Journal of Mathematical Sciences: Advances and Applications
Volume 46, 2017, Pages 17-30
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jmsaa_7100121841

COMPUTATION ON THE CAPACITY OF COMPOUND CHANNEL

LEI LIN

School of Mathematics and Systems Science
Guangdong Polytechnic Normal University
Guangzhou 510665
P. R. China
e-mail: linlei2@mail2.sysu.edu.cn

Abstract

We present an algorithm to compute the capacity of compound channel, which is specified by maximin of mutual informations. The algorithm utilizes the alternate optimization method similar to Arimoto-Blahut algorithm except for the more complex optimization method at the second phase. By transforming the original problem with non-differentiable objective function into a convex optimization problem with differentiable objective function, we obtain the necessary and sufficient conditions for the capacity-achieving input distributions. We also derive both upper and lower bounds of the maximin value, which provide a criterion to terminate any iterative methods. In addition, the convergence of the algorithm is proved. Finally, the numerical results are presented.

1. Introduction

In the early 1948, Shannon [1] introduced an important concept in information theory, channel capacity, which is to specify the asymptotic limit on the maximum rate at which information can be transmitted

2010 Mathematics Subject Classification: 68P30, 94A40, 94A15.
Keywords and phrases: Arimoto-Blahut algorithm, convex optimization, compound channel, maximin of mutual informations.
Received June 10, 2017

reliably over a channel. For the discrete memoryless channel with single input and single output, the capacity can be efficiently computed by the Arimoto-Blahut algorithm, which is an alternate optimization algorithm ([2], [3]).

In this paper, we focus on the computation of the capacity of discrete memoryless compound channel with m inputs and n outputs ([4], [5]), which can be specified by a collection of probability transition matrices $\{Q_k(j|i), 1 \le i \le m, 1 \le j \le n\}_{1 \le k \le K}$. The compound channel can also be considered as a special finite state channel with the channel state unknown initially and kept unchanged during the transmission. As is well known, the capacity of a compound channel is given by

$$C = \max_{p \in \mathbf{p}^m} \min_{1 \le k \le K} \{I(X; Y_k)\}, \tag{1}$$

where $I(X; Y_k)$ is the mutual information between the input and the output corresponding to the k-th channel probability transition matrix $[Q_k(j|i)]$ and

$$\mathbf{p}^m = \{p \in \mathbf{R}^m | p(i) \ge 0, \sum_i p(i) = 1\}, \tag{2}$$

is the set of all probability distributions on the channel input. Here we use X without the subscript k to indicate that the sender must design an encoder that is robust for all possible k, while we use Y_k to indicate that the receiver knows the channel law exactly. The maximin in (1) is also the maximum information rate at which common messages can be broadcasted reliably from one sender to multiple receivers [6]. However, note that the compound channel capacity is, in general, not equal to the minimum of the capacities of the different channels in the family, as the capacity-achieving input distributions may be different for different channels in the family.

The paper is organized as follows. Firstly, the analytical properties of the minimum mutual informations such as continuity, concavity and differentiability are analyzed. The necessary and sufficient conditions for the capacity-achieving input distributions are revealed. Both lower bounds and upper bounds are derived, which provide a criterion to terminate any iterative algorithms when the capacity has been approximated sufficiently closely. Secondly, an iterative method to compute the maximin value of K mutual informations is derived. Thirdly, the convergence of the proposed algorithm is proved. Finally, some numerical results are presented.

2. Necessary and Sufficient Conditions

Consider the compound channel described above. Let $Q_k = [Q_k(j|i)]$ be the k-th probability transition matrix. For any given distribution $p \in \mathbf{p}^m$, define

$$f(p) = \min_{1 \le k \le K} \{I(p, Q_k)\},$$

where

$$I(p, Q_k) \triangleq I(X; Y_k) = \sum_{i, j} p(i) Q_k(j|i) \log \frac{Q_k(j|i)}{q_k(j)}, \tag{3}$$

and

$$q_k(j) = \sum_i p(i) Q_k(j|i) \quad \text{for} \quad 1 \le k \le K. \tag{4}$$

Proposition 1. *The function* $f(p)$ *is continuous and concave over* $\mathbf{p}^m$.

Proof. The continuity is obvious, while the concavity follows from the concavity of $I(p, Q_k)$ and the fact that minimum operation preserves concavity ([7], p. 80). □

Hence, there exists p^* such that $f(p^*) = \max_{p \in \mathbf{p}^m} f(p)$. However, since $f(p)$ may be non-differentiable, the Karush-Kuhn-Tucker (KKT) conditions may not exist. Fortunately, we can verify that maximizing $f(p)$ is equivalent to

$$\max_{p,r}\{r\}$$

$$s.t.\begin{cases} r & \leq I(p,\, Q_k) \text{ for } 1 \leq k \leq K; \\ \sum_i p(i) & = 1; \\ p(i) & \geq 0, \end{cases} \tag{5}$$

which is a convex optimization problem since $r - I(p,\, Q_k)$ for $1 \leq k \leq K$ are convex in p [8] and r is linear. By defining

$$L(\lambda,\, \nu) = r + \lambda(\sum p(i) - 1) + \sum_k \nu_k (I(p,\, Q_k) - r),$$

where $\lambda,\, \nu_k$ are Lagrange multipliers, we have the following KKT conditions:

$$\frac{\partial L(\lambda,\, \nu)}{\partial p(i)}|_{p(i)=p^*(i)} = 0 \text{ for } p^*(i) > 0;$$

$$\frac{\partial L(\lambda,\, \nu)}{\partial p(i)}|_{p(i)=p^*(i)} \leq 0 \text{ for } p^*(i) = 0;$$

$$\frac{\partial L(\lambda,\, \nu)}{\partial r}|_{r=r^*} = 1 - \sum_k \nu_k = 0;$$

$$\nu_k(I(p^*,\, Q_k) - r^*) = 0;\ \nu_k \geq 0. \tag{6}$$

From (6), we obtain

$$\sum_k \nu_k T_k(i)_{p^*} = \log e - \lambda \quad \text{for } p^*(i) > 0;$$

$$\sum_k \nu_k T_k(i)_{p^*} \leq \log e - \lambda \quad \text{for } p^*(i) = 0, \tag{7}$$

where

$$T_k(i)_{p^*} = \sum_j Q_k(j|i) \log \frac{Q_k(j|i)}{\sum_i Q_k(j|i) p^*(i)}. \tag{8}$$

Multiplying both sides of (7) by $p^*(i)$ and summing over i, it can be seen that

$$\sum_k \nu_k I(p^*, Q_k) = \log e - \lambda. \tag{9}$$

Subtracting $r^* = \max_{p \in \mathbf{p}^m} f(p)$ from both sides of (9) and utilizing $\nu_k(I(p^*, Q_k) - r^*) = 0,$ we obtain

$$r^* = \log e - \lambda, \tag{10}$$

and Theorem 1.

Theorem 1. *For an input probability vector* p^* *satisfying that* $C = f(p^*) = \max_{p \in \mathbf{p}^m} f(p),$ *there exists a vector* $\nu = (\nu_1, \nu_2, \dots, \nu_k)$ $(\nu_k \geq 0)$ *satisfying* $\sum_k \nu_k = 1$ *such that*

$$\sum_k \nu_k T_k(i)_{p^*} = C \quad \text{for } p^*(i) > 0;$$

$$\sum_k \nu_k T_k(i)_{p^*} \leq C \quad \text{for } p^*(i) = 0. \tag{11}$$

Theorem 2. *For any probability vector* p *and nonnegative vector* $\nu = (\nu_1, \nu_2, \dots, \nu_k)$ $(\nu_k \geq 0)$ *satisfying* $\sum_k \nu_k = 1,$ *the following hold*:

$$I_L(p) \leq C \leq I_U(p), \tag{12}$$

where

$$I_L(p) = \min_k \{I(p, Q_k)\}, \tag{13}$$

$$I_U(p) = \min_\nu \max_{1 \leq i \leq m} \{\sum_k \nu_k T_k(i)_p\}. \tag{14}$$

Proof. The lower bound is obvious, while the proof of the upper bound is similar to Problem 4.17 in ([9], p. 524). Actually, for any ν $(\nu_k \geq 0)$ satisfying $\sum_k \nu_k = 1,$ it can be seen that

$$\max_{1 \leq i \leq m} \{\sum_k \nu_k T_k(i)_p\}$$

$$= \sum p^*(i) \max_{1\le i\le m}\{\sum_k \nu_k T_k(i)_p\}$$

$$\ge \sum p^*(i)(\sum_k \nu_k T_k(i)_p)$$

$$= \sum_k \nu_k(\sum p^*(i) T_k(i)_p)$$

$$\ge \sum_k \nu_k(\sum p^*(i) T_k(i)_{p^*}) = C. \tag{15}$$

□

Remark. If p^* achieve C, from Theorem 1 and Theorem 2 we get $I_L(p^*) = I_U(p^*)$.

3. The Iterative Optimizing Algorithm

The computation of C will be derived from the following theorem:

Theorem 3. *Suppose that the channel transition matrices* $Q_1, \ldots, Q_K$ *are* $m \times n$. *For any* $n \times m$ *transition matrices* $B_1, \ldots, B_K$, *let*

$$F(p, Q_k, B_k) = \sum_i \sum_j p(i) Q_k(j|i) \log \frac{B_k(i|j)}{p(i)},$$

$$J(p, B) = \min_{1\le k\le K}\{F(p, Q_k, B_k)\}.$$

Then the following is true:

(1)

$$C = \max_p \max_{B_1,\ldots,B_K} J(p, B). \tag{16}$$

(2) *For fixed* p, $J(p, B)$ *is maximized by*

$$B_k^* = \frac{p(i) Q_k(j|i)}{\sum_i p(i) Q_k(j|i)}, \ 1 \le k \le K. \tag{17}$$

(3) *For fixed* $B_1, \ldots, B_K$, $J(p, B)$ *is maximized by*

$$p^*(i) = \frac{\exp(\sum_k \mu_k \sum_j Q_k(j|i) \log B_k(i|j))}{\sum_i \exp(\sum_k \mu_k \sum_j Q_k(j|i) \log B_k(i|j))}, \tag{18}$$

where

$$\mu_k \geq 0, \sum_k \mu_k = 1,$$

and $\mu = (\mu_1, \ldots, \mu_K) = e_l$ *such that* p^* *satisfies that* $F_l(p^*) \leq F_j(p^*)$ *for all* $j \neq l$ *or* $\mu = \mu^0$ *such that* $F_l(p^*) = F_m(p^*)$ *for all* $\mu_l^0, \mu_m^0 \neq 0$ *and* $F_l(p^*) \leq F_j(p^*)$ *for all* $\mu_j^0 = 0$. *Hereafter, for fixed* B_k, *we use* $F_k(p) \triangleq F(p, Q_k, B_k)$ *for convenience and* e_l *states K-dimensional unit vector, that is, the l-th component of* e_l *is* 1 *and the rest is* 0.

Proof. (1) From Theorem 1 in [3], it can be seen that $F(p, Q_k, B_k) \leq F(p, Q_k, B_k^*)$ for fixed p. So it is obvious that

$$J(p, B) \leq J(p, B^*), \tag{19}$$

which means that $f(p) = \max_{B_1, \ldots, B_K} J(p, B)$.

(2) This fact is an immediate consequence of (19).

(3) Since $J(p, B)$ may be non-differentiable over p, Lagrange multiplier method can not be used. Fortunately, we can still transform the problem of maximizing $J(p, B)$ into an equivalent optimization problem

$$\max_{p, r}\{r\}$$

$$s.t. \begin{cases} r - F(p, Q_k, B_k) \leq 0 & \text{for } 1 \leq k \leq K, \\ \sum_i p(i) - 1 = 0. \end{cases} \tag{20}$$

Here, we ignore the constraint $p(i) \geq 0$ temporarily. Similar to Problem (5), it can be verified that Problem (20) is a convex optimization problem. By defining

$$L(\lambda, \mu) = r + \lambda(\sum p(i) - 1) + \sum \mu_k (F_k(p) - r),$$

the following KKT conditions:

$$\begin{aligned} \frac{\partial L(\lambda, \mu)}{\partial r}\Big|_{r=r^*} &= 0; \\ \frac{\partial L(\lambda, \mu)}{\partial p(i)}\Big|_{p(i)=p^*(i)} &= 0; \\ \mu_k (F_k(p^*) - r^*) &= 0 \text{ for } 1 \leq k \leq K; \\ \mu_k &\geq 0 \text{ for } 1 \leq k \leq K, \end{aligned} \tag{21}$$

are true. The Equation (18) can be easily derived from the second constraint of (21) and the condition $\sum_i p^*(i) = 1$. Solving from the first equation of the above KKT conditions, we get $\sum_k \mu_k = 1$. In addition, for μ, there are two kinds of cases. The one is $\mu = e_l$, that is, $\mu_l = 1$ and $\mu_j = 0$ for all $j \neq l$. From the condition $\mu_l(F_l(p^*) - r^*) = 0$, it is obvious that $r^* = F_l(p^*)$. Hence p^* should satisfy $r^* = F_l(p^*) \leq F_j(p^*)$ for all $j \neq l$. The other one is $\mu = \mu^0$. For all $\mu_l^0, \mu_m^0 \neq 0$ it is obvious that $F_l(p^*) = r^*$ and $F_m(p^*) = r^*$, which means that $F_l(p^*) = F_m(p^*)$. Meanwhile, for all $\mu_j^0 = 0$, $F_l(p^*) = r^* \leq F_j(p^*)$ should be satisfied. Moreover, substituting (18) into (16) it can be seen that

$$r^* = \log \sum_i \exp(c(i)). \tag{22}$$

Notice that in any case p^* is always positive so that the constraint $p^*(i) \geq 0$ is not operative.

□

Algorithm 1. Maximin of Mutual Information Algorithm

(1) **Initialization:** Choose an arbitrary probability vector $p(p(i) > 0)$ and a very small number ε.

(2) **Iteration:**

Step 2.1. Compute $T_k(i)_p$ and $I(p, Q_k)$ for $1 \leq k \leq K$; if $I_U(p) - I_L(p) < \varepsilon$, output $C = I_L(p)$ and exit the iterations; otherwise, compute B_k^* for $1 \leq k \leq K$ according to (17).

Step 2.2. Fix $B_k = B_k^*$ and find p^* according to (18) by identifying the true solution of μ :

Case 1. Let $\mu = e_l$ and compute p^* and $F_k(p^*)$ for $1 \leq k \leq K$; if $F_l(p^*) \leq F_j(p^*)$ for all $j \neq l$, set $p = p^*$ and go to Step 2.1.

Case 2. If Case 1 does not occur, search μ^0 such that p^* satisfies $F_l(p^*) = F_m(p^*)$ for all $\mu_l^0, \mu_m^0 \neq 0$ and $F_l(p^*) \leq F_j(p^*)$ for all $\mu_j^0 = 0$. Set $p = p^*$ and go to Step 2.1.

Remark. In this algorithm, we utilize *simplex method* to compute $I_U(p)$ and *Newton method* to search μ^0.

4. Proof of the Convergence of the Algorithm

In this section, we use a method similar to that of [2] and [10] to prove the convergence of the algorithm. For simplicity, let

$$C(t, t) = \max_{B_1, B_2} J(p^t, B),$$

$$C(t+1, t) = \max_{p} J(p, B^t), \tag{23}$$

where p^t stands for the input probability vector in the t-th iteration and $B_k^t(i|j) = \dfrac{p^t(i)Q_k(j|i)}{\sum_i p^t(i)Q_k(j|i)}.$

Theorem 4. *Let* $p^1 \in \mathbf{p}^m$ *with all components strictly positive. Then the values* $C(t, t)$ *converge monotonically from below to C as* $t \to \infty$.

Proof. According to the Algorithm 1, the following relations are obvious:

$$C(t, t) \le C(t+1, t) \le C(t+1, t+1) \le \ldots \le C. \tag{24}$$

From (22), it is obvious that $C(t+1, t) = \log \sum_i c^t(i)$, where

$$c^t(i) = \exp(\sum_k \mu_k \sum_j Q_k(j|i) \log B_k^t(i|j)).$$

Now suppose that p^* is an optimal probability vector, then

$$\sum_i p^*(i) \log \frac{p^{t+1}(i)}{p^t(i)}$$

$$= \sum_i p^*(i) \log(\frac{1}{p^t(i)} \cdot \frac{c^t(i)}{\sum_i c^t(i)})$$

$$= -C(t+1, t) + \sum_k \mu_k \sum_{i,j} p^*(i) Q_k(j|i) \log \frac{Q_k(j|i)}{q_k^t(j)}$$

$$= -C(t+1, t) + \sum_k \mu_k \sum_{i,j} p^*(i) Q_k(j|i) (\log \frac{Q_k(j|i)}{q_k^*(j)} + \log \frac{q_k^*(j)}{q_k^t(j)})$$

$$\overset{(a)}{\ge} C - C(t+1, t) + \sum_k \mu_k \sum_j q_k^*(j) \log \frac{q_k^*(j)}{q_k^t(j)}, \tag{25}$$

where (a) follows from $C \le I(p^*, Q_k)$ for all $1 \le k \le K$ and $\sum_k \mu_k = 1$. Because the last term of (25) are nonnegative (both μ_k and the relative entropies are nonnegative), for any t,

$$C - C(t+1, t) \le \sum_i p^*(i) \log(p^{t+1}(i) / p^t(i)), \tag{26}$$

is true. Hence, summing up (26) over t from 1 to N,

$$\sum_{t=1}^{N}[C - C(t+1, t)] \leq \sum_i p^*(i)\log(p^N(i)/p^1(i))$$

$$\leq \sum_i p^*(i)\log(p^*(i)/p^1(i)), \tag{27}$$

for any integer $N \geq 1$. Note that $\sum_i p^*(i)\log(p^*(i)/p^1(i))$ is a finite constant. Now assume that the limit of $C(t+1, t)$ is not equal to C. Since $C - C(t+1, t)$ is nonnegative and non-increasing with increasing t, the left hand of (27) would go to infinity, which is a contradiction. Therefore, we have

$$\lim_{t\to\infty} C(t+1, t) = \lim_{t\to\infty} C(t, t) = C.$$

$\square$

5. Numerical Results

In this section, we present some examples in the cases of $K = 2$ and $K = 3$.

Example 1. Consider four pairs of channels with respective probability transition matrices as below:

A.

$$Q_1 = \begin{pmatrix} 0.53 & 0.07 & 0.40 \\ 0.41 & 0.36 & 0.23 \\ 0.45 & 0.50 & 0.05 \end{pmatrix}, \; Q_2 = \begin{pmatrix} 0.37 & 0.59 & 0.04 \\ 0.20 & 0.12 & 0.68 \\ 0.37 & 0.32 & 0.31 \end{pmatrix};$$

B.

$$Q_1 = \begin{pmatrix} 0.18 & 0.31 & 0.51 \\ 0.45 & 0.24 & 0.31 \\ 0.65 & 0.20 & 0.15 \end{pmatrix}, \; Q_2 = \begin{pmatrix} 0.13 & 0.48 & 0.39 \\ 0.21 & 0.50 & 0.29 \\ 0.15 & 0.08 & 0.77 \end{pmatrix};$$

C.

$$Q_1 = \begin{pmatrix} 0.28 & 0.48 & 0.24 \\ 0.01 & 0.43 & 0.56 \\ 0.27 & 0.39 & 0.34 \end{pmatrix}, Q_2 = \begin{pmatrix} 0.09 & 0.24 & 0.67 \\ 0.51 & 0.05 & 0.44 \\ 0.34 & 0.33 & 0.33 \end{pmatrix}, Q_3 = \begin{pmatrix} 0.09 & 0.40 & 0.51 \\ 0.34 & 0.55 & 0.11 \\ 0.51 & 0.43 & 0.06 \end{pmatrix};$$

D.

$$Q_1 = \begin{pmatrix} 0.04 & 0.42 & 0.54 \\ 0.35 & 0.38 & 0.27 \\ 0.21 & 0.68 & 0.11 \end{pmatrix}, Q_2 = \begin{pmatrix} 0.37 & 0.08 & 0.55 \\ 0.05 & 0.50 & 0.45 \\ 0.35 & 0.10 & 0.55 \end{pmatrix}, Q_3 = \begin{pmatrix} 0.46 & 0.07 & 0.47 \\ 0.35 & 0.43 & 0.22 \\ 0.48 & 0.17 & 0.35 \end{pmatrix}.$$

In our computation, we set $\varepsilon = 10^{-6}$. The convergence of the algorithm can be seen from Figure 1 and Figure 2, in which C_A, C_B, C_C, and C_D stand for the maximin values and $I_L(*)$, $I_U(*)$ denote the lower and upper bounds of $*$, respectively.

6. Conclusion

In this paper, we have analyzed the analytical properties of the minimum mutual informations such as continuity, concavity, and differentiability. The necessary and sufficient conditions for the capacity-achieving input distributions have been characterized. Both lower and upper bounds are derived, which provide a criterion to terminate any iterative algorithms when the capacity has been approximated sufficiently closely. Based on these, an iterative method to compute the maximin value of K mutual informations has been derived. Finally, the convergence of the algorithm was proved.

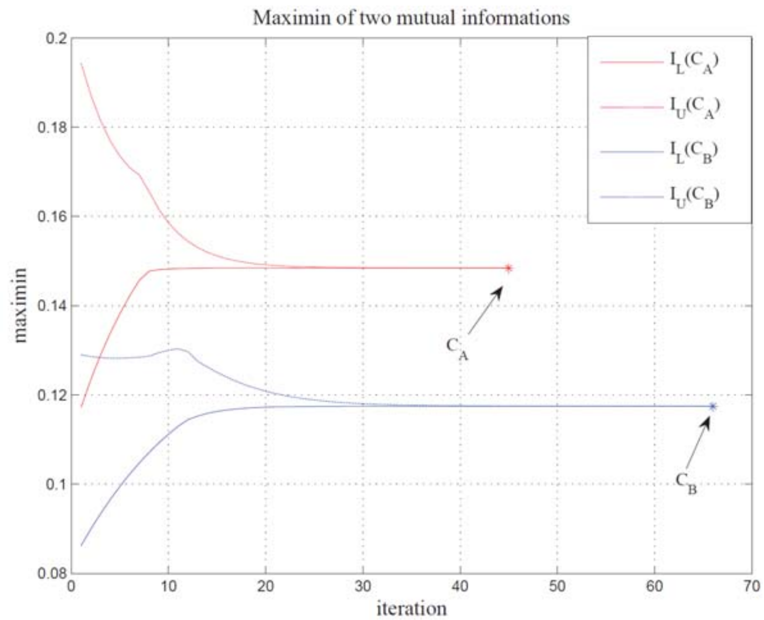

Figure 1. In the case of $K = 2$.

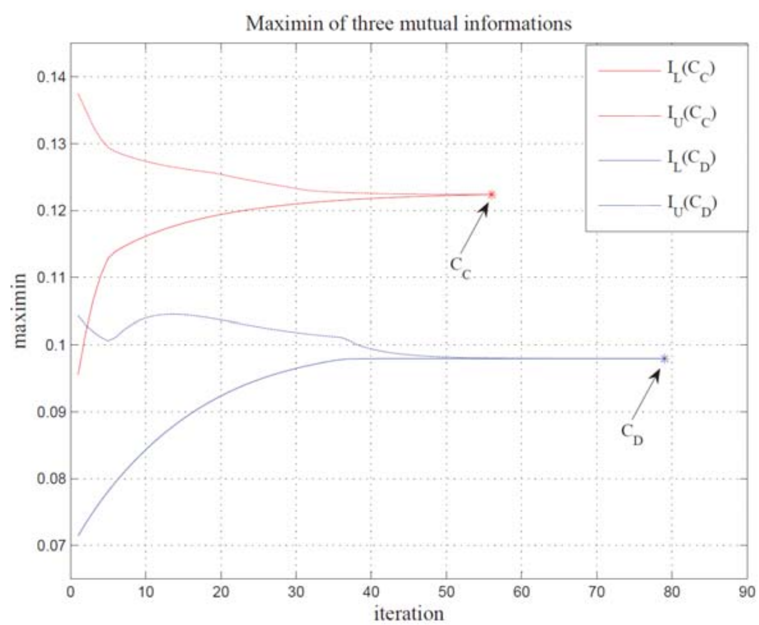

Figure 2. In the case of $K = 3$.

References

[1] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948), 379-426; 623-656.

[2] S. Arimoto, An algorithm for computing the capacity of arbitrary discrete memoryless channels, IEEE Trans. Inform. Theory IT-18(1) (1972), 14-20.

[3] R. E. Blahut, Computation of channel capacity and rate distortion functions, IEEE Trans. Inform. Theory IT-18(4) (1972), 460-473.

[4] D. Blackwell, L. Breiman and A. J. Thomasian, The capacity of a class of channels, Ann. Math. Statist. 30(4) (1959), 1229-1241.

[5] J. Wolfowitz, Coding Theorems of Information Theory, Springer-Verlag, Berlin, 1964.

[6] T. M. Cover, Broadcast channels, IEEE Trans. Inform. Theory IT-18(1) (1972), 2-14.

[7] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.

[8] T. M. Cover and J. A. Thomas, Elements of Information Theory, John Wiley and Sons, Inc., New York, 1991.

[9] R. G. Gallager, Information Theory and Reliable Communication, John Wiley and Sons, Inc., New York, 1968.

[10] I. Csiszár, On the computation for rate-distortion functions, IEEE Trans. Inform. Theory 20(1) (1974), 122-124.

■