(1,1,0), (1,0,1), and (0,1,1). In the first case s^* remains equal to 2, in the next three cases s^* is equal to 1, and in the last three cases s^* becomes 0.

If the triple (x_i, x_i, x_i) is (1,1,1), similar derivations can be made. As a result, we can construct the two matrices

$$M_0 = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & 3 & 1 \end{bmatrix} \qquad M_1 = \begin{bmatrix} 1 & 3 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

M_0 (resp. M_1) refers to the situation where (0,0,0) (resp. (1,1,1)) is transmitted by F. The rows (resp. columns) are indexed by the values 0, 1, or 2 of s (resp. s^*). In position (s, s^*) the entry of M_x is the number of triples that can be generated when the triple (x, x, x) is transmitted to the mailbox in state s and modified in such a way that the new state of the mailbox is s^*. The characteristic polynomial of $M := M_0 M_1$ is $\lambda^3 - 30\lambda^2 + 57\lambda - 1$ and the largest eigenvalue of M is $\lambda_M = 27.9629$. Using the same ideas as in the case $b = 1$, we see that, for any $\omega \in [0,1]$, the pair of rates $R_F = \mathcal{H}(\omega)/3$, $R_G = (\omega \log \lambda_M)/6$, is achievable. In the present case these pairs improve only slightly on the time-sharing strategy in the region of intermediate rates. Generalizing this strategy for larger values of b does not lead to any further improvement.

REFERENCES

[1] D. Blackwell, L. Breiman, and J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, pp. 1209–1220, 1958.
[2] T. W. Benjamin, "Coding for a noisy channel with permutation errors," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1975.
[3] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 310–314, May 1987.
[4] K. Kobayashi, "Combinatorial structure and capacity of the permuting relay channel," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 813–826, Nov. 1987.
[5] G. D. Forney, "Convolutional codes II: Maximum likelihood decoding," *Inform. Contr.*, vol. 25, pp. 222–266, 1974.
[6] W. W. Peterson and E. J. Weldon, *Error Correcting Codes*, 2nd ed. Cambridge, MA: MIT Press, 1972.
[7] E. Seneta, *Non-Negative Matrices*. Allen & Unwin, 1973.

Geometric Characterization of Capacity-Constraint Function

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

***Abstract*—We consider from a geometric point of view of maximizing the mutual information under a linear input constraint. Assuming a suitable regularity condition on the channel matrix, we find that the probability distribution (PD) equidistant from the row PD's of the channel matrix plays an important role, and the maximum is achieved by the projection of that PD onto the set of PD's satisfying the constraint.**

I. INTRODUCTION

We consider a discrete memoryless channel whose input and output alphabets are $X = \{x_1, \cdots, x_m\}$ and $Y = \{y_1, \cdots, y_n\}$, respectively, and whose channel matrix is

$$Q = (Q_{ij}), \qquad i = 1, \cdots, m, \ j = 1, \cdots, n$$

Manuscript received July 13, 1988; revised October 20, 1988. The material in this paper was presented at the IEEE International Symposium on Information Theory, Kobe, Japan, June 20, 1988.

The authors are with NTT Transmission Systems Laboratories, Take, Yokosuka-shi, Kanagawa-ken 238-03, Japan.

IEEE Log Number 8929771.

where $Q_{ij} \triangleq Q(y_j|x_i)$ is the conditional probability of y_j when an alphabet x_i was transmitted. We denote the ith row vector of the channel matrix Q by $Q^i = (Q_{i1}, \cdots, Q_{in})$. We denote the input probability distribution (PD) by $p = (p_i)$ and the output PD corresponding to p by $q = (q_j)$; i.e.,

$$q_j = \sum_{i=1}^{m} p_i Q_{ij} \text{ or } q = pQ.$$

For two PD's $q = (q_1, \cdots, q_n)$ and $\tilde{q} = (\tilde{q}_1, \cdots, \tilde{q}_n) \in \bar{\Delta}^n \triangleq \{(\alpha_1, \cdots, \alpha_n) | \sum_{j=1}^n \alpha_j = 1, \ \alpha_j \geq 0\}$, the Kullback–Leibler (K–L) divergence is defined as

$$D(q\|\tilde{q}) = \sum_{j=1}^{n} q_j \log(q_j/\tilde{q}_j).$$

It is known [2, p. 59] that the following "divergence geometry" holds in $\bar{\Delta}^n$.

Projection: Let V be a closed convex subset of $\bar{\Delta}^n$. We assume that, for $q \in \bar{\Delta}^n$, at least one $r \in V$ exists such that $D(r\|q) < \infty$. Then there is a unique $r \in V$ minimizing $D(r\|q)$. We call this r the projection of q onto V and denote it by $r = \pi(q|V)$.

Projection onto a Linear Set: We say that E is a linear set in $\bar{\Delta}^n$ if E can be represented as

$$E = \left\{ q = (q_1, \cdots, q_n) \Big| \sum_{j=1}^{n} a_{kj} q_j = b_k \right\},$$

where a_{kj}, b_k are some constants and k ranges over a finite index set. For example, a straight line, a plane, and so forth, are linear sets. Since a linear set E is a closed convex set of $\bar{\Delta}^n$, we can consider the projection onto E. For $q = (q_1, \cdots, q_n) \in \bar{\Delta}^n$, let us calculate the coordinate components of $q^* = \pi(q|E)$. Since q^* is the solution to the minimization problem $\min_{r \in E} D(r\|q)$, by Lagrange's method of indeterminate coefficients, introducing multipliers ξ_k, ζ, and letting

$$f(r) = D(r\|q) + \sum_k \xi_k \sum_{j=1}^{n} a_{kj} r_j + \zeta \sum_{j=1}^{n} r_j,$$

we obtain from $\partial f / \partial r_j = 0$

$$q_j^* = \alpha q_j \exp\left(- \sum_k \xi_k a_{kj} \right), \qquad j = 1, \cdots, n \tag{1}$$

where $\alpha = \exp(-1-\zeta)$ and ξ_k, ζ are the numbers determined by the conditions:

$$\sum_{j=1}^{n} a_{kj} q_j^* = b_k \qquad \sum_{j=1}^{n} q_j^* = 1.$$

Pythagorean Theorem: Let E be a linear set in $\bar{\Delta}^n$ and let $q^* = \pi(q|E)$ for $q \in \bar{\Delta}^n$. Then for any $r \in E$, the following equation holds;

$$D(r\|q^*) + D(q^*\|q) = D(r\|q).$$

Iterated Projection: Let E^0 be a linear set in $\bar{\Delta}^n$ and E^1 be a closed convex subset of E^0. Then for $q \in \bar{\Delta}^n$ we have

$$\pi(q|E^1) = \pi(\pi(q|E^0)|E^1).$$

Now, let E be a hyperplane in $\bar{\Delta}^n$, i.e., E is the intersection of a hyperplane in $\mathbb{R}^n$ and $\bar{\Delta}^n$. E separates $\bar{\Delta}^n$ into two disjoint subsets, say, E^- and E^+. Let E^+ include E. Thus E^+ is a closed set and E^- open. Here we consider a point $q \in E^-$ and a closed convex set $V \subset \bar{\Delta}^n$ and put $q^* = \pi(q|E)$. An easy calcula-

0018-9448/89/0700-0892$01.00 ©1989 IEEE

tion shows that E^+ is characterized by

$$E^+ = \{ p | D(p\|q^*) + D(q^*\|q) \le D(p\|q) \}.$$

Then we have the following Proposition.

Proposition 1: If $q^* = \pi(q|V)$, $V \subset E^+$.

Proof: See Csiszàr [3, theorem 2.2].

In the n-dimensional Euclidean space, let $\mathrm{con}(Q^1, \cdots, Q^m)$ denote the convex hull of $Q^1, \cdots, Q^m$, i.e., the minimum convex set including $Q^1, \cdots, Q^m$, and let $Q^1 \cup \cdots \cup Q^m$ denote the minimum subspace including $Q^1, \cdots, Q^m$.

We now consider the problem of maximizing the mutual information under a linear input constraint:

$$c(p) \triangleq \sum_{i=1}^{m} c_i p_i \le \Gamma$$

where $c_1, \cdots, c_m$ are nonnegative real numbers. We denote the maximum by $C(\Gamma)$ and call it a capacity-constraint function, i.e.,

$$C(\Gamma) \triangleq \max_{p:\, c(p) \le \Gamma} I(p, Q).$$

$C(\Gamma)$ exists only for $\Gamma \ge \Gamma_0 \triangleq \min_{1 \le i \le m} c_i$, otherwise the set of p's satisfying the constraint is empty. If Γ is greater than some Γ^*, $C(\Gamma)$ equals the capacity C of the channel Q. It is well-known [2, p. 137] that $C(\Gamma)$ has the following properties.

Proposition 2: $C(\Gamma)$ is a nondecreasing concave function. $C(\Gamma)$ is strictly increasing for $\Gamma_0 \le \Gamma \le \Gamma^*$ and differentiable at $\Gamma > \Gamma_0$ except for $\Gamma = \Gamma^*$.

Fig. 1 shows a typical graph of capacity-constraint function.

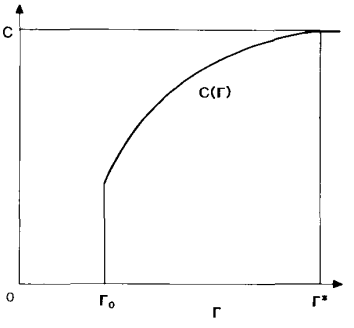

Fig. 1. Typical graph of capacity-constraint functions.

II. Preliminaries to the Theorems

We first consider a supportive maximization problem:

$$F(\gamma) \triangleq \max_{p \in \bar{\Delta}^n} \left(I(p, Q) - \gamma c(p) \right).$$

It is known [2, pp. 137–142] that the following properties hold.

Proposition 3: We obtain

$$C(\Gamma) = \min_{\gamma \ge 0} \left(F(\gamma) + \gamma \Gamma \right), \qquad \Gamma > \Gamma_0$$

If, for some $\gamma \ge 0$, a p maximizes $I(p, Q) - \gamma c(p)$ and $c(p) = \Gamma$, then $I(p, Q) = C(\Gamma)$.

Proposition 4: For any $\gamma \ge 0$,

$$F(\gamma) = \min_{q \in \bar{\Delta}^n} \max_{1 \le i \le m} \left(D(Q^i\|q) - \gamma c_i \right).$$

The minimum is achieved if and only if $q = pQ$ for a p that maximizes $I(p, Q) - \gamma c(p)$. This q is unique. Moreover, a p maximizes $I(p, Q) - \gamma c(p)$ if and only if there is a constant F such that

$$D(Q^i\|pQ) - \gamma c_i = F, \quad \text{if } p_i > 0$$
$$\le F, \quad \text{if } p_i = 0.$$

Proposition 5: We have

$$C(\Gamma) = \min_{q \in \bar{\Delta}^n} \min_{\gamma \ge 0} \max_{1 \le i \le m} \left(D(Q^i\|q) + \gamma(\Gamma - c_i) \right), \qquad \Gamma > \Gamma_0.$$

q achieves the minimum if and only if $q = pQ$ for a p which maximizes $I(p, Q)$ under the constraint $c(p) \le \Gamma$.

From the foregoing propositions, we notice that the output PD that achieves $C(\Gamma)$ can be represented as $q = pQ$ for a p that maximizes $I(p, Q) - \gamma c(p)$ for some $\gamma \ge 0$ and is unique. Therefore, rather than looking for an input PD, we look for an output PD that achieves $C(\Gamma)$. In other words, we use the Kullback–Leibler divergence to clarify in a geometric way where the PD that achieves $C(\Gamma)$ is in the output probability space $\bar{\Delta}^n$.

Let us first attempt to solve this problem directly using Lagrange's method of indeterminate coefficients. Introducing α, β as indeterminate coefficients and letting

$$f(p) = I(p, Q) + \alpha c(p) + \beta \sum_{i=1}^{m} p_i,$$

from $\partial f / \partial p_i = 0$ we have

$$\sum_{j=1}^{n} Q_{ij} \log Q_{ij} - \sum_{j=1}^{n} Q_{ij} \log \sum_{i=1}^{m} p_i Q_{ij} - 1 + \alpha c_i + \beta = 0,$$
$$i = 1, \cdots, m.$$

Although we can solve these equations to obtain $p = (p_i)$, if the solution does not satisfy $p_i \ge 0$ $(i = 1, \cdots, m)$, then it is evidently inadequate. What should we do when some p_i is negative? To answer this question, we must use the divergence geometry to look directly for $\hat{q}$ which achieves $C(\Gamma)$.

From no on, the channel matrix Q is assumed to satisfy the condition: $\mathrm{rank}(Q) = m$ $(m \le n)$. According to this condition, notice that there exists the right inverse matrix $R = (R_{ji})$ of Q which satisfies

$$\sum_{j=1}^{n} Q_{ij} R_{ji'} = \delta_{ii'} \tag{2}$$

where $\delta_{ii'}$ is Kronecker's delta and $\sum_{i=1}^{m} R_{ji} = 1$, $j = 1, \cdots, n$. Furthermore, a PD q^0 exists that satisfies $D(Q^1\|q^0) = \cdots = D(Q^m\|q^0)$.

III. Theorems

Now, since input PD p is in the set $\{ p | c(p) \le \Gamma \}$, the corresponding output PD q is in $V(\Gamma) \triangleq \{ q | q = pQ,\ c(p) \le \Gamma \}$. This condition of $V(\Gamma)$ can be rewritten in terms of q. From $q_j = \sum_{i=1}^{m} p_i Q_{ij}$, we have $p_i = \sum_{j=1}^{n} q_j R_{ji}$ (or $p = qR$), then

$$c(p) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} c_i R_{ji} \right) q_j.$$

Letting $s_j = \sum_{i=1}^{m} c_i R_{ji}$, and $s(q) \triangleq \sum_{j=1}^{n} s_j q_j$, we have $V(\Gamma) = \{ q | s(q) \le \Gamma \} \cap \mathrm{con}(Q^1, \cdots, Q^m)$. In addition, we have $c_i = \sum_{j=1}^{n} s_j Q_{ij}$. Furthermore, we define a linear set $E(\Gamma)$ by $E(\Gamma) = \{ q | s(q) = \Gamma \}$. Here we have the following theorem.

Theorem 1: Let q^0 be a probability distribution equidistant from $Q^1, \cdots, Q^m$, i.e., q^0 satisfies $D(Q^1\|q^0) = \cdots = D(Q^m\|q^0)$. If we denote $\hat{q} = \pi(q^0|V(\Gamma))$, then $\hat{q}$ achieves the capacity-constraint function $C(\Gamma)$.

Before this is fully proved, we show that it holds in a simple concrete example. If $m = n$ and $Q = I_n$ ($n \times n$ unit matrix), then $p = q$. In this case, since $q^0 = (1/n, \cdots, 1/n)$ and $D(q\|q^0) =$

$\log n - H(q)$, we have

$$\begin{aligned}\max_{c(p)\le\Gamma} I(p,Q) &= \max_{c(p)\le\Gamma} H(p)\\ &= \max_{q\in V(\Gamma)} \left(\log n - D(q\|q^0)\right)\\ &= \log n - \min_{q\in V(\Gamma)} D(q\|q^0)\\ &= \log n - D(\hat{q}\|q^0).\end{aligned}$$

Therefore, we see that $\hat{q}$ attains $C(\Gamma)$.

Proof: Denote $\hat{p} \triangleq \hat{q}R$. There are two possibilities: $c(\hat{p}) < \Gamma$ and $c(\hat{p}) = \Gamma$. First, if $c(\hat{p}) < \Gamma$, then for any $\Gamma' \ge \Gamma$ we obtain $\pi(q^0|V(\Gamma')) = \pi(q^0|V(\Gamma)) = \hat{q}$. Therefore, this is a constraint-free problem, namely, a problem of the channel capacity. In this case, we have already shown [1] that $\hat{q}$ achieves $C = C(\Gamma)$.

Next, if $c(\hat{p}) = \Gamma$, without loss of generality, we can assume that

$$\begin{aligned}&\hat{p}_1,\cdots,\hat{p}_k > 0\\ &\hat{p}_{k+1} = \cdots = \hat{p}_m = 0.\end{aligned} \qquad (k \ge 1)$$

If $k = 1$, this problem is trivial because $V(\Gamma)$ consists of only one point and therefore $C(\Gamma) = 0$. Let $k \ge 2$.

We calculate the coordinate of $\hat{q}$. Since $\hat{q} = \pi(q^0|V(\Gamma))$ is also the projection of q^0 onto the linear set

$$E(\Gamma)\cap \mathrm{con}(Q^1,\cdots,Q^k) = \left\{ q|s(q) = \Gamma,\ p_i = \sum_{j=1}^{n} q_j R_{ji} = 0\ (i = k+1,\cdots,m)\right\},$$

from (1) we have

$$\hat{q}_j = \alpha q_j^0 \exp\left(-\gamma s_j - \sum_{i=k+1}^{m} \xi_i R_{ji}\right), \qquad j = 1,\cdots,n \quad (3)$$

where $\alpha = \exp(-1-\zeta)$ and γ, ξ_k, ζ are determined by

$$s(\hat{q}) = \Gamma \qquad \sum_{j=1}^{n} \hat{q}_j R_{ji} = 0, \qquad i = k+1,\cdots,m \qquad \sum_{j=1}^{n} \hat{q}_j = 1.$$

Now, denote

$$\begin{aligned}s_j' &= s_j + (1/\gamma)\sum_{i=k+1}^{m} \xi_i R_{ji}, \qquad j = 1,\cdots,n\\ c_i' &= \sum_{j=1}^{n} s_j' Q_{ij}, \qquad i = 1,\cdots,m\end{aligned}$$

and define a hyperplane $E'(\Gamma)$ by

$$E'(\Gamma) = \left\{ q|s'(q) \triangleq \sum_{j=1}^{n} s_j' q_j = \Gamma\right\}.$$

Since $\hat{q}_j = \alpha q_j^0 \exp(-\gamma s_j')$, $j = 1,\cdots,n$, from (3), we obtain $\hat{q} = \pi(q^0|E'(\Gamma))$. Let 1,2 be two elements chosen arbitrarily from the set $\{1,\cdots,m\}$ (and renumbered for the sake of reducing the number of symbols). Since

$$\begin{aligned}&D(Q^1\|\hat{q}) - D(Q^2\|\hat{q})\\ &\quad= \sum_{j=1}^{n} Q_{1j}\log(Q_{1j}/\hat{q}_j) - \sum_{j=1}^{n} Q_{2j}\log(Q_{2j}/\hat{q}_j)\\ &\quad= D(Q^1\|q^0) - D(Q^2\|q^0) + \gamma\left(\sum_{j=1}^{n} Q_{1j}s_j' - \sum_{j=1}^{n} Q_{2j}s_j'\right)\\ &\quad= \gamma(c_1' - c_2')\end{aligned}$$

we have

$$D(Q^i\|\hat{q}) - \gamma c_i' = F, \qquad i = 1,\cdots,m$$

where F is a constant. Notice that, by Propositions 3–5, $\hat{q}$ achieves the maximum of the mutual information under the constraint $\sum_{i=1}^{m} c_i' p_i \le \Gamma$. The corresponding set of constraints in $\bar{\Delta}^n$ is $V'(\Gamma) \triangleq \{q|s'(q) \le \Gamma\}\cap \mathrm{con}(Q^1,\cdots,Q^m)$. Now since $\hat{q}$ attains the minimum of $D(q\|q^0)$ in $q \in V(\Gamma)$ and also in $q \in E'(\Gamma)$ at the same time, $V(\Gamma)$ must lie on the opposite side of q^0 with respect to $E'(\Gamma)$ by Proposition 1. Therefore, we have $V(\Gamma) \subset V'(\Gamma)$. Since $\hat{q} \in V(\Gamma) \subset V'(\Gamma)$, we conclude that $\hat{q}$ achieves the maximum of the mutual information in $V(\Gamma)$.

In the above proof, we assumed that we knew beforehand which $\hat{p}_i$ are 0. As shown below, we introduce a concrete algorithm for determining which i is equal to 0.

Theorem 2: For $\Gamma_0 \le \Gamma \le \Gamma^*$, we can obtain $\hat{q}$ which achieves $C(\Gamma)$ by at most $m-2$ projections onto linear sets.

At this point, it is necessary to explain Theorem 2 more concretely. For $\Gamma_0 \le \Gamma \le \Gamma^*$, $\hat{q}$ satisfies $s(\hat{q}) = \Gamma$. Then, let $E^0 = E(\Gamma)$, $q^1 = \pi(q^0|E^0)$ and put $p^1 = q^1 R$ be an input PD corresponding to q^1. If $p_i^1 \ge 0$ for all $i = 1,\cdots,m$, q^1 equals the $\hat{q}$ which achieves $C(\Gamma)$. However, if $p_i^1 < 0$ for some i, q^1 does not achieve $C(\Gamma)$. We must therefore look for another PD. Now, without loss of generality, assume that

$$\begin{aligned}&p_1^1,\cdots,p_{m_1}^1 > 0\\ &p_{m_1+1}^1,\cdots,p_m^1 \le 0.\end{aligned}$$

Then denote $E^1 = \{q|s(q) = \Gamma\}\cap(Q^1 \cup \cdots \cup Q^{m_1})$ and project q^1 onto E^1 to obtain $q^2 \triangleq \pi(q^1|E^1)$. In other words, we project q^1 onto the linear set E^1 determined by $E(\Gamma)$ and Q^i which correspond to positive coefficients p_i^1 of q^1. Next, let $p^2 = q^2 R$ and check whether or not $p_i^2 > 0$. If $p_i^2 \le 0$, exclude the corresponding Q^i and obtain q^3 by projection. We repeat this procedure until all coefficients become positive. Since the worst case is that one Q^i is excluded at each step, we obtain a maximum of $m-2$ projections onto linear sets.

Proof: The fundamental part of the proof is similar to that of Theorem 1. The reader may also refer to [1, theorem 2].

IV. Examples

Example 1

$$Q = \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \qquad c_1 = 0,\ c_2 = 1,\ c_3 = 2.$$

We look for the maximum value of $I(p,Q) = H(p)$, the entropy of p. In this case, since the $\hat{p}$ which achieves the maximum always lies in the interior of $\mathrm{con}(Q^1,Q^2,Q^3) = \bar{\Delta}^3$ we can obtain the following representation by Lagrange's method of indeterminate coefficients (Fig. 2):

$$C(\Gamma) = H(\hat{p}) = \log \alpha^{-1} + \gamma\Gamma, \qquad 0 \le \Gamma \le 1,$$

where

$$\gamma = \log\frac{1-\Gamma+\sqrt{1+6\Gamma-3\Gamma^2}}{2\Gamma} \qquad \alpha^{-1} = \frac{7-3\Gamma+\sqrt{1+6\Gamma-3\Gamma^2}}{2(2-\Gamma)^2}.$$

Example 2

$$Q = \begin{pmatrix} 2/3 & 1/6 & 1/6\\ 1/6 & 2/3 & 1/6\\ 1/6 & 1/6 & 2/3\end{pmatrix}, \qquad c_1 = 0,\ c_2 = 1,\ c_3 = 2.$$

Thus we have $q^0 = (1/3,1/3,1/3)$. Letting $\hat{q} = \pi(q^0|V(\Gamma))$ and $\hat{p} = \hat{q}R$, we find by a simple calculation that $\hat{p}_1 > \hat{p}_2 > \hat{p}_3$

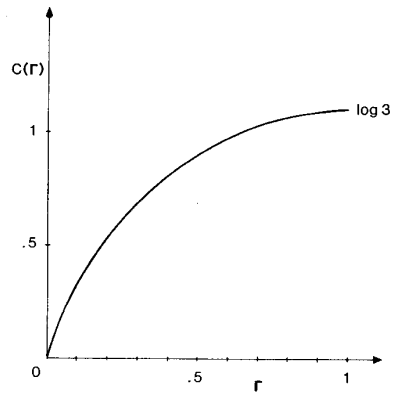

Fig. 2. Capacity-constraint function of Example 1.

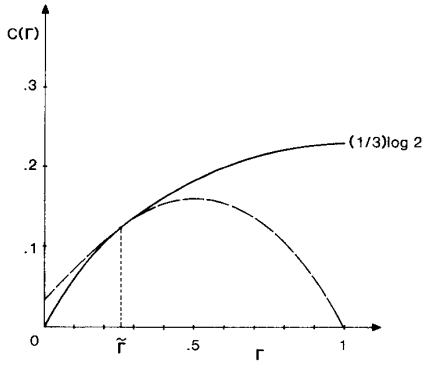

Fig. 3. Capacity-constraint function of Example 2. Solid curve is obtained by using (4) and (5). Dashed curve in $[0, \tilde{\Gamma}]$ (resp. $(\tilde{\Gamma}, 1]$) is obtained by applying (5) (resp. (4)) beyond limit.

always holds. Therefore, depending on whether $\hat{p}_3$ is positive or not, $\hat{q}$ lies in $\mathrm{con}(Q^1, Q^2, Q^3)$ or in $\mathrm{con}(Q^1, Q^2)$. Then we have the following representation (Fig. 3):

$$C(\Gamma) = \begin{cases} (2/3)\log(2/3) + (1/6)\log(1/6) \\ \quad -((1/2)\Gamma + (2/3))\log((-1/2)\Gamma + (2/3)) \\ \quad -((1/2)\Gamma + (1/6))\log((1/2)\Gamma + (1/6)), \\ \qquad\qquad 0 \le \Gamma \le \Gamma_1 \quad (4) \\ -H(2/3, 1/6, 1/6) + \log \alpha^{-1} + \gamma\Gamma, \\ \qquad\qquad \Gamma_1 < \Gamma \le 1 \quad (5) \end{cases}$$

where

$$\gamma = \frac{1}{2}\log\frac{1-\Gamma+\sqrt{13+6\Gamma-3\Gamma^2}}{2(1+\Gamma)}$$

$$\alpha^{-1} = \exp\gamma + \exp(-\gamma) + \exp(-3\gamma)$$

and Γ_1 is the solution in the interval $(0,1)$ of

$$\frac{-1+\Gamma+\sqrt{13+6\Gamma-3\Gamma^2}}{2(3-\Gamma)} = \frac{1+\sqrt{21}}{10}, \qquad \Gamma_1 \doteq 0.264.$$

$H(2/3, 1/6, 1/6)$ is the entropy of the PD $(2/3, 1/6, 1/6)$.

V. Conclusion

In this correspondence a geometric method for computing a capacity-constraint function $C(\Gamma)$ is considered. The K–L divergence, having properties like a metric between two PD's is used to look for the output PD that attains $C(\Gamma)$. It has been shown that $C(\Gamma)$ is attained by the projection of q^0 equidistant from each row vector of Q onto the set of PD's satisfying a given constraint condition. Moreover, the PD attaining $C(\Gamma)$ is obtained by using Lagrange's method of indeterminate coefficients at most $m-2$ times. We would like to be able to attack the case where the rank assumption is not valid. More study is needed to determine whether this geometric method can be applied to compute the rate-distortion function $R(D)$ directly.

Acknowledgment

The authors would like to thank the reviewers for their helpful suggestions and comments.

References

[1] K. Nakagawa and F. Kanaya, "A new geometric capacity characterization of a discrete memoryless channel," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 318–321, 1988.
[2] I. Csiszàr and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[3] I. Csiszàr, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.*, vol. 3, no. 1, pp. 146–158, 1975.
[4] S. Muroga, "On the capacity of a discrete channel, 1," *J. Phys. Soc. Japan*, vol. 8, pp. 484–494, 1953.
[5] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 14–20, 1972.
[6] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, 1972.
[7] G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

More on the Decoder Error Probability for Reed–Solomon Codes

KAR-MING CHEUNG

Abstract—**McEliece and Swanson offered an upper bound on $P_E(u)$, the decoder error probability for Reed–Solomon codes (more generally, linear maximum distance separable codes), given that u symbol errors occur. Their upper bound is slightly greater than Q, the probability that a completely random error pattern will cause a decoder error. We use a combinatorial technique similar to the principle of inclusion and exclusion to obtain an exact formula for $P_E(u)$. The $P_E(u)$'s for the (255,223) Reed–Solomon code used by NASA, and the (31,15) Reed–Solomon code (JTIDS code) are calculated using the exact formula and are observed to approach the Q's of the codes rapidly as u gets large. An upper bound for the expression $|(P_E(u)/Q)-1|$ is derived and subsequently shown to decrease nearly exponentially as u increases.**

I. Introduction

We begin with the following definitions. Let C be a linear code of length n, dimension k, and minimum distance d. Let q be a positive power of a prime. An (n, k, d) linear code C over GF(q) is maximum distance separable (MDS) if the Singleton bound is achieved; that is, $d = n - k + 1$. A code is t-error correcting if for some integer t, $2t \le d - 1$.

Manuscript received August 7, 1987; revised September 15, 1988. This paper was partially presented at the 1986 International Symposium on Information Theory, Ann Arbor, MI, October 8, and at the 1988 IEEE Military Communication Conference, San Diego, CA, October 23.

The author is with the Jet Propulsion Laboratory, 238-420, 4800 Oak Grove Drive, Pasadena, CA 91109, USA.

IEEE Log Number 8929772.