R. AHLSWEDE*, J. KÖRNER

On the connection between the entropies of input and output distributions of discrete memoryless channels

1. Introduction

Let A and B be finite sets and let $w(.|.)$ be a stochastic $|A|\times|B|$ — matrix. The transmission probabilities of a discrete memoryless channel (d.m.c.) $\mathfrak{D}$ are defined by

(1.1) $$P(b^n|a^n)=\prod_{t=1}^{n} w(b_t|a_t) \text{ for every } a^n=(a_1, \ldots, a_n)\in A^n=\prod_1^n A,$$
$$b^n=(b_1, \ldots, b_n)\in B^n=\prod_1^n B \text{ and every } n\in N=\{1, 2, 3, \ldots\}.$$

A is the input and B the output alphabet of this channel. The elements of A^n (resp. B^n) are the input (resp. output) words of length n.

Let p^n be a probability distribution (p.d.) on A^n and q^n the corresponding p.d. on B^n, that is,

(1.2) $$q^n(b^n)=\sum_{a^n\in A^n} p^n(a^n)\, P(b^n|a^n) \text{ for all } b^n\in B^n.$$

We shall denote by H the entropy function of random variables and we use the same letter also for the entropy of p.d.'s.

Finally, we frequently use the function h, defined by

(1.3) $$h(\lambda)=-\lambda\log\lambda-(1-\lambda)\log(1-\lambda),\ 0\le\lambda\le\frac{1}{2}.$$

In [1] Wyner and Ziv proved the very interesting Theorem below, which establishes a connection between the entropies $H(p^n)$ and $H(q^n)$ in case $\mathfrak{D}$ is a *binary symmetric channel*. In [2] this result was applied to various multi-user communication problems.

Theorem (Wyner and Ziv [1])

Let $\mathfrak{D}$ be a *binary symmetric channel* (b.s.c.) with "crossover" probability p_0, then for all $n\in N$ and all p.d. p^n on A^n:

(1.4) $$\frac{1}{n}H(p^n)\ge h(\lambda)\Rightarrow\frac{1}{n}H(q^n)\ge h(\lambda(1-p_0)+(1-\lambda)p_0)\ (0\le\lambda\le 1).$$

---

* Research of this author was supported by the National Science Foundation under Grant No. GK-40492.

The bound to the right in (1.4) is clearly best possible and is already obtained for $n=1$. Defining for any real number c, $0 \le c \le \log|A|$, and any $n \in N$ the function $f^n(c)$ by

$$(1.5) \qquad f^n(c) = \min_{p^n: \frac{1}{n} H(p^n) \ge c} \frac{1}{n} H(q^n)$$

we can state (1.4) in the following form:

(1.6) $f^n(c) = h[h^{-1}(c)(1-p_0) + (1-h^{-1}(c))p_0]$ where h^{-1} is the inverse of the function defined in (1.3),

$$0 \le c \le \log 2, \text{ and } n \in N.$$

Since it is to be expected that relations of a similar type as (1.6) shall play an increasing role in the analysis of multi-user channels and sources it seems desirable to have results for more general channels. One of our results (Theorem 2 in Section 3) says that the equality $f^n(c) = f^1(c)$, $n \in N$, holds for *all memoryless binary alphabet channels.* However, examples 1 and 2 in Section 3 show that the identity does not hold if $|A| \ge 3$ or $|B| \ge 3$. In order to prepare for results of a more general nature, we introduce now the concept of the *gerbator of a channel.* This concept is motivated by (1.4) and the name is choosen for historical reasons (see [1], footnote 1).

G is the *gerbator* of the d.m.c. $\mathfrak{D}$ if for any real $c > 0$, any $n \in N$ and any p^n:

(1.7) $\quad \frac{1}{n} H(p^n) \ge c \Rightarrow \frac{1}{n} H(q^n) \ge G(c)$ and $G(c)$ is *maximal* with this property.

Frequently we write f instead of f^1 and we call this function the *characteristic function* of the channel. In this terminology the result by Wyner and Ziv says that for the binary symmetric channel the gerbator equals the characteristic function. Our *Theorem 1* in Section 2 states that for general discrete memoryless channels the *gerbator is equal to the convex lower envelope of the characteristic function.*

Finally, in Section 4 we extend the concept of a gerbator to arbitrarily varying channels and to multiple-access channels, and we obtain simple characterizations (Theorem 3).

2. The gerbator of a discrete memoryless channel

We denote by f^* the convex lower envelope of the characteristic function $f(c) = \min_{p^1: H(p^1) \ge c} H(q^1)$. Equivalently we can define f^* by

$$(2.1) \qquad f^*(c) = \inf_{n \in N} \min_{\substack{p^n = p_1 \times \dots \times p_n \\ \frac{1}{n} H(p^n) \ge c}} \frac{1}{n} H(q^n).$$

THEOREM 1

For a discrete memoryless channel we have

$$(2.2)\qquad \inf_{n\in N} \min_{p^n:\ \frac{1}{n}H(p^n)\geq c} \frac{1}{n} H(q^n) = \inf_{n\in N} \min_{\substack{p^n=p_1\times\dots\times p_n \\ \frac{1}{n}H(p^n)\geq c}} \frac{1}{n} H(q^n)$$

or equivalently the gerbator G equals the lower envelope f^* of the characteristic function f.

For the proof we need a well-known identity, which was also used in [1].

Let $X^n=(X_1, \dots, X_n)$ be a vector valued random variable with values in A^n and distribution p^n. Furthermore, let $Y^n=(Y_1, \dots, Y_n)$ be a vector valued random variable with values in B^n and distribution q^n given by

$$(2.3)\qquad q^n(b^n)=\sum_{a^n\in A^n} p^n(a^n)\, P(b^n|a^n) \text{ for } b^n\in B^n, \text{ where } P(.|.)$$

is the transmission probability for a *memoryless* and not necessarily stationary channel.

Under that assumption we have

$$(2.4)\qquad H(Y_k|Y_1, \dots, Y_{k-1}, X_1, \dots, X_{k-1})=H(Y_k|X_1, \dots, X_{k-1}),$$

for $k=2, 3, \dots n$.

Proof of Theorem 1

We have to show that

$$(2.5)\qquad \min\left\{\frac{1}{n}\sum_{k=1}^{n} H(Y_k|Y_{k-1}, \dots, Y_1) \mid \frac{1}{n}\sum_{k=1}^{n} H(X_k|X_{k-1}, \dots, X_1)\geq c\right\}\geq f^*(c).$$

Since $H(Y_k|Y_{k-1}, \dots, Y_1)\geqslant H(Y_k|Y_{k-1}, \dots, Y_1, X_{k-1}, \dots, X_1)$ and since (2.4) holds, it suffices to show that

$$(2.6)\qquad \min\left\{\frac{1}{n}\sum_{k=1}^{n} H(Y_k|X_{k-1}, \dots, X_1) \mid \frac{1}{n}\sum_{k=1}^{n} H(X_k|X_{k-1}, \dots, X_1)\geq c\right\}\geq f^*(c).$$

One easily verifies that

$$(2.7)\qquad \text{Prob}(Y_k=b_k|X_{k-1}=a_{k-1}, \dots, X_1=a_1)=$$
$$=\sum_{a_k} \text{Prob}(X_k=a_k|X_{k-1}=a_{k-1}, \dots, X_1=a_1)\ W(b_k|a_k)$$

and therefore we have

$$(2.8)\qquad H(Y_k|X_{k-1}=a_{k-1}, \dots, X_1=a_1)\geq f^*(H(X_k|X_{k-1}=a_{k-1}, \dots, X_1=a_1)).$$

The convexity of f^* and (2.8) imply that

$$(2.9)\qquad H(Y_k|X_{k-1}, \dots, X_1)\geq f^*\ (H(X_k|X_{k-1}, \dots, X_1)).$$

This, and again the convexity of f^* yield (2.6).

3. Convexity of the characteristic function in case $|A| = |B| = 2$

In section 2 we proved that the gerbator G of a d.m.c. equals f^*, the convex lower envelope of the characteristic function f. It was shown in [1] that in case of a *binary symmetric channel* the function

(3.1) $$f(c)=h(h^{-1}(c)\cdot(1-p_0)+(1-h^{-1}(c))\,p_0),$$

$0 \le c \le \log 2$, $p_0 \in \left[0, \frac{1}{2}\right]$, is *strictly convex* (U) in c and that

(3.2) $$f(c)=G(c),\ 0 \le c \le \log 2.$$

In this Section we prove that (3.2) holds for *all binary* (not necessarily symmetric) channels and that (3.2) does *not* hold in general if $|A| \ge 3$ or $|B| \ge 3$

THEOREM 2

For a discrete memoryless channel with binary alphabets the following statements hold:

(1) *the characteristic function*

$$f(c)=\min_{p^1:\,H(p^1)\geqq c} H(q^1) \text{ is convex, that is, } f(c)=f^*(c).$$

(2) $$f^n(c)=\min_{p^n:\,\frac{1}{n}H(p^n)\geqq c} \frac{1}{n}H(q^n)=f(c)=G(c) \text{ for all } n \in N.$$

Proof

We know from Theorem 1 that $G(c)=f^*(c) \le f^n(c)$ and since also $f^n(c) \le f(c)$, (2) if a consequence of (1).

It remains to be shown that f is convex. The function

(3.3) $$F(c)=\min_{p:\,H(p)=c} H(q)$$

is equal to $f(c)$ for every c, $0 \le c \le \log 2$, if F is monotonically increasing. It suffices therefore to show that F has positive first and second derivatives. For this purpose we give now a suitable description of F.

For given c we find that one of the two distributions with entropy c, which minimizes $H(q)$. We denote the smaller of the two components of this probability vector by x and also the smaller of the components of the corresponding output vector by y. Thus to every $x, 0 \le x \le \frac{1}{2}$, we have associated a $y, 0 \le y \le \frac{1}{2}$. Denoting this mapping by l we can write

(3.4) $$F(c)=h(l(h^{-1}(c))), \text{ with } h \text{ as defined in (1.3).}$$

We describe now l analytically and show that $y=l(x)$ is linear and increasing. The transmission matrix w can be written in the form

(3.5) $$w=\begin{pmatrix}\alpha\bar{\alpha}\\ \beta\bar{\beta}\end{pmatrix}, \text{ where } \bar{r}=1-r \text{ for any real number } r.$$

W.l.o.g. we can assume that

(3.6) $$\alpha \ge \beta \ge \bar{\alpha}.$$

Clearly,

$$l(x)=\min\{x\alpha+\bar{x}\beta,\ x\bar{\alpha}+\bar{x}\bar{\beta},\ \bar{x}\alpha+x\beta,\ \bar{x}\bar{\alpha}+x\bar{\beta}\}=\min\{x(\alpha-\beta)+ +\beta,\ x(\bar{\alpha}-\bar{\beta})+\bar{\beta},\ x(\beta-\alpha)+\alpha,\ x(\bar{\beta}-\bar{\alpha})+\bar{\alpha}\}.$$

We claim that

(3.7) $$l(x)=x(\bar{\beta}-\bar{\alpha})+\bar{\alpha},\ 0\leq x\leq\frac{1}{2}.$$

The inequality $x(\alpha-\beta)+\beta\leq x(\beta-\alpha)+\alpha$ holds for all $x\in\left[0,\frac{1}{2}\right]$, because the lines intersect as $x=\frac{1}{2}$ and for $x=0$ we have $\beta\leq\alpha$ by assumption. Similarly,

$$x(\bar{\beta}-\bar{\alpha})+\bar{\alpha}\leq x(\bar{\alpha}-\bar{\beta})+\bar{\alpha}\ \text{for}\ 0\leq x\leq\frac{1}{2}.$$

(3.7) follows now from those two inequalities and the relations $\bar{\alpha}\leqslant\beta$ and $(\alpha-\beta)=(\bar{\beta}-\bar{\alpha})$. (The two lines $x(\alpha-\beta)+\beta$ and $x(\bar{\beta}-\bar{\alpha})+\bar{\alpha}$ are always parallel and they coincide exactly in case $\bar{\alpha}=\beta$, that is for the b.s.c.).

Set $x=h^{-1}(c)$ and $y=(\bar{\beta}-\bar{\alpha})x+\bar{\alpha}=\gamma x+\bar{\alpha}$, $\gamma>0$. Then for $0\leq x\leq\frac{1}{2}$

(3.8) $$F'(c)=h'(y)\frac{dy}{dx}\frac{dx}{dc}=\gamma\log\frac{1-y}{y}\cdot\log^{-1}\frac{1-x}{x}\geq 0$$

and also

(3.9) $$F''(c)=\gamma\log^{-3}\frac{1-x}{x}\left\{\frac{1}{(1-x)x}\log\frac{1-y}{y}-\frac{\gamma}{(1-y)y}\log\frac{1-x}{x}\right\}.$$

Since $\gamma\log^{-3}\frac{1-x}{x}\geq 0$ it remains to be shown that

(3.10) $$\left(\frac{1-y}{y}\right)^{(1-y)y}\geq\left(\frac{1-x}{x}\right)^{\gamma(1-x)x}.$$

If $y\leq x$, then $\frac{1-y}{y}\geq\frac{1-x}{x}\geq 1$, $1-y\geq 1-x$ and $\frac{y}{\gamma}=x+\frac{\bar{\alpha}}{\gamma}>x$.

In the remaining case we have $\frac{1}{2}\geq y\geq x$ and it turns out that the b.s.c. has an interesting extremal property, which makes it possible to reduce the above problem to the case of a b.s.c., where the solution is known ([1]).

We proceed as follows. For a given pair (x,y) $x\leq y\leq\frac{1}{2}$, we consider the class Wxy of all channels w^* such that

(3.11) $$y=\gamma^*x+\bar{\alpha}^*$$

Since $\frac{1-x}{x}\geq 1$ it suffices to prove (3.10) only for the maximal γ^* occurring in (3.11). We complete the proof of the theorem by showing that the maximal value of γ^* is assumed for a binary symmetric channel. For a b.s.c. satisfying (3.11) we have

(3.12) $$y=(1-2\delta)x+\delta,\ \text{where}\ \delta=\frac{y-x}{1-2x}$$

Suppose that $\gamma^*=1-2\delta$ is not maximal, then for an $\varepsilon>0$

$$y=(1-2\delta)x+\varepsilon x+\delta(\varepsilon)\ \text{and therefore}\ \delta(\varepsilon)=\delta-\varepsilon x.$$

For the corresponding matrix $\begin{pmatrix} 1-\delta+\varepsilon x & \delta-\varepsilon x \\ \delta-\varepsilon(1-x) & 1-\delta+\varepsilon(1-x) \end{pmatrix}$ we have to satisfy the constraints

$$1-\delta+\varepsilon x \geq \delta-\varepsilon(1-x) \geq \delta-\varepsilon x. \tag{3.13}$$

The last inequality however does not hold for $x<\frac{1}{2}$. For $x=\frac{1}{2}$ we have $y=\frac{1}{2}$ and (3.10) holds anyhow.

We show now by example that Theorem 2 does not generalize to non-binary alphabet channels.

Example 1 ($|A|\geq 3$, $|B|=2$).

$$w=\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \vdots & \vdots \\ \frac{3}{4} & \frac{1}{4} \\ 0 & 1 \end{pmatrix}, \ |A|=m+1$$

We shall show now that the characteristic function f is *not convex*, simply because it is 0 at $c=0$, increases then to the value $h\left(\frac{1}{4}\right)$ and stays constant over a certain interval.

Let $c=\sup\left\{c \mid f(c)<h\left(\frac{1}{4}\right)\right\}$. Since $\frac{1}{3}\left(\frac{3}{4},\frac{1}{4}\right)+\frac{2}{3}(0,1)=\left(\frac{1}{4},\frac{3}{4}\right)$ we obtain that $c=H(p)$, where $p=\left(\frac{1}{3m},\ldots,\frac{1}{3m},\frac{2}{3}\right)$

Hence, $c=h\left(\frac{1}{3}\right)+\frac{1}{3}\log m$. Set $\bar{c}=\sup\left\{c \mid f(c)=h\left(\frac{1}{4}\right)\right\}$. We have $\bar{c}=\log m$ and $\bar{c}>c$ for $m\geq 2$.

Example 2 ($|A|=2$, $|B|\geq 3$). Since example 1 shows already that Theorem 2 does not extend to general alphabets we comment on the present case only briefly and omit calculations.

Suppose that $w=\binom{p_1}{p_2}$ and that $H(p_1)<H(p_2)$. It can be shown that $f(c)=\min_{\substack{p=(\lambda,1-\lambda)\\ h(\lambda)\geq c}} H(q)=\min(f_1(c),f_2(c))$, where $f_1(c)=H(\lambda p_1+(1-\lambda)p_2)$ for $\lambda\leq\frac{1}{2}$ and $h(\lambda)=c$, $f_2(c)=H(\lambda p_1+(1-\lambda)p_2)$ for $\lambda>\frac{1}{2}$ and $h(1-\lambda)=c$.

Let now λ^* be such that $H(\lambda^* p_1+(1-\lambda^*)p_2)=\max_{0\leq\lambda\leq 1} H(\lambda p_1+(1-\lambda)p_2)$.

If $\lambda^*>\frac{1}{2}$ then $f_1(c)$ is not convex, $f_2(c)$ is convex but $f(c)=\min(f_1(c), f_2(c))$ is not convex. This occurs for instance if

$w=\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \end{pmatrix}$. If $\lambda^*\leq\frac{1}{2}$ then $f_1(c)\equiv f(c)$ and we have convexity of this function.

An open problem

Since $f(c)$ is in general not convex one might investigate whether for a given matrix w there exists a function $n_0(w)$ such that f^{n_0} is convex and equals therefore f^*. It should be possible to give a bound on $n_0(w)$, which depends only on $|A|$ and $|B|$.

At any rate it is desirable to have estimates on the speed with which $f^n(c)$ converges to $f^*(c)$ as n goes to infinity.

4. General results

The purpose of this last Section is to demonstrate that the concepts and results of Section 2 can be extended to various channels, which have previously been considered in the literature. Since by its very nature our discussion can be by no means exhaustive, we shall limit ourselves to two typical examples, arbitrarily varying channels and multiple-access channels (see [7] and [4], [5], [6]). We did not include compound channels, because the generalization of Theorem 1 to this case is straightforward. At the present state of the theory we cannot fully judge the importance of our results for existing coding problems, but since the results seem to be answers to natural questions, we are optimistic about their future role. We recall now briefly the definitions of the channels mentioned.

Let S be any set and let $W=\{w(.|.|s) \ |s\in S\}$ be a set of stochastic $|A|\times\times|B|$-matrices. For every $s^n=(s_1, \ldots, s_n)\in\prod_1^n S$ we define $P(.|.|s^n)$ by

$$(4.1) \qquad P(b^n|a^n|s^n)=\prod_{t=1}^n w(b_t|a_t|s_t) \text{ for every } a^n\in A^n,\ b^n\in B^n.$$

An arbitrarily varying channel (a.v.ch.) a is defined by a sequence $(a^n)_{n\in N}$, where

$$(4.2) \qquad a^n=\{P(.|.|s^n)\,|\,s^n\in S^n\},\ n\in N.$$

For the sake of notational simplicity we define multiple-access channels (see [4], [5], [6]) only in case of two senders even though our results below extend to more general cases.

Let A, D, and B be finite sets and set $A^n=\prod_1^n A$, $D^n=\prod_1^n D$, and $B^n=\prod_1^n B$. Furthermore let $w(.|.,.)$ be a stochastic $|A\times D|\times|B|$-matrix. The transmission probabilities of a multiple-access channel m are defined by

$$(4.3) \qquad P(b^n|a^n, d^n)=\prod_{t=1}^n w(b_t|a_t, d_t)$$

for every $a^n=(a_1, \ldots, a_n)\in A^n$, $d^n\in D^n$, $b^n\in B^n$, and $n\in N$.

We introduce now a notion of gerbator for the channels described. For p^n on A^n and $P(.|.|s^n)$ define $q^n(.|s^n)$ by

$$(4.4) \qquad q^n(b^n|s^n)=\sum_{a^n\in A^n} p^n(a^n)\,P(b^n|a^n|s^n) \text{ for all } b^n\in B^n.$$

G_a *is the gerbator of* a if for any real $c>0$, any n and any p^n:

$$(4.5)\qquad \frac{1}{n} H(p^n) \geq c \Rightarrow \frac{1}{n} \min_{s^n} H(q^n(.\,|\,s^n) \geq G_a(c)$$

and $G_a(c)$ is maximal with this property.

Let r^n be a p.d. on $A^n \times D^n$ and define q^n on B^n by $q^n(b^n) = \sum_{a^n, d^n} r^n(a^n, d^n) P(b^n \,|\, a^n, d^n)$ for all $b^n \in B^n$. Denote the marginal distribution of r^n on A^n (resp. D^n) by r_A^n (resp. r_D^n).

The *gerbator of* m is a function of two variables G_m: $G_m(c_1, c_2)$ is the largest number such that for all n and r^n

$$(4.6)\qquad \frac{1}{n} H(r_A^n) \geq c_1, \ \frac{1}{n} H(r_D^n) \geq c_1 \Rightarrow \frac{1}{n} H(q^n) \geq G_m(c_1, c_2).$$

In case we allow only product distributions $r^n = r_A^n \times r_D^n$ in (4.6) we are led to the definition of the I-gerbator G_m^I.

We denote the characteristic function of $w(.|.|s)$ by $f(c, s)$ and we define the characteristic functions f^I and f for m by

$$(4.7)\qquad f^I(c_1, c_2) = \min_{\substack{r = r_A \times r_D \\ H(r_A) \geq c_1,\, H(r_D) \geq c_2}} H(q)$$

$$(4.8)\qquad f(c_1, c_2) = \min_{\substack{r:\, H(r_A) \geq c_1 \\ H(r_D) \geq c_2}} H(q).$$

Finally, we denote the lower envelopes of these three functions by $f^*(c, s)$, $f^{*I}(c_1, c_2)$, and $f^*(c_1, c_2)$.

THEOREM 3 (Characterization of gerbators of a and m).

The following identities hold:

a.) $G_a(c)$ *equals the convex lower envelope* $\bar{f}$ *of* $\min_{s \in S} f(c, s)$,

b.) $G_m^I(c_1, c_2) = f^{*I}(c_1, c_2)$,

c.) $G_m(c_1, c_2) = f^*(c_1, c_2)$.

Proof

a.) By using time sharing we clearly get that $G_a(c) \leq \bar{f}(c)$.

We prove the converse inequality by looking at an individual $P(.|.|s^n)$, where $s^n = (s_1, \ldots, s_n)$. Since (2.4) holds in case of a nonstationary memoryless channel, (2.9) holds also in the present situation and we therefore have

$$(4.9)\qquad H(Y_k | X_{k-1}, \ldots, X_n; s_k) \geq f^*(H(X_1 | X_{k-1}, \ldots, X_1), s_k),$$

where the distribution of Y_k depends now of course on s_k.

Since $\bar{f}(c) \leq f^*(c, s)$ for all $s \in S$ and since $\bar{f}$ is convex

$$(4.10)\qquad \frac{1}{n} \sum_{k=1}^{n} (Y_k | X_{k-1}, \ldots, X_1) \geq \frac{1}{n} \sum_{k=1}^{n} \bar{f}(H(X_k | X_{k-1}, \ldots, X_1) \geq$$

$$\geq \bar{f}\left(\frac{1}{n} H(X_1, \ldots, X_n)\right),$$ which was to be proved.

b.) We have to show that

$$\min\left\{\frac{1}{n}\sum_{k=1}^{n} H(Y_k \mid Y_{k-1}, \ldots, Y_1) \mid \frac{1}{n}\sum_{k=1}^{n} H(X_k|X_{k-1}, \ldots, X_1) \geq c,\right.$$

$$\left.\frac{1}{n}\sum_{k=1}^{n} H(Z_k|Z_{k-1}, \ldots, Z_1) \geq c_2\right\} \geq f^{*I}(c_1, c_2).$$

By viewing $A \times D$ as a single input alphabet and (X_t, Z_t), $t=1, \ldots, n$; as random variable with values in $A \times D$ we can use the result stated in (2.4) and obtain

(4.11) $$H(Y_k|Y_{k-1}, \ldots, Y_1, X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1)=$$

$=H(Y_k|X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1)$. Since $H(Y_k|Y_{k-1}, \ldots, Y_1) \geq$

$\geq H(Y_k|Y_{k-1}, \ldots, Y_1, X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1)$ it suffices to show that

$$\min \frac{1}{n}\sum_{k=1}^{n} H(Y_k|X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1) \geq f^{*I}(c_1, c_2).$$

One easily verifies that

$$\text{Prob}\,(Y_k=b_k|X_{k-1}=a_{k-1}, \ldots, X_1=a_1, Z_{k-1}=d_{k-1}, \ldots, Z_1=d_1)=\sum_{a_k, d_k} \text{Prob}$$

$$(X_k=a_k, Z_k=d_k|X_{k-1}=a_{k-1}, \ldots, X_1=a_1, Z_{k-1}=d_{k-1}, \ldots, Z_1=d_1)\; w(b_k|a_k, d_k)$$

and therefore we have

(4.12) $$H(Y_k|X_{k-1}=a_{k-1}, \ldots, X_1=a_1, Z_{k-1}=d_{k-1}, \ldots, A_1=d_1) \geqslant$$

$$\geq f^{*I}(H(X_k|X_{k-1}=a_{k-1}, \ldots, X_1=a_1), H(Z_k|Z_{k-1}=d_{k-1}, \ldots, Z_1=d_1)).$$

The componentwise convexity of f^{*I} yields

(4.13) $$H(Y_k|X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1) \geq$$

$$\geq f^{*I}(H(X_k|X_{k-1}, \ldots, X_1), H(Z_k|Z_{k-1}, \ldots, Z_1))$$

This, and the convexity of f^{*I} imply

(4.14) $$\frac{1}{n}\sum_{k=1}^{n} H(Y_k|X_{k-1}, \ldots, X_1, Z_{k-1}, \ldots, Z_1)$$

$$\geq f^{*I}\left(\frac{1}{n}\sum_{k=1}^{n} H(X_k|X_{k-1}, \ldots, X_1), \frac{1}{n}\sum_{k=1}^{n} H(Z_k \mid Z_{k-1}, \ldots, Z_1)\right)$$

Thus, $G_m^I(c_1, c_2) \geq f^{*I}(c_1, c_2)$. The converse inequality is obtained again by time sharing.

c.) In the case of dependent input probability distributions (4.12) holds with f^{*I} replaced by f^*. Convexity then yields (4.13) for f^* and finally also (4.14). Thus, $G_m(c_1, c_2) \geq f^*(c_1, c_2)$. The converse relationship is again obvious.

REFERENCES

1. WYNER, A. D., ZIV, J., *A theorem on the entropy of certain binary sequences and applications: part I.* IEEE—Trans. Information Theory **19**, 769—772 (1973).
2. WYNER, A. D., *A theorem on the entropy of certain binary sequences and applications: part II.* IEEE—Trans. Information Theory **19**, 772—777 (1973).

3. AHLSWEDE, R., KÖRNER, J., *A source coding problem with side information and a converse for degraded broadcast channels.* Submitted to IEEE—Trans. Information Theory.
4. AHLSWEDE, R., *Multi-way communication channels.* In: 2nd International Symposium on Information Theory, Tsahkadsor, 1971, pp. 23-52, Budapest, Akadémiai Kiadó, 1973.
5. AHLSWEDE, R., *The capacity region of a channel with 2 senders and 2 receivers.* To appear in Ann. of Probability.
6. SLEPIAN, D., WOLF, J. K., *A coding theorem for multiple access channels with correlated sources.* To appear in Bell Syst. Tech. J.
7. AHLSWEDE, R., WOLFOWITZ, J., *The capacity of a channel with arbitrary varying channel probability functions and binary output alphabet.* Z. Wahrscheinlichkeitstheorie verw. Geb., **15**, 186–194 (1970).