# Information Theory II

Spring Semester 2017

Prof. Dr. A. Lapidoth / Prof. Dr. S. Moser

*Signal and Information Processing Laboratory*

*Institut für Signal- und Informationsverarbeitung*

## Model Answers to Exercise 3 of March 9, 2017

http://www.isi.ee.ethz.ch/teaching/courses/it2.html

---

### Problem 1 — *Gaussian Multipath Channel*

a) The channel can be simplified to

$$Y = 2X + Z,$$

where $Z = Z_1 + Z_2$. Since $Z_1$ and $Z_2$ are jointly Gaussian, we know that $Z$ is also a Gaussian random variable. Its variance is

$$\mathsf{Var}(Z) = \mathsf{Var}(Z_1 + Z_2) = \mathsf{Var}(Z_1) + \mathsf{Var}(Z_2) + 2\,\mathsf{Cov}[Z_1, Z_2] = 2(1+\rho)\sigma^2.$$

Introducing $X' = 2X$, this channel is equivalent to the Gaussian channel

$$Y = X' + Z,$$

with an average-power constraint $4\mathsf{E}_\text{s}$ and noise variance $2(1+\rho)\sigma^2$. Its capacity is therefore

$$\mathsf{C} = \frac{1}{2}\log\left(1 + \frac{2\mathsf{E}_\text{s}}{(1+\rho)\sigma^2}\right).$$

b) For $\rho = 0$, we have $\mathsf{C} = \frac{1}{2}\log\bigl(1 + \frac{2\mathsf{E}_\text{s}}{\sigma^2}\bigr)$.

For $\rho = 1$, we have $\mathsf{C} = \frac{1}{2}\log\bigl(1 + \frac{\mathsf{E}_\text{s}}{\sigma^2}\bigr)$, which is the same as the capacity of a channel with only one path.

For $\rho = -1$, we have $\mathsf{C} = \infty$ because the noise in the two paths cancels and the capacity of a noise-free channel with an infinite alphabet is infinite.

### Problem 2 — *Additive Noise Channel*

Since the channel is noise-free with probability 1/10, we guess that the capacity of the channel is infinite. In order for a coding scheme to achieve infinite capacity, we must show that for any $\epsilon > 0$, there exists an $n_0$ such that for any $n > n_0$, we can send an arbitrary number of bits using $n$ channel uses with probability of error less that $\epsilon$. We propose two different schemes:

**Coding scheme 1:** The encoder maps the message to a rational number with absolute value less than $\sqrt{\mathsf{E}_\text{s}}$ and transmits that number repeatedly. The decoder looks at the received sequence to see if there is a rational number. If yes, declare it to be the input. Otherwise, declare an error.

Since the output will be irrational with probability one if $Z$ is Gaussian, the probability of error is $(0.9)^n$. For any $\epsilon > 0$, we can find an $n_0$ such that $(0.9)^n < \epsilon$, for $n > n_0$. Note that there is an infinite number of rational numbers between $-\sqrt{\mathsf{E}_\text{s}}$ and $\sqrt{\mathsf{E}_\text{s}}$, so we can indeed send an infinite number of messages.

**Coding scheme 2:** The encoder maps the message to a real number with absolute value less than $\sqrt{\mathsf{E}_\mathrm{s}}$ and transmits the number repeatedly. The decoder looks at the received sequence to see if there are any repeated numbers. If yes, declare them to be the input. Otherwise, declare an error.

With probability one, the only repeated received symbols will be the noise free copies of the input. Thus, the probability of error for this scheme is the probability that $Z = 0$ for none or only one out of the $n$ transmissions: $(0.9)^n + \binom{n}{1}(0.1)(0.9)^{n-1}$. Again, we can choose $n_0$ large enough to ensure probability of error less than any $\epsilon > 0$, and there is an infinite number of real numbers between $-\sqrt{\mathsf{E}_\mathrm{s}}$ and $\sqrt{\mathsf{E}_\mathrm{s}}$.

## Problem 3 — *DMC with a Cost Constraint*

We want to show that the capacity of the channel $W$ under const constraint $\beta$ is given by

$$\mathsf{C}_{\mathrm{inf}}(\beta) = \max_{P_X \colon \mathsf{E}[b(X)] \le \beta} I(X;Y) = \max_{P_X \colon \mathsf{E}[b(X)] \le \beta} I(P_X, W). \tag{1}$$

We begin with some rather technical observations about $\mathsf{C}_{\mathrm{inf}}(\cdot)$:

- Let $\beta_{\min} = \min_{x \in \mathcal{X}} b(x)$. Since $\mathsf{E}[b(X)]$ cannot be smaller than $\beta_{\min}$, the function $\mathsf{C}_{\mathrm{inf}}(\beta)$ is only defined for $\beta \ge \beta_{\min}$.
- For every $\beta \ge \beta_{\min}$, the maximum in (1) is indeed achieved, i.e., there exists a PMF $P_X^*$ satisfying $\mathsf{E}[b(X)] \le \beta$ and $\mathsf{C}_{\mathrm{inf}}(\beta) = I(P_X^*, W)$. (Because the input alphabet is finite, the set $\{P_X : \mathsf{E}[b(X)] \le \beta\}$ is compact; and since $I(P_X, W)$ is continuous in $P_X$, the existence of the maximum follows from the extreme value theorem.)
- The function $\mathsf{C}_{\mathrm{inf}}(\beta)$ is continuous in $\beta$ for all $\beta \ge \beta_{\min}$. (This can be shown with a nontrivial topological argument.)
- The function $\mathsf{C}_{\mathrm{inf}}(\beta)$ is nondecreasing in $\beta$. This is easy to show: Let $\beta \ge \beta^*$. Let $Q^*$ be a PMF that achieves the maximum in (1), so $\mathsf{E}_{Q^*}[b(X)] \le \beta^*$ and $\mathsf{C}_{\mathrm{inf}}(\beta^*) = I(Q^*, W)$ hold. Then,

  $$\begin{aligned} \mathsf{C}_{\mathrm{inf}}(\beta) &= \max_{P_X \colon \mathsf{E}[b(X)] \le \beta} I(P_X, W) \\ &\stackrel{\text{(i)}}{\ge} I(Q^*, W) \\ &\stackrel{\text{(ii)}}{=} \mathsf{C}_{\mathrm{inf}}(\beta^*), \end{aligned}$$

  where (i) holds because $Q^*$ satisfies $\mathsf{E}[b(X)] \le \beta$; and (ii) holds because $Q^*$ achieves $\mathsf{C}_{\mathrm{inf}}(\beta^*)$.
- The function $\mathsf{C}_{\mathrm{inf}}(\beta)$ is concave in $\beta$. This can be shown as follows: Let $\beta_1 \ge \beta_{\min}$, $\beta_2 \ge \beta_{\min}$, and $\lambda \in [0, 1]$. Let $Q_1$ and $Q_2$ be PMFs that achieve the maximum in (1), so $\mathsf{E}_{Q_1}[b(X)] \le \beta_1$, $\mathsf{C}_{\mathrm{inf}}(\beta_1) = I(Q_1, W)$, $\mathsf{E}_{Q_2}[b(X)] \le \beta_2$, and $\mathsf{C}_{\mathrm{inf}}(\beta_2) = I(Q_2, W)$ hold. Then,

  $$\begin{aligned} \mathsf{C}_{\mathrm{inf}}(\lambda\beta_1 + \bar{\lambda}\beta_2) &= \max_{P_X \colon \mathsf{E}[b(X)] \le \lambda\beta_1 + \bar{\lambda}\beta_2} I(P_X, W) \\ &\stackrel{\text{(i)}}{\ge} I(\lambda Q_1 + \bar{\lambda} Q_2, W) \\ &\stackrel{\text{(ii)}}{\ge} \lambda I(Q_1, W) + \bar{\lambda} I(Q_2, W) \\ &\stackrel{\text{(iii)}}{=} \lambda \mathsf{C}_{\mathrm{inf}}(\beta_1) + \bar{\lambda} \mathsf{C}_{\mathrm{inf}}(\beta_2), \end{aligned}$$

where (i) holds because the mixture $\lambda Q_1+\bar{\lambda}Q_2$ satisfies $\mathsf{E}[b(X)]\leq\lambda\beta_1+\bar{\lambda}\beta_2$; (ii) follows from the concavity of the mutual information; and (iii) holds because $Q_1$ and $Q_2$ achieve $\mathsf{C}_{\text{inf}}(\beta_1)$ and $\mathsf{C}_{\text{inf}}(\beta_2)$, respectively.

(A remark for the interested reader: It can be shown that any function which is convex on some open interval $(a,b)$ is continuous on that interval. However, convexity does *not* imply that the function is continuous in the endpoints $a$ or $b$. The same result holds for concavity.)

We continue with the achievability part. Assume for now that $\beta>\beta_{\min}$ holds. For any $\epsilon$ satisfying $0<\epsilon\leq\beta-\beta_{\min}$, we will show the achievability of all rates $\mathsf{R}$ satisfying

$$\mathsf{R}<\mathsf{C}_{\text{inf}}(\beta-\epsilon)-3\epsilon. \tag{2}$$

By the continuity of $\mathsf{C}_{\text{inf}}(\cdot)$, the RHS of (2) tends to $\mathsf{C}_{\text{inf}}(\beta)$ as $\epsilon$ tends to zero; therefore for every rate $\mathsf{R}<\mathsf{C}_{\text{inf}}(\beta)$ there exists an $\epsilon$ such that (2) holds.
Let $\mathsf{R}$ and $\epsilon$ be such that (2) holds. Let $P_X^*$ be a PMF that achieves $\mathsf{C}_{\text{inf}}(\beta-\epsilon)$, so $\mathsf{E}_{P_X^*}[b(X)]\leq\beta-\epsilon$ and $\mathsf{C}_{\text{inf}}(\beta-\epsilon)=I(P_X^*,W)$ hold. Generate an $(n,\mathsf{R})$ codebook where the symbols in every codeword are drawn IID according to $P_X^*$. When a sequence $Y^n$ is received, the decoder looks in the codebook for a codeword that is jointly typical with $Y^n$ and satisfies the cost constraint. If there is only one such codeword, the decoder outputs the message corresponding to this codeword; if there is more than one or no such codeword, the decoder declares an error.
We now analyze the average error probability averaged over all such randomly generated codebooks. We list and analyze the possible error types as follows:

- The transmitted codeword does not satisfy

  $$\frac{1}{n}\sum_{i=1}^{n}b(X_i)\leq\beta$$

  and is thus rejected by the decoder. Since the input symbols are drawn IID according to $P_X^*$ and we have $\mathsf{E}_{P_X^*}[b(X)]\leq\beta-\epsilon$, the probability of this type of error tends to zero as $n$ tends to infinity by the law of large numbers.

- The transmitted codeword and the received sequence are not jointly typical. The probability of this type of error tends to zero as $n$ tends to infinity.

- There exists a codeword which is not transmitted but which is jointly typical with $Y^n$. The probability of this event tends to zero as $n$ tends to infinity because $\mathsf{R}$ satisfies (2).

We have thus shown that for $n$ large enough, the probability of error averaged over the above codebooks is smaller than $\epsilon$. It then follows that there must be at least one codebook whose average probability of error is smaller than $\epsilon$. We then throw away half of the codewords in this codebook, namely those with the highest error probabilities. The maximum error probability of the remaining codewords must be smaller than $2\epsilon$. Since all remaining codewords have an error probability less than one, they must satisfy the cost constraint; indeed, if they would not satisfy the cost constraint, the decoder would always declare an error and the error probability would be one. Note that by throwing away half of the codewords, the rate is reduced by $\frac{1}{n}$ bits, which tends to zero as $n$ tends to infinity.
We conclude the proof of the achievability part by considering the case $\beta=\beta_{\min}$. Now, let $P_X^*$ be a PMF that achieves $\mathsf{C}_{\text{inf}}(\beta)$, so $\mathsf{E}_{P_X^*}[b(X)]\leq\beta$ and $\mathsf{C}_{\text{inf}}(\beta)=I(P_X^*,W)$ hold. Let the symbols in every codeword in the codebook be drawn IID according to $P_X^*$. In this case, any $x$ with $P_X^*(x)>0$ satisfies $b(x)=\beta$, and therefore the cost constraint will always be satisfied. The rest of the analysis is the same as in the case $\beta>\beta_{\min}$.

We next prove the converse part. We will show that if a sequence of codebooks of rate $\mathsf{R} > \mathsf{C}_{\text{inf}}(\beta)$ satisfies the weaker cost constraint

$$\frac{1}{2^{n\mathsf{R}}} \sum_{m=1}^{2^{n\mathsf{R}}} \frac{1}{n} \sum_{i=1}^{n} b\bigl(x_i(m)\bigr) \leq \beta, \tag{3}$$

the average probability of error cannot tend to zero as $n$ tends to infinity. To that end, observe that

$$\begin{aligned}
I(X^n; Y^n) &= H(Y^n) - H(Y^n|X^n) \\
&= \sum_{i=1}^{n} \bigl[H(Y_i|Y^{i-1}) - H(Y_i|X^n, Y^{i-1})\bigr] \\
&= \sum_{i=1}^{n} \bigl[H(Y_i|Y^{i-1}) - H(Y_i|X_i)\bigr] \\
&\leq \sum_{i=1}^{n} [H(Y_i) - H(Y_i|X_i)] \\
&= \sum_{i=1}^{n} I(X_i; Y_i) \\
&\overset{\text{(i)}}{\leq} \sum_{i=1}^{n} \mathsf{C}_{\text{inf}}(\mathsf{E}[b(X_i)]) \\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathsf{C}_{\text{inf}}(\mathsf{E}[b(X_i)]) \\
&\overset{\text{(ii)}}{\leq} n \cdot \mathsf{C}_{\text{inf}}\left( \sum_{i=1}^{n} \frac{1}{n} \mathsf{E}[b(X_i)] \right) \\
&\overset{\text{(iii)}}{\leq} n \cdot \mathsf{C}_{\text{inf}}(\beta),
\end{aligned}$$

where (i) follows from the definition (1); (ii) holds because $\mathsf{C}_{\text{inf}}(\cdot)$ is concave; and (iii) holds because $\sum_{i=1}^{n} \frac{1}{n} \mathsf{E}[b(X_i)] = \mathsf{E}\bigl[\sum_{i=1}^{n} \frac{1}{n} b(X_i)\bigr] \leq \beta$ as the codebook is assumed to satisfy the cost constraint (3) and because $\mathsf{C}_{\text{inf}}(\cdot)$ is nondecreasing. Next, we have

$$\begin{aligned}
n\mathsf{R} &= H(M) \\
&= I(M; \hat{M}) + H(M|\hat{M}) \\
&\overset{\text{(i)}}{\leq} I(X^n; Y^n) + H(M|\hat{M}) \\
&\overset{\text{(ii)}}{\leq} I(X^n; Y^n) + 1 + P_{\text{e}} \cdot n\mathsf{R} \\
&\leq n \cdot \mathsf{C}_{\text{inf}}(\beta) + 1 + P_{\text{e}} \cdot n\mathsf{R},
\end{aligned}$$

where (i) follows from the data processing inequality; and (ii) follows from Fano's inequality. Solving for $P_{\text{e}}$, we obtain

$$P_{\text{e}} \geq \frac{\mathsf{R} - \mathsf{C}_{\text{inf}}(\beta) - \frac{1}{n}}{\mathsf{R}},$$

which for $\mathsf{R} > \mathsf{C}_{\text{inf}}(\beta)$ precludes the possibility that the average probability of error tends to zero as $n$ tends to infinity.