only linearly and not exponentially with the source block length. Since a trellis approaches a tree as the constraint length grows large, this work also suggests an alternate tree coding scheme and proof of the tree coding theorem of Jakatdar and Pearlman [6].

Appendix A

The generalized Gallager function $E_0^{jk}(\rho)$ is defined in (21). In the following we prove that given $R_{N_{jk}} > R_{N_{jk}}(D_\theta)$ for all j and k; or equivalently given (7a), that the per-letter "rate" associated with each code letter being always greater than the rate $r_l(d_\theta)$ induced by the rate-distortion function of the corresponding source letter u_l, will imply (22a), that is

$$\left[\frac{E_0^{jk}}{R_{N_{jk}}} - \rho\right] > 0, \qquad \text{for all } j \text{ and } k \text{ and } -1 < \rho < 0. \quad (27)$$

From the properties of the Gallager function [1, p. 394] we can write

$$\left[\frac{E_l(\rho)}{r} - \rho\right] > 0 \qquad -1 < \rho < 0, \quad \text{for } r > r_l(d_\theta) \quad (28)$$

where $E_l(\rho)$ and $r_l(d_\theta)$ are respectively the Gallager function and the rate-distortion function associated with the letter u_l. We will use the property (28) to establish (27) as follows:

$$\left[\frac{E_0^{jk}(\rho)}{R_{N_{jk}}} - \rho\right] = \frac{N_{jk}^{-1}\sum_{l=1}^{N_{jk}} E_{N_j+l}(\rho)}{N_{jk}^{-1} k \log q} - \rho$$

$$= \frac{1}{k \log q}\left[\sum_{l=1}^{N_{jk}} E_{N_j+l}(\rho) - \rho k \log q\right].$$

Since at depth $(j+m)$ of the trellis

$$r_{j+m} = \frac{\log q}{n_{j+m}} \quad \text{or} \quad \log q = n_{j+m} r_{j+m},$$

we can rewrite the previous right-hand term as

$$= \frac{1}{k \log q}\left[\sum_{l=1}^{N_{jk}} E_{N_j+l}(\rho) - \rho \cdot \sum_{m=1}^{k} n_{j+m} r_{j+m}\right]$$

$$= \frac{1}{k \log q}\left[\sum_{l=1}^{n_{j+1}} \left(E_{N_j+l}(\rho) - \rho r_{j+1}\right) + \cdots + \sum_{l=1}^{n_{j+k}} \left(E_{N_{j+k-1}+l}(\rho) - \rho r_{j+k}\right)\right]$$

$$= \frac{1}{k \log q}\left[r^{j+1} \sum_{l=1}^{n_{j+1}} \left(\frac{E_{N_j+l}(\rho)}{r_{j+1}} - \rho\right) + \cdots + r^{j+k} \sum_{l=1}^{n_{j+k}} \left(\frac{E_{N_{j+k-1}+l}(\rho)}{r_{j+k}} - \rho\right)\right]$$

$$= \frac{1}{k}\left[\frac{1}{n_{j+1}} \sum_{l=1}^{n_{j+1}} \left(\frac{E_{N_j+l}(\rho)}{r_{j+1}} - \rho\right) + \cdots + \frac{1}{n_{j+k}} \sum_{l=1}^{n_{j+k}} \left(\frac{E_{N_{j+k-1}+l}(\rho)}{r_{j+k}} - \rho\right)\right].$$

Given that r^{j+1} is greater than $r_l(d_\theta)$ for all indices l on the $(j+1)$ stage of the trellis (7a), that is for all $l \in \{N_j + 1, \cdots, N_j + n_{j+1}\}$, we can conclude that

$$\sum_{l=1}^{n_{j+1}} \left(\frac{E_{N_j+l}(\rho)}{r_{j+1}} - \rho\right) > 0.$$

Similarly, all the summation terms in the bracket above are positive, and therefore (22) is established and the proof is completed.

References

[1] A. J. Viterbi and J. K. Omura, "Trellis encoding of memoryless discrete-time sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 325–332, May 1974.

[2] ——, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[3] F. Jelinek, "Tree coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 584–590, Sept. 1969.

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

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

[6] P. Jakatdar and W. A. Pearlman, "Very low rate tree coding of Stationary Gaussian Sources," in *Proc. Sixteenth Annual Conf. Inform. Sci. and Syst.*, Princeton University, Princeton, NJ, pp. 559–563, Mar. 1982.

[7] J. B. Anderson and J. B. Bodie, "Tree encoding of speech," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 379–387, July 1976.

[8] W. A. Finamore and W. A. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 144–155, Mar. 1980.

[9] W. A. Fuller, *Introduction to Statistical Time Series*. New York: Wiley, 1976.

[10] H. H. Tan, "Tree coding of discrete-time abstract alphabet stationary block-ergodic sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 671–681, Nov. 1976.

[11] R. M. Gray, "Sliding-block source coding," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 357–368, July 1975.

[12] ——, "Time-invariant trellis encoding of ergodic discrete-time sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 71–83, Jan. 1977.

[13] L. C. Stewart, R. M. Gray, and Y. Linde, "The design of trellis waveform coders," *IEEE Trans. Commun.*, vol. COM-30, pp. 702–709, Apr. 1982.

[14] Y. Linde and R. M. Gray, "A fake process approach to data compression," *IEEE Trans. Commun.*, vol. COM-26, pp. 840–847, June 1978.

An Information-Theoretic Proof of Hadamard's Inequality

THOMAS M. COVER, FELLOW, IEEE, AND ABBAS EL GAMAL, SENIOR MEMBER, IEEE

***Abstract*—Hadamard's inequality follows immediately from inspection of both sides of the entropy inequality $h(X_1, X_2, \cdots, X_n) \leqslant \Sigma h(X_i)$, when $(X_1, X_2, \cdots, X_n)$ is multivariate normal.**

I. Introduction

The most familiar of Hadamard's inequalities is that the determinant of a matrix A is less than the product of the lengths of its rows, i.e., $|A| \leqslant \prod_i (\Sigma_j a_{ij}^2)^{1/2}$. An equivalent Hadamard inequality states that, for symmetric nonnegative definite matrices K, the determinant is less than the product of the diagonal elements, i.e., $|K| \leqslant \prod k_{ii}$. To see that the first inequality follows

Manuscript received September 24, 1982; revised April 22, 1983. This work was supported in part by the National Science Foundation under Contracts ECS78-23334 and JSEP DAAG29-81-K-0057, and ENG79-08948.

The authors are with the Department of Electrical Engineering, (the first author jointly with the Statistics Department), Stanford University, Stanford, CA 94305.

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

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:

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

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

If $\boldsymbol{X}$ 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 $\boldsymbol{X}$ 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 $(x, y) \in (\mathscr{X} \times \mathscr{Y})$, 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 $x_1 \neq x_2 \in \mathscr{X}$, then the number of elements $y \in \mathscr{Y}$ with $p(x_1, y) \cdot p(x_2, y) > 0$ is different from one.**

I. Introduction

Let $(X, Y) \in (\mathscr{X} \times \mathscr{Y})$ 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

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

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