CARDINALITY BOUNDS ON AUXILIARY VARIABLES IN MULTIPLE-USER THEORY VIA THE METHOD OF AHLSWEDE AND KÖRNER

by

Masoud Salehi

TECHNICAL REPORT NO. 33
AUGUST 1978

PREPARED UNDER THE AUSPICES
OR
NATIONAL SCIENCE FOUNDATION
GRANT ENG 76-03684

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Cardinality Bounds on Auxiliary Variables in Multiple-User Theory Via the Method of Ahlswede and Körner

Masoud Salehi†

ABSTRACT

It is necessary to bound the cardinality of the random variable U in expressions like

$$\max\{\lambda_1 I(Z;U) + \lambda_2 I(X;Y|U)\}$$

in order to show the computability of the capacity region in problems in multiple-user theory. The mappings suggested by Ahlswede and Körner, together with the standard theory of convexity, have previously been used to find bounds for source coding with side information. In this paper we apply these methods to some problems in multiple-user theory. Improved bounds are found using the theorem of Fenchel and Eggleston. Cardinality bounds are found for i) more capable broadcast channels, ii) broadcast channels with degraded messages, iii) multiple access channels with correlated sources, and iv) multiple access channels with feedback.

---

†Stanford University. This work was partially supported by the National Science Foundation Grant ENG76-03684.

I. Introduction

The description of a capacity region is in computable form if there exists an algorithm of finite length for computing it to any given degree of accuracy. An example of a computable description is Shannon's formula

$$C = \max_{p(x)} I(X;Y) ,$$

where $\underline{X}$ and $\underline{Y}$ are finite sets and $p(y|x)$ is given for all $x \varepsilon \underline{X}$, $y \varepsilon \underline{Y}$. Here C is expressed as the maximum of a continuous function over a compact set. The equally correct expression

$$C = \sup_{n} \sup_{P(X_1,\ldots,X_n)} \frac{1}{n} I(X_1,X_2,\ldots,X_n;Y_1,Y_2,\ldots,Y_n) \qquad (1)$$

is not in computable form as described above.

In many problems in multiple-user theory, capacity regions are described in terms of auxiliary random variables. In order to express these regions in computable form, it is necessary to bound the cardinality of these random variables.

Bounds have previously been found for degraded broadcast channels (Gallager [2]), source coding with side information (Ahlswede and Körner [3]), and the common information problem (Wyner [9]).

In this paper we will use a slightly modified version of the method of Ahlswede and Körner. This modification improves their result by 2. Then we will apply this technique to i) more capable broadcast channels, ii) broadcast channels with degraded messages, iii) multiple access channels with correlated sources and iv) multiple access channels with feedback.

II. A Modified Version of the Technique of Ahlswede and Körner

The key point in Ahlswede and Körner's method is a clever use of the Caratheodory's Theorem as stated below.

Caratheodory's Theorem: Let X be a subset of R^n, let $C(X)$ denote the convex hull of X, and let $\underline{y}$ be a point of $C(X)$; then there exists a set of s points $\underline{X}_1,\ldots,\underline{X}_s$, all belonging to X, with $s \leq n+1$, such that $\underline{y}$ is a point of the simplex whose vertices are $\underline{X}_1,\ldots,\underline{X}_s$.

Proof: See [4], page 35.

We will state below the closely related theorem of Fenchel and Eggleston. The use of this theorem enables us to improve Ahlswede and Körner's bound by 1.

Theorem (Fenchel-Eggleston): If in the conditions of Caratheodory's theorem it is also assumed that the set X is the union of at most n connected sets then $s \leq n$.

Proof: See [4], page 35.

Ahlswede and Körner establish the following lemma of crucial importance in finding cardinality bounds:

Lemma 1 (Ahlswede and Körner): Let P be any subset of R^n, and let $f_j(\underline{P})$, $j=1,2,\ldots,k$ be real valued functions on P. Then to any probability measure μ (on the Borel subsets of) P there exists $k+1$ elements $\underline{P}_i$ of P and constants $\alpha_i \geq 0$, $i=1,2,\ldots,k$, $\sum_{i=1}^{k+1} \alpha_i = 1$, such that

$$\int f_j(\underline{P})\, \mu(d\underline{P}) = \sum_{i=1}^{k+1} \alpha_i\, f_j(\underline{P}_i)\ , \quad j = 1,2,\ldots,k\ .$$

Proof: See [3], Lemma 3.

However, using the theorem of Fenchel-Eggleston, we can improve the bound in Lemma 1 from $k+1$ to k, under conditions that continue to hold true for the channel cardinality theorems.

Lemma 2: Let $\mathcal{P}$ be any subset of R^n consisting of at most k connected subsets. Let $f_j(\underline{P})$, $j = 1,2,\ldots,k$ be real valued continuous functions on $\mathcal{P}$. Then to any probability measure μ (on the Borel subsets of) $\mathcal{P}$ there exists k elements $\underline{P}_i$ of $\mathcal{P}$ and constants $\alpha_i \geq 0$, $i = 1,2,\ldots,k$, $\sum_{i=1}^{k} \alpha_i = 1$, such that

$$\int f_j(\underline{P})\, d\mu = \sum_{i=1}^{k} \alpha_i\, f_j(\underline{P}_i) , \quad j = 1,2,\ldots,k .$$

Proof: The image of a set S consisting of at most K connected subsets under a continuous mapping $\underline{f}: S \to \mathbb{R}^k$ is again a set consisting of at most k connected subsets. This shows that the set

$$\mathcal{D} \overset{\Delta}{=} \left\{ (f_1(\underline{P}), f_2(\underline{P}), \ldots, f_k(\underline{P})) \in R^k : \underline{P} \in \mathcal{P} \right\}$$

consists of at most k connected subsets.

For any measure μ on $\mathcal{P}$, the point

$$\left(\int f_1(\underline{P})\mu(d\underline{P}), \int f_2(\underline{P})\mu(d\underline{P}), \ldots, \int f_k(\underline{P})\mu(d\underline{P}) \right)$$

belongs to the convex hull of $\mathcal{D}$. Now we can apply the theorem of Fenchel and Eggleston and find k points in $\mathcal{P}$, namely $\underline{P}_i$, $i = 1,2,\ldots,k$, and k scalars α_i, $i = 1,2,\ldots,k$, $\alpha_i \geq 0$ and $\sum_{i=1}^{k} \alpha_i = 1$, such that

$$\int f_j(\underline{P})\mu(d\underline{P}) = \sum_{i=1}^{k} \alpha_i\, f_j(\underline{P}_i) , \quad j = 1,2,\ldots,k .$$

Application: In the usual Shannon channel, assume $X = \{1,2,\ldots,\|X\|\}$ is the input alphabet and $Y = \{1,2,\ldots,\|Y\|\}$ is the output alpahbet. Then in order to achieve capacity it is enough to put positive probability on at most $\min\{\|X\|,\|Y\|\}$ input letters.

Proof: It is obvious that the number of input letters with positive probability under the capacity-achieving distribution is less than or equal to $\|X\|$. It remains to prove this number is also less than or equal to $\|Y\|$.

To show this, let $\underline{P}^*$ be the probability vector on X, the input alphabet, that achieves capacity. Then

$$C = H^*(Y) - H^*(Y|X) .$$

We will prove that there exists another probability vector putting positive probability on at most $\|Y\|$ letters of X and achieving both $H^*(Y)$ and $H^*(Y|X)$.

Let P be the set consisting of all extreme points of the $\|X\|$-dimensional probability simplex. Let $\underline{P} \varepsilon C(P)$ be a probability distribution on X, and let $\underline{P}^* \varepsilon C(P)$ be the probability distribution achieving capacity. Then there exists a measure $\mu^*(d\underline{P})$ on P such that

$$\int_P \underline{P}\, \mu^*(d\underline{P}) = \underline{P}^* .$$

Define $H^*(Y)$ to be the entropy of Y when $\underline{P}^*$ is the probability distribution on X. For each $\underline{P} \varepsilon P$ define

$$f_1(\underline{P}) = P_Y(1)$$

$$f_2(\underline{P}) = P_Y(2)$$

$$\vdots$$

$$f_{\|Y\|-1}(\underline{P}) = P_Y(\|Y\|-1)$$

$$f_{\|Y\|}(\underline{P}) = H_{\underline{P}}(Y|X)$$

where $H_{\underline{P}}(Y|X)$ is the entropy of Y given X when $\underline{P}$ is the probability distribution on X. Note that

$$\int f_j(\underline{P})\mu^*(d\underline{P}) = P_Y^*(j) \ , \quad j = 1,2,\ldots,\|Y\|-1 \tag{2}$$

and

$$\int f_{\|Y\|}(\underline{P})\mu^*(d\underline{P}) = \int H_{\underline{P}}(Y|X)\mu^*(d\underline{P}) = H^*(Y|X) \tag{3}$$

Now applying Lemma 2 we see that there are $\|Y\|$ extreme points $\underline{P}_1, \underline{P}_2, \ldots, \underline{P}_{\|Y\|} \in \mathcal{P}$ and non-negative numbers α_i , $i = 1,2,\ldots,\|Y\|$, such that

$$P_Y^*(j) = \sum_{i=1}^{\|Y\|} \alpha_i \ f_j(\underline{P}_i) \ , \quad j = 1,2,\ldots,\|Y\|-1 \tag{4}$$

$$H^*(Y|X) = \sum_{i=1}^{\|Y\|} \alpha_i \ f_{\|Y\|} \ (\underline{P}_i) \ , \tag{5}$$

i.e., there exists a probability mass function $(\alpha_1, \alpha_2, \ldots, \alpha_{\|Y\|})$ putting positive probability on at most $\|Y\|$ letters of X and resulting in the same $H^*(Y|X)$ and $H^*(Y)$ as does $\underline{P}^*$. This completes the proof.

Another gain in the cardinality bound in multiple-user problems follows from the convexity of the capacity regions and the fact that a convex set can be described by its supporting hyperplanes. The following lemma clears this point.

Lemma 3: Let for $i = 1,2,\ldots,n$, $g_i : R^m \to R$ be a set of n positive valued functions, let $T \subseteq R^m$ be a closed set, and define

$$E \overset{\Delta}{=} \left\{ \underline{X} \in R^n : \underline{X} \geq \underline{0},\ \exists \underline{t} \in T\ ;\ x_i \leq g_i(\underline{t}),\ i = 1,2,\ldots,n \right\}.$$

Let $E' = C(E)$ denotes the convex hull of E, and let

$$E'' \overset{\Delta}{=} \left\{ \underline{X} \in R^n,\ \underline{X} \geq \underline{0} : \forall\ \underline{\lambda} \in R^n,\ \underline{\lambda} \geq \underline{0},\ \underline{\lambda}^t\ \underline{X} \leq G(\underline{\lambda}) \right\}, \qquad (6)$$

where $$G(\underline{\lambda}) \overset{\Delta}{=} \sup_{\underline{t} \in T} \underline{\lambda}^t\ [g_1(\underline{t}), g_2(\underline{t}), \ldots, g_n(\underline{t})]^t .$$

Then $E' = E''$.

Proof:

i) E'' is convex. This is easy to check from the definition of E''.

ii) $E \subset E''$. Let $\underline{X} \in E$. Then $\exists\ \underline{t}^* \in T$ such that $0 \leq x_i \leq g_i(\underline{t}^*)$, $i = 1,2,\ldots,n$. For $\underline{\lambda} \geq \underline{0}$, we have

$$\underline{\lambda}^t\ \underline{X} \leq \underline{\lambda}^t\ [g_1(\underline{t}^*), g_2(\underline{t}^*), \ldots, g_n(\underline{t}^*)]^t \leq G(\underline{\lambda}) ,$$

therefore $\underline{X} \in E''$.

From i) and ii) it follows that $E' \subset E''$.

iii) To prove $E'' \subset E'$ it is enough to show $\underline{y} \notin E'$ implies $\underline{y} \notin E''$.

It is easy to see that E' is a closed set lying in the first orthant and is closed under projection on the coordinate axes. It is bounded by n

coordinate hyperplanes and an upper surface. Now let $\underline{y}$ be in the first orthant, but $\underline{y} \notin E'$. Then the line joining $\underline{y}$ to the origin cuts the upper surface of E' at a point $\underline{X}$. By construction, and since E' is closed, $\underline{y} > \underline{X}$. Since E' is closed under projection on the coordinate axes, the supporting hyperplane at $\underline{X}$ is of the form $\underline{\lambda}^t \underline{X} = 1$, where $\underline{\lambda} > \underline{0}$, and for all $\underline{Z} \in E'$ we have $\underline{\lambda}^t \underline{Z} \leq 1$. From $\underline{y} > \underline{X}$, we have

$$\underline{\lambda}^t \underline{y} > \underline{\lambda}^t \underline{X} = 1 \geq \underline{\lambda}^t \underline{Z} \quad \text{for all } \underline{Z} \in E' . \tag{7}$$

Thus $\underline{\lambda}^t \underline{y} > G(\underline{\lambda})$. Therefore $\underline{y} \notin E''$.

<u>Corollary</u>: The function $G(\underline{\lambda})$ completely describes the region E' in Lemma 3 through Equation (6).

<u>Application</u>: Source coding with side information: $\|U\| \leq \|X\|$.

Let $\{(X_i, Y_i)\}_{i=0}^{\infty}$ be a sequence of i.i.d. drawings from a source with joint probability distribution $P(x,y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Wyner [1] and Ahlswede and Körner [3] have shown that the following is the achievable rate region for noiseless coding of the X and Y processes for error free reconstruction of the Y process

$$R = \left\{ (R_X, R_Y) : R_X \geq I(X;U),\ R_Y \geq H(Y|U) \right\}, \tag{8}$$

where U is a discrete random variable with $\|U\| \leq \|X\| + 2$ and

$$p(x,y,u) = p(u)p(x|u)p(y|x) . \tag{9}$$

We shall improve the bound as follows. By Lemma 3, we can write

$$R = \left\{ \underline{R} = (R_X, R_Y) : \forall\ \underline{\lambda} \in R^2,\ \underline{\lambda} \geq \underline{0},\ \underline{\lambda}^t \underline{R} \geq G(\underline{\lambda}) \right\} , \tag{10}$$

where $$G(\underline{\lambda}) = \inf \underline{\lambda}^t \, [I(X;U), H(Y|U)]^t \, , \tag{11}$$

and the infimum is over all discrete random variables U where $U \to X \to Y$ form a Markov chain. Knowledge of $G(\underline{\lambda})$ is sufficient information to get R, and we will prove for all $\underline{\lambda}$ that $G(\underline{\lambda})$ can be achieved by considering those U's with cardinality less than or equal to $\|X\|$.

Fix $\underline{\lambda} = (\lambda_1, \lambda_2) \geq 0$ and let $\mathcal{P}$ in Lemma 2 be the $\|X\|$- dimensional probability simplex. Let $\mathcal{X} = \{1,2,\ldots,\|X\|\}$ be the range of X and interpret

$$\underline{P} = \Big(P(X=1|U=u), p(X=2|U=u), \ldots, P(X=\|X\|\,|U=u)\Big)$$

as a point in $\mathcal{P}$. Then each probability distribution on U defines a measure $\mu(d\underline{P})$ on $\mathcal{P}$. Let $P_X^*(\cdot)$ achieve $G(\underline{\lambda})$ and let $\mu^*(d\underline{P})$ achieve $P_X^*(\cdot)$ (and thus $G(\underline{\lambda})$). Denote by $H^*(X)$ the entropy of X under probability distribution $P_X^*(\cdot)$, and define

$$f_j(\underline{P}) = P_X(j) \;, \quad j = 1,2,\ldots,\|X\|-1 \tag{12}$$

$$f_{\|X\|}\,(\underline{P}) = -\lambda_1 H_{\underline{P}}(X) + \lambda_2 H_{\underline{P}}(Y) \tag{13}$$

where $H_{\underline{P}}(X)$ and $H_{\underline{P}}(Y)$ are the entropies of X and Y respectively when the distribution of X is $\underline{P}$. Noting that

$$\int H_{\underline{P}}(X)\mu^*(d\underline{P}) = H^*(X|U) \tag{14}$$

$$\int H_{\underline{P}}(Y)\mu^*(d\underline{P}) = H^*(Y|U) \, , \tag{15}$$

and using Lemma 2 it is seen that there exist $\|X\|$ elements $\underline{P}_i \in \mathcal{P}$, $i = 1,2,\ldots,\|X\|$, and constants

$$\alpha_i \geq 0 \ , \ i = 1,2,\ldots,\|X\| \ , \ \sum_{i=1}^{\|X\|} \alpha_i = 1$$

such that

$$P_X^*(j) = \sum_{i=1}^{\|X\|} \alpha_i \ f_j(\underline{P}_i) \ , \ j = 1,2,\ldots,\|X\|-1 \tag{16}$$

$$\lambda_2 H^*(Y|U) - \lambda_1 \ H^*(X|U) = \sum_{i=1}^{\|X\|} \alpha_i \ f_{\|X\|}(\underline{P}_i) \ . \tag{17}$$

From $P_X^*(j)$, $j = 1,2,\ldots,\|X\|-1$, we can compute $H^*(X)$ and can form

$$\lambda_1 \ H^*(X) + \lambda_2 \ H^*(Y|U) - \lambda_1 \ H^*(X|U) = \lambda_1 \ I^*(X;U) + \lambda_2 \ H^*(Y|U)$$

$$= G(\underline{\lambda}) \ . \tag{18}$$

This means that in order to achieve $G(\underline{\lambda})$ it is enough to put positive probability on at most $\|X\|$ elements of P . This in turn shows it is enough to consider U's with $\|U\| \leq \|X\|$.

III. Some New Cardinality Bounds

In this section we apply results of Section II to some problems in multiple-user theory. In these cases

$$G(\underline{\lambda}) = \sup \underline{\lambda}^t \ \underline{I} \ , \tag{19}$$

where $\underline{I}$ and the set on which sup is defined are given in each case.

a) More capable broadcast channels, $\|U\| \leq \|X\|$:

The capacity region given in El Gamal [5] is

$$C = \Big\{ (R_0,R_1,R_2) : R_0 + R_1 + R_2 \le I(X;Y)$$

$$R_0 + R_1 + R_2 \le I(X;Y|U) + I(U;Z)$$

$$R_0 + R_2 \le I(U;Z) \Big\} \tag{20}$$

where $p(u,x,y,z) = p(u)p(x|u)p(y,z|x)$. By Lemma 3, all information necessary to describe C is in

$$G(\underline{\lambda}) = \sup \underline{\lambda}^t \, \underline{I} \quad \text{where} \quad \underline{I} = (I(X;Y), I(X;Y|U) + I(U;Z), I(U;Z))^t,$$

$$C = \Big\{ (R_0,R_1,R_2) \ge 0 : \forall\, \underline{\lambda} > \underline{0}\, , \underline{\lambda}^t(R_0+R_1+R_2, R_0+R_1+R_2, R_0+R_2)^t \le G(\underline{\lambda}) \Big\},$$

and sup in $G(\underline{\lambda})$ is over all r.v.'s U such that

$$p(u,x,y,z) = p(u)p(x|u)p(y,z|x) \ . \tag{21}$$

For each $\underline{P}$ in the $\|X\|$ – dimensional simplex $P_{\|X\|}$ define

$$f_j(\underline{P}) = P_X(j) \ , \quad j = 1,2,\ldots,\|X\|-1 \tag{22}$$

$$f_{\|X\|}(\underline{P}) = \lambda_1[I_{\underline{P}}(X;Y) - H_{\underline{P}}(Y)] + \lambda_2[I_{\underline{P}}(X;Y) - H_{\underline{P}}(Z)] - \lambda_3\, H_{\underline{P}}(Z) \tag{23}$$

then
$$\int f_j(\underline{P})\mu^*(d\underline{P}) = P_X^*(j) \ , \ j = 1,2,\ldots,\|X\|-1 \ .$$

From $P_X^*(j)$, $j = 1,2,\ldots,\|X\|-1$, we can calculate $H^*(X)$, $H^*(Y)$, $H^*(Z)$ and then form

$$\int f_{\|X\|}(\underline{P})\mu^*(d\underline{P}) + \lambda_1 H^*(Y) + \lambda_2 H^*(Z) + \lambda_3 H^*(Z)$$

$$= \lambda_1[I^*(X;Y|U) - H^*(Y|U) + H^*(Y)] + \lambda_2[I^*(X;Y|U) -$$

$$- H^*(Z|U) + H^*(Z)] + \lambda_3[H^*(Z) - H^*(Z|U)]$$

$$= \lambda_1 I^*(X;Y) + \lambda_2[I^*(X;Y|U) + I^*(Z;U)] + \lambda_3 I^*(Z;U)$$

$$= G(\underline{\lambda}) \ . \tag{24}$$

Applying Lemma 2, we conclude that $\|U\| \le \|X\|$ is enough to describe C .

b) Broadcast channels with degraded messages: $\|U\| \le \|X\|$

Körner and Marton [6] have shown for broadcast channels with degraded messages that the capacity region is given by

$$C = \Big\{(R_1,R_0) : \ R_1 \le I(X;Y|U)$$

$$R_0 \le I(U;Z)$$

$$R_0 + R_1 \le I(X;Y)\Big\}$$

where U is a random variable such that

$$p(u,x,y,z) = p(u)p(x|u)p(y,z|x) \ . \tag{25}$$

By Lemma 3, we have

$$C = \Big\{(R_1,R_0) : \forall\ \underline{\lambda} > \underline{0}\ ,\ \underline{\lambda}^t(R_1,R_0,R_1+R_0)^t \le G(\underline{\lambda})\Big\}$$

$$G(\underline{\lambda}) = \sup \underline{\lambda}^t \, [I(X;Y|U),\ I(U;Z),\ I(X;Y)]^t$$

where sup is over all r.v.'s U such that (25) is satisfied. For $\underline{P} \in \mathcal{P}_{\|X\|}$, define

$$f_j(\underline{P}) = P_X(j) \ , \quad j = 1,2,\ldots,\|X\|-1$$

$$f_{\|X\|}(\underline{P}) = \lambda_1 \, I_{\underline{P}}(X;Y) - \lambda_2 \, H_{\underline{P}}(Z) + \lambda_3[I_{\underline{P}}(X;Y) - H_{\underline{P}}(Y)] \ .$$

Noting that $\int f_j(\underline{P})\mu^*(d\underline{P}) = P_X^*(j) \ , \quad j = 1,2,\ldots,\|X\|-1$

and that knowledge of $P_X^*(j)$, $j = 1,2,\ldots,\|X\|-1$ is enough to compute $H^*(Z)$ and $H^*(Y)$, we can compute $G(\underline{\lambda})$, i.e.,

$$\lambda_2 H^*(Z) + \lambda_3 \, H^*(Y) + \int f_{\|X\|}(\underline{P})\mu^*(d\underline{P})$$

$$= \lambda_1 I^*(X;Y|U) + \lambda_2[H^*(Z) - H^*(Z|U)] + \lambda_3[I^*(X;Y|U) - H^*(Y|U) + H^*(Y)]$$

$$= \lambda_1 \, I^*(X;Y|U) + \lambda_2 \, I^*(Z;U) + \lambda_3 \, I^*(X;Y)$$

$$= G(\underline{\lambda}) \ . \qquad (26)$$

Then Lemma 2 applies, and $\|U\| \leq \|X\|$ is seen to be sufficient to achieve $G(\underline{\lambda})$.

c) <u>Multiple access channels with correlated sources: $\|U\| \le \min\{\|Y\|, \|X_1\| \cdot \|X_2\|\}$</u>

The capacity region for this case is given in Slepian and Wolf [7] by

$$C = \Big\{ (R_0, R_1, R_2) : R_1 \le I(X_1;Y|X_2,U)$$

$$R_2 \le I(X_2;Y|X_1,U)$$

$$R_1 + R_2 \le I(X_1,X_2;Y|U)$$

$$R_0 + R_1 + R_2 \le I(X_1,X_2;Y) \Big\} \qquad (27)$$

where

$$p(u,x_1,x_2,y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1,x_2) \ . \qquad (28)$$

By using Lemma 3 we have the equivalent description

$$C = \Big\{ (R_0,R_1,R_2) : \forall\ \underline{\lambda} \ge \underline{0},\ \underline{\lambda}^t (R_1, R_2, R_1 + R_2, R_0 + R_1 + R_2) \le G(\underline{\lambda}) \Big\},$$

where

$$G(\underline{\lambda}) = \sup \underline{\lambda}^t \Big[I(X_1;Y|X_2,U), I(X_2;Y|X_1,U), I(X_1,X_2;Y|U), I(X_1,X_2:Y) \Big]^t$$

and sup is over all r.v.'s U such that (28) is satisfied. We develop two bounds on $\|U\|$.

i) Let $\mathcal{X}_1 = \{1,2,\ldots,\|X_1\|\}$ and $\mathcal{X}_2 = \{1,2,\ldots,\|X_2\|\}$ be the input alphabets and $\mathcal{Y} = \{1,2,\ldots,\|Y\|\}$ be the output alphabet. Let $\mathcal{P}_{\|X_1\| \cdot \|X_2\|}$ be the subset of the $\|X_1\| \cdot \|X_2\|$ – dimensional simplex representing independent probability distribution on $\mathcal{X}_1$, $\mathcal{X}_2$. Then each $\underline{P} \in \mathcal{P}_{\|X_1\| \cdot \|X_2\|}$ defines a distribution on $\mathcal{X}_1$ and $\mathcal{X}_2$, say $P_{X_1X_2}(\cdot,\cdot)$. This distribution has $\|X_1\|\cdot\|X_2\|$ components. For $i = 1,2,\ldots,\|X_1\|\cdot\|X_2\|-1$, let $f_i(\underline{P})$ denote all components of $P_{X_1X_2}(\cdot,\cdot)$ except one, say $P_{X_1X_2}(1,1)$, and let

$$f_{\|X_1\|\cdot\|X_2\|}(\underline{P}) = \lambda_1 \; I_{\underline{P}}(X_1;Y|X_2) + \lambda_2 \; I_{\underline{P}}(X_1;Y|X_1) + \lambda_3 \; I_{\underline{P}}(X_1,X_2;Y) \; .$$

Note that knowledge of $P^*_{X_1X_2}(\cdot,\cdot)$ is enough to compute $I^*(X_1,X_2;Y)$ and that

$$\int f_j(\underline{P})\mu^*(d\underline{P}) \; , \; j = 1,2,\ldots,\|X_1\|\cdot\|X_2\|-1$$

are enough to compute $P^*_{X_1X_2}(\cdot,\cdot)$. Now we can form

$$\int f_{\|X_1\|\cdot\|X_2\|}(\underline{P})\mu^*(d\underline{P}) + \lambda_4 \; I^*(X_1,X_2;Y)$$

$$= \lambda_1 \; I^*(X_1;Y|X_2,U) + \lambda_2 \; I^*(X_2;Y|X_1,U) + \lambda_3 \; I^*(X_1,X_2;Y|U)$$

$$+ \lambda_4 \; I^*(X_1,X_2;Y)$$

$$= G(\underline{\lambda}) \; . \tag{29}$$

Applying Lemma 2 we conclude that $\|U\| \le \|X_1\|\cdot\|X_2\|$ is enough to achieve $H(\underline{\lambda})$.

ii) Now focus attention on the output Y . Let $\mathcal{P}_{\|X_1\|\cdot\|X_2\|}$ be as in part i). Then for each $\underline{P} \in \mathcal{P}_{\|X_1\|\cdot\|X_2\|}$, $P_Y(\cdot)$ is uniquely determined and is a continuous function of $\underline{P}$. Let

$$f_j(\underline{P}) = P_Y(j) \; , \; j = 1,2,\ldots,\|Y\|-1 \; ,$$

$$f_{\|Y\|}(\underline{P}) = \lambda_1 \; I_{\underline{P}}(X_1;Y|X_2) + \lambda_2 \; I_{\underline{P}}(X_2;Y|X_1) + \lambda_3 \; I_{\underline{P}}(X_1,X_2;Y) + $$
$$+ \lambda_4[I_{\underline{P}}(X_1,X_2;Y) - H_{\underline{P}}(Y)] \; .$$

Then

$$\int f_{\|Y\|}(\underline{P})\mu^*(d\underline{P}) + \lambda_4 \; H^*(Y) = \lambda_1 \; I^*(X_1;Y|X_2,U) + \lambda_2 \; I^*(X_2;Y|X_1,U) +$$
$$+ \lambda_3 \; I^*(X_1,X_2;Y|U) + \lambda_4[I^*(X_1,X_2;Y|U) - H^*(Y|U) + H^*(Y)]$$
$$= \lambda_1 \; I^*(X_1;Y|X_2,U) + \lambda_2 \; I^*(X_2;Y|X_1,U) + \lambda_3 \; I^*(X_1,X_2;Y|U)$$
$$+ \lambda_4 \; I^*(X_1,X_2;Y)$$
$$= G(\underline{\lambda}) \; . \qquad (30)$$

Now Lemma 2 applies and thus $\|U\| \le \|Y\|$ is sufficient.

From i) and ii) we conclude that

$$\|U\| \le \min\{\|Y\|, \|X_1\| \cdot \|X_2\|\}$$

is sufficient to achieve capacity.

d) <u>Multiple access channels with feedback: $\|U\| \le \min\{\|Y\|, \|X_1\| \cdot \|X_2\|\}$</u>

An achievable rate region for this problem as given in Cover and Leung-Yan-Cheong [8] is

$$R = \Big\{ (R_1,R_2) : \; R_1 \le I(X_1;Y|X_2,U)$$
$$R_2 \le I(X_2;Y|X_1,U)$$
$$R_1 + R_2 \le I(X_1,X_2;Y) \Big\}$$

where U is a random variable and

$$p(u,x_1,x_2,y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1,x_2) \ . \tag{31}$$

Using Lemma 3, the alternative description of $\mathcal{R}$ is

$$\mathcal{R} = \left\{(R_1,R_2) : \forall \ \underline{\lambda} \geq \underline{0} \ , \ \underline{\lambda}^t(R_1,R_2,R_1 + R_2)^t \leq G(\underline{\lambda}) \right\}$$

where

$$G(\underline{\lambda}) = \sup \underline{\lambda}^t \left[I(X_1;Y|X_2,U), \ I(X_2;Y|X_1,U), \ I(X_1,X_2;Y) \right]^t$$

and the supremum is over all r.v.'s U such that (31) is satisfied.

i) Let $\underline{P}$ and $f_j(\underline{P})$, $j = 1,2,\ldots,\|X_1\|\cdot\|X_2\|-1$, be as in part i) of c) and let

$$f_{\|X_1\|\cdot\|X_2\|}(\underline{P}) = \lambda_1 \ I_{\underline{P}}(X_1;Y|X_2) + \lambda_2 \ I_{\underline{P}}(X_2;Y|X_1) \ .$$

It is easy to see

$$\int f_{\|X_1\|\cdot\|X_2\|}(\underline{P})\mu^*(d\underline{P}) + \lambda_3 \ I^*(X_1,X_2;Y) = G(\underline{\lambda}) \ ,$$

which shows $\|U\| \leq \|X_1\|\cdot\|X_2\|$ is sufficient to achieve capacity.

ii) With $\underline{P}$ and $f_j(\underline{P})$ as in part ii) of c), let

$$f_{\|Y\|}(\underline{P}) = \lambda_1 \ I_{\underline{P}}(X_1;Y|X_2) + \lambda_2 \ I_{\underline{P}}(X_2;Y|X_1) + \lambda_3[I_{\underline{P}}(X_1,X_2;Y) - H_{\underline{P}}(Y)] \ .$$

Then we can see that

$$\int f_{\|Y\|}(\underline{P})\mu^*(d\underline{P}) + \lambda_3 H^*(Y) = \lambda_1 I^*(X_1;Y|X_2,U) + \lambda_2 I^*(X_2;Y|X_1,U)$$

$$+ \lambda_3[I^*(X_1,X_2;Y|U) + H^*(Y) - H^*(Y|U)]$$

$$= G(\underline{\lambda}) . \tag{32}$$

Now from i) and Eq. (32), we conclude that $G(\underline{\lambda})$ can be achieved with $\|U\| \leq \min\{\|Y\|, \|X_1\|\cdot\|X_2\|\}$.

IV. Conclusions

In this work, we have used a slight modification of the method of Ahlswede and Körner to get bounds on the cardinality of the auxiliary random variables in the definitions of capacity regions in multiple-user theory. The bounds obtained are natural in the sense that they involve only the cardinalities of input and output alphabets and no constants appear in them. The problem of the tightness of these bounds remains open.

The following table compares the new bounds with the bounds obtained previously.

Problem	Previous Bounds	New Bounds
Source coding with side information	$\|U\| \leq \|X\| + 2$ (Ahlswede and Körner)	$\|U\| \leq \|X\|$
Broadcast channels with degraded messages	$\|U\| \leq \|X\| + 2$ (Körner and Marton)	$\|U\| \leq \|X\|$
More capable broadcast channels	$\|U\| \leq \|X\| + 2$ (A. El Gamal)	$\|U\| \leq \|X\|$
Multiple access channels with correlated sources	$\|U\| \leq \|X_1\|\cdot\|X_2\| + 2$ (Slepian and Wolf)	$\|U\| \leq \min\{\|X_1\|\cdot\|X_2\|, \|Y\|\}$
Multiple access channels with feedback		$\|U\| \leq \min\{\|X_1\|\cdot\|X_2\|, \|Y\|\}$

Acknowledgement:

The author wishes to thank Professor Thomas M. Cover of Stanford University for various helpful suggestions and comments.

References

[1] Wyner, A.D., "On Source Coding with Side Information at the Decoder," IEEE Trans. on Information Theory, Vol. IT-21, pp. 294-300, May 1975.

[2] R.G. Gallager, "Capacity and Coding for Degraded Broadcast Channels," Problemy Peredachi Informatsii, Vol. 10, pp. 3-14, July-Sept. 1974.

[3] R.F. Ahlswede and J. Körner, "Source Coding with Side Information and a Converse for Degraded Broadcast Channels," IEEE Trans. on Information Theory, Vol. IT-21, pp. 629-632, November 1975.

[4] H.G. Eggleston, Convexity, Cambridge University Press, N.Y. 1963.

[5] A. El Gamal, "The Capacity of a Class of Broadcast Channels," Technical Report No. 24, Dept. of Statistics, Stanford University, CA.

[6] J. Körner and K. Marton, "General Broadcast Channels with Degraded Message Sets," IEEE Trans. on Information Theory, Vol. IT-23, pp. 60-64, January 1977.

[7] D. Slepian and J.K. Wolf, "A Coding Theorem for Multiple Access Channels with Correlated Sources," Bell Systems Technical Journal, Vol. 52, pp. 1037-1076, September 1973.

[8] T.M. Cover and S.K. Leung-Yan-Cheong, "A Rate Region for Multiple Access Channel with Feedback," submitted to IEEE Trans. on Information Theory.

[9] A.D. Wyner, "The Common Information of Two Dependent Random Variables," IEEE Trans. on Information Theory, Vol. IT-21, pp. 163-179, March 1975.