of possible non-adjacent vertices. (In graph theory this number is called the independence number.) A choice for a codebook are, e.g., the vertices 0 and 1. (By the way, because the adjacency graph for $n = 1$ for this channel looks like a pentagon, this channel is sometimes called the “pentagon channel”.)

In the case of block length $n = 2$ we proceed similarly. But now, for each sequence of two input symbols we draw a vertex and there is an edge between two vertices if the two input sequences corresponding to these two vertices can be confused (i.e., if they can lead to the same output.

© Amos Lapidoth, 2015/2016
sequence). In our example, 00 and 20 are adjacent because they have the possible common output 11 (or 14). A codebook consisting of five codewords leading to error-free transmission is drawn in Figure 3.

Problem 3

Symbol Error Rate vs. Message Error Rate

a) We have to show that if \( \frac{1}{n} \sum_{k=1}^{n} \Pr[X_k \neq \hat{X}_k] \) does not tend to zero as \( n \) tends to infinity, then neither does \( \Pr[(X_1, X_2, \ldots, X_n) \neq (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)] \).

To this end, we demonstrate that the probability of a message error is lower bounded by the average probability of a symbol error. We have

\[
\Pr[(X_1, X_2, \ldots, X_n) \neq (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)] = \Pr\left(\bigcup_{k=1}^{n} \{X_k \neq \hat{X}_k\}\right) \geq \max_k \Pr[X_k \neq \hat{X}_k] \geq \frac{1}{n} \sum_{k=1}^{n} \Pr[X_k \neq \hat{X}_k],
\]

where (2) follows because the union is a larger set, and (3) follows because the maximum of a set of numbers is always larger than the average of the set. Hence, if \( \frac{1}{n} \sum_{k=1}^{n} \Pr[X_k \neq \hat{X}_k] \) is bounded away from zero, then so is \( \Pr[(X_1, X_2, \ldots, X_n) \neq (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)] \).

b) We state the proof again:

\[
\frac{1}{n} H(U_1, \ldots, U_n|Z_1, \ldots, Z_n) \overset{(i)}{=} \frac{1}{n} \sum_{i=1}^{n} H(U_i|U_1, \ldots, U_{i-1}, Z_1, \ldots, Z_n) \overset{(ii)}{\leq} \frac{1}{n} \sum_{i=1}^{n} H(U_i|Z_1, \ldots, Z_n) \overset{(iii)}{\leq} \frac{1}{n} \sum_{i=1}^{n} H_b(P_{e,i}) + \frac{1}{n} \sum_{i=1}^{n} P_{e,i} \log(K - 1) \overset{(iv)}{\leq} H_b\left(\frac{1}{n} \sum_{i=1}^{n} P_{e,i}\right) + \log(K - 1) - \frac{1}{n} \sum_{i=1}^{n} P_{e,i} \overset{(v)}{=} H_b(P_b) + P_b \log(K - 1).
\]

(i) follows by the chain rule;

(ii) follows because conditioning reduces entropy;

(iii) follows by upper bounding

\[H(U_i|Z_1, \ldots, Z_n) \leq H(U_i|Z_i)\]

(which follows because conditioning reduces entropy) and by applying Fano’s inequality (Theorem 2.10.1 in Cover & Thomas) with \( U_i \) replacing \( X \) and with \( Z_i \) replacing \( Y \);

(iv) follows from Jensen’s inequality (Theorem 2.6.2 in Cover & Thomas) and because entropy is concave in the input distribution;
(v) holds by the definition of $P_b$.

c) Let $(X_1, \ldots, X_n) = f(M)$ and $(\hat{X}_1, \ldots, \hat{X}_n) = f(\hat{M})$, where $\hat{M}$ is the message guessed by the decoder, and let the sequence $X_1, \ldots, X_n$ take value in a finite set of cardinality $|\mathcal{X}|$. Assume that $f(\cdot)$ is invertible. We prove the channel coding theorem for memoryless channels (though the converse holds more generally). Let $C$ equal the capacity of the channel. Then

$$R = \frac{1}{n}H(M) = \frac{1}{n}H(X_1, \ldots, X_n)$$
$$= \frac{1}{n} \left( H(X_1, \ldots, X_n) + H(X_1, \ldots, X_n | \hat{X}_1, \ldots, \hat{X}_n) ight)$$
$$= \frac{1}{n}H(X_1, \ldots, X_n | \hat{X}_1, \ldots, \hat{X}_n) + \frac{1}{n}I(X_1, \ldots, X_n; \hat{X}_1, \ldots, \hat{X}_n)$$
$$\leq H_b(P_b) + P_b \log(|\mathcal{X}| - 1) + \frac{1}{n}I(X_1, \ldots, X_n; Y_1, \ldots, Y_n) \quad (4)$$
$$\leq H_b(P_b) + P_b \log(|\mathcal{X}| - 1) + \frac{1}{n}nC \quad (5)$$
$$= H_b(P_b) + P_b \log(|\mathcal{X}| - 1) + C,$$

where we have used in (4) the stronger version of Fano’s inequality, which we proved in Part b), with the definition

$$P_b = \frac{1}{n} \sum_{k=1}^n \Pr[X_k \neq \hat{X}_k]; \quad (6)$$

and where Equation (5) follows from Lemma 7.9.2 in Cover & Thomas. Thus we have

$$R - C \leq H_b(P_b) + P_b \log(|\mathcal{X}| - 1) \quad (7)$$

from which follows that if $R > C$ then $P_b$ cannot tend to 0.
Problem 4

An Elementary Converse for the Binary Erasure Channel

a) Fix an arbitrary $s \in \mathcal{F}_{nk}$ and suppose $S^n_1 = s$. Since $S^n_1 = s$, the channel output sequence $Y^n_1$ takes value in the set

$$\mathcal{K}_n = \{ y \in \mathcal{Y}^n : y_i = \_ \text{ if } s_i = 1 \text{ and } y_i \in \{0, 1\} \text{ if } s_i = 0 \}. $$

In particular, the channel output sequence $Y^n_1$ cannot assume more than $|\mathcal{K}_n|$ different values. We therefore conclude that for $S^n_1 = s$ the decoded message $\phi(Y^n_1)$ cannot assume more than $|\mathcal{K}_n|$ different values so that no more than $|\mathcal{K}_n|$ messages can be decoded correctly. Hence, the average probability of a decoding error satisfies the lower bound

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 = s] \geq \frac{|\mathcal{M}| - |\mathcal{K}_n|}{|\mathcal{M}|} = \frac{2^{nR} - 2^{n(1-\kappa)}}{2^{nk}}.$$ 

Since the last result holds for any binary sequence $s$ in $\{0, 1\}^n$ with $\sum_{i=1}^n s_i = nk$, we conclude

$$\frac{2^{nR} - 2^{n(1-\kappa)}}{2^{nk}} \leq \frac{1}{|\mathcal{F}_{nk}|} \sum_{s \in \mathcal{F}_{nk}} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 = s]$$

(i) $$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{F}_{nk}|} \sum_{s \in \mathcal{F}_{nk}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 = s] \Pr[S^n_1 = s | X^n_1 = g(m), S^n_1 \in \mathcal{F}_{nk}]$$

(ii) $$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{F}_{nk}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 = s] \cdot \Pr[S^n_1 = s | X^n_1 = g(m), S^n_1 \in \mathcal{F}_{nk}]$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 \in \mathcal{F}_{nk}] = P_e(\mathcal{F}_{nk}),$$

where (i) holds since the sequence of random variables $S^n_1$ is independent of the channel input sequence $X^n_1$ and since the probability $\Pr[S^n_1 = s]$ is the same for all $s \in \mathcal{F}_{nk}$, i.e.,

$$\Pr[S^n_1 = s | X^n_1 = g(m)] = \Pr[S^n_1 = s] = \rho^{nk} (1 - \rho)^{n-nk}$$

holds for every message $m \in \mathcal{M}$ and every binary sequence $s \in \mathcal{F}_{nk}$; and (ii) is true because $s \in \mathcal{F}_{nk}$ and $S^n_1 = s$ imply $S^n_1 \in \mathcal{F}_{nk}$.

b) Fix a number $\alpha$ so that $1 - R < \alpha < \rho$. Then,

$$P_e^{(n)} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m)]$$

(iii) $$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\kappa \in \{0, 1\}} \Pr[S^n_1 \in \mathcal{F}_{nk} | X^n_1 = g(m)] \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 \in \mathcal{F}_{nk}]$$

(iv) $$= \sum_{\kappa \in \{0, 1\}} \Pr[S^n_1 \in \mathcal{F}_{nk}] \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr[\phi(Y^n_1) \neq m | X^n_1 = g(m), S^n_1 \in \mathcal{F}_{nk}]$$

$$\geq \sum_{\kappa \in \{0, 1\}} \Pr[S^n_1 \in \mathcal{F}_{nk}] \frac{2^{nR} - 2^{n(1-\alpha)}}{2^{nk}}$$

© Amos Lapidoth, 2015/2016
$$\Pr \left[ \sum_{i=1}^{n} S_i \geq \alpha n \right] \left( 1 - 2^{-n(R-(1-\alpha))} \right),$$

where (iii) follows from the Law of Total Probability, and (iv) is true because the sequence of random variables $S^n_1$ is independent of the channel input sequence $X^n_1$. The assumption $1 - R < \alpha$ implies $R > 1 - \alpha$. Therefore,

$$1 - 2^{-n(R-(1-\alpha))} \to 1, \text{ as } n \to \infty.$$ 

By the Weak Law of Large Numbers and since $\rho - \alpha > 0$

$$\Pr \left[ \sum_{i=1}^{n} S_i \geq \alpha n \right] = \Pr \left[ \frac{1}{n} \sum_{i=1}^{n} S_i \geq \rho - (\rho - \alpha) \right] \to 1, \text{ as } n \to \infty.$$ 

**Problem 5**

**Source-Channel Separation and Average Bit-Error Probability**

a) We upper-bound $I(U^k_1; \hat{U}^k_1)$ using the data processing inequality:

$$I(U^k_1; \hat{U}^k_1) \leq I(X^n_1; Y^n_1) \leq nC,$$

where the second inequality follows from Lemma 7.9.2 in Cover & Thomas.

b) In the computation, (a) follows from the chain-rule; (b) holds because of the chain rule and since $U^{i-1}$ is independent of $U_i$; (c) is true because $H(U_i) = 1$ and because

$$H(U_i|\hat{U}^k, U^{i-1}) \leq H(U_i|\hat{U}_i) \leq H_b\left( \Pr[U_i \neq \hat{U}_i] \right),$$

where we used that conditioning cannot increase entropy and that entropy is concave; and (d) holds since binary entropy is concave.

c) Combining the previous parts, we find that

$$nC \geq I(U^k_1; \hat{U}^k_1) \geq k \left( 1 - H_b\left( \frac{1}{k} \sum_{i=1}^{k} \Pr[U_i \neq \hat{U}_i] \right) \right).$$

Since $H_b(\cdot)$ is invertible on $[0, 1/2]$, we have

$$\frac{1}{k} \sum_{i=1}^{k} \Pr[U_i \neq \hat{U}_i] \geq H_b^{-1}\left( 1 - \frac{n}{k}C \right) > 0,$$

where the last inequality holds by assumption that $k/n > C$. 

© Amos Lapidoth, 2015/2016 7