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 5 of March 23, 2017

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

Problem 1 — *Waterfilling Example*

The function $f(b_1, \ldots, b_n) = \sum_{i=1}^{n} \sqrt{a_i + b_i}$ is concave in the tuple $(b_1, \ldots, b_n)$ because $x \mapsto \sqrt{a+x}$ is a concave function (its second derivative is $-\frac{1}{4}(a+x)^{-3/2}$, which is always negative). Since f is concave, since its partial derivatives are well-defined, and since the maximization is over all probability vectors $(b_1, \ldots, b_n)$, the Karush–Kuhn–Tucker conditions are necessary and sufficient for $(b_1, \ldots, b_n)$ to maximize f, i.e., $(b_1, \ldots, b_n)$ maximizes f if and only if for some $\lambda \in \mathbb{R}$

$$\frac{\partial f}{\partial b_k} = \lambda \quad \forall k \text{ such that } b_k > 0, \tag{1}$$

$$\frac{\partial f}{\partial b_k} \leq \lambda \quad \forall k \text{ such that } b_k = 0. \tag{2}$$

The partial derivative of f with respect to b_k is

$$\frac{\partial f}{\partial b_k} = \frac{1}{2\sqrt{a_k + b_k}},$$

so we can assume $\lambda \geq 0$ and (1) is equivalent to

$$\begin{aligned} \frac{1}{2\sqrt{a_k + b_k}} &= \lambda, \\ \frac{1}{4(a_k + b_k)} &= \lambda^2, \\ a_k + b_k &= \frac{1}{4\lambda^2}, \\ b_k &= \frac{1}{4\lambda^2} - a_k \quad \forall k \text{ such that } b_k > 0. \end{aligned} \tag{3}$$

Similarly, (2) is equivalent to

$$b_k \geq \frac{1}{4\lambda^2} - a_k \quad \forall k \text{ such that } b_k = 0. \tag{4}$$

We now distinguish between two cases: if $\frac{1}{4\lambda^2} - a_k \leq 0$, then (3) cannot hold (because $b_k > 0$), so (4) has to hold, i.e., $b_k = 0$; and if $\frac{1}{4\lambda^2} - a_k > 0$, then (4) cannot hold (because $b_k = 0$), so (3) has to hold. Substituting $\nu = \frac{1}{4\lambda^2}$, we conclude that

$$b_k = \max\{0, \nu - a_k\} = (\nu - a_k)^+$$

has to hold for all k, where ν has to be chosen such that $\sum_{k=1}^{n} b_k = 1$ is satisfied. Thus, the optimal solution for this maximization problem can be obtained by waterfilling.

Problem 2 *Parallel Channels and Waterfilling*

Because Z_1 and Z_2 are uncorrelated, the Gaussian channels are independent. We know that the optimal power allocation for independent Gaussian channels can be obtained by waterfilling, i.e.,

$$\mathsf{E}_j = (\nu - \sigma_j^2)^+,$$

where ν is chosen such that $\mathsf{E}_1 + \mathsf{E}_2 = \mathsf{E}$. Hence, because $\sigma_1^2 > \sigma_2^2$, the first channel is not used if E is very small. More precisely, the first channel is not used if and only if $\nu \le \sigma_1^2$ holds. This is equivalent to

$$\mathsf{E} = \mathsf{E}_2 = (\nu - \sigma_2^2)^+ = \underbrace{\nu}_{\le \sigma_1^2} - \sigma_2^2 \le \sigma_1^2 - \sigma_2^2.$$

We conclude that both channels are used as soon as

$$\mathsf{E} > \sigma_1^2 - \sigma_2^2.$$

Problem 3 *Dependent Channels and Waterfilling*

a) The optimal power allocation uses waterfilling on the eigenvalues of the covariance matrix, and the inputs will be rotated by the eigenvectors of the covariance matrix. Since the channels are not independent, the eigenvectors are not $(0,1)^\mathsf{T}$ and $(1,0)^\mathsf{T}$, so even for small values of E, both channels will be used.

b) To perform waterfilling on the eigenvalues of the covariance matrix, we need to compute the eigenvalues. Since the eigenvalues are the roots of the characteristic polynomial, we solve

$$\det(\mathsf{K_{ZZ}} - \lambda \mathsf{I}) = (8-\lambda)(5-\lambda) - 2\cdot 2 = \lambda^2 - 13\lambda + 36 = (\lambda - 9)(\lambda - 4) \overset{!}{=} 0,$$

so the eigenvalues are $\lambda_1 = 9$ and $\lambda_2 = 4$. From Problem 2 we know that only the second eigenvalue will be used for $\mathsf{E} \le \lambda_1 - \lambda_2 = 5$. For $\mathsf{E} > 5$, we have $\mathsf{E} = (\nu - 9) + (\nu - 4)$, so $\nu = \frac{\mathsf{E}+13}{2}$, $\mathsf{E}_1 = \nu - 9 = \frac{\mathsf{E}-5}{2}$, and $\mathsf{E}_2 = \nu - 4 = \frac{\mathsf{E}+5}{2}$. The capacity of the parallel channel is therefore

$$\mathsf{C}(\mathsf{E}) = \begin{cases} \frac{1}{2}\log\bigl(1 + \frac{\mathsf{E}}{4}\bigr) & \text{if } 0 \le \mathsf{E} \le 5, \\ \frac{1}{2}\log\Bigl(1 + \frac{(\mathsf{E}-5)/2}{9}\Bigr) + \frac{1}{2}\log\Bigl(1 + \frac{(\mathsf{E}+5)/2}{4}\Bigr) & \text{if } \mathsf{E} > 5. \end{cases}$$

c) To achieve this capacity, design an encoder and decoder pair for the independent parallel Gaussian channel $\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} + \tilde{\mathbf{Z}}$ with total average-power constraint E and covariance matrix

$$\mathsf{\Lambda} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}.$$

Because $\mathsf{K_{ZZ}}$ is positive definite, there exists an orthogonal matrix U be such that $\mathsf{K_{ZZ}} = \mathsf{U}\mathsf{\Lambda}\mathsf{U}^\mathsf{T}$ holds, which is e.g. satisfied by

$$\mathsf{U} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}.$$

(The columns of U are the normalized eigenvectors of $\mathsf{K_{ZZ}}$.) Rotate the channel input and the channel output, i.e., use $\mathbf{X}_i = \mathsf{U}\tilde{\mathbf{X}}_i$ as the channel input and pass $\tilde{\mathbf{Y}}_i = \mathsf{U}^\mathsf{T}\mathbf{Y}_i$ to the decoder. Since U is orthogonal, the rotated channel input still satisfies the total average-power constraint E. From the lecture we know that the channel with covariance matrix $\mathsf{\Lambda}$ has the same capacity as the channel with the covariance matrix $\mathsf{K_{ZZ}}$, so this scheme is indeed capacity-achieving.