Statistical proofs of some matrix inequalities

C. Radhakrishna Rao

Department of Statistics, 325 Joab L. Thomas Bldg., Pennsylvania State University, University Park, PA 16802-2111, USA

Received 12 March 1999; accepted 16 November 1999

Submitted by S. Puntanen

Abstract

Matrix algebra is extensively used in the study of linear models and multivariate analysis (see for instance Refs. [18,21]). During recent years, there have been a number of papers where statistical results are used to prove some matrix theorems, especially matrix inequalities (Refs. [5,7,8,10,14,15]). In this paper, a number of matrix results are proved using some properties of Fisher information and covariance matrices. A unified approach is provided through the use of Schur complements. It may be noted that the statistical results used are derivable without using matrix theory. © 2000 Elsevier Science Inc. All rights reserved.

AMS classification: 15A45; 62B10; 62G05

Keywords: Carlen’s inequality; Cauchy–Schwarz inequality; Generalized inverse; Harmonic mean inequality; Information inequalities; Kronecker product; Miline’s inequality; Parallel sum of matrices; Schur complement; Schur product

1. Introduction

This paper was motivated by a lecture given by Kagan [9] on statistical proofs of some matrix results. I have shown how some of the important results in matrix algebra can be derived in an elegant manner from statistical results which can be proved without using matrix theory. The results might be of interest to teachers and students of statistics. The statistical results used in the proofs are briefly reviewed in this section. We consider matrices and vectors with elements from a real field. Some
of the results can be easily extended to the case where the elements are complex numbers.

**Notation:** A symmetric \( n \times n \) matrix \( B \) is said to be nonnegative definite (nnd) if \( a' Ba \geq 0 \) for all \( n \)-vectors \( a \), which is indicated by writing \( B \geq 0 \). \( B \) is said to be positive definite (pd) if \( a' Ba > 0 \) for all non-null \( n \)-vectors \( a \), which is indicated by writing \( B > 0 \). The notation \( C : p \times k \) is used to denote a matrix \( C \) with \( p \) rows and \( k \) columns.

**Information matrix:** Let \( X \) be a \( p \)-vector random variable (rv) with the density function \( f(x, \theta) \), where \( \theta'' = (\theta_1, \ldots, \theta_q) \) is a \( q \)-vector parameter. Fisher information matrix on \( \theta \) based on \( X \) is defined by

\[
I(\theta) = \left( i_{rs} \right), \quad i_{rs} = E \left[ \frac{\partial \log f}{\partial \theta_r} \cdot \frac{\partial \log f}{\partial \theta_s} \right].
\]

Let \( T(X) \) be a function of \( X \). Then, under some conditions, we have the following results (see [18, pp. 329–332]) for proofs):

1. \( I_T(\theta) \leq I_X(\theta) \). \hspace{1cm} (1.1)

If \( X_1 \) and \( X_2 \) are independent rv’s, whose densities involve the same parameter \( \theta \), then

\[
I_{X_1, X_2}(\theta) = I_{X_1}(\theta) + I_{X_2}(\theta).
\] \hspace{1cm} (1.2)

Let \( X \sim N_p(B \mu, A) \), i.e., distributed as \( p \)-variate normal with mean \( B \mu \), where \( B : p \times k \), and \( \mu \) is a \( k \)-vector parameter, and a pd covariance matrix \( A \). Further, let \( G : q \times p \) of rank \( q \), and \( Y = GX \). Then

\[
I_X(\mu) = B' A^{-1} B, \hspace{1cm} (1.3)
\]

\[
I_Y(\mu) = B' G' (GAG')^{-1} GB, \hspace{1cm} (1.4)
\]

Further, if \( G \) is invertible, then

\[
I_X(\mu) = I_Y(\mu). \hspace{1cm} (1.5)
\]

**Covariance matrix:** Let \( X \) be a \( p \)-vector rv and \( Y \) be a \( q \)-vector rv with zero mean values. We denote the covariance matrix between \( X \) and \( Y \) by

\[
C(X, Y) = E(XY') = C(Y, X)',
\]

which is a \( p \times q \) matrix. Then:

1. \( C(X, X) : p \times p \) is an nnd matrix. \hspace{1cm} (1.6)

2. If the rank of \( C(X, X) \) is \( k < p \), then there exists a matrix \( D : p \times (p - k) \) such that

\[
Y = D' X = 0 \quad \text{a.s.} \hspace{1cm} (1.7)
\]
3. Given an nnd matrix, $A: p \times p$, there exists a $p$-vector rv $X$ such that

$$C(X, X) = A.$$ (1.8)

Results (1.6) and (1.7) are easy to demonstrate.

Note: Result (3) follows if we assume the existence of an nnd matrix $A^{1/2}$ such that $(A^{1/2})(A^{1/2}) = A$ (or a matrix $B$ such that $A = BB'$). Then we need only choose an rv $Z$ with $I$ as its covariance matrix and define $X = A^{1/2}Z$ (or $X = BZ$). However, we prove (3) without using the matrix result on the existence of $A^{1/2}$ (or the factorization $A = BB'$).

We assume that there exists a $(p - 1)$-vector rv with any given $(p - 1) \times (p - 1)$ nnd matrix as its covariance matrix and show that there exists a $p$-vector rv associated with any given $p \times p$ nnd matrix. Since an rv exists when $p = 1$, there exists one when $p = 2$ and so on.

Let us consider the nnd matrix $A: p \times p$ in the partitioned form

$$A = \begin{pmatrix} A_{11} & a \\ a' & \beta \end{pmatrix},$$

where $A_{11} : (p - 1) \times (p - 1)$. Then, since $A$ is nnd, the quadratic form

$$a' A_{11} a + 2c(a'a) + \beta \geq 0$$

for all $a$ and $c$. (1.9)

If $\beta = 0$, then $a'a = 0$ for all $a$ which implies that $a = 0$. If $X_1$ is an rv associated with the nnd matrix $A_{11}$, then $(X_1', 0)'$ is an rv associated with $A$.

If $\beta \neq 0$, then, choosing $c = -(a'a)/\beta$, (1.9) reduces to

$$a'(A_{11} - \beta^{-1}a'a)a \geq 0$$

for all $a$.

i.e., $A_{11} - \beta^{-1}a'a$ is nnd. Let $X_1$ be an rv associated with $(A_{11} - \beta^{-1}a'a) : (p - 1) \times (p - 1)$ and $z$ be a univariate rv with variance $\beta$ and independent of $X_1$. Then it is easy to verify that the rv

$$u = \begin{pmatrix} X_1 + a\beta^{-1}z \\ z \end{pmatrix}$$

has $A$ as its covariance matrix.

Generalized inverse of a matrix: Given a matrix $B: m \times n$, there exists a matrix $G: n \times m$ not necessarily unique such that $BGB = B$. We denote any solution of $G$ by $B^-$ and call it a generalized inverse. If $C$ is any matrix and $\mu(C) \subset \mu(B)$, i.e., the column space of $C$ is contained in the column space of $B$, then

$$BB^-C = C,$$ (1.11)

i.e., $BB^-$ behaves like a unit matrix in such a case (see [18, Section 16.5] and [21, Chapter 8]).

A result on estimation: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are the minimum variance unbiased (MVU) estimators of $\theta_1$ and $\theta_2$, then
Lemma 1.1. Let $A : p \times p$ be an nnd matrix partitioned as

\[
A = \begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{pmatrix}.
\]

Then the following are necessary and sufficient for $A$ to be nnd

(a) $A_{11}$ is nnd,
(b) $A_{22} - A_{21}A_{11}^{-1}A_{12} = A_{21}$ (Schur complement of $A_{11}$) is nnd,
(c) $\mu(A_{12}) \subset \mu(A_{11}).$

The lemma was proved by Albert [1] using purely matrix methods. A statistical proof is given by Dey et al. [5]. However, we sketch the proof for sufficiency as it plays a crucial role in our investigations. (a) implies that there exists an rv $X$ with mean zero and covariance matrix $A_{11}$. (b) implies that there exists an rv $Z$ with mean zero and covariance matrix $A_{21}$. Further, $Z$ can be chosen to be independent of $X$. Define $Y = Z + A_{21}A_{11}^{-1}X$ and consider the random variable $(X', Y')$. Verify that $A$ is the covariance matrix of $(X', Y')$. This implies that $A$ is nnd.

In the following sections, we prove a number of matrix results using the results of the present section. The elegance of proofs suggests that the statistical approach may offer a powerful method for solving complicated matrix problems. For simplicity, we consider matrices with real elements throughout the paper. Extension to matrices with complex numbers can be made in a similar way.

2. Cauchy–Schwarz (CS) and related inequalities

A simple (statistical) proof of the usual CS inequality is to use the fact that the raw second moment of a random variable is nonnegative. Consider a bivariate rv, $(x, y)$ such that

\[
0 < b_{11} = E(x^2), \quad b_{12} = E(xy), \quad b_{22} = E(y^2).
\]

Then

\[
E \left( y - \frac{b_{12}}{b_{11}} x \right)^2 = b_{22} - \frac{b_{12}^2}{b_{11}} \geq 0,
\]

which written in the form

\[
b_{12}^2 \leq b_{11} b_{22},
\]

is the usual CS inequality.
In particular, if $(x, y)$ is discrete bivariate distribution with $(x_i, y_j)$ having probability $p_{ij}$, then
\[(\Sigma p_{ij} x_i y_j)^2 \leq (\Sigma p_{ij} x_i^2)(\Sigma p_{ij} y_j^2)\] (2.2)
since $b_{12} = \Sigma p_{ij} x_i y_j$, $b_{11} = \Sigma p_{ij} x_i^2$ and $b_{22} = \Sigma p_{ij} y_j^2$.

The matrix version is obtained by considering vector variables $X$ and $Y$ such that
\[\Sigma_{11} = C(X, X), \quad \Sigma_{12} = C(X, Y), \quad \Sigma_{22} = C(Y, Y),\]
and noting that
\[C(Y - \Sigma_{21} \Sigma_{11}^{-1} X, Y - \Sigma_{21} \Sigma_{11}^{-1} X) = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \geq 0.\] (2.3)

As a corollary to (2.3), we have
\[AA' \geq AB'(BB')^{-1}BA'\] (2.4)
for matrices $A$ and $B$ having the same number of columns.

If $A$ is an nnd matrix, then there is an rv $X$ whose covariance matrix is $A$. Now consider rv's, $Y_1 = b'X$ and $Y_2 = c'X$, and use the CS inequality $V(Y_1)V(Y_2) \geq [\text{Cov}(Y_1, Y_2)]^2$. We have
\[(c'Ac)(b'Ab) \geq (b'Ac)^2 \quad \text{for all vectors } b \text{ and } c,\] (2.5)
\[(c'A^{-1}c)(b'Ab) \geq (b'c)^2 \quad \text{for all } c \subset \mu(A).\] (2.6)
(Inequality (2.6) follows from (2.5) by writing $c = Ad$ and using the relation $AA^{-1}A = A$.) Expressions (2.1)–(2.6) are the different forms of the CS inequality, all derivable using the fact that the second moment of an rv is nonnegative.

3. A matrix equality

The equality given in the following theorem is useful in the distribution theory of quadratic forms [18, p. 77]. We give a statistical proof of the equality.

**Theorem 3.1.** Let $\Sigma : p \times p$ be a pd matrix and $B : k \times p$ of rank $k$, $C : p - k \times p$ of rank $p - k$ be matrices such that $BC' = 0$. Then
\[C'(C\Sigma C')^{-1}C + \Sigma^{-1}B'(B\Sigma^{-1}B')B\Sigma^{-1} = \Sigma^{-1}.\] (3.1)

**Proof.** Consider the random variable $X \sim N_p(\mu, \Sigma)$ and the transformation
\[Y_1 = CX \sim N_{p-k}(C\mu, C\Sigma C'),\]
\[Y_2 = B\Sigma^{-1}X \sim N_k(B\Sigma^{-1}\mu, B\Sigma^{-1}B').\]
Here $Y_1$ and $Y_2$ are independent, and using results (1.1)–(1.5) on Fisher information we have

\[ \Sigma^{-1} = I_x(\mu) = I_{Y_1}(\mu) + I_{Y_2}(\mu). \]

Note that \( I_{Y_1}(\mu) \) is the first term and \( I_{Y_2}(\mu) \) is the second term on the left-hand side of (3.1). See [18, p. 77] for an algebraic proof. □

4. Schur and Kronecker product of matrices

The following theorems are well known.

**Theorem 4.1.** Let \( A \) and \( B \) be ndd matrices of order \( p \times p \). Then the Schur product \( A \circ B \) is ndd. [\( A \circ B = (a_{ij}b_{ij}) \), where \( A = (a_{ij}) \) and \( B = (b_{ij}) \).]

**Proof.** Let \( X \) be a \( p \)-vector rv such that \( C(X;X) = A \) and \( Y \) be an independent \( p \)-vector rv such that \( C(Y;Y) = B \). Then it is seen that

\[ C(X \circ Y, X \circ Y) = A \circ B \]

so that \( A \circ B \) is the covariance matrix of an rv, which proves the result. □

**Theorem 4.2.** Let \( A : p \times p \) be an ndd matrix and \( B : q \times q \) be an ndd matrix. Then the Kronecker product \( A \otimes B \) is ndd. [\( A \otimes B = (a_{ij}B) \), where \( A = (a_{ij}) \).]

**Proof.** Let \( X \) be a \( p \)-vector rv such that \( C(X, X) = A \) and \( Y \) be an independent \( q \)-vector rv such that \( C(Y, Y) = B \). Then

\[ C(X \otimes Y, X \otimes Y) = A \otimes B \]

so that \( A \otimes B \) is the covariance matrix of an rv, which proves the desired result. □

Note that Theorem 4.1 is a consequence of Theorem 4.2 as \( A \circ B \) is a principal submatrix of \( A \otimes B \).

5. Milne’s inequality

Milne [12] proved the following result.

**Theorem 5.1.** For any constants \( w_i > 0, i = 1, \ldots, n \), such that \( w_1 + \cdots + w_n = 1 \) and given \( p_i, (|p_i| < 1), i = 1, \ldots, n \),

\[ \left( \sum \frac{w_i}{1 - p_i^2} \right)^2 \geq \left( \sum \frac{w_i}{1 - p_i} \right) \left( \sum \frac{w_i}{1 + p_i} \right) \geq \sum \frac{w_i}{1 - p_i^2}. \quad (5.1) \]
Proof. We give a statistical proof to Milne’s inequalities. The left-hand side inequality is easily proved by expanding and comparing the terms of $w_i^2$ and $w_i w_j$.

To prove the right-hand side inequality, consider the bivariate discrete distribution of $(X, Y)$:

$$
\begin{align*}
\text{Probability} & \quad w_1, \ldots, w_n, \\
\text{Value of } X & \quad (1 - \rho_1)^{-1}, \ldots, (1 - \rho_n)^{-1}, \\
\text{value of } Y & \quad (1 + \rho_1)^{-1}, \ldots, (1 + \rho_n)^{-1}.
\end{align*}
$$

We use the well-known formula

$$
2C(X, Y) = E(X_1 - X_2)(Y_1 - Y_2),
$$

where $(X_1, Y_1)$ and $(X_2, Y_2)$ are independent samples from distribution (5.2). A typical term on the right-hand side of (5.3) is

$$
\left( \frac{1}{1 - \rho_i} - \frac{1}{1 - \rho_j} \right) \left( \frac{1}{1 + \rho_i} - \frac{1}{1 + \rho_j} \right) = -\frac{(\rho_i - \rho_j)^2}{(1 - \rho_i^2)(1 - \rho_j^2)} < 0.
$$

Hence $0 \geq C(X, Y) = E(XY) - E(X)E(Y)$. The right-hand side inequality follows by observing that

$$
E(XY) = \sum \frac{w_i}{1 - \rho_i^2}, \quad E(X) = \sum \frac{w_i}{1 - \rho_i}
$$

and

$$
E(Y) = \sum \frac{w_i}{1 + \rho_i}.
$$

Corollary 1. For any r.v $X$ with $\Pr[|X| < A] = 1$,

$$
E \left( \frac{1}{A - X} \right) E \left( \frac{1}{A + X} \right) \geq E \left( \frac{1}{A^2 - X^2} \right).
$$

We use the same arguments as in the main theorem exploiting (5.3), choosing $A - X$ and $A + X$ as the variables $X$ and $Y$.

Corollary 2 (Matrix version of Milne’s inequality). Let $A, V_1, \ldots, V_n$ be symmetric commuting matrices such that $A > 0$, $-A < V_i < A$, $i = 1, \ldots, n$. Then

$$
\left( \sum w_i (A - V_i)^{-1} \right) \left( \sum w_i (A + V_i)^{-1} \right) \geq \sum w_i (A^2 - V_i^2)^{-1},
$$

where $w_i > 0$ and $w_1 + \cdots + w_n = 1$.

In proving Colloary 2, we use the algebraic result that under the conditions on $A$ and $V_i$’s, they can be simultaneously reduced to a diagonal form by pre- and post-multiplying by the same orthogonal matrices. (It would be nice to derive the spectral decomposition of a matrix using statistical arguments.) Kagan [9] gave an alternative proof of (5.1).
6. Convexity of some matrix functions

Let

\[ A_i = \begin{pmatrix} A_{11i} & A_{12i} \\ A_{21i} & A_{22i} \end{pmatrix}, \quad i = 1, \ldots, n, \]

be n by n matrices. Further let

\[ A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \sum_{i=1}^{n} w_i A_i, \quad w_i > 0 \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1, \quad (6.1) \]

\[ B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = A_1 \circ A_2 \circ \cdots \circ A_n = H^n_1 \circ A_1. \quad (6.2) \]

Since \( A \) and \( B \) are n by n matrices, we have

\[ A_{22} - A_{21}A_{11}A_{12} \geq 0, \quad A_{11} - A_{12}A_{22}A_{21} \geq 0, \quad (6.3) \]

\[ B_{22} - B_{21}B_{11}B_{12} \geq 0, \quad B_{11} - B_{12}B_{22}B_{21} \geq 0, \quad (6.4) \]

As a consequence of (6.3) and (6.4), we have the following theorems.

**Theorem 6.1.** Let \( C_i, \quad i = 1, \ldots, n, \) be symmetric matrices of the same order. Then

\[ \sum_{i=1}^{n} w_i C_i^2 \geq \left( \sum_{i=1}^{n} w_i C_i \right)^2, \quad (6.5) \]

\[ \prod_{i=1}^{n} \circ C_i^2 \geq \left( \prod_{i=1}^{n} \circ C_i \right)^2. \quad (6.6) \]

**Proof.** Choose \( A_{11i} = C_i^2, \quad A_{12i} = A_{21i} = C_j \) and \( A_{22i} = I. \) Then \( A_j \) is n by n by the sufficiency part of Lemma 1.1. Hence using (6.3) and (6.4), we have (6.5) and (6.6), respectively. \( \Box \)

**Theorem 6.2.** Let \( C_i, \quad i = 1, \ldots, n, \) be p by p matrices. Then

\[ \sum_{i=1}^{n} w_i C_i^{-1} \geq \left( \sum_{i=1}^{n} w_i C_i \right)^{-1}, \quad (6.7) \]

\[ \prod_{i=1}^{n} \circ C_i^{-1} \geq \left( \prod_{i=1}^{n} \circ C_i \right)^{-1}. \quad (6.8) \]

**Proof.** Choose \( A_{11i} = C_i^{-1}, \quad A_{12i} = A_{21i} = C_j, \quad A_{22i} = I. \) Applying the same argument as in Theorem 6.1, results (6.7) and (6.8) follow from (6.3) and (6.4), respect-
tively. Inequality (6.7) is proved by Olkin and Pratt [16]. See also [11, Chapter 16]. In particular, if \( C_1 \) and \( C_2 > 0 \), then \( (C_1 \circ C_2) \geq (C_1^{-1} \circ C_2^{-1})^{-1} \). □

**Theorem 6.3.** Let \( C_i : s_i \times s_i, \; i = 1, \ldots, n \) and \( B_i : s_i \times m, \; i = 1, \ldots, n, \) be matrices such that \( \Sigma B_iB_i^\prime = I_m \). Then

\[
\sum B_i^\prime C_iB_i \geq \left( \sum B_i^\prime C_iB_i \right)^2.
\]

(6.9)

In particular, when \( C_i \) are symmetric for all \( i \), (6.9) reduces to

\[
\sum B_i^\prime C_i^2B_i \geq \left( \sum B_i^\prime C_iB_i \right)^2.
\]

(6.10)

**Proof.** Since the matrix

\[
\begin{pmatrix}
B_i^\prime C_iC_i^\prime B_i & B_i^\prime C_iB_i \\
B_i^\prime C_iB_i & B_i^\prime B_i
\end{pmatrix}
\]

is nnd, we get (6.9) and (6.10) by applying (6.3). Some of these results are proved in [8,10] using information inequalities (1.4) and (1.5). □

**Theorem 6.4.** Let \( A \) and \( B \) be pd matrices of order \( n \times n \) and \( X \) and \( Y \) be \( m \times n \) matrices. Then

\[
(X \circ Y)'(A \circ B)^{-1}(X \circ B) \leq (X'A^{-1}X) \circ (Y'B^{-1}Y).
\]

(6.11)

**Proof.** The result follows by considering the Schur product of the nnd matrices

\[
\begin{pmatrix}
X'A^{-1}X \\
X
\end{pmatrix}
\text{ and }
\begin{pmatrix}
Y'B^{-1}Y \\
Y
\end{pmatrix}
\]

and taking the Schur complement. □

**Theorem 6.5 [6].** Let \( A \) be a pd matrix. Then

\[
A \circ A^{-1} \geq I.
\]

(6.12)

**Proof.** The Schur product

\[
\begin{pmatrix}
A & I \\
I & A^{-1}
\end{pmatrix} \circ \begin{pmatrix}
A^{-1} & I \\
I & A
\end{pmatrix} = \begin{pmatrix}
A \circ A^{-1} & I \\
I & A^{-1} \circ A
\end{pmatrix}
\]

(6.13)

is nnd, since each matrix on the left-hand side of (6.13) is nnd. Multiplying the right-hand side matrix of (6.13) by the vector \( b' = (a', -a') \) on the left and by \( b \) on the right, where \( a \) is an arbitrary vector, we have

\[
a'(A \circ A^{-1})a - a'a \geq 0 \; \forall a \Rightarrow (A \circ A^{-1}) \geq I.
\]

Purely matrix proofs of some of the results in this section can be found in [2,3]. □
7. Inequalities on harmonic mean and parallel sum of matrices

The following theorem is a matrix version of the inequality on harmonic means given in [20].

**Theorem 7.1.** Let $A_1$ and $A_2$ be random pd matrices of the same order, and $E$ denote expectation. Then

$$E \left[ (A_1^{-1} + A_2^{-1})^{-1} \right] \leq \left[ (E(A_1))^{-1} + (E(A_2))^{-1} \right]^{-1}.$$  \hfill (7.1)

**Proof.** Let $C$ be any pd matrix. Then

$$
\begin{pmatrix}
A_1C^{-1}A_1 & A_1 \\
A_1 & C
\end{pmatrix}
$$  \hfill (7.2)

is nd by the sufficiency part of Lemma 1. Taking expectation of (7.2), and considering the Schur complement, we have

$$E(A_1C^{-1}A_1) \geq E(A_1)(E(C))^{-1}E(A_1).$$  \hfill (7.3)

Now choosing $C = A_1 + A_2$ and noting that

$$(A_1^{-1} + A_2^{-1})^{-1} = A_1 - A_1(A_1 + A_2)^{-1}A_1$$

the result (7.1) follows by an application of (7.3). The same result was proved by Rao [17] using a different method. \Box

**Corollary 3.** Let $A_1, \ldots, A_p$ be random pd matrices and $M_1, \ldots, M_p$ be their expectations. Then

$$E(\tilde{A}) \leq \tilde{M},$$  \hfill (7.4)

where

$$\tilde{A} = (A_1^{-1} + \cdots + A_p^{-1})^{-1}, \quad \tilde{M} = (M_1^{-1} + \cdots + M_p^{-1})^{-1}.$$  

**Proof.** Result (7.4) is proved by repeated applications of (7.1). \Box

**Corollary 4.** Parallel sum inequality. Let $A_{ij}$, $i = 1, \ldots, p$ and $j = 1, \ldots, q$ be pd matrices. Denote

$$\tilde{A}_i = (A_{i1}^{-1} + \cdots + A_{ip}^{-1})^{-1},$$  

$$A_j = A_{j1} + \cdots + A_{jq}, \quad \tilde{A}_j = (A_{j1}^{-1} + \cdots + A_{jq}^{-1})^{-1}.$$

Then

$$\sum \tilde{A}_j \leq \tilde{A}_\cdot,$$  \hfill (7.5)
The result follows from (7.4) by considering \( A_{ij}, \ i = 1, \ldots, p \), as possible values of \( A_1, \ldots, A_p \), giving uniform probabilities to index \( j \), and taking expectation over \( j \).

We can get a weighted version of inequality (7.5) by attaching weights \( w_j \) to index \( j, \ j = 1, \ldots, q \).

**Theorem 7.2.** Let \( A_1, \ldots, A_n \) be \( n \times d \) matrices of the same order \( p \) and \( B : p \times k \) of rank \( k \) and \( \mu(B) \subseteq \mu(A_i) \) for all \( i \). Then

\[
\left( \sum (B'A_i B)^{-1} \right)^{-1} \geq B' \left( \sum A_i \right)^{-1} B
\]

(7.6)

where \( A_i^{-} \) denotes any g-inverse of \( A_i \).

**Proof.** Let \( A_i = X'_i X_i \) and consider independent linear models

\[
Y_i = X_i \theta + \epsilon_i, \quad E(\epsilon_i) = 0, \quad \text{and} \quad C(\epsilon_i, \epsilon_i) = I, \quad i = 1, \ldots, n.
\]

(7.7)

The best estimate of \( B' \theta \) (in the sense of having minimum dispersion error matrix) from the entire model is

\[
B' \left( \sum X'_i X_i \right)^{-1} \sum X'_i Y_i
\]

with the covariance matrix

\[
B' \left( \sum A_i \right)^{-1} B.
\]

(7.8)

The estimate of \( B' \theta \) from the \( i \)th part of the model, \( Y_i = X_i \theta + \epsilon_i \), is

\[
\hat{\theta}_i = B' (X'_i X_i)^{-1} X'_i Y_i
\]

(7.9)

with the covariance matrix

\[
B'A_i^{-} B.
\]

(7.10)

Combining the estimates (7.9) in an optimum way have the estimate of \( B' \theta \)

\[
(\sum (B'A_i^{-} B)^{-1} (\sum (B'A_i^{-} B)^{-1} \hat{\theta}_i
\]

with the covariance matrix

\[
(\sum (B'A_i^{-} B)^{-1})^{-1}.
\]

(7.11)

Obviously (7.11) is not smaller than (7.8) which yields the desired inequality.

The results of Theorems 7.1 and 7.2 are similar to those of Dey et al. [5].

**8. Inequalities on the elements of an inverse matrix**

The following theorem is of interest in estimation theory.
Theorem 8.1. Let
\[ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}^{-1} \end{pmatrix}, \]
where \( \Sigma : (p+q) \times (p+q) \) is a pd matrix and \( \Sigma_{11} \) is of the order \( p \times p \). Then
\[ \Sigma_{11}^{-1} \geq (\Sigma_{11})^{-1}. \] (8.1)

Proof. Consider the rv
\[ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q} \left( \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) , \]
where \( X \) is \( p \)-variate and \( Y \) is \( q \)-variate. Using formulae (1.3) and (1.4) on information matrices, we have
\[ I_{X,Y}(\mu) = \Sigma_{11}^{-1}, \quad I_X(\mu) = \Sigma_{11}^{-1} \] (8.2)
and the information inequality (1.1)
\[ I_{X,Y}(\mu) \geq I_X(\mu) \]
yields the desired result. \( \square \)

9. Carlen’s superadditivity of Fisher information

Let \( f(u) \) be the probability density of a \( p + q \) vector variable \( u' = (u_1, \ldots, u_{p+q}) \) and define
\[ J_f = (j_{rs}) = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}, \]
where
\[ j_{rs} = E \left( \frac{\hat{\partial} \log f}{\hat{\partial} u_r} \cdot \frac{\hat{\partial} \log f}{\hat{\partial} u_s} \right) \]
and \( f_{11} \) and \( f_{22} \) are matrices of orders \( p \) and \( q \), respectively. Let \( g(u_1, \ldots, u_p) \) and \( h(u_{p+1}, \ldots, u_{p+q}) \) be the marginal probability densities of \( u_1, \ldots, u_p \) and \( u_{p+1}, \ldots, u_{p+q} \), respectively with the corresponding \( J \) matrices, \( J_g \) of order \( p \times p \) and \( J_h \) of order \( q \times q \). Then Carlen [4] proved that
\[ \text{Trace} \ J_f \geq \text{Trace} \ J_g + \text{Trace} \ J_h. \] (9.1)

We prove result (9.1), using Fisher information inequalities. Let us introduce a \((p + q)\)-vector parameter
\[ \theta' = (\theta_1', \theta_2') = (\theta_1, \ldots, \theta_p, \theta_{p+1}, \ldots, \theta_{p+q}), \]
where \( \theta_1 \) is a \( p \)-vector and \( \theta_2 \) is a \( q \)-vector, and consider the probability densities
\[ f(u_1 + \theta_1, \ldots, u_{p+q} + \theta_{p+q}), \]
\[ g(u_1 + \theta_1, \ldots, u_p + \theta_p) \quad \text{and} \quad h(u_{p+1} + \theta_{p+1}, \ldots, u_{p+q} + \theta_{p+q}). \]
It is seen that in terms of information

\[ I_f(\theta_1) = f_{11} \geq I_g(\theta_1) = J_g, \]
\[ I_f(\theta_2) = f_{22} \geq I_h(\theta_2) = J_h, \]

which shows that

\[ f_{11} \geq J_g, \quad f_{22} \geq J_h. \]

Taking traces, we get the desired result (9.1). A similar proof is given by Kagan and Landsman [7].

10. Inequalities on principal submatrices of an nnd matrix

A \( k \times k \) principal submatrix of an nnd matrix \( A : n \times n \) is a matrix \([A] : k \times k \) obtained by retaining columns and rows indexed by the sequence \((i_1, \ldots, i_k)\), where \( 1 \leq i_1 < i_2 \cdots < i_k \leq n \). It is clear that if \( A \succ 0 \), then \([A] \succeq 0\).

**Theorem 10.1.** If \( A \) is nnd matrix, then

(i) \([A^2] \succeq [A]^2\), and if \( A \) is pd,

(ii) \([A^{-1}] \succeq [A]^{-1}\).

(iii) \([X'[A]^{-1}[X] \succeq [X'A^{-1}X]\), where \( X \) has the same number of rows as \( A \).

**Proof.** All the results are derived by observing that

\[
\begin{pmatrix}
A & X \\
X' & X'A^{-1}X
\end{pmatrix} \succeq 0,
\]  

(10.1)

which implies

\[
\begin{pmatrix}
[A] & [X] \\
[X'] & [X'A^{-1}X]
\end{pmatrix} \succeq 0
\]

(10.2)

and taking Schur complement for special choices of \( X \). For (i), choose \( A^2 \) for \( A \) and \( X = A \). For (ii), choose \( X = I \). (iii) Follows directly from (10.2). Different proofs and additional results are given by Zhang [22].

**Acknowledgement**

The research work of the paper is supported by a grant from the ARO, DAAH04-96-1-0082. The author would like to thank the referees for the valuable suggestions they have made.
References