Extension of Some Results for Channel Capacity Using A Generalized Information Measure

Aharon Ben-Tal*' & Marc Teboulle**
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109

Technical Report 86-38

September 1986

*Technion, Israel Institute of Technology and Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, U.S.A.

**Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5

'Research supported by NSF Grant ECS-8604354

EXTENSION OF SOME RESULTS FOR CHANNEL CAPACITY USING A GENERALIZED INFORMATION MEASURE

Aharon Ben-Tal*' & Marc Teboulle**

September 1986

*Technion, Israel Institute of Technology and Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, U.S.A.

**Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5

'Research supported by NSF Grant ECS-8604354

Abstract

A new formulation for the channel capacity problem is derived by using the duality theory of convex programming. The simple nature of this dual representation is suitable for computational purposes. The results are derived in a unified way by formulating the channel capacity problem as a special case of a general class of concave programming problems involving a generalized information measure recently introduced by Burbea and Rao [10].

Key words: Information Theory, Channel capacity, convex programming duality, generalized information measure.

1. INTRODUCTION

Channel capacity, a basic concept in information theory, was introduced by Shannon [27] to specify the maximum rate at which information can be reliably conveyed by the channel. Roughly speaking, the basic theorem of information theory, the so-called "noisy channel coding theorem", states that if a given noisy channel has capacity C, it is possible to transmit over this channel messages of sufficiently large length and still be able to decode them with an arbitrary small probability of error, provided the rate of transmission is less than C. Methods for computing the capacity C of a discrete channel have been studied by Muroga [21], Cheng [12], Takano [28]. The best known algorithm however is the one introduced independently by Arimoto [2] and Blahut [7]. A somewhat similar iterative procedure based on the method of quasi-concave programming was proposed by Meister and Oettli [20]. In these previously mentioned works, the computational schemes are derived using the primal formulation of the channel capacity problem.

In this paper the classical channel capacity problem is embedded in a family (P_ϕ) of linearly constrained concave programming problems, each member of which is determined by a choice of a convex function ϕ; the classical case corresponding to $\phi(t) = t \log t$. The objective function in (P_ϕ) is the generalized average mutual information measure, recently introduced by Burbea and Rao [10], and the optimal value of (P_ϕ) is our generalized channel capacity C_ϕ.

A duality theory is developed for (P_ϕ) resulting in a dual representation of C_ϕ. As a special case, a new formulation of the classical channel capacity is obtained. The dual of (P_ϕ) (denoted (D_ϕ)) is a specially structured unconstrained minimax problem, and thus rendering itself to efficient computational methods. The dual formulation is also very useful to obtain upper bounds for C_ϕ . The paper is organized as follows. In section 2 the formulation of the classical channel capacity problem is given and the iterative method of Arimoto [2], Blahut [7] is briefly reviewed. In section 3 we formulate the generalized capacity C_ϕ and develop the theory leading to two different dual representations, the first (Theorem 3.1) is suitable for computations and the second (Corollary 3.1) is useful to derive upper bounds. The bounds are given in section 4, it is also shown there that for symmetric channels, the bound is attained, i.e. an explicit formula for C_ϕ is obtained. Section 5 contains concluding remarks and a brief discussion on possible extensions.

2. THE CHANNEL CAPACITY PROBLEM

Consider a communication channel described by an input alphabet $A = \{1,2,\ldots,m\}$, an output alphabet $B = \{1,2,\ldots,n\}$ and by a probability transition matrix $Q = \{Q_{kj}\}$, where Q_{kj} is the probability of receiving the output letter $k \in B$ when input letter $j \in A$ was transmitted, i.e. $\sum_{k=1}^{n} Q_{kj} = 1$ for all $j \in A$ and $Q_{kj} \geq 0$ for all $k \in B$, $j \in A$. The capacity of the channel is defined as

$$C := \max_{p \in \mathbf{P}_m} I(p,Q) := \max_{p \in \mathbf{P}_m} \sum_{j=1}^{m} \sum_{k=1}^{n} p_j Q_{kj} \log \frac{Q_{kj}}{\sum_{\ell=1}^{m} p_\ell Q_{K\ell}} \tag{2.1}$$

where

$$\mathbf{P}_m := \left\{ p \in \mathbb{R}^m : p_j \geq 0 \ \forall \ j \in A \ , \ \sum_{j=1}^{m} p_j = 1 \right\} \tag{2.2}$$

is the set of all discrete finite probability measures on the channel input, and $I(p,Q)$ is known as the <u>average mutual information</u> between the channel input and channel output, considered here as a function of p. The utility of the concept of capacity is widely discussed in the literature and for more details the reader is referred to Shannon [26], Gallager [15], Jelineck [16] and to the more recent book of Csiszar and Korner [14]. For a given probability transition matrix Q, it is shown in Gallager [15] that $I(,Q)$ is a concave function of p and therefore problem (2.1) is a <u>concave programming problem</u> over the simplex $\mathbf{P}_m$, then any of a number of readily available nonlinear programming codes can be used to compute C. However as reported by Blahut [7], computational experience with nonlinear programming codes applied to problem (2.1) have proved to be inefficient even for small alphabets sizes and to be impractical for the larger alphabet sizes. This motivates, independently Arimoto [2] and Blahut [7] to develop a systematic iterative method for computing the capacity. This was done by exploiting the special structure of the objective function $I(\cdot,Q)$. More specifically, let $P = (P_{jk})$ denote a transition matrix from the channel output alphabet to the channel input alphabet, then

$$I(P,Q) = \max_{P \in T}\left\{J(p,P;Q) := \sum_{j=1}^{m} \sum_{k=1}^{n} P_j Q_{kj} \log \frac{P_{jk}}{P_j}\right\} \tag{2.3}$$

where

$$T := \{P \in \mathbf{R}^m \times \mathbf{R}^n : \sum_{j=1}^{m} P_{jk} = 1 \text{ , all } k=1,\ldots,n, P_{jk} \geq 0 \text{ all } j \text{ and } k\} .$$

This can be verified by noting that the maximized J is given by

$$P^*_{jk} = \frac{Q_{kj}p_j}{\sum_{\ell=1}^{m} Q_{k\ell}P_\ell}$$

The Arimoto-Blahut algorithm can be summarized as follows:

(0) Choose an initial probability vector $p^{(0)} \in \mathbf{P}_m$.

At iteration r , where $p^{(r)}$ is given,

(i) Compute $P^{(r)} = \underset{P \in T}{\text{argmax}}\ J(p^{(r)}, P; Q)$

(ii) Update $p^{(r+1)} = \underset{p \in \mathbf{P}_m}{\text{argmax}}\ J(p, P^{(r)}; Q)$ (2.4)

(iii) Iterate $r \leftarrow r + 1$

The solution of (2.4) is explicitly given by

$$p_j^{(r+1)} = p_j^{(r)} \frac{c_j(p^{(r)})}{\sum_j p_j c_j(p^{(r)})} \tag{2.5}$$

where for any $p \in \mathbf{P}_m$

$$c_j(p) = \exp\left\{ \sum_k Q_{kj} \log \frac{Q_{kj}}{\sum_\ell p_\ell Q_{k\ell}} \right\} .$$

That the method (2.5) converges, i.e. $\lim_{r \to \infty} I(p^{(r)}, Q) = C$, see Arimoto [2] and Blahut [7] . The amount of Computation involved depends upon the size of the channel matrix. Some conditions under which the amount of computation can be reduced are discussed in Cheng [12] and Takano [28].

In this paper we suggest a dual formulation to the channel capacity problem. The simple nature of this dual problem opens the possibility to apply many recent numerical schemes available in the Mathematical programming literature, in particular for large alphabet size (see e.g. [27]). This will be discussed in the next sections.

3. THE GENERALIZED CAPACITY PROBLEM AND ITS DUALS

In this section we derive in a unified way a dual representation for the channel capacity problem. Recall that the channel capacity C is given as the optimal value of the following optimization problem:

$$\sup_{p\in\mathbf{P}_m} \sum_j \sum_k p_j Q_{kj} \log \frac{Q_{kj}}{\sum_{\ell=1}^{m} p_\ell Q_{k\ell}} \quad . \tag{3.1}$$

We denote by q_k the output probabilities, then $q_k := \sum_{j=1}^{m} p_j Q_{kj}$, $q_k \geq 0$ for all $k=1,\ldots,n$, $\sum_{k=1}^{n} q_k = 1$, i.e. with our notation $q \in \mathbf{P}_n$. In the decision variables (p_j, q_k), problem (3.1) can be written equivalently as

$$(P) \qquad \sup_{p\in\mathbf{P}_m} \sup_{q\in\mathbf{P}_n} \sum_j \sum_k p_j Q_{kj} \log Q_{kj} - \sum_{k=1}^{n} q_k \log q_k \tag{3.2}$$

$$\text{s.t.} \qquad \sum_{j=1}^{m} p_j Q_{kj} = q_k \quad \forall \; k=1,\ldots,n \; . \tag{3.2}$$

The objective function in (3.2) is concave in (p,q), therefore problem (P) is a linearly constrained concave program. Also note that the feasible set of (P) is a compact convex polyhedron in $\mathbb{R}^{m+n}$, hence the sup in (P) is actually attained. The special structure of the objective function formulated in (3.2) (linear in p minus strictly convex in q)

motivates us to consider the following general class of concave programming problems

$$(P_\phi) \qquad \max_{p\in\mathbf{P}_m} \max_{q\in\mathbf{P}_n} \sum_{j=1}^{m} p_j \sum_{k=1}^{m} \phi(Q_{kj}) - \sum_{k=1}^{n} \phi(q_k) \tag{3.4}$$

$$\text{s.t.} \qquad \sum_{j=1}^{m} p_j Q_{kj} = q_k \ , \ k = 1,\dots,n \ .$$

Throughout the rest of this paper, we assume that ϕ is a given twice continuously differentiable strictly convex function defined on an interval containing $(0,1]$, normalized such that $\phi(0) = \phi(1)$, $\phi'(1) < \infty$, and satisfying the additional assumption $\lim_{t\to 0^+} \phi'(t) = -\infty$. Note that the latter assumption holds if ϕ is essentially smooth in $[0,+\infty)$ (see e.g. Rockafellar [24]). We denote the class of such ϕ by Φ, accordingly C_ϕ will denote the optimal value of problem (P_ϕ). An important example of functions $\phi \in \Phi$ is provided by the family Φ_α of functions (parametrized by α):

$$\phi_\alpha(t) = \begin{cases} \frac{1}{1-\alpha}(t-t^\alpha) & \text{for } 0 < \alpha < 1 \\ t \log t & \text{for } \alpha = 1 \ . \end{cases}$$

Clearly with $\phi_1(t) = t \log t$, problem (P_ϕ) is just the classical channel capacity problem (P).

The objective function used in (P_ϕ) is exactly the generalized average mutual information measure introduced and studied by Burbea and Rao [10]. A related generalized measure of information was also recently introduced by Ben-Tal and Teboulle [6] and the associated rate distortion function was studied. For additional generalizations and applications of generalized information measures the reader is referred to Aczel [1], Arimoto [3], Burbea [8], [9], Burbea and Rao [11], Csiszar [11], Rao and Yayak [22], Renyi [23], and Ziv and Zakai [30].

The dual representation of ($P\phi$) will be derived via Langrangian duality. Before stating the main result of this section we introduce the following notations and definitions. For any $\phi \in \Phi$, let $f: \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$f(v) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \sum_{k=1}^{n} \phi^*(v_k - \eta) \right\} \tag{3.5}$$

where $\phi^*(y) = \sup_t \{ty - \phi(t)\}$ denotes the usual convex conjugate of ϕ. Also let $\ell: \mathbb{R}^n \to \mathbb{R}$ be defined by:

$$\ell(v) = \max_{1 \le j \le m} \{\ell_j(v) + b_j\} \tag{3.6}$$

where $\{\ell_j\}$ are the linear functions given by $\ell_j(v) = -\sum_{k=1}^{m} v_k Q_{kj}$ and b_j denote the constants $b_j = \sum_{k=1}^{m} \phi(Q_{kj})$.

Theorem 3.1. The dual problem of (P_ϕ), for $\phi \in \Phi$, is given by

$$(D_\phi) \qquad \inf_{v \in \mathbb{R}^n} \{f(v) + \ell(v)\} . \tag{3.7}$$

If (P_ϕ) is feasible the minimum in (D_ϕ) is attained and the optimal values coincide:

$$C_\phi = \max(P_\phi) = \min(D_\phi) .$$

Proof. The Lagrangian for problem (Pϕ) is

$$L(p,q,v) = -\sum_{k=1}^{n} \phi(q_k) + \sum_{j=1}^{m} p_j \sum_{k=1}^{n} \phi(Q_{kj}) + \sum_{k=1}^{n} q_k v_k - \sum_{j,k} p_j v_k Q_{kj}$$

and is separable in the decision variables (p,q). The dual objective function is then:

$$h(v) = \max_{p\in\mathbf{P}_m} \sum_{j=1}^{m} p_j\left(\sum_{k=1}^{n} \phi(Q_{kj}) - v_k Q_{kj}\right) + \max_{q\in\mathbf{P}_n} \sum_{k=1}^{n} q_k v_k - \phi(q_k) \tag{3.8}$$

and the dual problem is defined as $\min_{v\in\mathbf{R}^n} h(v)$. The first "max" in (3.8) is easily computed:

$$\max_{p\in\mathbf{P}_m} \sum_j p_j \left(\sum_k \phi(Q_{kj}) - v_k Q_{kj}\right) = \max_{1\le j\le m} \left\{\sum_{k=1}^{n} \phi(Q_{kj}) - v_k Q_{kj}\right\} , \tag{3.9}$$

To evaluate the second "max" in (3.8) we note that a Langrangian dual of

$$\alpha := \max_{q\in\mathbf{P}_n} \sum_{k=1}^{n} q_k v_k - \phi(q_k) \tag{3.10}$$

is given by

$$\beta := \min_{\eta\in\mathbf{R}} \left\{\eta + \sum_k \max_{q_k\ge 0} \{q_k(v_k-\eta) - \phi(q_k)\}\right\} . \tag{3.11}$$

where η is the Lagrange multiplier for the equation $\Sigma q_k = 1$. Problem (3.10) is a linearly constrained concave program satisfying trivially the Slater constraint qualification and hence by standard duality arguments [23] we have $\alpha = \beta$.

To compute β , we consider the problem $e(y) = \sup_{t\ge 0}\{ty-\phi(t)\}$. By the monotonocity of ϕ' , it follows that

$$e(y) = \begin{cases} \sup_{t\in\mathbb{R}} \{ty-\phi(t)\} = \phi^*(y) & \text{if} \quad y \geq \phi'(0) \\ -\phi(0) & \text{if} \quad y < \phi'(0) \end{cases} .$$

But since we assumed that $\phi'(0) = -\infty$, we conclude in fact that $e(y) = \phi^*(y)$. Using this result in (3.11) we get

$$\beta = \min_{\eta\in\mathbb{R}} \left\{ \eta + \sum_{k=1}^{n} \phi^*(v_k-\eta) \right\} . \tag{3.12}$$

Substituting (3.9) and (3.12) in (3.8) and using the notations (3.5), (3.6) we thus obtain $h(v) = f(v) + \ell(v)$ and hence the dual representation (3.7) is proved. Further, for $\phi \in \Phi$, ϕ is convex and then $\eta + \sum_k \phi^*(v_k-\eta)$ is jointly convex in $(\eta,v) \in \mathbb{R} \times \mathbb{R}^n$, hence (by Rockafellar [25] Theorem 1) $f(v)$ is convex. Finally since (P_ϕ) is assumed feasible and it is linearly constrained the Slater regularity condition holds trivially then it follows from standard duality results (see e.g. [24]) that the $\inf\,(D_\phi)$ is attained and $\max(P_\phi) = \min(D_\phi)$.

□

The next result shows how to obtain the optimal solution of the primal problem (P_ϕ) from an optimal solution of the dual.

<u>Theorem 3.2</u>. Let $\bar{v}$ be the optimal solution of (D_ϕ). Then the optimal solution $(\bar{p}, \bar{q})$ of the primal problem (P_ϕ) is computed as follows:

$$\bar{q}_k = (\phi')^{-1}(\bar{v}_k - \bar{\eta}) \quad , \quad k = 1,\ldots,n \tag{3.13}$$

where $\bar{\eta}$ is the unique solution of the equation:

$$\sum_k (\phi')^{-1}(\bar{v}_k - \bar{\eta}) = 1 \tag{3.14}$$

and $\bar{p}$ is the optimal solution of the linear program

$$(L_\phi) \qquad \max \sum_j p_j \left(\sum_k \phi(Q_{kj})\right)$$

$$\text{s.t.} \qquad \sum_j p_j Q_{kj} = \bar{q}_k \quad , \quad k = 1,\dots,n$$

$$\sum_j p_j = 1 \ , \quad p_j \geq 0 \ , \ j = 1,\dots,m \quad .$$

<u>Proof</u>: The expressions (3.13) for $\bar{q}_k$ is just the optimality conditions for $q_k = \bar{q}_k$ to solve the inner maximization in (3.11) (recall that the optimal q_k cannot be zero as explained in the proof of Theorem 3.1). The optimality condition for $\eta = \bar{\eta}$ to solve the convex unconstrained problem (3.12) is:

$$\sum_k (\phi^*)'(v_k - \eta) = 1 \ . \tag{3.15}$$

But $(\phi^*)' = (\phi')^{-1}$ (see e.g. [24] section 26) and thus (3.15) is exactly (3.14). Now since $\phi \in \Phi : \phi'(0) = -\infty$ and $\phi'(1) < \infty$ implying $(\phi')^{-1}(-\infty) = 0$ and $(\phi')^{-1}(\infty) > 1$. Thus equation (3.14) has a solution $\bar{\eta} = \eta(\bar{v})$ for all $\bar{v}$, which is also unique since ϕ' (and hence $(\phi')^{-1}$) is strictly monotone. The statement concerning $\bar{p}$ follows immediately from (3.4). □

The dual problem (3.7) is an unconstrained discrete minimax problem. For such problems, many algorithms have been proposed in the non smooth optimization literature, see e.g. Wolfe [29], Lemarechal [18], [19] and more recently Kiwiel [17]. Alternatively, the dual problem (3.7) can be reformulated as a linearly constrainted convex program in $\mathbb{R}^{n+1}$. Indeed, by defining $v_{n+1} := \max_{1\le j\le n} \{\ell_j(v) + b_j\}$, problem (3.7) is equivalent to

$$\min_{v\in\mathbb{R}^{n+1}} f(v) + v_{n+1} \tag{3.16}$$

$$\text{s.t.} \qquad \ell_j(v) + b_j - v_{n+1} \le 0 \quad \text{for all} \quad j = 1,\dots,m$$

in which case many nonlinear programming codes are readily available to solve the dual formulation (3.16) (see e.g. Shittowsky [26] or the recent generalized reduced gradient code of Lasdon).

The special structure of the dual problem

$$\min_{v\in\mathbb{R}^n} \{f(v) + \max_{1\le j\le m} \{\ell_j(v) + b_j\}\}$$

suggests a method in which f is approximated at the r-th iteration by a polyhedral function (i.e. pointwise maximum of finitely many affine functions) $\pi_r(v)$, and the next iteration point v^{r+1} is the optimal solution of

$$\min_v\{\pi_r(v) + \max\{\ell_j(v) + b_j\}\} \quad . \tag{3.17}$$

Since (3.17) is a linear minimax problem it can be solved efficiently with simplex like algorithms (e.g. [4], [5]). The new polyhedral approximation $\pi_{r+1}(v)$ is obtained by

$$\pi_{r+1}(v) = \max\{\pi_r(v), s_{r+1}(v)\}$$

where $s_{r+1}(\cdot)$ is the affine support of f at v^{r+1}.

Let us derive the dual representation of the classical channel capacity C. This is done simply by substituting $\phi(t) = t \log t$ in problem (D_ϕ). The conjugate is $\phi^*(t^*) = e^{t^*-1}$ and so

$$f(v) = \min_{\eta\in\mathbf{R}}\left\{\eta + \sum_k e^{v_k-\eta-1}\right\} = \log \sum_{k=1}^n e^{v_k}$$

by simple calculus.

Then, using Theorem 3.1 a dual representation of C is given by

$$C = \min_{v\in\mathbf{R}^n}\left\{\log \sum_{k=1}^n e^{v_k} + \max_{1\le j\le m}\left\{\sum_k Q_{kj}\log Q_{kj} - \sum_k Q_{kj}v_k\right\}\right\} . \qquad (3.18)$$

In the last part of this section we present an alternative dual problem to (P_ϕ) which is given directly in terms of the problem's data.

Corollary 3.1. Let $\phi \in \Phi$, then

$$C = \min_{y\in\mathbf{P}_n} \max_{1\le j\le m}\left\{\sum_k \phi'(y_k)(y_k-Q_{kj}) + \phi(Q_{kj}) - \phi(y_k)\right\} . \qquad (3.19)$$

Proof: From the proof of theorem 3.2 it follows that $f(v)$ defined in (3.5) is given by:

$$f(v) = \eta(v) + \sum_k \phi^*(v_k - \eta(v)) \qquad (3.20)$$

where $\eta(v)$ is the unique solution of the equation

$$\sum_k (\phi^*)'(v_k-\eta) = 1 \ .$$

Define new variables $y_k = (\phi^*)'(v_k-\eta(v))$. Then $\sum_k y_k = 1$ and $y_k \geq 0$ since $(\phi^*)'(-\infty) = 0$ and $(\phi^*)'$ is increasing, hence $y \in \mathbf{P}_n$. From the definition of y_k , and using $(\phi^*)' = (\phi')^{-1}$:

$$v_k - \eta(v) = \phi'(y_k) \tag{3.21}$$

Substituting (3.21) in (3.7) and using (3.20) we obtain

$$C_\phi = \eta(v) + \Sigma\phi^*(\phi'(y_k)) + \max_{1\leq j\leq m} \sum_k \phi(Q_{kj}) - Q_{kj}(\eta(v) + \phi'(y_k)) \ .$$

But $\sum_k Q_{kj} = 1$, hence

$$C_\phi = \sum_k \phi^*(\phi'(y_k)) + \max_{1\leq j\leq m} \sum_k \phi(Q_{kj}) - Q_{kj}\phi'(y_k) \ . \tag{3.22}$$

Finally using in (3.22) a simple fact concerning conjugate functions:

$$\phi^*(\phi'(t)) = t\phi'(t) - \phi(t)$$

we get the derived result (3.19). □

Applying Corollary 3.1 to the classical case $\phi(t) = t \log t$ a little algebra show that a dual representation of C is

$$C = \min_{y\in\mathbf{P}_n} \max_{1\leq j\leq m} \sum_k Q_{kj} \log \frac{Q_{kj}}{y_k}$$

and we recover here a result given in Meister and Oettli [20]. □

4. UPPER BOUND FOR C_ϕ

The dual representation derived in corollary 3.2 is also useful to derive upper bound on C_ϕ .

Theorem 4.1. For any $\phi \in \Phi$,

$$C_\phi \le \sum_k \{\phi'(\sum_j \frac{Q_{kj}}{m}) \sum_j \frac{Q_{kj}}{m} - \phi(\sum_j \frac{Q_{kj}}{m})\} + \max_{1\le j\le m} \sum_k \phi(Q_{kj}) - Q_{kj}\phi'(\sum_j \frac{Q_{kj}}{m}). \quad (4.1)$$

Proof: From the dual representation (3.19)

$$C_\phi \le \max_{1\le j\le m} \left\{ \sum_{k=1}^{n} \phi'(y_k)(y_k - Q_{kj}) + \phi(Q_{kj}) - \phi(y_k) \right\} \quad (4.2)$$

for every y satisfying $\sum_{k=1}^{n} y_k = 1$ and $y_k \ge 0$, $k = 1,\dots,n$, in particular for $y_k = \frac{\sum_{j=1}^{m} Q_{kj}}{m}$. Substituting this special choice in (4.2), and rearrange term, we obtain the upper bound in the theorem. □

Example 4.1: Let us consider the family Φ_α with functions ϕ_α given by

$$\phi_\alpha(t) = \begin{cases} \frac{1}{1-\alpha}(t - t^\alpha) & 0 < \alpha < 1 \\ t \log t & \alpha = 1 \end{cases} .$$

We denote respectively by C_α and U_α the corresponding capacity and its upper bound. Using Theorem 4.1 we get

$$C_\alpha \le U_\alpha$$

with

$$U_\alpha = \begin{cases} \frac{1}{m^\alpha} \sum_k (\sum_j Q_{kj})^\alpha + \frac{1}{1-\alpha} \max_{1\le j\le m} \frac{\alpha}{m^\alpha - 1} \sum_k Q_{kj} (\sum_j Q_{kj})^{\alpha-1} - \sum_k Q_{kj}^\alpha & \text{if } 0<\alpha<1 \\ \log m + \max_{1\le j\le m} \sum_k Q_{kj} \log \frac{Q_{kj}}{\sum_\ell Q_{k\ell}} & \text{if } \alpha = 1 . \end{cases}$$

In particular we see that the classical lower bound derived in Arimoto [2] is recovered i.e.

$$C \equiv C_1 \le U_1 \quad .$$

An interesting special case for which the upper bound is attained is in the case of symmetric channel i.e. those with the same set of entries in columns and rows with possible permutations. In that case we have $\sum_{j=1}^{m} Q_{kj} = \text{const.} = \delta$ for all k .

Theorem 4.2. If the channel is symetric, then for any $\phi \in \Phi$, C_ϕ is equal to the upper bound, i.e.

$$C\phi = -n\phi(\tfrac{1}{n}) + \frac{1}{m} \sum_{j,k} \phi(Q_{kj}) \ . \tag{4.3}$$

Proof: From the primal formulation of C_ϕ (see (3.4)) a lower bound is given by

$$C_\phi \ge - \sum_k \phi(q_k) + \sum_{j=1}^{m} p_j \sum_{k=1}^{n} \phi(Q_{kj}) \tag{4.4}$$

for any $(p,q) \in \mathbf{P}_m \times \mathbf{P}_n$ satisfying:

$$\sum_j p_j Q_{kj} = q_k \quad , \quad k = 1,\dots,n \ . \tag{4.5}$$

Since the channel is symetric $\sum_j Q_{kj} = \delta$ for all k and this implies that $m = n\delta$. Thus $p_j^* = \frac{1}{m}$ and $q_k^* = \frac{1}{n}$ satisfy (4.5). Substituting (p^*,q^*) in (4.4) we obtain the lower bound

$$C_\phi \geq -n\phi(\tfrac{1}{n}) + \frac{1}{m} \sum_{j,k} \phi(Q_{kj}) \ . \tag{4.6}$$

From Theorem 4.1, using the fact $m = n\delta$ we have the upper bound:

$$C_\phi \leq -n\phi(\tfrac{1}{n}) + \max_j \sum_k \phi(Q_{kj}) \ . \tag{4.7}$$

Since the channel is symetric, it has the same set of entries in each column, thus $\sum_k \phi(Q_{kj}) = \gamma_j = \text{const.}$ for all j and hence,

$$\max_{1 \leq j \leq m} \sum_k \phi(Q_{kj}) = \frac{1}{m} \sum_{j,k} \phi(Q_{kj}) \ . \tag{4.8}$$

Therefore substituting (4.8) in (4.7), we see that the upper bound for C_ϕ coincides with its lower bound given in (4.6). □

<u>Example 4.2</u>. Consider the Binary Symetric Channel (BSC) defined by $Q = \begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}$. Using Theorem 4.2 we obtain. For any $\phi \in \Phi$,

$$C_\phi^{BSC} = -2\phi(\tfrac{1}{2}) + \phi(1-\beta) + \phi(\beta) \ .$$

In particular for $\phi(t) = t \log t$ we get the well-known result (see e.g. [15], [16]):

$$C^{BSC} = \log 2 + (1-\beta)\log(1-\beta) + \beta \log \beta \ .$$

5. CONCLUSIONS AND EXTENSION

A new formulation of the channel capacity problem has been obtained by using the duality theory of convex programming. This new dual representation seems useful for computational purposes and derivation of bounds. Furthermore the results in the paper demonstrate that the new information measure of Burbea and Rao [10] can be used successfully to develop a generalized channel capacity theory.

Finally we remark that our duality framework can be easily extended to the multiple constrained channel capacity problem (i.e. which include additional linear inequality constraints on the input probability p, see [5]) to produce simple dual formulation. Also, at the price of some additional technicalities, the continuous alphabet channel problem [7], [15] can be considered via a duality theory for infinite dimensional optimization problems framework to obtain the continuous version of Theorem 3.1 and its corollary.

References

[1] Aczel, J. (1984). "Measuring information beyond communication theory - Why some generalized information measures may be useful, others not", Aequationes Mathematicae, 27, pp. 1-19.

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

[3] Arimoto, S. (1977). "Information measures and capacity of order α for discrete memoryless channels", in Topics in Information Theory, Edited by I. Csiszar and P. Elias, North-Holland Publishing Company, Amsterdam, pp. 41-52.

[4] Armstrong, R.D. and Godfrey, J.R. (1979). "Two linear programming algorithms for the discrete ℓ_1 norm problem", Mathematics of Computation 33, pp. 289-300.

[5] Barrodale, I. and Roberts, F.D.K. (1973). "An improved algorithm for discrete ℓ_1-linear approximation", SIAM J. Numer. Anal., 10, pp. 839-848.

[6] Ben-Tal, A. and Teboulle, M. (1985). "Rate distortion theory with generalized information measures via convex programming duality", to appear in IEEE Transactions of Inform. Theory.

[7] Blahut, R.E. (1972). "Computation of channel capacity and rate distortion functions", IEEE Trans. on Information Theory, Vol. IT 28, pp. 489-495.

[8] Burbea, J. (1983). "J-Divergence and related topics". Encycl. Statist. Sci., 4, pp. 290-296, Wiley, New York.

[9] Burbea, J. (1984). "The Bose-Einstein Entropy of degree α and its Jensen difference", Utilitas Mathematica, 25, pp. 225-240.

[10] Burbea, J. and Rao, C.R. (1982). "On the convexity of some divergence measures based on entropy functions", IEEE Trans. on Information Theory, Vol. IT28, pp. 489-495.

[11] Burbea, J. and Rao, C.R. (1982). "Entropy differential metric, distance and divergence measures in probability spaces - a unified approach", J. of Multivariate Analysis, 12, pp. 575-596.

[12] Cheng, M.C. (1979). "On the computation of capacity of a discrete memoryless channel", Inform. Control. 24, pp. 292-298.

[13] Csiszar, I. (1967). "Information-Type Measures of Difference of Probability Distributions and Indirect Observations", Studia Sci. Mat. Hungar., 2, pp. 299-318.

[14] Csiszar, I. and Korner, J. (1981). "Information theory: coding theorems for discrete memoryless systems", New York, Academic Press.

[15] Gallager, R.G. (1968). "Information theory and reliable communication, New York, Wiley.

[16] Jelinek, F. (1968). "Probalistic information theory", New York, McGraw-Hill.

[17] Kiwiel, K.C. (1985). "Methods of descent for nondifferentiable Optimization", Lecture Notes in Math., No. 1133, Springer Verlag, Berlin.

[18] Lemarechal, C. (1975). "An extension of Davidon methods to nondifferentiable problems", Math. Progr. Study, 3, pp. 95-109.

[19] Lemarechal, C. (1980). "Nondifferentiable optimization" in Dixon, Spedicato and Szego (eds.) Nonlinear Optimization Theory and Algorithms, Birkhauser, Boston.

[20] Meister, B. and Oettli, W. (1967). "On the capacity of a discrete constant channel", Inform. Control 11, pp. 341-351.

[21] Muroga, S. (1953). "On the capacity of a discrete channel", J. Phys. Soc., Jpaan 8, pp. 484-494.

[22] Rao, C.R. and Nayak, T.K. (1985). "Cross entropy, dissimilarity measures, and characterizations of quadratic entropy", IEEE Trans. on Information Theory, 31, pp. 589-593.

[23] Renyi, A. (1961). "On Measures of Entropy and Information", in Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Vol. 1, pp. 547-561.

[24] Rockafellar, R.T. (1970). Convex Analysis, Princeton University Press, Princeton, N.J.

[25] Rockafellar, R.T. (1974). Conjugate Duality and Optimization, Regional Conference Series in Applied Mathematics, No. 16, SIAM.

[26] Shannon, C.E. (1948). "A mathematical theory of communication", Bell Syst. Tech. Journal, Vol. 27, pp. 379-423, pp. 623-656.

[27] Schittowski, K. (1980). "Nonlinear programming codes", Lecture Notes in Economics and Mathematical Systems, Vol. 183, Springer Verlag, Berlin.

[28] Takano, S. (1975). "On a method of calculating the capacity of a discrete memoryless channel", Inform. Control 29, pp. 327-336.

[29] Wolfe, P. (1975). "A method of conjugate subgradients for minimizing nondifferentiable convex functions functions", Math. Program. Study 3, pp. 145-173.

[30] Ziv., J., Zakai, M. (1973). "On functionals satisfying a data-processing theorem", IEEE Trans. on Information Theory, IT19, pp. 275-283.