J. of Inequal. & Appl., 1999, Vol. 3, pp. 143–152
Reprints available directly from the publisher
Photocopying permitted by license only

A Stronger Version of Matrix Convexity as Applied to Functions of Hermitian Matrices

ABRAM KAGAN* and PAUL J. SMITH

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

(Received 14 January 1998; Revised 17 February 1998)

A stronger version of matrix convexity, called hyperconvexity is introduced. It is shown that the function A^2 is hyperconvex on the set of Hermitian matrices A and A^{-1} is hyperconvex on the set of positive definite Hermitian matrices. The new concept makes it possible to consider weighted averages of matrices of different orders. Proofs use properties of the Fisher information matrix, a fundamental concept of mathematical statistics.

Keywords: Matrix convexity; Hyperconvexity; Fisher information

AMS 1991 Subject Classification: Primary 15A45; Secondary 62B10, 15A09

1. INTRODUCTION

It is known that the function $f_1(A) = A^2$ is matrix convex on the set $\mathcal{H}_s$ of Hermitian $s \times s$ matrices and $f_2(A) = A^{-1}$ is matrix convex on the set $\mathcal{H}_s^+$ of positive definite Hermitian $s \times s$ matrices. For analytical proofs of these and related results see [4, Ch. 16].

The authors have recently suggested in [3] two purely statistical proofs of the above results, one based on the Gauss–Markov theorem for least squares estimators and the other based on properties of the Fisher information matrix.

* Corresponding author.

It turns out that the statistical viewpoint is not only of a methodological interest but also prompts a stronger version of matrix convexity where weighted averages

$$w_1 A_1 + \cdots + w_n A_n$$

with

$$w_1 > 0, \ldots, w_n > 0, \quad w_1 + \cdots + w_n = 1 \tag{1.1}$$

are replaced with

$$B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n. \tag{1.2}$$

In (1.2), A_j is an $s_j \times s_j$ matrix, B_j is an $s_j \times m$ matrix, $j = 1, 2, \ldots, n$ and the role of (1.1) is played by

$$B_1^{\mathrm{T}} B_1 + \cdots + B_n^{\mathrm{T}} B_n = I_m, \tag{1.3}$$

I_m being the identity $m \times$ m matrix.

In Section 2 it is proved that

$$B_1^{\mathrm{T}} A_1^2 B_1 + \cdots + B_n^{\mathrm{T}} A_n^2 B_n \geq (B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n)^2 \tag{1.4}$$

for any Hermitian $s_j \times s_j$ matrices A_j, $j = 1, \ldots, n$. (As usual, $A \geq B$ means that $A - B$ is positive semidefinite.) We call the property (1.4) *hyperconvexity* of $f_1(A) = A^2$ on the set $\mathcal{H} = \cup_{s=1}^{\infty} \mathcal{H}_s$ of *all* Hermitian matrices.

If $s_j = s$ and $B_j = v_j I_s, j = 1, \ldots, n$ with $v_1^2 + \cdots + v_n^2 = 1$ then for $w_j = v_j^2$, (1.4) becomes

$$w_1 A_1^2 + \cdots + w_n A_n^2 \geq (w_1 A_1 + \cdots + w_n A_n)^2, \tag{1.5}$$

i.e., hyperconvexity of A^2 implies convexity. Similarly, it is proved in Section 2 that $f_2(A) = A^{-1}$ is hyperconvex on the set $\mathcal{H}^+ = \cup_{s=1}^{\infty} \mathcal{H}_s^+$ of *all* positive definite Hermitian matrices.

The proofs are based on basic properties of the Fisher information matrix for normal random vectors. In Section 3 analytical proofs of hyperconvexity of A^2 on $\mathcal{H}$ and A^{-1} on $\mathcal{H}^+$ are given. These proofs were suggested by Israel Gohberg and are published here with his kind consent. Now a few comments on statistical approach to matrix inequalities.

In Carlen [1] it was proved analytically that if $\mathbf{X}$ is an s-variate random vector decomposed as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \tag{1.6}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are r- and q-variate subvectors then the matrices of Fisher information $I(\mathbf{X})$, $I(\mathbf{X}_1)$ and $I(\mathbf{X}_2)$ (on s-, r-, and q-dimensional location parameter) contained in $\mathbf{X}$, $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively (see Section 2 for the definition of the Fisher information matrix), satisfy the inequality

$$\operatorname{tr} I(\mathbf{X}) \geq \operatorname{tr} I(\mathbf{X}_1) + \operatorname{tr} I(\mathbf{X}_2). \tag{1.7}$$

In [2] a simple statistical interpretation and proof of (1.7) is given.

Let now A be an arbitrary positive definite Hermitian $s \times s$ matrix decomposed as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{1.8}$$

where A_{11} and A_{22} are $r \times r$ and $q \times q$ matrices.

Consider an s-variate normal random vector $\mathbf{X}$ decomposed as in (1.6) with mean

$$E\mathbf{X} = \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

and variance–covariance matrix

$$V(\mathbf{X}) = A.$$

For such $\mathbf{X}$, $\mathbf{X}_1$ and $\mathbf{X}_2$ the Fisher information matrices $I(\mathbf{X})$, $I(\mathbf{X}_1)$ and $I(\mathbf{X}_2)$, are easily calculated,

$$I(\mathbf{X}) = A^{-1}, \quad I(\mathbf{X}_1) = A_{11}^{-1}, \quad I(\mathbf{X}_2) = A_{22}^{-1}$$

so that (1.7) results in

$$\operatorname{tr} A^{-1} \geq \operatorname{tr} A_{11}^{-1} + \operatorname{tr} A_{22}^{-1}. \tag{1.9}$$

Certainly (1.9) can be proved analytically but the statistical interpretation sheds new light on (1.9) and similar inequalities and prompts possible extensions.

2. HYPERCONVEXITY OF A^2 AND A^{-1}

THEOREM 1 *Let A_j be a Hermitian $s_j \times s_j$ matrix, B_j an $s_j \times m$ matrix, $j = 1, \ldots, n$. If*

$$B_1^{\mathrm{T}} B_1 + \cdots + B_1^{\mathrm{T}} B_n = I_m, \tag{2.1}$$

the identity $m \times m$ matrix, then

$$B_1^{\mathrm{T}} A_1^2 B_1 + \cdots + B_n^{\mathrm{T}} A_n^2 B_n \geq (B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n)^2. \tag{2.2}$$

Remark If $A_1, \ldots, A_n$ are of different orders the standard weighted average, $w_1 A_1 + \cdots + w_n A_n$, where $w_1 > 0, \ldots, w_n > 0$, $w_1 + \cdots + w_n = 1$ does not make sense while the matrix weighted average $B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n$, where $B_1, \ldots, B_n$ are subject to (2.1), is a well defined $m \times m$ matrix.

Proof of Theorem 1 Consider an s_j-variate normal random vector $\mathbf{X}_j$ with mean vector $E\mathbf{X}_j = A_j B_j \boldsymbol{\theta}$ and the identity variance–covariance matrix, $V(\mathbf{X}_j) = I_{s_j}$. Here

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$$

is an m-dimensional parameter.

The probability density function of $\mathbf{X}_j$ is

$$\varphi_j(\mathbf{x}; \boldsymbol{\theta}) = (2\pi)^{-s_j/2} \exp\left[-\frac{1}{2}(\mathbf{x} - A_j B_j \boldsymbol{\theta})^{\mathrm{T}} (\mathbf{x} - A_j B_j \boldsymbol{\theta})\right], \quad \mathbf{x} \in \mathbf{R}^{s_j}. \tag{2.3}$$

For an arbitrary random vector $\mathbf{X}$ with density $f(\mathbf{x};\boldsymbol{\theta})$ depending on $\boldsymbol{\theta}$, the $m\times m$ matrix of Fisher information on $\boldsymbol{\theta}$ contained in $\mathbf{X}$ is defined as

$$\begin{aligned} I(\mathbf{X};\boldsymbol{\theta}) &= E\Big[(D\log f)(D\log f)^{\mathrm{T}}\Big] \\ &= \int (D\log f)(D\log f)^{\mathrm{T}} f(\mathbf{x};\boldsymbol{\theta})\,\mathrm{d}\mathbf{x} \end{aligned} \tag{2.4}$$

where

$$D = \begin{bmatrix} \partial/\partial\theta_1 \\ \vdots \\ \partial/\partial\theta_m \end{bmatrix}.$$

The definition requires certain regularity conditions from $f(\mathbf{x};\boldsymbol{\theta})$ that are fulfilled for $\varphi_j(\mathbf{x};\boldsymbol{\theta})$ (see, e.g. [5, Ch. 5].) Simple calculations lead to

$$I(\mathbf{X}_j;\boldsymbol{\theta}) = B_j^{\mathrm{T}} A_j^2 B_j. \tag{2.5}$$

Assuming $\mathbf{X}_1,\ldots,\mathbf{X}_n$ independent, consider the following statistic $\mathbf{T}:\mathbf{R}^{s_1+\cdots+s_n}\to\mathbf{R}^m$:

$$\mathbf{T} = B_1^{\mathrm{T}}\mathbf{X}_1 + \cdots + B_n^{\mathrm{T}}\mathbf{X}_n. \tag{2.6}$$

As a linear combination of jointly normal random vectors, $\mathbf{T}$ has m-variate normal distribution with mean

$$E\mathbf{T} = B_1^{\mathrm{T}}E\mathbf{X}_1 + \cdots + B_n^{\mathrm{T}}E\mathbf{X}_n = (B_1^{\mathrm{T}}A_1B_1 + \cdots + B_n^{\mathrm{T}}A_nB_n)\boldsymbol{\theta}$$

and variance–covariance matrix

$$V(\mathbf{T}) = B_1^{\mathrm{T}}V(\mathbf{X}_1)B_1 + \cdots + B_n^{\mathrm{T}}V(\mathbf{X}_n)B_n = I_m$$

due to (2.1).

Applying (2.4) to $\mathbf{T}$ results in

$$I(\mathbf{T};\boldsymbol{\theta}) = (B_1^{\mathrm{T}}A_1B_1 + \cdots + B_n^{\mathrm{T}}A_nB_n)^2. \tag{2.7}$$

Two fundamental properties of the Fisher information matrix are:

(i) For independent $\mathbf{X}_1, \ldots, \mathbf{X}_n$,

$$I(\mathbf{X}_1, \ldots, \mathbf{X}_n; \boldsymbol{\theta}) = I(\mathbf{X}_1; \boldsymbol{\theta}) + \cdots + I(\mathbf{X}_n; \boldsymbol{\theta}); \tag{2.8}$$

(ii) For any statistic $\mathbf{T} = \mathbf{T}(X_1, \ldots, X_n)$,

$$I(\mathbf{T}; \boldsymbol{\theta}) \leq I(\mathbf{X}_1, \ldots, \mathbf{X}_n; \boldsymbol{\theta}). \tag{2.9}$$

Combining (2.5), (2.7)–(2.9) proves (2.2).

THEOREM 2 *Let A_j be a positive definite Hermitian $s_j \times s_j$ matrix, B_j an $s_j \times m$ matrix, $j = 1, \ldots, n$. If $B_1, \ldots, B_n$ are subject to* (2.1) *then*

$$B_1^{\mathrm{T}} A_1^{-1} B_1 + \cdots + B_n^{\mathrm{T}} A_n^{-1} B_n \geq (B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n)^{-1}. \tag{2.10}$$

In other words, the function A^{-1} is hyperconvex on $\mathcal{H}^+$.

Before proving Theorem 2 notice that $m \times m$ matrix $B = B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n$ is invertible. In fact, let $\mathbf{x} \in \mathbf{R}^m$ be such that $\mathbf{x}^{\mathrm{T}} B \mathbf{x} = 0$. It means that

$$\mathbf{x}^{\mathrm{T}} B_j^{\mathrm{T}} A_j B_j \mathbf{x} = 0, \quad j = 1, \ldots, n$$

or that

$$\mathbf{y}_j^{\mathrm{T}} A_j \mathbf{y}_j = 0, \quad j = 1, \ldots, n$$

for $\mathbf{y}_j = B_j \mathbf{x}$.

Since A_j is positive definite, $\mathbf{y}_j = \mathbf{0}$, the null vector is $\mathbf{R}^{s_j}$. Then

$$0 = \mathbf{y}_1^{\mathrm{T}} \mathbf{y}_1 + \cdots + \mathbf{y}_n^{\mathrm{T}} \mathbf{y}_n = \mathbf{x}^{\mathrm{T}} (B_1^{\mathrm{T}} B_1 + \cdots + B_n^{\mathrm{T}} B_n) \mathbf{x} = \mathbf{x}^{\mathrm{T}} \mathbf{x}.$$

Thus, $\mathbf{x} = \mathbf{0}$, the null vector is $\mathbf{R}^n$.

Proof of Theorem 2 The idea is the same as in the proof of Theorem 1, though the choice of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is different.

Namely, let $\mathbf{X}_j$ be an s_j-variate normal random vector with $E\mathbf{X}_j = B_j \boldsymbol{\theta}$ and $V(\mathbf{X}_j) = A_j$. The probability density function of $\mathbf{X}_j$ is

$$\varphi_j(\mathbf{x}; \boldsymbol{\theta}) = (2\pi)^{-s_j/2} (\det A_j^{-1/2}) \exp\left[-\frac{1}{2} (\mathbf{x} - B_j \boldsymbol{\theta})^{\mathrm{T}} A_j^{-1} (\mathbf{x} - B_j \boldsymbol{\theta}) \right],$$
$$\mathbf{x} \in \mathbf{R}^{s_j}. \tag{2.11}$$

Applying (2.4) to the density (2.11) leads to

$$I(\mathbf{X}_j;\boldsymbol{\theta}) = B_j^{\mathrm{T}} A_j^{-1} B_j. \tag{2.12}$$

Again assuming $\mathbf{X}_1, \ldots, \mathbf{X}_n$ independent one gets for $\mathbf{T}$ defined in (2.6):

$$E\mathbf{T} = B_1^{\mathrm{T}} B_1 \boldsymbol{\theta} + \cdots + B_n^{\mathrm{T}} B_n \boldsymbol{\theta} = \boldsymbol{\theta}$$

due to (2.1), and

$$V(\mathbf{T}) = B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n.$$

The matrix of Fisher information on $\boldsymbol{\theta}$ in $\mathbf{T}$ is

$$I(\mathbf{T};\boldsymbol{\theta}) = (B_1^{\mathrm{T}} A_1 B_1 + \cdots + B_n^{\mathrm{T}} A_n B_n)^{-1}. \tag{2.13}$$

Combining (2.8), (2.9), (2.12), and (2.13) proves (2.10).

3. ANALYTICAL PROOFS OF THEOREMS 1 AND 2

Let P be the orthogonal projector from $\mathbf{R}^N$ into a subspace $\mathbf{R}^m$. One may always assume that

$$P = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.1}$$

LEMMA 1 *For any Hermitian $N \times N$ matrix A,*

$$PA^2P \geq (PAP)^2. \tag{3.2}$$

Proof of Lemma 1 Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{3.3}$$

where A_{11}, A_{22} are $m \times m$ and $(N-m) \times (N-m)$ matrices respectively, $A_{21} = A_{12}^{\mathrm{T}}$. Then

$$(PAP)^2 = \begin{bmatrix} A_{11}^2 & 0 \\ 0 & 0 \end{bmatrix}, \qquad PA^2P = \begin{bmatrix} A_{11}^2 + A_{12}A_{12}^{\mathrm{T}} & 0 \\ 0 & 0 \end{bmatrix}$$

whence (3.2), since $A_{12}A_{12}^{\mathrm{T}}$ is positive semidefinite.

For $A_1,\dots,A_n$, $B_1,\dots,B_n$ from Theorem 1 let

$$A = \operatorname{diag}(A_1,\dots,A_n) = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{bmatrix},$$

an $N\times N$ matrix with $N=s_1+\cdots+s_n$ and

$$B = \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix},$$

an $N\times m$ matrix. One has

$$B^{\mathrm{T}}AB = B_1^{\mathrm{T}}A_1B_1+\cdots+B_n^{\mathrm{T}}A_nB_n$$

and

$$B^{\mathrm{T}}B = I_m$$

whence, by the way, $m\le N$.

Let $\mathbf{b}_{m+1},\dots,\mathbf{b}_N$ be (column) vectors in $\mathbf{R}^N$ that, together with the columns $\mathbf{b}_1,\dots,\mathbf{b}_m$ of B form an orthonormal basis in $\mathbf{R}^N$. The matrix

$$\tilde{B} = [\mathbf{b}_1\cdots\mathbf{b}_m\mathbf{b}_{m+1}\cdots\mathbf{b}_N]$$

is an orthogonal $N\times N$ matrix.

If P is the projector into the m-dimensional subspace $Sp(\mathbf{b}_1,\dots,\mathbf{b}_m)$ then

$$P\tilde{B} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$$

so that

$$\tilde{B}^{\mathrm{T}}PA^2P\tilde{B} = \begin{bmatrix} B^{\mathrm{T}}A^2B & 0 \\ 0 & 0 \end{bmatrix} \tag{3.4}$$

and

$$(\tilde{B}^{\mathrm{T}}PAP\tilde{B})^2 = \begin{bmatrix} (B^{\mathrm{T}}AB)^2 & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.5}$$

On the other side,

$$(\tilde{B}^{\mathrm{T}} PAP\tilde{B})^2 = \tilde{B}^{\mathrm{T}}(PAP)\tilde{B}\tilde{B}^{\mathrm{T}}(PAP)\tilde{B} = \tilde{B}^{\mathrm{T}}(PAP)^2\tilde{B} \tag{3.6}$$

since $\tilde{B}\tilde{B}^{\mathrm{T}} = I_N$.

By Lemma 1, $PA^2P \geq (PAP)^2$ implying

$$\tilde{B}^{\mathrm{T}}(PA^2P)\tilde{B} \geq \tilde{B}^{\mathrm{T}}(PAP)^2\tilde{B}. \tag{3.7}$$

Combining (3.4)–(3.7) proves (2.2).

To prove analytically Theorem 2, we need the following analog of Lemma 1.

LEMMA 2 *If P is the projector from Lemma* 1 *and A a positive definite Hermitian $N \times N$ matrix then*

$$PA^{-1}P \geq (PAP)^{-1}, \tag{3.8}$$

the left and right hand side of (3.8) *being considered $m \times m$ matrices.*

Proof Represent the matrix (3.3) as

$$A = \begin{bmatrix} I_m & 0 \\ C & I_{N-m} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I_m & C^{\mathrm{T}} \\ 0 & I_{N-m} \end{bmatrix} \tag{3.9}$$

where

$$C = A_{21}A_{11}^{-1}, \quad C^{\mathrm{T}} = A_{11}^{-1}A_{12}, \quad D = A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

Then

$$A^{-1} = \begin{bmatrix} I_m & -C^{\mathrm{T}} \\ 0 & I_{N-m} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I_m & 0 \\ -C & I_{N-m} \end{bmatrix} \tag{3.10}$$

whence

$$PA^{-1}P = A_{11}^{-1} + C^{\mathrm{T}}D^{-1}C, \quad (PAP)^{-1} = A_{11}^{-1}$$

proving (3.8).

Starting now with (3.8) and repeating the arguments used in the analytical proof of Theorem 1, one comes to

$$B^{\mathrm{T}}A^{-1}B \geq (B^{\mathrm{T}}AB)^{-1} \tag{3.11}$$

which is exactly (2.10).

References

[1] E.A. Carlen, Superadditivity of Fisher's information and logarithmic Sobolev inequalities, *J. Funct. Anal.*, **101** (1991), 194–211.
[2] A. Kagan and Z. Landsman, Statistical meaning of Carlen's superadditivity of the Fisher information, *Stat. Prob. Letters*, **32** (1997), 175–179.
[3] A. Kagan and P.J. Smith, *Least squares, Fisher information and matrix convexity*, preprint.
[4] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.
[5] C.R. Rao, *Linear Statistical Inference and Its Applications*, 2nd edn., Wiley, New York, 1973.