$(R_2')_{n+1}$ is defined by

$$\frac{I_{2m}+\epsilon_x+(R_2')_{n+1}}{(R_2)_n-\epsilon_n}=\frac{I_{1m}+\epsilon_x}{(R_1)_n-\epsilon_n}.$$

Since $(R_2')_{n+1}$ tends to $(R_2^*)_{n+1}$ if ϵ_x and ϵ_n tend to zero, it follows that the rate of $C_t^{(n)}$ tends to $\boldsymbol{R}_n$ defined in (11).

We split the information into p pairs of $\{\lfloor p(\overline{B}+2\epsilon)\rfloor, \lfloor p(\overline{H}+2\epsilon)\rfloor\}$ bit and a_p-D pairs of $\{0, \lfloor p(q_m-\delta)(R_2')_{n+1}\rfloor\}$ bit. The first part is transmitted in packets $\mathscr{P}_i$, with the Tolhuizen scheme. In $\mathscr{P}_i$, $D<i\le a_p$, we encode the m states with $\{\lceil (I_{1m}+\epsilon_x)(q_m+\delta)p\rceil, \lceil (I_{2m}+\epsilon_x)(q_m+\delta)p\rceil\}$ symbols. User 2 adds $\lfloor p(q_m-\delta)(R_2')_{n+1}\rfloor$ new information symbols to these and fills up until he has $\lceil (I_{2m}+\epsilon_x+(R_2')_{n+1})(q_m+\delta)p\rceil$. All of them are transmitted together, using the (asymmetrical) code $C_{(t')}^{(n)}$ with

$$t' := \left\lceil \frac{(I_{1m}+\epsilon_x)(q_m+\delta)p}{(R_1)_n-\epsilon_n}\right\rceil$$

$$= \left\lceil \frac{(I_{2m}+\epsilon_x+(R_2')_{n+1})(q_m+\delta)p}{(R_2)_n-\epsilon_n}\right\rceil.$$

Now the proof can be finished in the same way as in Theorem 5.

One final remark concerns the convergence of the sequence $\boldsymbol{R}_n$. In the symmetrical case the points (R_n, R_n) are all on the line $R_1=R_2$, and their distance to the origin increases monotonically. In the asymmetrical case, however, the points $((R_1)_n,(R_2)_n)$ are not on a straight line. Here we must show that the sequence $(\eta_n)_{n\in\mathbb{N}}$ is monotonically decreasing (or increasing, depending on $\eta_1>1$ or $\eta_1<1$); hence $(R_1)_n$ is increasing (decreasing) and has a limit R_1. Although $(R_2)_n$ is not monotonic, it can be shown that this sequence converges, too. We find that the limit (R_1, R_2) satisfies (1), which is what we had to show.

Acknowledgment

The author wishes to thank Prof. J. P. M. Schalkwijk, P. J. Hazewindus, and F. M. J. Willems for their contributions to this research.

References

[1] R. Ash, *Information Theory*. New York: Interscience, 1965.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kiado, 1981.

[3] J. P. M. Schalkwijk, "The binary multiplying channel—A coding scheme that operates beyond Shannon's inner bound region," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 107–110, Jan. 1982.

[4] ——, "On an extension of an achievable rate region for the binary multiplying channel," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 445–448, May 1983.

[5] J. P. M. Schalkwijk, J. E. Rooyackers, and B. J. M. Smeets, "Generalized Shannon strategies for the binary multiplying channel," in *Proc. 4th Benelux Symp. Information Theory*, Haasrode, Belgium, May 1983, pp. 207–210.

[6] J. P. M. Schalkwijk, "The capacity region of the binary multiplying channel—A converse," in *Proc. NATO-ASI Performance Limits in Communication Theory and Practice*, Il Ciocco, Italy, July 1986.

[7] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, 1961, pp. 611–644. Reprinted in *Key Papers in the Development of Information Theory*, D. Slepian, Ed. New York: IEEE, 1974, pp. 339–372.

[8] D. Slepian and J. K. Wolf, "Noiseless coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[9] L. Tolhuizen, "Discrete coding for the B.M.C., based on Schalkwijk's strategy," in *Proc. 6th Benelux Symp. Information Theory*, Mierlo, The Netherlands, May 1985, pp. 207–210.

A New Geometric Capacity Characterization of a Discrete Memoryless Channel

KENJI NAKAGAWA, ASSOCIATE MEMBER, IEEE, AND FUMIO KANAYA, MEMBER, IEEE

***Abstract*—A novel geometrical characterization of capacity of a discrete memoryless channel is proposed according to Csiszàr's theorem, which represents the capacity using the Kullback–Leibler discrimination information. As a result, a new geometrical capacity computing method is given.**

I. Introduction

Computational methods for the capacity of a discrete memoryless channel proposed to date may be divided into direct computing methods [1]–[3] and sequential computing methods [4]–[6]. In the direct computing method the capacity C is calculated according to linear equation theory and the Kuhn–Tucker condition of convex programming. In the sequential computing method a sequence $\{p^n\}_{n=0}^{\infty}$ starting at an appropriate initial probability distribution p^0 is defined, and it is shown that the sequence converges to a probability distribution attaining C. Furthermore, the convergence speed is estimated.

The present correspondence belongs to the direct computing category. Since Muroga's method [1] is based on the linear equation theory, it is not easy to understand the relation between the row probability vectors of a channel matrix and the probability vector attaining C. After characterizing C geometrically, we present a new computational method based on Csiszàr's theorem which describes C as the solution of a minimax problem using the Kullback–Leibler discrimination information.

II. Definitions

Let $X=\{x_1,\cdots,x_m\}$ and $Y=\{y_1,\cdots,y_n\}$ be the input and output alphabets, respectively. Let $Q(y_j|x_i)$ be the conditional probability of y_j when x_i is given. We treat a discrete memoryless channel whose channel matrix is

$$Q=\left(Q(y_j|x_i)\right), \qquad i=1,\cdots,m,\ j=1,\cdots,n.$$

(The (i,j) entry of Q is $Q(y_j|x_i)$.) If there is no ambiguity, we also call Q itself a channel. Let

$$Q^i=(Q(y_1|x_i),\cdots,Q(y_n|x_i))$$

be the probability distribution on Y when x_i is transmitted. We denote a probability distribution on X by $p(x_i)$ and that on Y corresponding to $p(x_i)$ by $q(y_j)$; i.e.,

$$q(y_j)=\sum_{i=1}^{m} p(x_i)Q(y_j|x_i), \qquad \text{or } q=pQ.$$

We define two sets of probability distributions:

$$\Delta^n=\left\{(\alpha_1,\cdots,\alpha_n)\,\middle|\, \sum_{j=1}^{n}\alpha_j=1,\ \alpha_j>0\ (j=1,\cdots,n)\right\}$$

$$\overline{\Delta}^n=\left\{(\alpha_1,\cdots,\alpha_n)\,\middle|\, \sum_{j=1}^{n}\alpha_j=1,\ \alpha_j\ge 0\ (j=1,\cdots,n)\right\}.$$

We define the Kullback–Leibler information by

$$D(q^1\|q^2)\triangleq\sum_{j=1}^{n} q_j^1\log\left(q_j^1/q_j^2\right)$$

Manuscript received February 2, 1987; revised May 18, 1987.
The authors are with NTT Laboratories, 1-2356, Take, Yokosuka-shi, Kanagawa-ken, 238-03 Japan.
IEEE Log Number 8820328.

0018-9448/88/0300-0318$01.00 ©1988 IEEE

where $q^1=(q_1^1,\cdots,q_n^1)$, $q^2=(q_1^2,\cdots,q_n^2)\in\bar{\Delta}_n$. It is well-known that $D(q^1\|q^2)$ has the following properties:

1) $D(q^1\|q^2)\geq 0$: equality holds if and only if $q^1=q^2$;
2) in general, neither $D(q^1\|q^2)=D(q^2\|q^1)$ nor $D(q^1\|q^2)+D(q^2\|q^3)\geq D(q^1\|q^3)$ holds;
3) convexity: if $q=(1-\lambda)q^1+\lambda q^2$ and $\bar{q}=(1-\lambda)\bar{q}^1+\lambda\bar{q}^2 (0\leq\lambda\leq 1)$, then $D(q\|\bar{q})\leq(1-\lambda)D(q^1\|\bar{q}^1)+\lambda D(q^2\|\bar{q}^2)$.

For k points $q^1,\cdots,q^k$ in $\mathbb{R}^n$, let $C(q^1,\cdots,q^k)$ be their convex hull and $q^1\smile\cdots\smile q^k$ be the minimum linear subspace they span.

III. Lemmas and Known Theorems

Lemma 1: Any $q\in C(q^1,\cdots,q^k)$ is contained in a simplex with vertices belonging to $\{q^1,\cdots,q^k\}$ (see [7, p. 15]).

Lemma 2: If three distinct points $q^1,q^2,q^3\in\Delta^n$ are located on a line in this order, the following inequality holds:

$$D(q^1\|q^2)+D(q^2\|q^3)<D(q^1\|q^3).$$

Proof: From this condition, a positive number α exists such that $q^1-q^2=\alpha(q^2-q^3)$. Thus if $q^i=(q_1^i,\cdots,q_n^i)$ $(i=1,2,3)$ we have

$$\begin{aligned}D(q^1\|q^3)-D(q^1\|q^2)-D(q^2\|q^3)&=\sum_{j=1}^{n}(q_j^1-q_j^2)\log(q_j^2/q_j^3)\\&=\alpha\sum_{j=1}^{n}(q_j^2-q_j^3)\log(q_j^2/q_j^3)\\&=\alpha\{D(q^2\|q^3)+D(q^3\|q^2)\}>0.\end{aligned}$$

Q.E.D.

Lemma 3: For $q^1,q^2\in\Delta^n (q^1\neq q^2)$, $\lambda\geq 0$,

$$f(\lambda)=D(q^1\|(1-\lambda)q^1+\lambda q^2)$$

and

$$g(\lambda)=D((1-\lambda)q^1+\lambda q^2\|q^1)$$

are both increasing functions of λ.

Proof: Lemma 3 is trivial by Lemma 2. Q.E.D.

Lemma 4: Let V be a closed convex subset of $\bar{\Delta}^n$. For $q\in\bar{\Delta}^n$, if there is some $r\in V$ such that $D(r\|q)<\infty$, then a unique $r^0\in V$ exists minimizing

$$D(r\|q),\qquad (r\in V)$$

(see [8, p. 59]).

We call this r^0 the projection of q onto V, and denote it by

$$r^0=\mathrm{pr.}_V(q).$$

We define a linear set E in $\bar{\Delta}^n$ by

$$E=\left\{q=(q_1,\cdots,q_n)\in\bar{\Delta}^n\,\middle|\,\sum_{j=1}^{n}a_{kj}q_j=b_k,\ a_{kj},b_k \text{ are constants and } k \text{ ranges over a finite index set}\right\}.$$

For example, a straight line connecting two probability distributions in $\bar{\Delta}^n$, a plane determined by three distributions, and so forth, are linear sets.

Pythagoras Theorem: Let E be a linear set in $\bar{\Delta}^n$. For $r\in\bar{\Delta}^n$, let $q=\mathrm{pr.}_E(r)$. Then for any $s\in E$,

$$D(s\|q)+D(q\|r)=D(s\|r)$$

holds (see [8, p. 59]).

Theorem (Csiszàr): The capacity of a discrete memoryless channel Q is equal to

$$C=\min_{q\in\bar{\Delta}^n}\max_{1\leq i\leq m}D(Q^i\|q).$$

Furthermore, $q^0\in\bar{\Delta}^n$ which achieves the minimum is unique and $q^0=p^0Q$, where p^0 is any probability distribution that maximizes the mutual information $I(p,Q)$ (see [8, p. 142, 147]).

Theorem (Kuhn–Tucker): An input probability distribution p maximizes $I(p,Q)$ if and only if a constant C exists satisfying

$$D(Q^i\|pQ)\begin{cases}=C, & \text{if } p(x_i)>0\\ \leq C, & \text{if } p(x_i)=0\end{cases}$$

(see [9, p. 91]).

IV. Geometric Characterization of Capacity

Without loss of generality we let $Q^1,\cdots,Q^k$ be the extreme points of $V=C(Q^1,\cdots,Q^m)$.

Theorem 1: If a probability distribution q^0 satisfying

$$D(Q^i\|q^0)=C\ (\text{constant for } i=1,\cdots,k)$$

is in V, C is the capacity of the channel Q.

Proof: From Lemma 1, a subset of $\{Q^1,\cdots,Q^k\}$ exists, say, $\{Q^1,\cdots,Q^r\}$, such that

$$q^0=\sum_{i=1}^{r}\alpha_iQ^i$$

$$\sum_{i=1}^{r}\alpha_i=1,\ \alpha_i>0,\ i=1,\cdots,r.$$

Defining an input probability distribution p^0 by

$$p^0(x_i)=\begin{cases}\alpha_i, & i=1,\cdots,r\\ 0, & i=r+1,\cdots,k,\end{cases}$$

we obtain $q^0=p^0Q$. From the theorem assumption, for $i=1,\cdots,r$, we have

$$D(Q^i\|p^0Q)=D(Q^i\|q^0)=C.$$

On the other hand, for $i=r+1,\cdots,m$, from Lemma 1 a subset of $\{Q^1,\cdots,Q^k\}$ exists, say, $\{Q^1,\cdots,Q^s\}$, such that

$$Q^i=\sum_{h=1}^{s}\beta_hQ^h$$

$$\sum_{h=1}^{s}\beta_h=1,\ \beta_h>0,\qquad h=1,\cdots,s.$$

Therefore, using the convexity of D, we have

$$\begin{aligned}D(Q^i\|q^0)&=D\left(\sum_{h=1}^{s}\beta_hQ^h\|q^0\right)\\&\leq\sum_{h=1}^{s}\beta_hD(Q^h\|q^0)\\&=C,\qquad i=r+1,\cdots,m.\end{aligned}$$

Consequently, p^0 satisfies the Kuhn–Tucker condition, and q^0 attains the capacity. These results are independent of the choice of points representing q^0 and Q^i as convex linear combinations. Q.E.D.

A distribution $q\in\bar{\Delta}^n$ satisfying

$$D(Q^1\|q)=\cdots=D(Q^k\|q)$$

is called an equidistant point from $Q^1,\cdots,Q^k$. According to the previous theorem, we find that when we try to compute C, it is not necessary to consider the probability vectors that are not the extreme points of V.

Even if the point equidistant from the extreme points of V exists, it may not be in V. Since the unique q^0 that attains the capacity must be in V, any equidistant point outside V does not attain it. The following theorem specifies the relation between an equidistant point $q \notin V$ and the capacity-achieving point q^0.

Theorem 2: If q is equidistant from $Q^1, \cdots, Q^k$, then $q^0 = \mathrm{pr.}_V(q)$ achieves the capacity.

Proof: By Lemma 1 a subset of $\{Q^1, \cdots, Q^k\}$ exists, say, $\{Q^1, \cdots, Q^t\}$, such that

$$q^0 = \sum_{i=1}^{t} \gamma_i Q^i$$

$$\sum_{i=1}^{t} \gamma_i = 1, \ \gamma_i > 0, \qquad i = 1, \cdots, t.$$

Thus denoting $E = Q^1 \smile \cdots \smile Q^t$ we have

$$q^0 = \mathrm{pr.}_V(q) = \mathrm{pr.}_E(q).$$

The Pythagoras theorem shows that

$$D(Q^i \| q^0) = C \text{ (constant for } i = 1, \cdots, t).$$

If we can show

$$D(Q^i \| q^0) \le C, \qquad i = t+1, \cdots, k,$$

the proof is completed. Let L_i be the line connecting Q^i $(i = t+1, \cdots, k)$ and q^0. First, we show that on the line L_i, $q^i = \mathrm{pr.}_{L_i}(q)$, q^0, and Q^i are located in this order. Suppose q^i is between q^0 and Q^i. Since the line segment connecting q^0 and Q^i is included in V, we have

$$D(q^0 \| q) \le D(q^i \| q) \tag{1}$$

because of the minimality of $D(q^0 \| q)$. On the other hand, by the Pythagoras theorem,

$$D(q^0 \| q^i) + D(q^i \| q) = D(q^0 \| q)$$

holds, and therefore,

$$D(q^i \| q) < D(q^0 \| q).$$

However, this contradicts (1).

Next, suppose q^0, Q^i, and q^i are located in this order. By the Pythagoras theorem and Lemma 3, we have

$$\begin{aligned} D(q^0 \| q) &= D(q^0 \| q^i) + D(q^i \| q) \\ &> D(Q^i \| q^i) + D(q^i \| q) \\ &= D(Q^i \| q). \end{aligned}$$

This also contradicts the minimality of $D(q^0 \| q)$. Therefore, it has been shown that q^i, q^0, and Q^i are in this order on the line L_i. Now when $i = t+1, \cdots, k$, by the theorem assumption, we have

$$D(Q^1 \| q) = D(Q^i \| q). \tag{2}$$

Furthermore, by the Pythagoras theorem, we have

$$D(Q^1 \| q^0) + D(q^0 \| q) = D(Q^1 \| q) \tag{3}$$

$$D(Q^i \| q^i) + D(q^i \| q) = D(Q^i \| q) \tag{4}$$

$$D(q^0 \| q^i) + D(q^i \| q) = D(q^0 \| q). \tag{5}$$

Therefore, from Lemma 2 and (2)–(5) we have

$$\begin{aligned} D(Q^i \| q^0) &\le D(Q^i \| q^i) - D(q^0 \| q^i) \\ &= D(Q^i \| q) - D(q^0 \| q) \\ &= D(Q^1 \| q) - D(q^0 \| q) \\ &= D(Q^1 \| q^0) \\ &= C. \end{aligned}$$

Q.E.D.

From now on, we assume that

$$\dim(Q^1 \smile \cdots \smile Q^k) = k-1,$$

i.e., k points $Q^1, \cdots, Q^k$ are in the general position. In this case, the point q equidistant from $Q^1, \cdots, Q^k$ always exists and it is represented as

$$q = \sum_{i=1}^{k} \lambda_i Q^i.$$

If $\lambda_i \ge 0$ for all $i = 1, \cdots, k$, the $q \in V$, and so q achieves the capacity by Theorem 1. Muroga [1] indicates that "if $\lambda_i < 0$ for at least one i, choose $k-1$ points arbitrarily from $Q^1, \cdots, Q^k$ and represent the point equidistant from these $k-1$ points as a linear combination shown above. Repeat this calculation for all possible choices of $k-1$ points. If there exist cases where all the coefficients are nonnegative, the maximum transmission rate among them is the capacity. Otherwise, reduce the number of points to $k-2$, $k-3, \cdots$, and do a similar calculation until we have some cases where all the coefficients are nonnegative."

This method is correct but contains much redundancy. A more effective method is proposed below.

Theorem 3: We can obtain the q^0 which achieves the capacity by a maximum of $k-2$ projections onto linear sets.

Since $q^0 = \mathrm{pr.}_V(q)$ according to Theorem 2, in principle we can obtain q^0 by one projection. However, it is difficult to calculate q^0 using this method. In fact, when we want to project q onto V, we must solve the minimum problem

$$\min_{r \in V} D(r \| q).$$

However, in general, q^0 is on the boundary of V, so we cannot use Lagrange's method of indeterminate coefficients to solve it. Theorem 3 offers an algorithmic method to obtain q^0 which circumvents the difficulty at the sacrifice of a possible increase in number of iterations. Here "algorithmic" means the iterative projections onto linear sets, in which case we can use Lagrange's method.

Proof: Represent the point q equidistant from the extreme points of V as

$$q = \sum_{i=1}^{k} \lambda_i Q^i,$$

$$\sum_{i=1}^{k} \lambda_i = 1, \ \lambda_1, \cdots, \lambda_{k_1} > 0, \ \lambda_{k_1+1}, \cdots, \lambda_k \le 0.$$

Denoting $E^1 = Q^1 \smile \cdots \smile Q^{k_1}$ and $q^1 = \mathrm{pr.}_{E^1}(q)$, we have

$$D(Q^i \| q^1) \begin{cases} = C_1, & \text{constant for } i = 1, \cdots, k_1 \\ \le C_1, & i = k_1+1, \cdots, k. \end{cases}$$

In fact, the equality for $i = 1, \cdots, k_1$ holds by the Pythagoras theorem. For $i = k_1+1, \cdots, k$, let L_{1i} be the line connecting q^1 and Q^i, and let $r^{1i} = \mathrm{pr.}_{L_{1i}}(q)$. Then we find, as previously mentioned in proving Theorem 2, that Q^i, q^1, and r^i are located on L_{1i} in this order. Thus we have

$$\begin{aligned} D(Q^i \| q^1) &\le D(Q^1 \| q^1) \\ &= C_1. \end{aligned}$$

Now suppose q^1 is represented as

$$q^1 = \sum_{i=1}^{k_1} \mu_i Q^i$$

$$\sum_{i=1}^{k_1} \mu_i = 1,\ \mu_1,\cdots,\mu_{k_2} > 0,\ \mu_{k_2+1},\cdots,\mu_{k_1} \le 0.$$

Denoting $E^2 = Q^1 \cup \cdots \cup Q^{k_2}$ and $q^2 = \mathrm{pr.}_{E^2}(q^1)$, we have

$$D(Q^i \| q^2) \begin{cases} = C_2, & \text{constant for } i = 1,\cdots,k_2 \\ \le C_2, & i = k_2+1,\cdots,k_1 \end{cases}$$

in the same way as before. For $i = k_1+1,\cdots,k$, it can be shown that $D(Q^i\|q^2) \le C_2$ holds as follows. Let L_{2i} be the line connecting q^2 and Q^i $(i = k_1+1,\cdots,k)$, and let $r^{1i} = \mathrm{pr.}_{L_{1i}}(q)$. Then it is evident from the previous argument that on the line L_{2i}, Q^i, q^2, r^{2i} are in this order. Therefore, we have

$$\begin{aligned} D(Q^i\|q^2) &\le D(Q^i\|r^{2i}) - D(q^2\|r^{2i}) \\ &= D(Q^i\|q^1) - D(q^2\|q^1) \\ &\le C_1 - D(q^2\|q^1) \\ &= D(Q^1\|q^2) \\ &= C_2, \qquad i = k_1+1,\cdots,k. \end{aligned}$$

We iterate this procedure to obtain $q^1, q^2, q^3,\cdots$ until all coefficients are positive. Let k_{i+1} be the number of Q^j having positive coefficients in the representation of q^i. The worst case is that $k_{i+1} = k_i - 1$ holds for all $i = 1,2,\cdots$. Since the point equidistant from two points always belongs to the line segment connecting them, we can obtain q^0 that achieves the capacity by a maximum of $k-2$ iterative projections. Q.E.D.

Now we assume

$$\dim(Q^1 \cup \cdots \cup Q^k) = d - 1 < k - 1.$$

In this case, a point equidistant from $Q^1,\cdots,Q^k$ does not exist. However, d points chosen arbitrarily from $Q^1,\cdots,Q^k$ are in the general position. Therefore, there exists a point equidistant from these d points. Thus by using the foregoing method we obtain ${}_kC_d$ values of channel capacity for all combinations. The following theorem ensures that the maximum among these ${}_kC_d$ values is the true capacity.

Theorem 4: If $\dim(Q^1 \cup \cdots \cup Q^k) = d-1$, the maximum value among the ${}_kC_d$ values of "capacity" computed for d points chosen arbitrarily from $Q^1,\cdots,Q^k$ is the true capacity.

Proof: Let $Q^1,\cdots,Q^h$ be the points in $Q^1,\cdots,Q^k$ such that $D(Q^1\|q^0) = \cdots = D(Q^h\|q^0)$ is the greatest value among $D(Q^1\|q^0),\cdots,D(Q^k\|q^0)$, where q^0 is the capacity-achieving point. Since $q^0 \in C(Q^1,\cdots,Q^h)$, by Lemma 1 a subset of $\{Q^1,\cdots,Q^h\}$ exists, say, $\{Q^1,\cdots,Q^{h_1}\}$, such that q^0 is contained in the simplex having $Q^1,\cdots,Q^{h_1}$ vertices. Then $h_1 \le d$, and if we solve Csiszàr's minimax problem for any d points including those h_1 points, we obtain the true capacity C. The rates for the other choices of d points are, of course, not greater than C. Q.E.D.

V. Examples

1) *2×2 Channel Matrix:* The capacity C of a channel

$$Q = \begin{pmatrix} Q^1 \\ Q^2 \end{pmatrix} = \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix}, \qquad 0 \le a, b \le 1,$$

is

$$C = \begin{cases} \log(1+e^A) - (1-b)H^1/(a-b) \\ \quad + (1-a)H^2/(a-b), & a \ne b \\ 0, & a = b, \end{cases}$$

where

$$H^1 = -a\log a - (1-a)\log(1-a)$$
$$H^2 = -b\log b - (1-b)\log(1-b)$$

and

$$A = (H^1 - H^2)/(a-b).$$

In this case, the convex hull V of Q^1, Q^2 is the line segment connecting Q^1 and Q^2. Since the equidistant point q^0 from Q^1 and Q^2 always exists in V, $C = D(Q^1\|q^0)$ is the capacity by Theorem 1.

2) *3×3 Channel Matrix:* Next, we consider a channel

$$Q = \begin{pmatrix} Q^1 \\ Q^2 \\ Q^3 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

$$\left(\sum_{i=1}^{3} a_i = \sum_{i=1}^{3} b_i = \sum_{i=1}^{3} c_i = 1,\ a_i, b_i, c_i \ge 0,\ i = 1,2,3 \right).$$

For two probability distributions Q^i, Q^j $(i \ne j)$, the "midpoint" of Q^i and Q^j is defined by a point M^{ij} which satisfies the following:

1) M^{ij} is on the line segment connecting Q^i and Q^j;
2) $D(Q^i\|M^{ij}) = D(Q^j\|M^{ij})$.

Further, we call $D(Q^i\|M^{ij})$ the "half-length" of the line segment connecting Q^i and Q^j and denote it by $d(Q^i,Q^j)$. By definition, we have $d(Q^i,Q^j) = d(Q^j,Q^i)$. Without loss of generality, we may assume that $d(Q^1,Q^2)$ is the greatest value among $d(Q^1,Q^2)$, $d(Q^2,Q^3)$, and $d(Q^3,Q^1)$. Let q^0 be the equidistant point from Q^1,Q^2,Q^3; i.e., q^0 is the unique solution of the following equation:

$$D(Q^1\|q^0) = D(Q^2\|q^0) = D(Q^3\|q^0).$$

Then we have

$$C = \begin{cases} D(Q^1\|q^0), & \text{if } D(Q^3\|M^{12}) > D(Q^1\|M^{12}) \\ D(Q^1\|M^{12}), & \text{if } D(Q^3\|M^{12}) \le D(Q^1\|M^{12}). \end{cases}$$

References

[1] S. Muroga, "On the capacity of a discrete channel, 1," *J. Phys. Soc. Japan*, vol. 8, pp. 484–494, 1953.
[2] M. C. Cheng, "On the computation of capacity of a discrete memoryless channel," *Inform. Contr.*, vol. 24, pp. 292–298, 1974.
[3] S. Takano, "On a method of calculating the capacity of a discrete memoryless channel," *Inform Contr.*, vol. 29, pp. 327–336, 1975.
[4] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 14–20, 1972.
[5] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, 1972.
[6] M. Jimbo and S. Kunisawa, "An iteration method for calculating the relative capacity," *Inform. Contr.*, vol. 43, pp. 216–223, 1979.
[7] F. A. Valentine, *Convex Sets*. New York: McGraw-Hill, 1964.
[8] I. Csizàr and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[9] G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.