from the second, let $K = AA^t$. Then AA^t is nonnegative definite and

$$|A|^2 = |AA^t| \leqslant \prod (AA^t)_{ii} = \prod_i \left(\sum_j a_{ij}^2 \right). \tag{1}$$

The implication of the second inequality from the first follows from the fact that every nonnegative definite matrix K can be factored as $K = AA^t$. A typical proof of Hadamard's inequality is by induction (see, for example, Bellman [1]) and involves a determinant decomposition followed by an inspection of the resulting quadratic forms. A recent proof based on convexity arguments is given in Marshall and Olkin [2].

We offer here an information–theoretic proof.

II. Preliminaries

If X is a vector valued random variable having probability density function $f(x)$, define the (differential) entropy h of the random vector X by $h(X) = -\int f(x) \ln f(x)\, dx$.

From elementary information theory [3], we have the inequality

$$h(X_1, \cdots, X_n) \leqslant \sum_{i=1}^{n} h(X_i), \tag{2}$$

with equality if and only if $X_1, X_2, \cdots, X_n$ are independent random variables. The proof follows from Jensen's inequality as follows:

$$\begin{aligned} h(X_1, \cdots, X_n) - \sum_{i=1}^{n} h(X_i) &= -\int f(x_1, \cdots, x_n) \ln f(x_1, \cdots, x_n) \\ &\quad + \int f(x_1, \cdots, x_n) \ln \prod_i f_i(x_i) \\ &= \int f \ln \frac{\prod f_i}{f} \\ &\leqslant \ln \int f \frac{\prod f_i}{f} \\ &= \ln \int \prod f_i = \ln 1 = 0, \end{aligned} \tag{3}$$

with equality if and only if $f = \prod f_i$, by the strict concavity of the logarithm.

If X is an n-variate normal random vector with mean 0 and covariance matrix K, then a direct calculation [4, th. 4.5.1] establishes

$$\begin{aligned} h(X_1, \cdots, X_n) &= -\int f \ln f \\ &= -\int \frac{1}{(2\pi)^{n/2}|K|^{1/2}} e^{-(1/2)x^t K^{-1} x} \cdot \left[-\ln (2\pi)^{n/2}|K|^{1/2} - \frac{1}{2} \sum_{i,j} x_i (K^{-1})_{ij} x_j \right] dx \\ &= \ln (2\pi)^{n/2}|K|^{1/2} + \frac{1}{2} \sum_{i,j} (K^{-1})_{ij} E X_i X_j \\ &= \ln (2\pi)^{n/2}|K|^{1/2} + \frac{n}{2} \\ &= \frac{1}{2} \ln (2\pi e)^n |K|. \end{aligned} \tag{4}$$

Letting $n = 1$, we have

$$h(X_i) = \frac{1}{2} \ln 2\pi e k_{ii}. \tag{5}$$

III. Theorem and Proof

Theorem (Hadamard's Inequality): If K is nonnegative definite, then

$$|K| \leqslant \prod_i k_{ii}, \tag{6}$$

with equality if and only if $k_{ij} = 0$, for all $i \neq j$.

Proof: If the determinant $|K| = 0$, the inequality is trivially true. Let $|K| > 0$, and consider X to be normally distributed with mean 0 and covariance matrix K. Then from (2),

$$h(X_1, X_2, \cdots, X_n) \leqslant \sum h(X_i).$$

Substituting from (4) and (5) yields

$$\frac{1}{2} \ln (2\pi e)^n |K| \leqslant \sum \frac{1}{2} \ln 2\pi e k_{ii}. \tag{7}$$

Exponentiating preserves the inequality and yields the desired result.

Moreover, we have equality only if the X_i's are independent, hence uncorrelated. Thus equality holds only if K is diagonal.

References

[1] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed. New York: McGraw-Hill, 1970, pp. 126–130.

[2] A. Marshall and I. Olkin, "A convexity proof of Hadamard's inequality," *Amer. Math. Monthly*, vol. 89, no. 9, pp. 687–688, Nov. 1982.

[3] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley, 1968.

[4] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice Hall, 1971.

A Simple Proof of the Ahlswede–Csiszár One-Bit Theorem

ABBAS EL GAMAL, SENIOR MEMBER, IEEE

***Abstract*—It is proved that if (X, Y) are two finite alphabet correlated sources with $p(x, y) > 0$ for all $(x, y) \in (\mathcal{X} \times \mathcal{Y})$, and if a function $F(X, Y)$ is α-sensitive, then the rate R of transmission from X to Y necessary to compute $F(X, Y)$ reliably must be greater than $H(X|Y)$. The same result holds if the function is highly sensitive and for every $x_1 \neq x_2 \in \mathcal{X}$, then the number of elements $y \in \mathcal{Y}$ with $p(x_1, y) \cdot p(x_2, y) > 0$ is different from one.**

I. Introduction

Let $(X, Y) \in (\mathcal{X} \times \mathcal{Y})$ be two finite alphabet sources with joint probability mass function $p(x, y)$, and let (X_i, Y_i), $i = 1, 2, \cdots, n$, be n independent copies of (X, Y). Consider a function

$$F: \bigcup_{n=1}^{\infty} (\mathcal{X}^n \times \mathcal{Y}^n) \to \mathcal{R}.$$

Manuscript received July 6, 1982; revised April 12, 1983. This work was supported in part by the National Science Foundation under NSF Grant 80-26102 and in part by the Air Force under Contract 49620-79-C-0058.

The author is with the Information Systems Laboratory, Durand 137, Stanford University, Stanford, CA 04305.

0018-9448/83/1100-0931$01.00 ©1983 IEEE

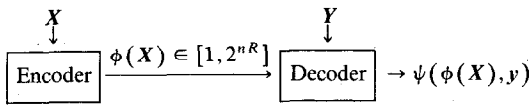

Fig. 1.

Let Φ_n: $\mathcal{X}^n \rightarrow [1, 2^{nR}]$ be an encoding function, and ψ_n: $[1, 2^{nR}] \times \mathcal{Y}^n \rightarrow \mathcal{R}$ be a decoding function (see Fig. 1). A rate R is said to achieve a reliable computation of $F(x, y)$ if there exists a sequence of encoding and decoding functions, indexed by n, such that the probability of decoding error

$$P_e \triangleq P\{\psi_n(\Phi_n(X), Y) \neq F(X, Y)\} \rightarrow 0.$$

The problem is to find the infimum of the set of achievable rates R.

Definition 1 [1]: A function $F(x, y)$ is said to be *sensitive*, if whenever $x \in \mathcal{X}^n$, $x' \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$ are such that x and x' differ in the ith component and $F(x, y) = F(x', y)$, then there exists a $y' \in \mathcal{Y}^n$, different from y only in the ith component, such that $F(x, y') \neq F(x', y')$.

In [1], it was proved that if $F(x, y)$ is sensitive then it can be reliably computed if and only if the rate of transmission $R > H(X|Y)$. In this note we give a simple proof of this result for the following larger class of functions.

Definition 2: A function $F(x, y)$ is said to be *α-sensitive* if for some $1 > \alpha > 0$, whenever $x \in \mathcal{X}^n$, $x' \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$ are such that $F(x, y) = F(x', y)$, and the Hamming Distance between x and x' is $d(x, x')$, then there exists at least $l = \min\{[\alpha n], d(x, x')\}$ distinct sequences $y_1, y_2, \cdots, y_l$ each differing from y in exactly one component, such that $F(x, y_i) \neq F(x', y_i)$, $i = 1, 2, \cdots, l$.

It is easy to see that if F is sensitive then it is α-sensitive for any α. It is equally easy to see that functions such as the Hamming distance between y and any fixed cyclic shift of x or the joint type of y and any fixed cyclic shift of x are α-sensitive, but not sensitive. For α-sensitive functions, we prove the following.

Theorem 1: Let (X, Y) have joint probability mass function $p(x, y) > 0$, for all $(z, y) \in \mathcal{X} \times \mathcal{Y}$, and $F(x, y)$ be α-sensitive. Then the rate R is achievable if and only if $R > H(X|Y)$.

II. Proof of Theorem

The proof of Theorem 1 needs the following lemma.

Lemma: Let (ϕ_n, ψ_n) be encoding and decoding functions for $F(x, y)$, such that $P_e < \epsilon$, for some $\epsilon > 0$. For $i \in [1, 2^{nR}]$, define $\Phi_n'(i) = \{x: \Phi_n(x) = i, x \in A_\epsilon(X)\}$, where $A_\epsilon(x)$ is the set of ϵ-typical x sequences [2]. There must exist an $i \in [1, 2^{nR}]$ such that

1) $P\{X \in \Phi_n'(i)\} \geq \dfrac{(1-\epsilon)}{6} \cdot 2^{-nR}$,

2) $P\{\psi_n(i, Y) \neq F(X, Y) | X \in \Phi_n'(i)\} < \dfrac{2\epsilon}{(1-\epsilon)}$,

3) $P\{(X, Y) \in A_\epsilon(X, Y) | X \in \Phi_n'(i)\} > (1 - 4\epsilon)$,

where $A_\epsilon(X, Y)$ is the set of jointly ϵ-typical sequences [2].

Proof: Define $p_i = P\{X \in \Phi_n'(i)\}$ and $a_i = P\{\Psi_n(i, Y) \neq F(X, Y) | X \in \Phi'(i)\}$. Then

$$\sum_{i=1}^{2^{nR}} p_i \geq (1 - \epsilon),$$

and

$$\sum_{i=1}^{2^{nR}} a_i p_i < \epsilon. \tag{1}$$

Rewrite (1) as

$$\sum_{i:\, a_i < 2\epsilon/(1-\epsilon)} a_i p_i + \sum_{i:\, a_i \geq 2\epsilon/(1-\epsilon)} a_i p_i < \epsilon.$$

Clearly,

$$\bar{p} \triangleq \sum_{i:\, a_i < 2\epsilon/(1-\epsilon)} p_i \geq \frac{1}{2} \cdot (1 - \epsilon).$$

Therefore,

$$\frac{1}{\bar{p}} \sum_{i:\, a_i < 2\epsilon/(1-\epsilon)} a_i p_i < \frac{2\epsilon}{(1-\epsilon)}. \tag{2}$$

Now, for every $i \in [1, 2^{nR}]$, define the set

$$B_i = \{(x, y): x \in \Phi_n'(i), (x, y) \in A_\epsilon(X, Y)\}.$$

Rewrite (2) as

$$\frac{1}{\bar{p}} \sum_{\substack{i:\, a_i < 2\epsilon/(1-\epsilon), \\ P\{B_i\} > (1-(3\epsilon/(1-\epsilon))) \cdot p_i}} a_i p_i + \frac{1}{\bar{p}} \sum_{\substack{i:\, a_i < 2\epsilon/(1-\epsilon), \\ P\{B_i\} \leq (1-(3\epsilon/(1-\epsilon))) \cdot p_i}} a_i p_i < \frac{2\epsilon}{(1-\epsilon)}.$$

Now,

$$\bar{\bar{p}} \triangleq \sum_{\substack{i:\, a_i < 2\epsilon/(1-\epsilon), \\ P\{B_i\} > (1-(3\epsilon/(1-\epsilon))) \cdot p_i}} p_i \geq \frac{\bar{p}}{3} \geq \frac{1}{6}(1 - \epsilon).$$

Therefore, there must exist an $i \in [1, 2^{nR}]$, such that

$$p_i \geq \frac{(1-\epsilon)}{6} \cdot 2^{-nR}, a_i < \frac{2\epsilon}{(1-\epsilon)},$$

and

$$\frac{P\{B_i\}}{p_i} > \frac{(1-4\epsilon)}{(1-\epsilon)},$$

and the lemma is proved.

Proof of Theorem 1: Achievability follows from the Slepian–Wolf source coding theorem [3]. To prove the converse consider a given sequence of encoding and decoding functions $\{\Phi_n, \psi_n\}$ such that $P_e \rightarrow 0$, and $R = H(X|Y) - a$, for some $a > 0$. For $\epsilon > 0$, and sufficiently large n, $P_e < \epsilon$. From the lemma there must exist an $i \in [1, 2^{nR}]$ satisfying conditions 1)–3).

Construct a binary matrix with $|\Phi_n'(i)|$ rows and $|\mathcal{Y}^n|$ columns (see Fig. 2) such that each row corresponds to a distinct $x \in \Phi_n'(i)$, and each column corresponds to a distinct $y \in \mathcal{Y}^n$. An entry $a(x, y)$ of this matrix is a one if $\psi(i, y) = F(x, y)$, otherwise it is a zero. Denote by $P(0)$ the probability of the set of sequences (x, y) with $a(x, y) = 0$. From the lemma, it follows that

$$P(0) < \frac{2\epsilon}{(1-\epsilon)} \cdot P\{X \in \Phi_n'(i)\}. \tag{3}$$

Using the fact that F is α-sensitive we now establish a contradiction by showing that $P(0)$ must be much larger than the right-hand side of (3).

Fix a $\beta > 0$, such that $\epsilon < \beta < \alpha$. Fix a column $y \in A_\epsilon(Y)$ of the matrix and define

$$A_y = \{x: a(x, y) = 1, (x, y) \in A_\epsilon(X, Y)\}.$$

Now, order the sequences (x, y), $x \in A_y$, in descending order according to their probabilities. Denote these probabilities by $p(x_1, y) \geq p(x_2, y) \geq \cdots \geq p(x_{|A_y|}, y)$. Select distinct pairs as follows.

First take $x_1' = x_1$ and pair it with an x_1'' that satisfies $d(x_1', x_1'') \geq [\beta n]$ and has the largest $p(x, y)$. Clearly $p(x_1'', y)$

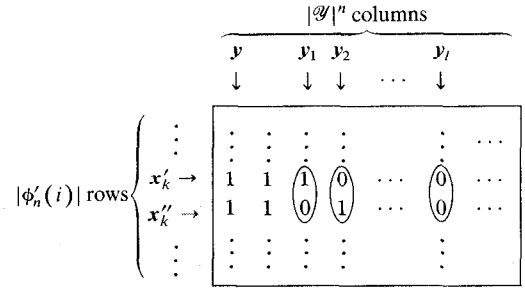

Fig. 2. Binary matrix used in the proof of the theorem.

$\geq p(x_v, y)$, where $v = \sum_{i=1}^{[\beta n]} {}^{1}(|\mathcal{X}| - 1)^i \binom{n}{i}$. Next, select the $x_2' \in (A_y - \{x_1', x_1''\})$ with the largest $p(x, y)$ and pair it with an unselected x_2'' satisfying $d(x_2', x_2'') \geq [\beta n]$ and that has the largest $p(x, y)$. Clearly $p(x_2', y) \geq p(x_3, y)$ and $p(x_2'', y) \geq p(x_{v+2}, y)$. Repeat this process until no more pairing is possible. By the definition of α-sensitivity, the two rows corresponding to the kth selected pair must have at least $[\beta n]$ zeros each of probability strictly greater than $p' \cdot p(x_{v+2k-1}, y)$, where $p' = \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\{p(x, y)\}$.

Denote, for every $y \in A_\epsilon(Y)$, the total probability of the zeros resulting from all the pairs by $P(y, 0)$. Then

$$P(y,0) > [\beta n] \cdot p' \cdot \sum_{k=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2k-1}, y),$$

where

$$\left[\frac{1}{2}(|A_y| - v)\right]^+ = \begin{cases} 0, & \text{if } \frac{1}{2}(|A_y| - v) < 1 \\ \text{integer part of } \frac{1}{2}(|A_y| - v), & \text{otherwise.} \end{cases}$$

Repeating the above argument for every $y \in A_\epsilon(Y)$, we conclude that the total probability of the zeros in the matrix is

$$P(0) \geq \sum_{y\in A_\epsilon(Y)} P(y,0)$$

$$> \frac{[\beta n]\cdot p'}{n\cdot|\mathcal{Y}|} \cdot \sum_{y\in A_\epsilon(Y)} \sum_{k=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2k-1}, y). \quad (4)$$

The factor $1/n|\mathcal{Y}|$ appears because every zero may appear at most $(n|\mathcal{Y}| - 1)$ times in the counting for different y's.

Now, from the third condition on $i \in [1, 2^{nR}]$, it follows that

$$\sum_{y\in A_\epsilon(Y)} \sum_{j=1}^{|A_y|} p(x_j, y) \geq (1 - 4\epsilon)\cdot P\{X \in \Phi_n'(i)\}. \quad (5)$$

For every $y \in A_\epsilon(Y)$, we have

$$\sum_{j=1}^{|A_y|} p(x_j, y)$$

$$\leq (v+1)\cdot 2^{-n(H(X,Y)-\epsilon)} + \sum_{j=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2j-1}, y)$$

$$+ \sum_{j=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2j}, y)$$

$$\leq (v+1)\cdot 2^{-n(H(X,Y)-\epsilon)} + 2\cdot \sum_{j=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2j-1}, y). \quad (6)$$

Substituting from (6) into (5) we obtain

$$\sum_{y\in A_\epsilon(Y)} \sum_{j=1}^{[(1/2)(|A_y|-v)]^+} p(x_{v+2j-1}, y)$$

$$\geq \frac{1}{2}\Big((1-4\epsilon)\cdot P\{X \in \Phi_n'(i)\} - (v+1)\cdot 2^{-n(H(X,Y)-2\epsilon)}\Big). \quad (7)$$

Substituting from (7) and part 1) of the lemma into (4) we conclude that, for sufficiently large n,

$$P(0) > \frac{[\beta n]\cdot p'}{n\cdot|\mathcal{Y}|}\cdot\frac{1}{4}(1-5\epsilon)\cdot P\{X \in \Phi_n'(i)\}. \quad (8)$$

For sufficiently small ϵ, (3) and (8) give the desired contradiction.

Corollary: (Second part of Theorem 3 in [1]) Let $F(x, y)$ be highly sensitive, in the sense of [1]. If for every $x_1 \neq x_2 \in \mathcal{X}$ the number of elements $y \in \mathcal{Y}$ with $p(x_1, y)\cdot p(x_2, y) > 0$ is different from one, then R is achievable if and only if $R > H(X|Y)$.

The proof follows from the proof of Theorem 1 by observing that the condition in the corollary guarantees that, for every $y \in A_\epsilon(Y)$, every pair of rows selected will have at least $[\beta n]$ zeros, and replacing p' by $p'' = \min_{(x,y):p(x,y)>0}\{p(x, y)\}$.

ACKNOWLEDGMENT

The author is indebted to I. Csiszar for his presentation of the results in [1]. Several discussions with K. F. Pang have helped improve the presentation of this correspondence.

REFERENCES

[1] R. Ahlswede and I. Csiszár, "To get a bit of information may be as hard as to get full information", *IEEE Trans. Inform. Theory*, vol. IT-27, no. 4, pp. 398–408, July 1981.

[2] A. El Gamal and T. Cover, "Multiple user information theory," in *Proc. IEEE*, vol. 68, no. 12, pp. 1466–1483, Dec. 1980.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 409–413, 1973.

Some Results on the Existence of Binary Linear Codes

MICHELE ELIA, MEMBER, IEEE

***Abstract*—A lower bound is given on the size of a linear code of given length and distance which improves the Varshamov bound in certain cases, though not asymptotically. Specific triples (n, k, d) are given with values larger than those guaranteed by the Varshamov bound.**

I. INTRODUCTION

An (n, k, d) binary linear code contains 2^k binary vectors of dimension n with minimum Hamming distance d. Varshamov showed [1] that an (n, k, d) binary code exists provided that

$$S_{n-1}^{d-2} < 2^{n-k}, \quad (1)$$

where

$$S_m^h = \sum_{j=0}^{h} \binom{m}{j}.$$

Manuscript received October 26, 1981; revised July 26, 1982. This work was presented at the Colloquium on Information Theory, Budapest, Hungary, August 1981.

The author is with the Istituto di Elettronica e Telecomunicazioni, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy.

0018-9448/83/1100-0933$01.00 ©1983 IEEE