MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A
ON THE MATRIX CONVEXITY OF THE MOORE-PENROSE INVERSE AND SOME APPLICATIONS

Demetrios G. Kaffes
M. Bhaskara Rao
Thomas Mathew
K. Subramanyam

Center for Multivariate Analysis
University of Pittsburgh
ON THE MATRIX CONVEXITY OF THE MOORE-PENROSE INVERSE AND SOME APPLICATIONS

Demetrios G. Kaffes
M. Bhaskara Rao*
Thomas Mathew*
K. Subramanyam*

July 1985

Technical Report No. 85-28

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*Research sponsored by the Air Force Office of Scientific Research (AFSC) under contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
ON THE MATRIX CONVEXITY OF THE MOORE–PENROSE INVERSE AND SOME APPLICATIONS

Demetrios G. Kaffes
University of Sheffield
Sheffield S3 7RH, UK

Thomas Mathew
University of Pittsburgh
Pittsburgh, PA 15260, USA

M. Bhaskara Rao
University of Sheffield
Sheffield S3 7RH, UK

K. Subramanyam
University of Pittsburgh
Pittsburgh, PA 15260, USA

Abstract

It is well known that if \( A \) and \( B \) are two positive definite matrices of the same order and \( 0 < \lambda < 1 \), then

\[
[\lambda A + (1-\lambda)B]^{-1} \leq \lambda A^{-1} + (1-\lambda)B^{-1}.
\]

It is easy to construct an example consisting of two positive semi-definite matrices for which the above inequality is not true when one replaces the inverse operation by Moore–Penrose inverse operation. In this paper, we give necessary and sufficient conditions for the validity of the inequality for every \( 0 < \lambda < 1 \). As an application, we give a sufficient condition under which the inequality \( (EA)^+ \leq E(A^+) \) is valid, where \( A \) is a square matrix of random variables which is almost surely positive semi-definite, generalizing the well-known result \( (EA)^{-1} \leq EA^{-1} \) when \( A \) is almost surely positive definite.

AMS subject classification: Primary 15A09, 15A45 Secondary 62H99

Key Words and Phrases: g-inverse, Moore–Penrose inverse, optimal designs, random matrices

1. Introduction:

Fedorov (1972, Theorem 1.1.12, p. 19) states that if $A$ and $B$ are positive definite matrices and $0 \leq \lambda \leq 1$ is any real number, then

$$[\lambda A + (1-\lambda)B]^{-1} \leq \lambda A^{-1} + (1-\lambda)B^{-1}. \quad (1)$$

(We say that, for two square matrices $C$ and $D$ of the same order, $C \leq D$ if $D - C$ is positive semi-definite.) For a proof of this inequality, see Moore (1973, p. 408) or Marshall and Olkin (1979, pp. 469-471).

The above inequality is useful in optimal designs and, especially, in linear optimal designs. This inequality is used in Lemma 2.9.1 of Fedorov (1972, p. 123).

Let $\mathcal{D}$ be the collection of all square matrices of order $q$ and $L$ a real linear functional on $\mathcal{D}$, i.e.,

$$L(A+B) = L(A) + L(B) \quad \text{for every } A, B \in \mathcal{D}, \quad (2)$$

and

$$L(cA) = cL(A) \quad \text{for every } c \text{ real and } A \in \mathcal{D}. \quad (3)$$

Assume, further, that

$$L(A) \geq 0 \text{ if } A \text{ is positive semi-definite.} \quad (4)$$

Consider an optimal design problem involving $q$ parameters. Let $\mathcal{M}$ be the collection of all information matrices and $\mathcal{M}_1$ the collection of all non-singular information matrices. Let $L$ be a linear functional satisfying the above three conditions. Lemma 2.9.1 of Fedorov [1972, p. 123] says that the function

$$L_1: \mathcal{M}_1 \to \mathbb{R}$$

defined by $L_1(M) = L(M^{-1})$, $M \in \mathcal{M}_1$ is a convex function on $\mathcal{M}_1$, i.e.
L_1[λM_1 + (1-λ)M_2] ≤ λL_1(M_1) + (1-λ)L_1(M_2),

for any \( M_1, M_2 \in \mathcal{M} \) and \( 0 < \lambda < 1 \). This is a simple consequence of the inequality (1) and the conditions (2), (3) and (4).

In Remark 1 on page 124, Fedorov comments that if \( M \in \mathcal{M} \) is singular, one can consider Moore-Penrose inverse \( M^+ \) of \( M \) and define \( L_1(M) = L(M^+) \). See also Remark 1 to Theorem 2.7.1. In other words, if we define \( L_1: \mathcal{M} \to \mathbb{R} \) by

\[
L_1(M) = L(M^+), \quad M^+ \text{ being the Moore-Penrose of } M, \ M \in \mathcal{M},
\]

his remarks seem to mean that \( L_1 \) is a convex function on \( \mathcal{M} \). We show that this is not true in general.

In this connection, we ask the following question. Let \( A \) and \( B \) be two positive semi-definite matrices of the same order and \( 0 < \lambda < 1 \) be any real number. Is the inequality

\[
[\lambda A + (1-\lambda)B]^+ \leq \lambda A^+ + (1-\lambda)B^+,
\]  

(5)

analogous to (1), true?

The organization of this paper is as follows. In Section 2, we give a necessary and sufficient condition for (5) to be valid for every \( 0 < \lambda < 1 \). In Section 3, we study this inequality in the context of a collection of positive semi-definite matrices indexed by a probability space. In particular, we examine under what conditions \( (EA)^+ \leq EA^+ \) when \( A \) is a symmetric matrix of random variables such that \( A \) is almost surely positive semi-definite.

For any matrix \( A \), range of \( A \) is defined to be the linear space spanned by the columns of \( A \) and it is denoted by \( R(A) \). \( A^- \) denotes an arbitrary g-inverse of \( A \), i.e. a matrix satisfying \( AA^-A = A \). For basic ideas concerning Moore-Penrose inverse, see Rao and Mitra (1971, pp 50-53).
2. Convexity of the Moore-Penrose Inverse:

The following result gives conditions under which (5) holds for every $0 \leq \lambda \leq 1$.

**Theorem 1.** Let $A$ and $B$ be two real positive semi-definite matrices of the same order. Then the following are equivalent:

(i) $R(A) = R(B)$.

(ii) There exist positive semi-definite $g$-inverses $A^-$ and $B^-$ of $A$ and $B$ respectively such that $[\lambda A^{-} + (1-\lambda)B^{-}] < \lambda A^{-} + (1-\lambda)B^{-}$ for some positive semi-definite $g$-inverse $[\lambda A^{-} + (1-\lambda)B^{-}]$ of $\lambda A^{-} + (1-\lambda)B^{-}$ and for every $0 < \lambda < 1$.

(iii) $[\lambda A^{-} + (1-\lambda)B^{-}] < [\lambda A^{-} + (1-\lambda)B^{-}]$ for every $0 < \lambda < 1$.

**Proof:** The proof of (i) $\Rightarrow$ (iii) is similar to the proof of the remark in Giovagnoli and Wynn (1985, p. 129). Let $P$ be an orthogonal matrix such that $A = P \text{diag}(A_1,0)P^T$, where $A_1$ is a diagonal positive definite matrix. When $R(A) = R(B)$, we then have $B = P \text{diag}(B_1,0)P^T$ where $B_1$ is positive definite. To show that (iii) holds, we have to show that $[\lambda A_1^{-} + (1-\lambda)B_1^{-}] < \lambda A_1^{-} + (1-\lambda)B_1^{-}$, which is true since $A_1$ and $B_1$ are positive definite. (iii) $\Rightarrow$ (ii) is obvious. We shall now prove (ii) $\Rightarrow$ (i).

Suppose $[\lambda A^{-} + (1-\lambda)B^{-}] < [\lambda A^{-} + (1-\lambda)B^{-}]$ for positive semi-definite $g$-inverses $A^{-}$ and $B^{-}$ (independent of $\lambda$) and a positive semi-definite $g$-inverse $[\lambda A^{-} + (1-\lambda)B^{-}]$ for every $\lambda$ as specified in the theorem. Premultiply and postmultiply the above by $\lambda A^{-} + (1-\lambda)B^{-}$ yielding $\lambda A^{-} + (1-\lambda)B^{-} < [\lambda A^{-} + (1-\lambda)B^{-}] [\lambda A^{-} + (1-\lambda)B^{-}] [\lambda A^{-} + (1-\lambda)B^{-}]$.

If $R(A) \neq R(B)$, assume without loss of generality that $R(A)$ is not contained in $R(B)$, in which case there exists a vector $b$ satisfying $Ab \neq 0$ and $Bb = 0$. Premultiplying and postmultiplying the above inequality by $b^T$ and $b$ respectively lead to $\lambda b^T Ab \leq \lambda^2 b^T Ab + \lambda^2 (1-\lambda) b^T Ab - \lambda^2 b^T Ab$ or equivalently $b^T Ab \leq \lambda (b^T Ab - b^T Ab)$, when $0 < \lambda < 1$.

But, since $b^T Ab > 0$, the above inequality cannot hold for all $0 < \lambda < 1$. This completes the proof.

**Remark 1.** From Theorem 1 it follows that when $A$ is positive semi-definite the function $A^+ A^+$ is matrix convex iff $A$ varies over a set of positive semi-definite matrices
with the same range. The 'if' part of this assertion is proved in Giovagnoli and Wynn (1985).

**Remark 2.** For a given \( \lambda \) with \( 0 < \lambda < 1 \), one can always find positive semi-definite g-inverses \([\lambda A + (1-\lambda)B]^{-}\), \(A^{-}\) and \(B^{-}\) (depending on \( \lambda \)) and satisfying \([\lambda A + (1-\lambda)B]^{-} \leq \lambda A^{-} + (1-\lambda)B^{-}\) even though \( R(A) \) and \( R(B) \) are different. This could be seen as follows. Let \( P \) be a nonsingular matrix satisfying \( A = P \text{diag}(\Delta_1, \Delta_2, Q)^P \) and \( B = P \text{diag}(D_1, 0, D_2, 0)^P \) where \( \Delta_1, \Delta_2, D_1, D_2 \) are diagonal positive definite matrices.

The existence of such a \( P \) is guaranteed by Theorem 6.2.3 in Rao and Mitra (1971).

For a given \( 0 < \lambda < 1 \), consider the g-inverses \([\lambda A + (1-\lambda)B]^{-} - (PT)^{-1} \text{diag}((\lambda A + (1-\lambda)B)^{-1})^{-1} \) where \( M \) and \( N \) are positive semi-definite matrices satisfying \((\lambda A + (1-\lambda)B)^{-1} \leq M + \lambda A^{-} + (1-\lambda)B^{-}\) and \((\lambda A + (1-\lambda)B)^{-1} \leq N + \lambda A^{-} + (1-\lambda)B^{-}\). With such a choice of \( M \) and \( N \), it can be verified that \([\lambda A + (1-\lambda)B]^{-} \leq \lambda A^{-} + (1-\lambda)B^{-}\).

**Corollary 1.** Let \( A_1, \ldots, A_k \) be \( k \) real positive semi-definite matrices. Then
\[
(\lambda_1 A_1 + \cdots + \lambda_k A_k)^{+} \leq \lambda_1 A_1^{+} + \cdots + \lambda_k A_k^{+}
\]
for every \( \lambda_i \) satisfying \( 0 < \lambda_i < 1 \) \((i = 1, 2, \ldots, k)\) and \( \prod_{i=1}^{k} \lambda_i = 1 \) iff \( R(A_i) = R(A_j) \) for \( i, j = 1, 2, \ldots, k \).

**Proof.** We shall prove the result for \( k = 3 \). The proof in the general case follows along similar lines, by induction. Assume \( \lambda_1 < 1 \), \( R(A_1) = R(A_2) = R(A_3) \). Then
\[
(\lambda_1 A_1 + \lambda_2 A_2 + \lambda_3 A_3)^{+} = (\lambda_1 A_1^{+} (1-\lambda_1) (1-\lambda_2) A_2^{+} + \lambda_3 A_3^{+})^{+} \leq \lambda_1 A_1^{+} + (1-\lambda_1) (1-\lambda_2) A_2^{+} + \lambda_3 A_3^{+}
\]
(applying Theorem 1). Since \( \frac{\lambda_2}{1-\lambda_1} + \frac{\lambda_3}{1-\lambda_1} = 1 \), applying Theorem 1 again, we get
\[
(\lambda_1 A_1 + \lambda_2 A_2 + \lambda_3 A_3)^{+} \leq \lambda_1 A_1^{+} + (1-\lambda_1) \left[ \frac{\lambda_2}{1-\lambda_1} A_2^{+} + \frac{\lambda_3}{1-\lambda_1} A_3^{+} \right] = \lambda_1 A_1^{+} + \lambda_2 A_2^{+} + \lambda_3 A_3^{+},
\]
which concludes the proof of the 'if' part. To prove the 'only if' part, choose \( \lambda_3 = 0 \). Then from Theorem 1, we get \( R(A_1) = R(A_2) \). Similarly \( R(A_i) = R(A_j) \) for all \( i, j \).

**Corollary 2.** Let \( A_n, n \geq 1 \) be a sequence of positive semi-definite matrices and \( \lambda_n, n \geq 1 \) be a sequence of nonnegative real numbers such that \((i) \sum \lambda_i = 1 \) \((ii) \sum \lambda_i A_1 \) converges \((iii) \sum \lambda_i A_1^{+} \) converges. Then \( \sum \lambda_i A_1^{+} \leq \sum \lambda_i A_1 \) for every such sequence \( \lambda_n \), \( n \geq 1 \) iff \( R(A_n) \) is the same for all \( n \geq 1 \).
Proof. The 'only if' part is proved as in Corollary 1. To prove the 'if' part, assume that \( R(A_i) \) is the same for all \( i \). Let \( B_n = \sum_{i=1}^{n} \lambda_i A_i \) and \( B = \sum_{i \geq 1} \lambda_i A_i \). Let \( C \) be a matrix having the same range as the \( A_i \)'s and let \( \delta = 1 - \sum_{i=1}^{n} \lambda_i \). If we assume that at least one \( \lambda_i \) is positive (1\( \leq i \leq n \)) in \( B_n \), then \( R(B_n + \delta C) = R(B) \). Since \( B_n + \delta C \to B \) as \( n \to \infty \), following the argument given in Stewart (1969, p. 34) (see also Campbell and Meyer, 1979, Chapter 10), we see that \( (B_n + \delta C)^+ \to B^+ \) as \( n \to \infty \). Applying Corollary 1, we get \( (B_n + \delta C)^+ \leq \lambda_1 A_1^+ + \cdots + \lambda_n A_n^+ + \delta C^+ \). The result now follows by taking limits as \( n \to \infty \).

Remark 3. The results in this section proved for real symmetric positive semi-definite matrices are also valid for complex hermitian positive semi-definite matrices, with obvious modifications in the proofs.

3. Some Extensions and Applications:

In this section, we consider the problem of extending the inequality specialized in Section 2 for a collection of positive semi-definite matrices indexed by a probability space.

Let \((Y, \mathcal{B}, u)\) be a probability space and \( A_y, y \in Y \) a collection of positive semi-definite matrices of the same order. Let \( A_y = ((a_{ijy})) \), \( 1 \leq i, j \leq n \) and \( y \in Y \). Assume that \( a_{ijy} \) as a function of \( y \) is measurable for every \( 1 \leq i, j \leq n \). There are three basic questions one can ask in this connection.

(a) Let \( A_y^+ = (b_{ijy}) \), \( 1 \leq i, j \leq n \), \( y \in Y \). Is \( b_{ijy} \) as a function of \( y \) measurable for every \( 1 \leq i, j \leq n \)?

(b) If the answer to (a) is affirmative and each \( a_{ijy} \) is integrable with respect to the measure \( u \), is each \( b_{ijy} \) integrable with respect to \( u \)?

(c) If the answers to (a) and (b) are affirmative, is the inequality

\[
\left( \int_y A_y u(dy) \right)^+ \leq \int_y A_y^+ u(dy)
\]

true?
We, first, tackle (a). We give two sets of sufficient conditions under which \( A_y^+ \) as a function of \( y \) is measurable.

**Theorem 2.**

(a) Suppose \( R(A_y) \) is the same for all \( y \epsilon D \epsilon B \) with \( m(D) = 1 \). Then \( A_y^+ \) as a function of \( y \) is measurable.

(b) Suppose \( Y \) is a topological space and \( B \) is some \( \sigma \)-field on \( Y \) containing all open subsets of \( Y \). Suppose \( R(A_y) \) is the same for all \( y \) in \( Y \).

If \( A_y \) as a function of \( y \) is continuous, then \( A_y^+ \) as a function of \( y \) is continuous.

**Proof:** Let \( A \) be any symmetric matrix with \( R(A_y) = R(A) \) for every \( y \) in \( D \). There exists an orthogonal matrix \( P \) such that \( P A P^T = \text{diag}(A_\times, 0) \), where \( A_\times \) is a diagonal matrix with diagonal entries being the non-zero eigen values of \( A \). Since Range \( (A_y) = \text{Range} (A) \), \( y \epsilon D \), \( P A_y P^T = \text{diag} (A_\times, 0) \) for some nonsingular matrix \( A_\times \) which is of same order as \( A_\times \). Note that \( A_y^+ = P^T \text{diag} ((A_\times)^{-1}, 0) P \). If \( A_y \) as a function of \( y \) is measurable (continuous) so is \( A_\times \) as a function of \( y \). Consequently, \( (A_\times)^{-1} \) as a function of \( y \) is measurable (continuous). Hence \( A_y^+ \) as a function of \( y \) is measurable (continuous).

**Theorem 3.**

(a) Suppose there exists a set \( D \epsilon B \) such that \( m(D) = 1 \) and \( A_{y_1} A_{y_2} = A_{y_2} A_{y_1} \) for every \( y_1, y_2 \epsilon D \). Then \( A_y^+ \) as a function of \( y \) is measurable.

(b) Suppose \( Y \) is a topological space and \( B \) is a \( \sigma \)-field on \( Y \) containing all open subsets of \( Y \). Suppose \( A_{y_1} A_{y_2} = A_{y_2} A_{y_1} \) for all \( y_1, y_2 \epsilon Y \). If \( A_y \) as a function of \( y \) is continuous, then \( A_y^+ \) as a function of \( y \) is continuous.

We need the following lemma in the proof of the above theorem.

**Lemma 1.** Let \( \{ A_y : y \epsilon Y \} \) be a family of pairwise commuting symmetric matrices of order \( n \times n \). Then there exists an orthogonal matrix \( C \) such that
\[ C \mathcal{A}_y C = \text{diag}(\lambda_{1y}, \lambda_{2y}, \ldots, \lambda_{ny}) , \]

where \( \lambda_{1y}, \lambda_{2y}, \ldots, \lambda_{ny} \) are the eigenvalues of \( \mathcal{A}_y \).

**Proof:** This result is well known when \( Y \) is finite. See, for example, Rao (1973, Exercise 15, p. 72). Let \( \Omega \) be the least ordinal number corresponding to the cardinal number of \( Y \). \( \Omega \) is obviously a limit ordinal. See Kamke (1950).

Let us identify \( Y \) with \([0, \Omega)\). In other words, the given family of matrices can be written down as a generalized sequence \( \mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_\alpha, \ldots, \alpha < \Omega \).

We claim that there exists a vector \( z^{(1)} \) of unit length such that it is an eigenvector for every \( \mathcal{A}_\alpha, \alpha < \Omega \). For this, we proceed as follows.

Let \( \lambda_{10} \) be any eigenvalue of \( \mathcal{A}_0 \). Let \( 0 < \alpha < \Omega \) be any ordinal number. Then, there exists a vector \( x^{(\alpha)} \) of unit length and real numbers \( \lambda_{18}, \beta < \alpha \), satisfying the following properties.

(i) \( \lambda_{18} \) is an eigenvalue of \( \mathcal{A}_\beta \) for every \( 0 \leq \beta < \alpha \).

(ii) \( \mathcal{A}_\beta x^{(\alpha)} = \lambda_{18} x^{(\alpha)} \) for every \( 0 \leq \beta < \alpha \).

(iii) \( \lambda_{18}, 0 \leq \beta < \alpha \) does not depend on \( \alpha \).

This \( x^{(\alpha)} \) is obtained by transfinite induction as follows.

There exists a vector \( x^{(1)} \) of unit length such that \( \mathcal{A}_0 x^{(1)} = \lambda_{10} x^{(1)} \). Note that \( \mathcal{A}_1 x^{(1)}, \mathcal{A}_2 x^{(1)}, \ldots \) are eigenvectors of \( \mathcal{A}_0 \) corresponding to the same eigenvalue \( \lambda_{10} \). Consequently, every vector in the linear manifold spanned by \( \{ x^{(1)}, \mathcal{A}_1 x^{(1)}, \mathcal{A}_2 x^{(1)}, \ldots \} \) is an eigenvector of \( \mathcal{A}_0 \) corresponding to the eigenvalue \( \lambda_{10} \). This linear manifold contains an eigenvector of \( \mathcal{A}_1 \). See Rao (1973, p. 39). Let us assume that this eigenvector \( x^{(2)} \), say, is of unit length and the corresponding eigenvalue for \( \mathcal{A}_1 \) be \( \lambda_{11} \). Thus at the second stage, we have

\[ \mathcal{A}_0 x^{(2)} = \lambda_{10} x^{(2)} \]
\[ \mathcal{A}_1 x^{(2)} = \lambda_{11} x^{(2)} . \]
Now, since $A_0$ and $A_2$ commute and also $A_1$ and $A_2$ commute, every vector in the linear manifold spanned by \(\{x^{(2)}, A_2x^{(2)}, A_2^2x^{(2)}, \ldots\}\) is an eigenvector of $A_0$ corresponding to the same eigenvalue $\lambda_{10}$ and also is an eigenvector of $A_1$ corresponding to the same eigenvalue $\lambda_{11}$. This manifold contains an eigenvector $x^{(3)}$ of $A_2$. Assume $x^{(3)}$ to be of unit length and $\lambda_{12}$ to be the corresponding eigenvalue of $A_2$. Thus we have

\[
A_0x^{(3)} = \lambda_{10}x^{(3)},
A_1x^{(3)} = \lambda_{11}x^{(3)},
A_2x^{(3)} = \lambda_{12}x^{(3)}.\]

Continuing this procedure for every $n < w$, where $w$ is the first infinite ordinal number, we find a sequence $x^{(n)}$, $1 \leq n < w$, of vectors of unit length and a sequence $\lambda_{1k}$, $0 \leq k < w$ of real numbers satisfying the following property:

\[
A_kx^{(n)} = \lambda_{1k}x^{(n)}, \quad 0 \leq k < n.
\]

It is important to note that once an eigenvalue enters into the system, it remains in the system at every stage of the induction process.

Since each $x^{(n)}$, $1 \leq n < w$ is of unit length, by compactness argument, this sequence admits a convergent subsequence converging to $x^{(w)}$, say. Obviously, this vector is of unit length. Further,

\[
A_kx^{(w)} = \lambda_{1k}x^{(w)} \text{ for } 0 \leq k < w.
\]

Now, every vector in the linear manifold spanned by \(\{x^{(w)}, A_wx^{(w)}, A_w^2x^{(w)}, \ldots\}\) is an eigenvector of $A_k$, $0 \leq k < w$, corresponding to the eigenvalue $\lambda_{1k}$, $0 \leq k < w$. But this manifold contains an eigenvector $x^{(w+1)}$ of $A_w$. Let us assume this vector to be of unit length and let the corresponding eigenvalue of $A_w$ be $\lambda_{1w}$. Thus we have
\[ A_k^{x(w+1)} = \chi_k x^{(w+1)}, \quad 0 < k < w+1. \]

This process is continued arguing separately for the case of limit ordinals and the case of non-limit ordinals.

Now, by compactness argument, \( x'(a), \alpha < \omega \) admits a subnet converging to a vector \( z(1) \) of unit length. This vector is the desired one.

Now, we claim that there exists a vector \( z(2) \) of unit length such that \( z(2) \bot z(1) \) and \( z(2) \) is a common eigenvector for each \( A_\alpha, 0 < \alpha < \omega \). Let \( \lambda_2 \) be an eigenvalue of \( A_0 \) admitting an eigenvector \( y^{(1)} \) such that \( y^{(1)} \) is of unit length and \( y^{(1)} \bot z(1) \).

Let \( 0 < \alpha < \omega \). We claim that there exists a vector \( y^{(\alpha)} \) of unit length and real numbers \( \lambda_2 \), \( 0 \leq \beta < \alpha \), satisfying the following properties.

(i) \( \lambda_2 \) is an eigenvalue of \( A_\beta \).

(ii) \( A_\beta y^{(\alpha)} = \lambda_2 y^{(\alpha)} \) for every \( 0 \leq \beta < \alpha \).

(iii) \( y^{(\alpha)} \bot z(1) \).

(iv) \( \lambda_2 \), \( 0 \leq \beta < \alpha \) is independent of \( \alpha \).

The \( y^{(\alpha)} \)'s and \( \lambda_2 \)'s are obtained by transfinite induction as follows. At the first step, for \( \alpha = 1 \), we have \( y^{(1)} \) and \( \lambda_2 \) satisfying (i) through (iv). Let \( \alpha = 2 \). The linear manifold spanned by \( \{y^{(1)}, A_1 y^{(1)}, A_1^2 y^{(1)}, \ldots\} \) contains an eigenvector \( y^{(2)} \) for \( A_1 \) with the corresponding eigenvalue, say, \( \lambda_{21} \). Since \( A_0 \) and \( A_1 \) commute, every vector in this manifold is an eigenvector of \( A_0 \) corresponding to the eigenvalue \( \lambda_2 \).

Without loss of generality we can assume \( y^{(2)} \) to be of unit length. Further, \( y^{(2)} \bot z(1) \). To prove this, consider \( A_1^n y^{(1)} \). We have

\[
(A_1^n y^{(1)})^T z(1) = y^{(1)} T A_1^n z(1) = y^{(1)} T (\lambda_{11})^n z(1) = (\lambda_{11})^n y^{(1)} T z(1) = 0
\]

Consequently, every vector in the linear manifold spanned by \( \{y^{(1)}, A_1 y^{(1)}, A_1^2 y^{(1)}, \ldots\} \) is orthogonal to \( z(1) \). Hence \( y^{(2)} \bot z(1) \). Thus, we have a vector \( y^{(2)} \) of unit length.
satisfying
\[ A_\beta y^{(2)} = \lambda_{2\beta} y^{(2)}, \quad 0 < \beta < \alpha \]
and
\[ y^{(2)} \perp z^{(1)}. \]

This process is continued as in the first part of this proof noting that once an eigenvalue \( \lambda_{2\beta} \) enters the system it remains in the system. By compactness of the unit ball of \( \mathbb{R}^n \), we can find a subnet of \( y^{(2)}, 0 < q < \Omega \) converging to a vector, say, \( z^{(2)} \). This \( z^{(2)} \) is the desired vector.

Thus, we can obtain \( n \) vectors \( z^{(1)}, z^{(2)}, \ldots, z^{(n)} \) satisfying the following properties

(a) \( \| z^{(i)} \| = 1, \ i = 1 \to n. \)

(b) \( z^{(i)} \perp z^{(j)}, \ i \neq j. \)

(c) \( A_\alpha z^{(i)} = \lambda_{ia} z^{(i)}, \ 0 < \alpha < \Omega, \ i = 1 \to n. \)

Define \( C = (z^{(1)}, z^{(2)}, \ldots, z^{(n)}) \). C is the required orthogonal matrix.

**Proof of Theorem 3:** By Lemma 1, there exists an orthogonal matrix \( C = (z^{(1)}, z^{(2)}, \ldots, z^{(n)}) \) such that

\[ C^T A_y C = \text{Diag}(\lambda_{1y}, \lambda_{2y}, \ldots, \lambda_{ny}), \ y \in Y, \]

where \( \lambda_{1y}, \lambda_{2y}, \ldots, \lambda_{ny} \) are the eigenvalues of \( A_y \). Let \( f_i : Y \to \mathbb{R} \) be defined by

\[ f_i(y) = \lambda_{iy}, \ y \in Y, \ i = 1 \to n. \]

It is easy to check that each \( f_i \) is measurable. For, \( f_i(y) = z^{(1)} y^T A_y z^{(1)}, \) a linear combination of the elements of \( A_y \). Let \( g_i : Y \to \mathbb{R} \) be defined by

\[ g_i(y) = \begin{cases} \frac{1}{f_i(y)} & \text{if } f_i(y) \neq 0 \\ 0 & \text{if } f_i(y) = 0, \ y \in Y, \ i = 1 \to n. \end{cases} \]
\(g_1, g_2, \ldots, g_n\) are, obviously, measurable functions. Now,

\[
A_y = C \text{Diag}(f_1(y), f_2(y), \ldots, f_n(y))C^T.
\]

Then \(A_y^+ = C \text{Diag}(g_1(y), g_2(y), \ldots, g_n(y))C^T\) (see Rao and Mitra 1971, p. 69).

Consequently, the elements of \(A_y^+\) as functions on \(Y\) are measurable.

Now, we come to the question raised in (b). \(A_y^+\) as a function of \(y\) need not be integrable. The following is a simple example. Let \(Y = (0,1)\), \(B = \text{Borel}\ \sigma\)-field on \(Y\), \(\mu = \text{Lebesgue measure on } B\), and \(A_y = (y), y \in Y\), is of order \(1 \times 1\). \(A_y\) as a function of \(y\) is integrable with respect to \(\mu\) but \(A_y^+\) is not.

The following result generalizes the inequality expounded in Section 2 and answers the query raised in (c).

**Theorem 5** Let \(R(A_y)\) be the same for all \(y \in D \in B\) with \(\mu(D) = 1\). Suppose \(A_y\) and \(A_y^+\) as functions of \(y\) are integrable with respect to \(\mu\). Then

\[
\left[ \int_Y A_y \, \mu(dy) \right]^+ \leq \int_Y A_y^+ \, \mu(dy).
\]

**Proof:** Let \(D\) be the collection of all positive semi-definite matrices of the same order \(n \times n\) as that of \(A_y\) and range the same as that of \(A_y\), \(y \in D\). Then \(D\) is a closed convex subset of an appropriate finite-dimensional Euclidean space and the map \(y \mapsto A_y\) from \(Y\) to \(D\) is measurable. By Theorem 1, the map \(A \mapsto A^+\) from \(D\) to \(D\) is convex. Let \(C \in \mathbb{R}^n\) be an arbitrary but fixed vector. Then the map \(f:D \to \mathbb{R}\) defined by \(f(A) = C^TA^+C\) is convex. By Jensen's inequality (see Ferguson (1967, p. 76)),

\[
f(\mathbb{E}A_{(\cdot)}) \leq \mathbb{E}f(A_{(\cdot)}), \text{ i.e.,}
\]

\[
C^T \left( \int_Y A_y \, \mu(dy) \right)^+ C \leq \int_Y C^T A_y^+ C \, \mu(dy)
\]

\[
= C^T \left( \int_Y A_y^+ \, \mu(dy) \right) C.
\]
This implies that, as $C$ is arbitrary,

$$\left( \int_{Y} A_{y} \mu(dy) \right)^{+} \leq \int_{Y} A_{y}^{+} \mu(dy).$$

This completes the proof.

The condition on the range in the above theorem, in a certain sense, is necessary for the inequality to be valid. If the above inequality is valid for all probability measures for which the concerned integrals are finite, then the above condition on the range is necessary.

The above result can be couched in the language of random matrices as follows.

**Corollary 3.** Let $A$ be a symmetric matrix of random variables such that $A$ is positive semi-definite almost surely and $R(A)$ is the same almost surely. Assume that $EA$ and $EA^{+}$ exist. Then

$$(EA)^{+} \leq EA^{+}.$$  

The above inequality is an analogue of the usual Harmonic-Arithmetic inequality, namely, if $f$ is an almost surely positive random variables with $Ef$ and $Ef^{-1}$ finite then $(Ef)^{-1} \leq Ef^{-1}$.

We also obtain as a corollary the following result due to Groves and Rothenberg (1969, p. 690). See also Srivastava (1970, p. 236).

**Corollary 4.** Let $A$ be a symmetric matrix of random variables such that $A$ is positive definite almost surely, and $EA$ and $EA^{-1}$ exist. Then

$$(EA)^{-1} \leq EA^{-1}.$$  

**Corollary 5.** Let $Y_{1}, Y_{2}, \ldots, Y_{N}$ be a random sample of size $N$ form a multivariate normal distribution with a singular variance covariance matrix $\Sigma$. Let $\overline{Y}$ be the sample mean and $S = \sum_{i=1}^{N} (Y_{i} - \overline{Y})(Y_{i} - \overline{Y})^{T}$. If $r = \text{rank}(\Sigma)$ and $N > r$, then $(ES)^{+} \leq E(S^{+})$. 
Proof: It is known that $R(S) \subseteq R(\Sigma)$ and rank($S$) = rank($\Sigma$) with probability 1 when $N > r$. Hence $R(S) = R(\Sigma)$ almost surely. The result now follows from Corollary 3.

REFERENCES

ON THE MATRIX CONVEXITY OF THE MOORE-PENROSE INVERSE AND SOME APPLICATIONS

Demetrios G. Kaffes, M. Bhaskara Rao, Thomas Mathew and K. Subramanyam

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh, Pittsburgh, PA 15260

Air Force Office of Scientific Research
Department of the Air Force
Bolling Air Force Base, DC 20332

Approved for public release; distribution unlimited

It is well known that if A and B are two positive definite matrices of the same order and $0 \leq \lambda \leq 1$, then $[(\lambda A + (1-\lambda)B)^{-1} < \lambda A^{-1} + (1-\lambda)B^{-1}]$. It is easy to construct an example consisting of two positive semi-definite matrices for which the above inequality is not true when one replaces the inverse operation by Moore-Penrose inverse operation. In this paper, we give necessary and sufficient conditions for the validity of the inequality $[(\lambda A + (1-\lambda)B)^+ < \lambda A^+ + (1-\lambda)B^+]$ for every $0 \leq \lambda \leq 1$. As an application, we give a sufficient condition under which...
the inequality $(EA)^+ \leq E(A^+)$ is valid, where $A$ is a square matrix of random variables which is almost surely positive semi-definite, generalizing the well-known result $(EA)^{-1} \leq EA^{-1}$ when $A$ is almost surely positive definite.