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 7 of April 6, 2017

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

Problem 1 *Cooperation on the Gaussian MAC*

Since both users have access to both messages, the Gaussian MAC reduces to a single-user Gaussian channel $Y_i = \tilde{x}_i + Z_i$ with input $\tilde{x}_i = x_{1,i} + x_{2,i}$, message set $\tilde{\mathcal{M}} = \mathcal{M}_1 \times \mathcal{M}_2$, and rate $\tilde{\mathsf{R}} = \mathsf{R}_1 + \mathsf{R}_2$. Since the average power of both transmitters is limited, $\tilde{x}_i$ cannot be chosen freely. In any coding scheme for which the average power of X_1 does not exceed E_1 and the average power of X_2 does not exceed E_2, the average power of $\tilde{X}$ is bounded for all messages $\tilde{m} \in \tilde{\mathcal{M}}$ as

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\bigl(\tilde{x}_i(\tilde{m})\bigr)^2 &= \frac{1}{n}\sum_{i=1}^{n}\bigl(x_{1,i}(\tilde{m}) + x_{2,i}(\tilde{m})\bigr)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\bigl(x_{1,i}(\tilde{m})\bigr)^2 + \frac{2}{n}\sum_{i=1}^{n} x_{1,i}(\tilde{m})x_{2,i}(\tilde{m}) + \frac{1}{n}\sum_{i=1}^{n}\bigl(x_{2,i}(\tilde{m})\bigr)^2 \\
&\le \mathsf{E}_1 + \mathsf{E}_2 + \frac{2}{n}\sum_{i=1}^{n} x_{1,i}(\tilde{m})x_{2,i}(\tilde{m}) \\
&= \mathsf{E}_1 + \mathsf{E}_2 + 2\sum_{i=1}^{n}\frac{x_{1,i}(\tilde{m})}{\sqrt{n}}\cdot\frac{x_{2,i}(\tilde{m})}{\sqrt{n}} \\
&\overset{\text{(i)}}{\le} \mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\sum_{i=1}^{n}\frac{\bigl(x_{1,i}(\tilde{m})\bigr)^2}{n}}\cdot\sqrt{\sum_{i=1}^{n}\frac{\bigl(x_{2,i}(\tilde{m})\bigr)^2}{n}} \\
&\le \mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2},
\end{aligned}$$

where (i) follows from the Cauchy–Schwarz inequality. The converse for the single-user Gaussian channel implies that for all achievable rate pairs $(\mathsf{R}_1, \mathsf{R}_2)$

$$\mathsf{R}_1 + \mathsf{R}_2 = \tilde{\mathsf{R}} \le \frac{1}{2}\log\left(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2}}{\sigma^2}\right) \tag{1}$$

has to hold. We now argue that if (1) holds with strict inequality, then the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable. If (1) holds with strict inequality, then it is possible to design single-user encoders and decoders for which the average power of $\tilde{X}$ does not exceed $\mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2}$ and whose probability of error tends to zero as the block length tends to infinity. Choosing the input symbols as

$$\begin{aligned}
x_{1,i}(\tilde{m}) &= \frac{\sqrt{\mathsf{E}_1}}{\sqrt{\mathsf{E}_1} + \sqrt{\mathsf{E}_2}}\,\tilde{x}_i(\tilde{m}), \\
x_{2,i}(\tilde{m}) &= \frac{\sqrt{\mathsf{E}_2}}{\sqrt{\mathsf{E}_1} + \sqrt{\mathsf{E}_2}}\,\tilde{x}_i(\tilde{m})
\end{aligned}$$

ensures $\tilde{x}_i = x_{1,i} + x_{2,i}$. In addition, the average power of X_1 does not exceed E_1 since

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\bigl(x_{1,i}(\tilde{m})\bigr)^2 &= \frac{\mathsf{E}_1}{\mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2}} \cdot \frac{1}{n}\sum_{i=1}^{n}\bigl(\tilde{x}_i(\tilde{m})\bigr)^2 \\
&\leq \frac{\mathsf{E}_1}{\mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2}} \cdot \Bigl(\mathsf{E}_1 + \mathsf{E}_2 + 2\sqrt{\mathsf{E}_1\mathsf{E}_2}\Bigr) \\
&= \mathsf{E}_1.
\end{aligned}$$

Similarly, the power constraint on X_2 is met. Therefore, any rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ satisfying (1) with strict inequality can be achieved.

Problem 2 *Gaussian MAC*

We first prove the achievability. For a given rate pair in the interior of the capacity region, choose $\epsilon > 0$ such that

$$\mathsf{R}_1 < \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_1 - \epsilon}{\sigma^2}\biggr) - 3\epsilon, \tag{2}$$

$$\mathsf{R}_2 < \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_2 - \epsilon}{\sigma^2}\biggr) - 3\epsilon, \tag{3}$$

$$\mathsf{R}_1 + \mathsf{R}_2 < \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2 - 2\epsilon}{\sigma^2}\biggr) - 4\epsilon. \tag{4}$$

Generate a rate-R_1 blocklength-n codebook $\mathcal{C}_1$ for the first user by drawing IID $\mathcal{N}(0, \mathsf{E}_1 - \epsilon)$, and generate a rate-$\mathsf{R}_2$ blocklength-n codebook $\mathcal{C}_2$ for the second user by drawing IID $\mathcal{N}(0, \mathsf{E}_2 - \epsilon)$. With these input distributions, we obtain

$$\begin{aligned}
I(X_1; Y|X_2) &= h(Y|X_2) - h(Y|X_1, X_2) \\
&= h(X_1 + X_2 + Z|X_2) - h(X_1 + X_2 + Z|X_1, X_2) \\
&= h(X_1 + Z|X_2) - h(Z|X_1, X_2) \\
&\overset{\text{(i)}}{=} h(X_1 + Z) - h(Z) \\
&= \frac{1}{2}\log\Bigl(2\pi e\bigl(\mathsf{E}_1 - \epsilon + \sigma^2\bigr)\Bigr) - \frac{1}{2}\log\bigl(2\pi e\sigma^2\bigr) \\
&= \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_1 - \epsilon}{\sigma^2}\biggr),
\end{aligned}$$

where (i) holds because X_1, X_2, and Z are independent. Similarly, it is possible to show that

$$\begin{aligned}
I(X_2; Y|X_1) &= \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_2 - \epsilon}{\sigma^2}\biggr), \\
I(X_1, X_2; Y) &= \frac{1}{2}\log\biggl(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2 - 2\epsilon}{\sigma^2}\biggr).
\end{aligned}$$

Combining these results with (2)–(4), we see that

$$\mathsf{R}_1 < I(X_1; Y|X_2) - 3\epsilon, \tag{5}$$

$$\mathsf{R}_2 < I(X_2; Y|X_1) - 3\epsilon, \tag{6}$$

$$\mathsf{R}_1 + \mathsf{R}_2 < I(X_1, X_2; Y) - 4\epsilon. \tag{7}$$

We use a joint weak typicality decoder: if there exists a unique pair of codewords which satisfies the average-power constraints and which is jointly typical with the received sequence Y^n, the decoder outputs the messages that correspond to the pair of codewords, and otherwise, the decoder declares an error. By the union bound, the probability of error is upperbounded by the sum of the following error probabilities:

- The transmitted codewords do not satisfy the average-power constraints. By the law of large numbers, the probability of this type of error tends to zero as n tends to infinity.
- The transmitted codewords and the received sequence are not jointly typical. The probability of this type of error also tends to zero as n tends to infinity.
- A codeword from the first codebook which was not transmitted, the codeword from the second transmitter, and the received sequence are jointly typical. The probability of this type of error tends to zero as n tends to infinity because of (5).
- The codeword from the first transmitter, a codeword from the second codebook which was not transmitted, and the received sequence are jointly typical. The probability of this type of error tends to zero as n tends to infinity because of (6).
- A codeword from the first codebook which was not transmitted, a codeword from the second codebook which was not transmitted, and the received sequence are jointly typical. The probability of this type of error tends to zero as n tends to infinity because of (7).

Thus, for n large enough, the probability of error averaged over all realizations of the codebooks tends to zero as n tends to infinity. For a given n, there must be at least one pair of codebooks whose probability of error is not larger than the averaged probability of error. To ensure that the average-power constraint is always satisfied, we modify the encoders in the following way: if the codeword associated with the message does not satisfy the average-power constraint, the all-zero codeword is sent instead. This modification does not change the probability of error, since in this case, the decoder would not have produced the correct messages anyway.
The proof of the converse is similar to the proof of the converse for the single-user Gaussian channel. Observe that

$$\begin{aligned}
I(M_1;\hat{M}_1) &\overset{\text{(i)}}{\leq} I(X_1^n;Y^n) \\
&\leq I(X_1^n;Y^n) + I(X_1^n;X_2^n|Y^n) \\
&= I(X_1^n;X_2^n,Y^n) \\
&= I(X_1^n;Y^n|X_2^n) + \underbrace{I(X_1^n;X_2^n)}_{=0} \\
&= h(Y^n|X_2^n) - h(Y^n|X_1^n,X_2^n) \\
&= \sum_{i=1}^{n} \Big[h(Y_i|X_2^n,Y^{i-1}) - h(Y_i|X_1^n,X_2^n,Y^{i-1}) \Big] \\
&= \sum_{i=1}^{n} \Big[h(Y_i|X_2^n,Y^{i-1}) - h(Y_i|X_{1,i},X_{2,i}) \Big] \\
&\leq \sum_{i=1}^{n} \Big[h(Y_i|X_{2,i}) - h(Y_i|X_{1,i},X_{2,i}) \Big] \\
&= \sum_{i=1}^{n} \Big[h(X_{1,i}+X_{2,i}+Z_i|X_{2,i}) - h(X_{1,i}+X_{2,i}+Z_i|X_{1,i},X_{2,i}) \Big] \\
&= \sum_{i=1}^{n} \Big[h(X_{1,i}+Z_i) - h(Z_i) \Big] \\
&\overset{\text{(ii)}}{\leq} \sum_{i=1}^{n} \bigg[\frac{1}{2}\log\Big(2\pi e\big(\mathsf{Var}(X_{1,i})+\sigma^2\big)\Big) - \frac{1}{2}\log\big(2\pi e\sigma^2\big) \bigg] \\
&= n\cdot\frac{1}{2}\sum_{i=1}^{n}\frac{1}{n}\log\bigg(1+\frac{\mathsf{Var}(X_{1,i})}{\sigma^2}\bigg)
\end{aligned}$$

$$
\begin{aligned}
&\leq n \cdot \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n} \log\left(1 + \frac{\mathsf{E}\big[X_{1,i}^2\big]}{\sigma^2}\right) \\
&\overset{\text{(iii)}}{\leq} n \cdot \frac{1}{2} \log\left(1 + \frac{\sum_{i=1}^{n} \frac{1}{n} \mathsf{E}\big[X_{1,i}^2\big]}{\sigma^2}\right) \\
&\leq n \cdot \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_1}{\sigma^2}\right),
\end{aligned}
$$

where (i) follows from the data processing inequality; (ii) follows from the maximum differential entropy principle and holds because $X_{1,i}$ and Z_i are independent, so the variance of their sum is the sum of their variances; and (iii) follows from Jensen's inequality since the logarithm is a concave function. Consequently,

$$
\begin{aligned}
\mathsf{R}_1 &= \frac{1}{n} H(M_1) \\
&= \frac{1}{n} I(M_1; \hat{M}_1) + \frac{1}{n} H(M_1|\hat{M}_1) \\
&\overset{\text{(i)}}{\leq} \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_1}{\sigma^2}\right) + \frac{H_\text{b}\big(\Pr\big[M_1 \neq \hat{M}_1\big]\big) + \Pr\big[M_1 \neq \hat{M}_1\big] \log 2^{n\mathsf{R}_1}}{n} \\
&= \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_1}{\sigma^2}\right) + \delta_n,
\end{aligned}
$$

where (i) follows from the previous computations and from Fano's inequality. Since $\Pr\big[M_1 \neq \hat{M}_1\big]$ must tend to zero as n tends to infinity, δ_n tends to zero as n tends to infinity, which proves the first part of the converse. Similarly,

$$
\mathsf{R}_2 \leq \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_2}{\sigma^2}\right)
$$

can be shown. The proof for

$$
\mathsf{R}_1 + \mathsf{R}_2 \leq \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2}{\sigma^2}\right)
$$

works along the same lines; the key difference is that

$$
h(X_{1,i} + X_{2,i} + Z_i) \leq \frac{1}{2} \log\Big(2\pi e\big(\mathsf{Var}(X_{1,i}) + \mathsf{Var}(X_{2,i}) + \sigma^2\big)\Big)
$$

is used, which holds because $X_{1,i}$, $X_{2,i}$, and Z_i are independent.

Problem 3 — *Gaussian MAC and TDMA*

Since the rate pair (R, R) is in the interior of the capacity region, $\mathsf{R} + \mathsf{R} < \frac{1}{2} \log\big(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2}{\sigma^2}\big)$ holds, which is equivalent to $2\mathsf{R} < \frac{1}{2} \log\big(1 + \frac{2\mathsf{E}}{\sigma^2}\big)$. This implies that for every $\epsilon > 0$ and n large enough, there exists a codebook of blocklength $n/2$ and rate $2\mathsf{R}$ whose probability of error does not exceed ϵ and for which the average-power of every codeword does not exceed $2\mathsf{E}$.
During the first $n/2$ transmissions, the first transmitter uses such a codebook to transmit the message $m_1 \in \{1, \ldots, 2^{\frac{n}{2} 2\mathsf{R}}\}$, while the second transmitter stays silent. During the second $n/2$ transmissions, the second transmitter uses the codebook to transmit its message, while the first transmitter stays silent. The first transmitter satisfies the average-power constraint because

$$
\frac{1}{n} \sum_{i=1}^{n} x_{1,i}^2(m_1) = \frac{1}{2} \cdot \frac{1}{n/2} \sum_{i=1}^{n/2} x_{1,i}^2(m_1) \leq \frac{1}{2} \cdot 2\mathsf{E} = \mathsf{E}.
$$

Similarly, the second transmitter satisfies the average-power constraint, too. By the union bound, the probability of error of this scheme does not exceed 2ϵ, which shows that any rate pair (R, R) in the interior of the capacity region can be achieved with time-division multiplexing if $\mathsf{E}_1 = \mathsf{E}_2$. (Note that many points in the capacity region cannot be achieved with time-division multiplexing; so except in some special cases, it is strictly suboptimal to use time-division multiplexing.)