SIAM J. MATRIX ANAL. APPL.
Vol. 9, No. 3, July 1988

010

DETERMINANT INEQUALITIES VIA INFORMATION THEORY*

THOMAS M. COVER† AND JOY A. THOMAS‡

Abstract. Simple inequalities from information theory prove Hadamard's inequality and some of its generalizations. It is also proven that the determinant of a positive definite matrix is log-concave and that the ratio of the determinant of the matrix to the determinant of its principal minor $|K_n|/|K_{n-1}|$ is concave, establishing the concavity of minimum mean squared error in linear prediction. For Toeplitz matrices, the normalized determinant $|K_n|^{1/n}$ is shown to decrease with n.

Key words. inequalities, entropy, Hadamard, determinants

AMS(MOS) subject classification. 94A15

1. Introduction. The entropy inequalities of information theory have obvious intuitive meaning. For example, the entropy (or uncertainty) of a collection of random variables is less than or equal to the sum of their entropies. Letting the random variables be multivariate normal will yield Hadamard's inequality [1], [2]. We shall find many such determinant inequalities using this technique. We use throughout the fact that if

$$\phi_K(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} e^{-(1/2)\mathbf{x}^t K^{-1}\mathbf{x}} \tag{1}$$

is the multivariate normal density with mean $\mathbf{0}$ and covariance matrix K, then the entropy $h(X_1, X_2, \cdots, X_n)$ is given by

$$h(X_1,X_2,\cdots,X_n) = -\int \phi_K \ln \phi_K = \frac{1}{2}\ln (2\pi e)^n |K|, \tag{2}$$

where $|K|$ denotes the determinant of K, and ln denotes the natural logarithm. This equality is verified by direct computation with the use of

$$\int \phi_K(\mathbf{x})\mathbf{x}^t K^{-1}\mathbf{x}\, d\mathbf{x} = \sum_i \sum_j K_{ij}(K^{-1})_{ij} = n = \ln e^n. \tag{3}$$

First we give some information theory preliminaries, then the determinant inequalities.

2. Information inequalities. In this section, we introduce some of the basic information theoretic quantities and prove a few simple inequalities using convexity. We assume throughout that the vector $(X_1, X_2, \cdots, X_n)$ has a probability density

$$f(x_1, x_2, \cdots, x_n).$$

We need the following definitions.

DEFINITION. The *entropy* $h(X_1, X_2, \cdots, X_n)$, sometimes written $h(f)$, is defined by

$$h(X_1,X_2,\cdots,X_n) = -\int f \ln f. \tag{4}$$

* Received by the editors October 28, 1987; accepted for publication November 2, 1987. This work was partially supported by Bell Communications Research and by National Science Foundation grant ECS85-20136. A preliminary version of this paper appears as Bellcore Technical Memo TM-ARH-010-203.

† Departments of Electrical Engineering and Statistics, Stanford University, Stanford, California 94305.
‡ Department of Electrical Engineering, Stanford University, Stanford, California 94305.

DEFINITION. The functional $D(f\|g) = \int f(\mathbf{x}) \ln (f(\mathbf{x})/g(\mathbf{x}))\, d\mathbf{x}$ is called the *relative entropy*, where f and g are probability densities.

The relative entropy $D(f\|g)$ is also known as the Kullback–Leibler information number, information for discrimination, and information distance. We also note that $D(f\|g)$ is the error exponent in the hypothesis test of f versus g.

DEFINITION. The *conditional entropy* $h(X\,|\,Y)$ of X, given Y, is defined by

$$h(X\,|\,Y) = -\int f(x, y) \ln f(x\,|\,y)\, dx\, dy. \tag{5}$$

We now observe certain natural properties of these information quantities.

LEMMA 1. $D(f\|g) \geqq 0$.

Proof. Let A be the support of f. Then by Jensen's inequality, $-D(f\|g) = \int_A f \ln (g/f) \leqq \ln \int_A f(g/f) = \ln \int_A g \leqq \ln 1 = 0$. □

LEMMA 2. *If (X, Y) have a joint density, then $h(X\,|\,Y) = h(X, Y) - h(Y)$.*

Proof. $h(X\,|\,Y) = -\int f(x, y) \ln f(x\,|\,y)\, dx\, dy = -\int f(x, y) \ln (f(x, y)/f(y))\, dx\, dy = -\int f(x, y) \ln f(x, y)\, dx\, dy + \int f(y) \ln f(y)\, dy = h(X, Y) - h(Y)$. □

LEMMA 3. $h(X\,|\,Y) \leqq h(X)$, *with equality if and only if X and Y are independent.*

Proof.

$$h(X) - h(X\,|\,Y) = \int f(x, y) \ln (f(x\,|\,y)/f(x)) = \int f(x, y) \ln (f(x, y)/f(x)f(y)) \geqq 0,$$

by $D(f(x, y)\|f(x)f(y)) \geqq 0$. Equality implies $f(x, y) = f(x)f(y)$ almost everywhere by strict concavity of the logarithm. □

LEMMA 4 (Chain Rule). $h(X_1, X_2, \cdots, X_n) = \sum_{i=1}^n h(X_i\,|\,X_{i-1}, X_{i-2}, \cdots, X_1) \leqq \sum_{i=1}^n h(X_i)$ *with equality if and only if $X_1, X_2, \cdots, X_n$ are independent.*

Proof. The equality is the chain rule for entropies, which we get by repeatedly applying Lemma 2. The inequality follows from Lemma 3, and we have equality if and only if $X_1, X_2, \cdots, X_n$ are independent. □

LEMMA 5. *If X and Y are independent, then $h(X + Y) \geqq h(X)$.*

Proof. $h(X + Y) \geqq h(X + Y\,|\,Y) = h(X\,|\,Y) = h(X)$. □

We will also need the entropy maximizing property of the multivariate normal.

LEMMA 6. *Let the random vector $\mathbf{X} \in \mathbf{R}^n$ have zero mean and covariance $K = E\mathbf{X}\mathbf{X}^t$, i.e., $K_{ij} = EX_iX_j$, $1 \leqq i, j \leqq n$. Then $h(\mathbf{X}) \leqq \frac{1}{2} \ln (2\pi e)^n |K|$, with equality if and only if $f(\mathbf{x}) = \phi_K(\mathbf{x})$.*

Proof. Let $g(\mathbf{x})$ be any density satisfying $\int g(\mathbf{x}) x_i x_j\, d\mathbf{x} = K_{ij}$, for all i, j. Then

$$\begin{aligned} 0 &\leqq D(g\|\phi_K) \\ &= \int g \ln (g/\phi_K) \\ &= -h(g) - \int g \ln \phi_K \\ &= -h(g) - \int \phi_K \ln \phi_K \\ &= -h(g) + h(\phi_K), \end{aligned} \tag{6}$$

where the substitution $\int g \ln \phi_K = \int \phi_K \ln \phi_K$ follows from the fact that g and ϕ_K yield the same moments of the quadratic form $\ln \phi_K(\mathbf{x})$. □

Motivated by a desire to prove Szasz's generalization of Hadamard's inequality in the next section, we develop a new inequality on the entropy rates of random subsets of random variables. Let $(X_1, X_2, \cdots, X_n)$ have a density and for every $S \subseteqq \{1, 2, \cdots, n\}$, denote by $X(S)$ the subset $\{X_i: i \in S\}$.

DEFINITION. Let

$$h_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:\,|S|=k} \frac{h(X(S))}{k}. \tag{7}$$

Here $h_k^{(n)}$ is the average entropy in bits per symbol of a randomly drawn k-element subset of $\{X_1, X_2, \cdots, X_n\}$. The following lemma states that the average entropy decreases monotonically in the size of the subset.

LEMMA 7.

$$h_1^{(n)} \geqq h_2^{(n)} \geqq \cdots \geqq h_n^{(n)}. \tag{8}$$

Proof. We will first prove the last inequality, i.e., $h_n^{(n)} \leqq h_{n-1}^{(n)}$. We write

$$h(X_1, X_2, \cdots, X_n) = h(X_1, X_2, \cdots, X_{n-1}) + h(X_n | X_1, X_2, \cdots, X_{n-1})$$

$$\begin{aligned} h(X_1, X_2, \cdots, X_n) &= h(X_1, X_2, \cdots, X_{n-2}, X_n) + h(X_{n-1} | X_1, X_2, \cdots, X_{n-2}, X_n) \\ &\leqq h(X_1, X_2, \cdots, X_{n-2}, X_n) + h(X_{n-1} | X_1, X_2, \cdots, X_{n-2}) \\ &\vdots \end{aligned}$$

$$h(X_1, X_2, \cdots, X_n) \leqq h(X_2, X_3, \cdots, X_n) + h(X_1).$$

Adding these n inequalities and using the chain rule, we obtain

$$nh(X_1, X_2, \cdots, X_n) \leqq \sum_{i=1}^{n} h(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) + h(X_1, X_2, \cdots, X_n) \tag{9}$$

or

$$\frac{1}{n} h(X_1, X_2, \cdots, X_n) \leqq \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)}{n-1}, \tag{10}$$

which is the desired result $h_n^{(n)} \leqq h_{n-1}^{(n)}$.

We now prove that $h_k^{(n)} \leqq h_{k-1}^{(n)}$ for all $k \leqq n$, by first conditioning on a k-element subset, then taking a uniform choice over its $(k-1)$-element subsets. For each k-element subset, $h_k^{(k)} \leqq h_{k-1}^{(k)}$, and hence the inequality remains true after taking the expectation over all k-element subsets chosen uniformly from the n elements. □

COROLLARY. *Let $r > 0$, and define*

$$g_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:\,|S|=k} e^{rh(X(S))/k}. \tag{11}$$

Then

$$g_1^{(n)} \geqq g_2^{(n)} \geqq \cdots \geqq g_n^{(n)}. \tag{12}$$

Proof. Starting from (10) in the proof of Lemma 7, we multiply both sides by r, exponentiate, and then apply the arithmetic mean geometric mean inequality to obtain

(13)

$$\exp\left(\frac{1}{n} rh(X_1, X_2, \cdots, X_n)\right) \leqq \exp\left(\frac{1}{n}\sum_{i=1}^{n} \frac{rh(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)}{n-1}\right)$$
$$\leqq \frac{1}{n}\sum_{i=1}^{n} \exp\left(\frac{rh(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)}{n-1}\right)$$
$$\text{for all } r \geqq 0,$$

which is equivalent to $g_n^{(n)} \leqq g_{n-1}^{(n)}$. Now we use the same arguments as in Lemma 7, taking an average over all subsets to prove the result that for all $k \leqq n$, $g_k^{(n)} \leqq g_{k-1}^{(n)}$. □

Finally, we have the entropy power inequality, the only result we do not prove.

LEMMA 8. *If* **X** *and* **Y** *are independent random n-vectors with densities, then*

$$\exp\left(\frac{2}{n} h(\mathbf{X}+\mathbf{Y})\right) \geqq \exp\left(\frac{2}{n} h(\mathbf{X})\right) + \exp\left(\frac{2}{n} h(\mathbf{Y})\right). \tag{14}$$

Proof. See Shannon [3] for the statement and Stam [4] and Blachman [5] for the proof. Unlike the previous results, the proof is not elementary. □

3. Determinant inequalities. Throughout we will assume that K is a nonnegative definite symmetric $n \times n$ matrix. Let $|K|$ denote the determinant of K.

We first prove a result due to Ky Fan [6].

THEOREM 1. ln $|K|$ *is concave.*

Proof. Let X_1 and X_2 be normally distributed n-vectors, $\mathbf{X}_i \sim \phi_{K_i}(\mathbf{x})$, $i = 1, 2$. Let the random variable θ have distribution $\Pr\{\theta = 1\} = \lambda$, $\Pr\{\theta = 2\} = 1 - \lambda$, $0 \leqq \lambda \leqq 1$. Let θ, $\mathbf{X}_1$, and $\mathbf{X}_2$ be independent and let $\mathbf{Z} = \mathbf{X}_\theta$. Then $\mathbf{Z}$ has covariance $K_Z = \lambda K_1 + (1 - \lambda)K_2$. However, $\mathbf{Z}$ will not be multivariate normal. By first using Lemma 6, followed by Lemma 3, we have

$$\begin{aligned} &\tfrac{1}{2} \ln (2\pi e)^n |\lambda K_1 + (1-\lambda)K_2| \geqq h(\mathbf{Z}) \geqq h(\mathbf{Z}|\theta) \\ &\qquad = \lambda \tfrac{1}{2} \ln (2\pi e)^n |K_1| + (1-\lambda) \tfrac{1}{2} \ln (2\pi e)^n |K_2|. \end{aligned} \tag{15}$$

Thus

$$|\lambda K_1 + (1-\lambda)K_2| \geqq |K_1|^\lambda |K_2|^{1-\lambda}, \tag{16}$$

as desired. □

The next theorem, used in [7], is too easy to require a new proof, but we provide it anyway.

THEOREM 2. $|K_1 + K_2| \geqq |K_1|$.

Proof. Let **X**, **Y** be independent random vectors with $\mathbf{X} \sim \phi_{K_1}$ and $\mathbf{Y} \sim \phi_{K_2}$. Then $\mathbf{X} + \mathbf{Y} \sim \phi_{K_1+K_2}$ and hence $\frac{1}{2} \ln (2\pi e)^n |K_1 + K_2| = h(\mathbf{X} + \mathbf{Y}) \geqq h(\mathbf{X}) = \frac{1}{2} \ln (2\pi e)^n |K_1|$, by Lemma 5. □

We now give Hadamard's inequality using the proof in [2]. See also [1] for an alternative proof.

THEOREM 3 (Hadamard). $|K| \leqq \prod K_{ii}$, *with equality if and only if* $K_{ij} = 0$, $i \neq j$.

Proof. Let $\mathbf{X} \sim \phi_K$. Then

$$\frac{1}{2} \ln (2\pi e)^n |K| = h(X_1, X_2, \cdots, X_n) \leqq \sum h(X_i) = \sum_{i=1}^{n} \frac{1}{2} \ln 2\pi e |K_{ii}|, \tag{17}$$

with equality if and only if $X_1, X_2, \cdots, X_n$ are independent, i.e., $K_{ij} = 0$, $i \neq j$.

We now prove a generalization of Hadamard's inequality due to Szasz [9]. Let $K(i_1, i_2, \cdots, i_k)$ be the k-rowed principal submatrix of K formed by the rows and columns with indices $i_1, i_2, \cdots, i_k$.

THEOREM 4 (Szasz). *If K is a positive definite $n \times n$ matrix and P_k denotes the product of all the principal k-rowed minors of K, i.e.,*

$$P_k = \prod_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |K(i_1, i_2, \cdots, i_k)|, \tag{18}$$

then

$$P_1 \geqq P_2^{1/\binom{n-1}{1}} \geqq P_3^{1/\binom{n-1}{2}} \geqq \cdots \geqq P_n. \tag{19}$$

Proof. Let $\mathbf{X} \sim \phi_K$. Then the theorem follows directly from Lemma 7, with the identification $h_k^{(n)} = (1/n) \ln P_k + \frac{1}{2} \ln 2\pi e$.

We can also prove a related theorem.

THEOREM 5. *Let K be a positive definite $n \times n$ matrix and let*

$$S_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |K(i_1, i_2, \cdots, i_k)|^{1/k}. \tag{20}$$

Then

$$\frac{1}{n} \operatorname{tr}(K) = S_1^{(n)} \geqq S_2^{(n)} \geqq \cdots \geqq S_n^{(n)} = |K|^{1/n}. \tag{21}$$

Proof. This follows directly from the corollary to Lemma 7, with the identification $g_k^{(n)} = (2\pi e) S_k^{(n)}$, and $r = 2$ in (11) and (12). □

We now prove a property of Toeplitz matrices, which are important as the covariance matrices of stationary random processes. A Toeplitz matrix K is characterized by the property that $K_{ij} = K_{rs}$ if $|i - j| = |r - s|$. Let K_k denote the principal minor $K(1, 2, \cdots, k)$. For such a matrix, the following property can be proved easily from the properties of the entropy function.

THEOREM 6. *If the positive definite $n \times n$ matrix K is Toeplitz, then*

$$|K_1| \geqq |K_2|^{1/2} \geqq \cdots \geqq |K_{n-1}|^{1/(n-1)} \geqq |K_n|^{1/n} \tag{22}$$

and $|K_k|/|K_{k-1}|$ is decreasing in k.

Proof. Let $(X_1, X_2, \cdots, X_n) \sim \phi_{K_n}$. Then the quantities $h(X_k | X_{k-1}, \cdots, X_1)$ are decreasing in k, since

$$\begin{aligned} h(X_k | X_{k-1}, \cdots, X_1) &= h(X_{k+1} | X_k, \cdots, X_2) \\ &\geqq h(X_{k+1} | X_k, \cdots, X_2, X_1), \end{aligned} \tag{23}$$

where the equality follows from the Toeplitz assumption and the inequality from the fact that conditioning reduces entropy. Thus the running averages

$$\frac{1}{k} h(X_1, \cdots, X_k) = \frac{1}{k} \sum_{i=1}^{k} h(X_i | X_{i-1}, \cdots, X_1) \tag{24}$$

are decreasing in k. The theorem then follows from

$$h(X_1, X_2, \cdots, X_k) = \tfrac{1}{2} \ln (2\pi e)^k |K_k|. \qquad \square$$

Since $h(X_n|X_{n-1}, \cdots, X_1)$ is a decreasing sequence, it has a limit. Hence by the Cesáro Mean Theorem,

$$\begin{aligned}\lim_{n\to\infty}\frac{h(X_1,X_2,\cdots,X_n)}{n} &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} h(X_k|X_{k-1},\cdots,X_1)\\ &= \lim_{n\to\infty} h(X_n|X_{n-1},\cdots,X_1).\end{aligned} \tag{25}$$

Translating this to determinants, we obtain the following result:

$$\lim_{n\to\infty}|K_n|^{1/n} = \lim_{n\to\infty}\frac{|K_n|}{|K_{n-1}|}, \tag{26}$$

which is one of the simple limit theorems for determinants that can be proved using information theory.

In problems connected with maximum entropy spectrum estimation, we would like to maximize the value of the determinant of a Toeplitz matrix, subject to constraints on the values in a band around the main diagonal. Choi and Cover [10] use information theoretic arguments to show that the matrix maximizing the determinant under these constraints is the Yule–Walker extension of the values along the band.

The proof of the next inequality (Oppenheim [11], Marshall and Olkin [12, p. 475]) follows immediately from the entropy power inequality, but because of the complexity of the proof of the entropy power inequality, is not offered as a simpler proof.

THEOREM 7 (Minkowski inequality [13]).

$$|K_1+K_2|^{1/n} \geqq |K_1|^{1/n} + |K_2|^{1/n}. \tag{27}$$

Proof. Let $\mathbf{X}_1$, $\mathbf{X}_2$ be independent with $\mathbf{X}_i \sim \phi_{K_i}$. Noting that $\mathbf{X}_1 + \mathbf{X}_2 \sim \phi_{K_1+K_2}$ and using the entropy power inequality (Lemma 8) yields

$$\begin{aligned}(2\pi e)|K_1+K_2|^{1/n} &= e^{(2/n)h(\mathbf{X}_1+\mathbf{X}_2)}\\ &\geqq e^{(2/n)h(\mathbf{X}_1)} + e^{(2/n)h(\mathbf{X}_2)}\\ &= (2\pi e)|K_1|^{1/n} + (2\pi e)|K_2|^{1/n}.\end{aligned} \tag{28}$$

□

4. Inequalities for ratios of determinants. We first prove a stronger version of Hadamard's theorem due to Ky Fan [8].

THEOREM 8. *For all* $1 \leqq p \leqq n$,

$$\frac{|K|}{|K(p+1,p+2,\cdots,n)|} \leqq \prod_{i=1}^{p}\frac{|K(i,p+1,p+2,\cdots,n)|}{|K(p+1,p+2,\cdots,n)|}. \tag{29}$$

Proof. We use the same idea as in Theorem 3, except that we use the conditional form of Lemma 3:

(30)

$$\begin{aligned}\frac{1}{2}\ln(2\pi e)^p\frac{|K|}{|K(p+1,p+2,\cdots,n)|} &= h(X_1,X_2,\cdots,X_p|X_{p+1},X_{p+2},\cdots,X_n)\\ &\leqq \sum h(X_i|X_{p+1},X_{p+2},\cdots,X_n)\\ &= \sum_{i=1}^{p}\frac{1}{2}\ln 2\pi e\frac{|K(i,p+1,p+2,\cdots,n)|}{|K(p+1,p+2,\cdots,n)|}.\end{aligned}$$

□

Before developing Theorem 9, we make an observation about minimum mean squared error linear prediction. If $(X_1, X_2, \cdots, X_n) \sim \phi_{K_n}$, we know that the conditional density of X_n given $(X_1, X_2, \cdots, X_{n-1})$ is univariate normal with mean linear in $X_1, X_2, \cdots, X_{n-1}$ and conditional variance σ_n^2. Here σ_n^2 is the minimum mean squared error $E(X_n - \hat{X}_n)^2$ over all linear estimators $\hat{X}_n$ based on $X_1, X_2, \cdots, X_{n-1}$.

LEMMA 9. $\sigma_n^2 = |K_n|/|K_{n-1}|$.

Proof. Using the conditional normality of X_n, Lemma 2 results in

$$\begin{aligned}\tfrac{1}{2}\ln 2\pi e\sigma_n^2 &= h(X_n|X_1,X_2,\cdots,X_{n-1})\\ &= h(X_1,X_2,\cdots,X_n)-h(X_1,X_2,\cdots,X_{n-1})\\ &= \tfrac{1}{2}\ln(2\pi e)^n|K_n| - \tfrac{1}{2}\ln(2\pi e)^{n-1}|K_{n-1}|\\ &= \tfrac{1}{2}\ln 2\pi e|K_n|/|K_{n-1}|.\end{aligned} \tag{31}$$

□

Minimization of σ_n^2 over a set of allowed covariance matrices $\{K_n\}$ is aided by the following theorem.

THEOREM 9. $\ln(|K_n|/|K_{n-p}|)$ *is concave in* K_n.

Proof. We remark that Theorem 1 cannot be used because $\ln(|K_n|/|K_{n-p}|)$ is the difference of two concave functions. Let $\mathbf{Z} = \mathbf{X}_\theta$, where $\mathbf{X}_1 \sim \phi_{S_n}(\mathbf{x})$, $\mathbf{X}_2 \sim \phi_{T_n}(\mathbf{x})$, $\Pr\{\theta = 1\} = \lambda = 1 - \Pr\{\theta = 2\}$, and $\mathbf{X}_1, \mathbf{X}_2, \theta$ are independent. The covariance matrix K_n of $\mathbf{Z}$ is given by

$$K_n = \lambda S_n + (1-\lambda)T_n. \tag{32}$$

The following chain of inequalities proves the theorem:

$$\begin{aligned}&\lambda\frac{1}{2}\ln(2\pi e)^p|S_n|/|S_{n-p}| + (1-\lambda)\frac{1}{2}\ln(2\pi e)^p|T_n|/|T_{n-p}|\\ &\stackrel{(a)}{=} \lambda h(X_{1n},X_{1,n-1},\cdots,X_{1,n-p+1}|X_{11},\cdots,X_{1,n-p})\\ &\qquad + (1-\lambda)h(X_{2n},X_{2,n-1},\cdots,\\ &\qquad\qquad X_{2,n-p+1}|X_{21},\cdots,X_{2,n-p})\\ &= h(Z_n,Z_{n-1},\cdots,Z_{n-p+1}|Z_1,\cdots,Z_{n-p},\theta)\\ &\stackrel{(b)}{\leqq} h(Z_n,Z_{n-1},\cdots,Z_{n-p+1}|Z_1,\cdots,Z_{n-p})\\ &\stackrel{(c)}{\leqq} \frac{1}{2}\ln(2\pi e)^p\frac{|K_n|}{|K_{n-p}|},\end{aligned} \tag{33}$$

where (a) follows from

$$h(X_n,X_{n-1},\cdots,X_{n-p+1}|X_1,\cdots,X_{n-p}) = h(X_1,\cdots,X_n) - h(X_1,\cdots,X_{n-p}),$$

(b) follows from the conditioning lemma, and (c) follows from a conditional version of Lemma 6. □

The above theorem for the case $p = 1$ is due to Bergstrøm [14]. However, for $p = 1$, we can prove an even stronger theorem, also due to Bergstrøm [14].

THEOREM 10. $|K_n|/|K_{n-1}|$ *is concave in* K_n.

Proof. Again we use the properties of Gaussian random variables. Let us assume that we have two independent Gaussian random vectors, $\mathbf{X} \sim \phi_{A_n}$ and $\mathbf{Y} \sim \phi_{B_n}$. Let

$\mathbf{Z} = \mathbf{X} + \mathbf{Y}$. Then

$$
\begin{aligned}
\frac{1}{2}\ln 2\pi e\frac{|A_n+B_n|}{|A_{n-1}+B_{n-1}|} &\overset{(a)}{=} h(Z_n|Z_{n-1},Z_{n-2},\cdots,Z_1)\\
&\overset{(b)}{\geqq} h(Z_n|Z_{n-1},Z_{n-2},\cdots,Z_1,X_{n-1},X_{n-2},\cdots,X_1,\\
&\qquad\qquad Y_{n-1},Y_{n-2},\cdots,Y_1)\\
&\overset{(c)}{=} h(X_n+Y_n|X_{n-1},X_{n-2},\cdots,X_1,Y_{n-1},Y_{n-2},\cdots,Y_1)\\
&\overset{(d)}{=} E\frac{1}{2}\ln[2\pi e \operatorname{Var}(X_n+Y_n|X_{n-1},X_{n-2},\cdots,X_1,\\
&\qquad\qquad Y_{n-1},Y_{n-2},\cdots,Y_1)]\\
&\overset{(e)}{=} E\frac{1}{2}\ln[2\pi e(\operatorname{Var}(X_n|X_{n-1},X_{n-2},\cdots,X_1)\\
&\qquad\qquad +\operatorname{Var}(Y_n|Y_{n-1},Y_{n-2},\cdots,Y_1))]\\
&\overset{(f)}{=} E\frac{1}{2}\ln\left(2\pi e\left(\frac{|A_n|}{|A_{n-1}|}+\frac{|B_n|}{|B_{n-1}|}\right)\right)\\
&= \frac{1}{2}\ln\left(2\pi e\left(\frac{|A_n|}{|A_{n-1}|}+\frac{|B_n|}{|B_{n-1}|}\right)\right).
\end{aligned}
\tag{34}
$$

In the above derivation, (a) follows from Lemma 9, (b) from the fact the conditioning decreases entropy, and (c) from the fact that Z is a function of X and Y. $X_n + Y_n$ is Gaussian conditioned on $X_1, X_2, \cdots, X_{n-1}, Y_1, Y_2, \cdots, Y_{n-1}$, and hence we can express its entropy in terms of its variance, obtaining (d). Then (e) follows from the independence of X_n and Y_n conditioned on the past $X_1, X_2, \cdots, X_{n-1}, Y_1, Y_2, \cdots, Y_{n-1}$, and (f) follows from the fact that for a set of jointly Gaussian random variables, the conditional variance is constant, independent of the conditioning variables (Lemma 9). In general, by setting $A = \lambda S$ and $B = \bar{\lambda}T$, we obtain

$$
\frac{|\lambda S_n+\bar{\lambda}T_n|}{|\lambda S_{n-1}+\bar{\lambda}T_{n-1}|}\geqq\lambda\frac{|S_n|}{|S_{n-1}|}+\bar{\lambda}\frac{|T_n|}{|T_{n-1}|},
\tag{35}
$$

i.e., $|K_n|/|K_{n-1}|$ is concave. Simple examples show that $|K_n|/|K_{n-p}|$ is not necessarily concave for $p \geqq 2$. □

5. Remarks. Concavity and Jensen's inequality play a role in all the proofs. The inequality $D(f\|g) = \int f\ln(f/g) \geqq 0$ is at the root of most of them.

Acknowledgment. We thank A. Bernardi and S. Pombra for their contributions.

REFERENCES

[1] A. MARSHALL AND I. OLKIN, *A convexity proof of Hadamard's inequality*, Amer. Math. Monthly, 89 (1982), pp. 687–688.

[2] T. COVER AND A. EL GAMAL, *An information theoretic proof of Hadamard's inequality*, IEEE Trans. Inform. Theory, IT-29 (1983), pp. 930–931.

[3] C. E. SHANNON, *A mathematical theory of communication*, Bell System Tech. J., 27 (1948), pp. 623–656.

[4] A. STAM, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Inform. and Control, 2 (1959), pp. 101–112.

[5] N. BLACHMAN, *The convolution inequality for entropy powers*, IEEE Trans. Inform. Theory, IT-11 (1965), pp. 267–271.

[6] KY FAN, *On a theorem of Weyl concerning the eigenvalues of linear transformations*, II, Proc. Nat. Acad. Sci. U.S.A., 36 (1950), pp. 31–35.

[7] S. POMBRA AND T. COVER, *Gaussian Feedback Capacity*, Technical Report 63, Dept. of Statistics, Stanford Univ., Stanford, CA, September 1987; IEEE Trans. Inform. Theory, to appear.

[8] KY FAN, *Some inequalities concerning positive-definite matrices*, Proc. Cambridge Philos. Soc., 51 (1955), pp. 414–421.

[9] L. MIRSKY, *On a generalization of Hadamard's determinantal inequality due to Szasz*, Arch. Math., VIII (1957), pp. 274–275.

[10] B. S. CHOI AND T. M. COVER, *An information-theoretic proof of Burg's maximum entropy spectrum*, Proc. IEEE, 72 (1984), pp. 1094–1095.

[11] A. OPPENHEIM, *Inequalities connected with definite Hermitian forms*, J. London Math. Soc., 5 (1930), pp. 114–119.

[12] A. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[13] H. MINKOWSKI, *Diskontinuitätsbereich für arithmetische Äquivalenz*, Journal für Math., 129 (1950), pp. 220–274.

[14] R. BELLMAN, *Notes on matrix theory*—IV: *an inequality due to Bergstrøm*, Amer. Math. Monthly, 62 (1955), pp. 172–173.