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 4 of March 16, 2017

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

Problem 1 ***"Double"-Gaussian Channel***

a) The capacity of this channel is given by

$$C = \frac{1}{2}\log\left(1+\frac{1}{N}\right).$$

The corresponding plot is shown in Figure 1.

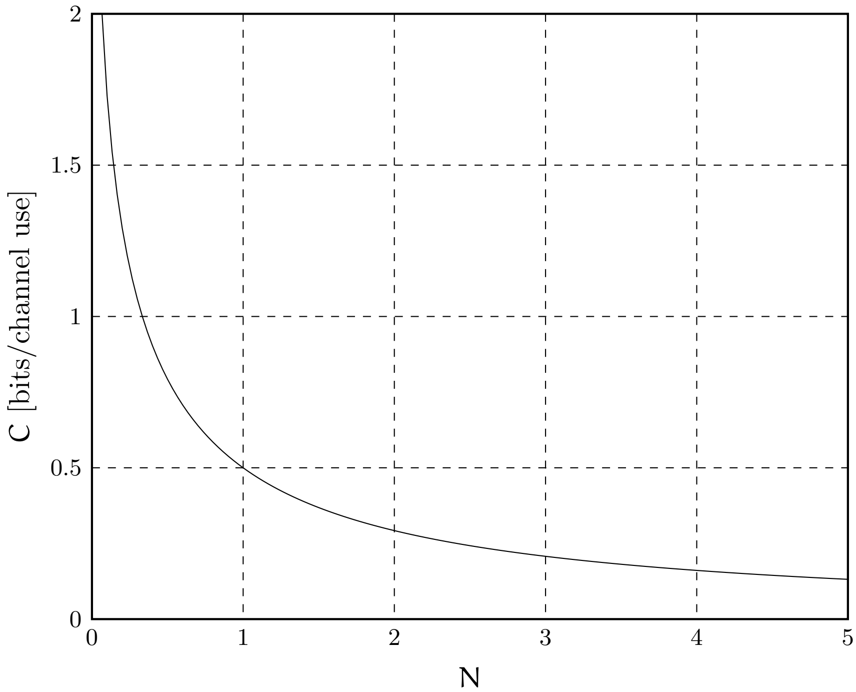

Figure 1: Channel capacity C (in bits per channel use) versus noise variance N.

b) Note that the capacity of two independent parallel Gaussian channels with fixed average-power constraints is

$$C = \frac{1}{2}\log\left(1+\frac{E_1}{N_1}\right) + \frac{1}{2}\log\left(1+\frac{E_2}{N_2}\right).$$

The capacity can be achieved with rate-splitting, where one part of the message is transmitted over one channel and the other part of the message is transmitted over the other channel. The converse is nearly the same as for the parallel Gaussian channel with a total average-power constraint and is omitted.

i) Since the total power is divided equally between the two channels, we have

$$\mathsf{C}_{\mathrm{A}} = 2 \cdot \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}}\right) = \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}}\right)$$

for scheme A and

$$\mathsf{C}_{\mathrm{B}} = \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}_1}\right) + \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}_2}\right)$$

for scheme B. From Figure 1, one can guess that the function $f\colon x \mapsto \log\bigl(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2x}\bigr)$ is convex. Indeed, its second derivative satisfies

$$\frac{\mathsf{E}_{\mathrm{s}} \cdot (4x + \mathsf{E}_{\mathrm{s}})}{(2x^2 + x \cdot \mathsf{E}_{\mathrm{s}})^2} \cdot \log e > 0$$

for all positive x. Therefore, using that f is convex,

$$\begin{aligned}
\mathsf{C}_{\mathrm{A}} &= \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}}\right) \\
&= \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2(\mathsf{N}_1/2 + \mathsf{N}_2/2)}\right) \\
&= f\left(\frac{1}{2}\mathsf{N}_1 + \frac{1}{2}\mathsf{N}_2\right) \\
&\leq \frac{1}{2} f(\mathsf{N}_1) + \frac{1}{2} f(\mathsf{N}_2) \\
&= \mathsf{C}_{\mathrm{B}}.
\end{aligned}$$

Thus, scheme B is better.

ii) Denote the capacity of scheme A with an optimal power allocation by $\mathsf{C}_{\mathrm{A}}^*$, let E_1 denote the power that is assigned to the first channel and let E_2 denote the power that is assigned to the second channel. By the concavity of the logarithm,

$$\begin{aligned}
\mathsf{C}_{\mathrm{A}}^* &= \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_1}{\mathsf{N}}\right) + \frac{1}{2} \log\left(1 + \frac{\mathsf{E}_2}{\mathsf{N}}\right) \\
&\leq \log\left[\frac{1}{2}\left(1 + \frac{\mathsf{E}_1}{\mathsf{N}}\right) + \frac{1}{2}\left(1 + \frac{\mathsf{E}_2}{\mathsf{N}}\right)\right] \\
&= \log\left(1 + \frac{\mathsf{E}_1 + \mathsf{E}_2}{2\mathsf{N}}\right) \\
&= \log\left(1 + \frac{\mathsf{E}_{\mathrm{s}}}{2\mathsf{N}}\right) \\
&= \mathsf{C}_{\mathrm{A}}.
\end{aligned}$$

Denoting the capacity of scheme B with an optimal power allocation by $\mathsf{C}_{\mathrm{B}}^*$, we have

$$\mathsf{C}_{\mathrm{B}}^* \overset{\text{(i)}}{\geq} \mathsf{C}_{\mathrm{B}} \overset{\text{(ii)}}{\geq} \mathsf{C}_{\mathrm{A}} \overset{\text{(iii)}}{\geq} \mathsf{C}_{\mathrm{A}}^*,$$

where (i) holds because optimizing the power allocation can only improve the capacity; (ii) follows from Part b-i); and (iii) is shown above. We conclude that also in the case of optimal power allocation, scheme B is better.

Problem 2 *Bandlimited Gaussian Channel*

If the signal-to-noise ratio P/N_0 is small compared to the bandwidth W, then

$$W \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}_0 W}\right) \approx \frac{\mathsf{P}}{\mathsf{N}_0} \log e, \tag{1}$$

so in this regime doubling the bandwidth has very little effect while doubling the power almost doubles the capacity. On the other hand, if W is small compared to P/N_0, then

$$W \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}_0 W}\right) \approx W \log\left(\frac{\mathsf{P}}{\mathsf{N}_0 W}\right), \tag{2}$$

so in this regime the capacity grows only logarithmically in P but almost linearly in W. (The approximation (1) is a first-order Taylor approximation around $\mathsf{P}/\mathsf{N}_0 = 0$, and in (2), the first term of the sum has been neglected.)

It is also possible to perform an exact analysis. Denote by C_{2W} the capacity obtained by doubling the bandwidth and by $\mathsf{C}_{2\mathsf{P}}$ the capacity obtained by doubling the power, so

$$\begin{aligned} \mathsf{C}_{2W} &= 2W \log\left(1 + \frac{\mathsf{P}}{2\mathsf{N}_0 W}\right) = W \log\left(1 + \frac{\mathsf{P}}{2\mathsf{N}_0 W}\right)^2 = W \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}_0 W} + \frac{\mathsf{P}^2}{4\mathsf{N}_0^2 W^2}\right), \\ \mathsf{C}_{2\mathsf{P}} &= W \log\left(1 + \frac{2\mathsf{P}}{\mathsf{N}_0 W}\right) = W \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}_0 W} + \frac{\mathsf{P}}{\mathsf{N}_0 W}\right). \end{aligned}$$

Thus, $\mathsf{C}_{2\mathsf{P}}$ is larger than C_{2W} if and only if $\frac{\mathsf{P}}{\mathsf{N}_0 W}$ is larger than $\frac{\mathsf{P}^2}{4\mathsf{N}_0^2 W^2}$, i.e., if and only if $W > \frac{\mathsf{P}}{4\mathsf{N}_0}$.

Problem 3 *Exponential Noise Channel*

We begin with the converse. We have

$$\begin{aligned} I(X^n; Y^n) &= h(Y^n) - h(Y^n|X^n) \\ &= \sum_{i=1}^{n} \Big[h(Y_i|Y^{i-1}) - h(Y_i|X^n, Y^{i-1}) \Big] \\ &= \sum_{i=1}^{n} \Big[h(Y_i|Y^{i-1}) - h(Y_i|X_i) \Big] \\ &\leq \sum_{i=1}^{n} \Big[h(Y_i) - h(Y_i|X_i) \Big] \\ &= \sum_{i=1}^{n} \Big[h(X_i + Z_i) - h(X_i + Z_i|X_i) \Big] \\ &= \sum_{i=1}^{n} \Big[h(X_i + Z_i) - h(Z_i|X_i) \Big] \\ &= \sum_{i=1}^{n} \Big[h(X_i + Z_i) - h(Z_i) \Big]. \end{aligned}$$

We use the maximum differential entropy principle to bound $h(X_i + Z_i)$. Since $\mathsf{E}[X_i] \leq \lambda$ and since $\mathsf{E}[Z_i] = \mu$, we see that $\mathsf{E}[X_i + Z_i] \leq \lambda + \mu$. The distribution that maximizes differential entropy among all nonnegative random variables with mean α is the exponential distribution because it has the form

$$f^*(x) = e^{\gamma_0 + \gamma_1 x} \cdot \mathrm{I}\{x \geq 0\}.$$

The exponential distribution with mean α has density

$$f(x) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}} \cdot \mathrm{I}\{x \geq 0\}$$

and differential entropy $\log(e\alpha)$ (see Exercise 1). Since the logarithm is a nondecreasing function and $\alpha \leq \lambda + \mu$, we obtain $h(X_i + Z_i) \leq \log(e(\lambda + \mu))$. Therefore,

$$\begin{aligned} I(X^n; Y^n) &\leq \sum_{i=1}^{n} \Big[h(X_i + Z_i) - h(Z_i) \Big] \\ &= \sum_{i=1}^{n} \Big[h(X_i + Z_i) - \log(e\mu) \Big] \\ &\leq \sum_{i=1}^{n} \Big[\log(e(\lambda + \mu)) - \log(e\mu) \Big] \\ &= n \log\left(1 + \frac{\lambda}{\mu}\right), \end{aligned}$$

which, together with Fano's inequality, shows that $\mathsf{C} \leq \log\big(1 + \frac{\lambda}{\mu}\big)$.
We only provide a sketch of the achievability part (for the interested reader). In order to achieve equality in the converse, we want the channel output $X + Z$ to have an exponential distribution with mean $\lambda + \mu$. Such an input distribution exists: if X is zero with probability $\frac{\mu}{\mu+\lambda}$ and exponentially distributed with mean $\mu + \lambda$ otherwise, then $X + Z$ has indeed the desired distribution. (This can be verified with the help of characteristic functions since the characteristic function of the sum of two independent random variables is the product of the characteristic functions). In order to avoid the technical issues that arise when the input distribution is a mixture between a discrete and a continuous distribution, we skip the rest of the achievability proof.