NORTH-HOLLAND

Some Further Matrix Extensions of the Cauchy-Schwarz and Kantorovich Inequalities, With Some Statistical Applications

Josip E. Pečarić
Faculty of Textile Technology
University of Zagreb
Prilaz Baruna Filipovića
41000 Zagreb, Croatia

Simo Puntanen
Department of Mathematical Sciences
University of Tampere
P.O. Box 607
FIN-33101 Tampere, Finland

and

George P. H. Styan
Department of Mathematics and Statistics
McGill University, Burnside Hall
805 ouest, rue Sherbrooke Street West
Montréal, Québec, Canada H3A 2K6

Submitted by Hans Joachim Werner

ABSTRACT

The well-known Cauchy-Schwarz and Kantorovich inequalities may be expressed in terms of vectors and a positive definite matrix. We consider what happens to these inequalities when the vectors are replaced by matrices, the positive definite matrix is allowed to be positive semidefinite singular, and the usual inequalities are replaced by Löwner partial orderings. Some examples in the context of linear statistical models are presented.

LINEAR ALGEBRA AND ITS APPLICATIONS 237/238:455–476 (1996)

0024-3795/96/$15.00
SSDI 0024-3795(95)00679-6

1. INTRODUCTION AND PRELIMINARIES

Let $\mathbf{x}$ and $\mathbf{y}$ be $n \times 1$ nonnull real vectors. Then

$$(\mathbf{x}'\mathbf{y})^2 \leqslant (\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y}) \tag{1.1}$$

is the vector version of the well-known Cauchy-Schwarz inequality (prime denotes transpose). Equality holds in (1.1) if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent, i.e.,

$$(\mathbf{x}'\mathbf{y})^2 = (\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y}) \Leftrightarrow \mathbf{x} \propto \mathbf{y}. \tag{1.2}$$

Let $\mathbf{A}$ be an $n \times n$ positive definite symmetric real matrix—throughout this paper all vectors and matrices are assumed to be real (but our results may be readily extended to complex vectors and matrices). Then there exists an $n \times n$ nonsingular matrix $\mathbf{F}$ such that

$$\mathbf{A} = \mathbf{F}\mathbf{F}'. \tag{1.3}$$

Let $\mathbf{t}$ be an $n \times 1$ vector. Then substituting $\mathbf{x} = \mathbf{F}'\mathbf{t}$ and $\mathbf{y} = \mathbf{F}^{-1}\mathbf{t}$ in (1.1) gives

$$(\mathbf{t}'\mathbf{t})^2 \leqslant (\mathbf{t}'\mathbf{A}\mathbf{t})(\mathbf{t}'\mathbf{A}^{-1}\mathbf{t}). \tag{1.4}$$

Equality holds in (1.4) if and only if $\mathbf{A}\mathbf{t} \propto \mathbf{t}$, i.e., $\mathbf{t}$ is an eigenvector of $\mathbf{A}$.

When $\mathbf{t}'\mathbf{t} = 1$, we may express (1.4) as

$$\mathbf{t}'\mathbf{A}^{-1}\mathbf{t} \geqslant (\mathbf{t}'\mathbf{A}\mathbf{t})^{-1}. \tag{1.5}$$

A "reversal" to (1.5) is provided by

$$\mathbf{t}'\mathbf{A}^{-1}\mathbf{t} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(\mathbf{t}'\mathbf{A}\mathbf{t})^{-1}, \tag{1.6}$$

where λ_1 and λ_n are the largest and smallest eigenvalues of $\mathbf{A}$. Equality holds in (1.6) when $\mathbf{t} = (\mathbf{h}_1 + \mathbf{h}_n)/\sqrt{2}$, where $\mathbf{h}_1$ and $\mathbf{h}_n$ are orthonormal eigenvectors of $\mathbf{A}$ corresponding to λ_1 and λ_n; when the eigenvalues λ_1 and λ_n are both simple (i.e., each has multiplicity 1), then this condition is also necessary.

The inequality (1.6) is a vector version of the well-known Kantorovich inequality (cf. Marcus and Minc, 1992, pp. 110, 117); see also Wang and Shao

(1992), who considered a constrained version of (1.6) with the vector $\mathbf{t}$ being restricted to a specific subspace. We note that the scalar multiplier on the right-hand side of (1.6) is the square of the ratio of the arithmetic and geometric means of λ_1 and λ_n.

Marshall and Olkin (1990) extended (1.5) and (1.6) by replacing the vector $\mathbf{t}$ with an $n \times t$ matrix $\mathbf{T}$ and the usual scalar inequality with the Löwner partial ordering. The $n \times n$ matrices $\mathbf{K}$ and $\mathbf{L}$ satisfy the Löwner partial ordering whenever the difference $\mathbf{L} - \mathbf{K} = \mathbf{FF}'$ for some matrix $\mathbf{F}$, i.e., $\mathbf{L} - \mathbf{K}$ is nonnegative definite and symmetric; cf., e.g., Marshall and Olkin (1979, p. 462). We then say that $\mathbf{K}$ is below $\mathbf{L}$ (with respect to the Löwner partial ordering) and write $\mathbf{K} \leqslant_{\mathrm{L}} \mathbf{L}$. Note that this ordering has usually been applied (particularly in statistics) when $\mathbf{K}$ and $\mathbf{L}$ are symmetric; this is, however, not necessary.

Further extensions of (1.5) and (1.6) were obtained by Baksalary and Puntanen (1991a), with the matrix $\mathbf{A}$ positive semidefinite singular and the inverse $\mathbf{A}^{-1}$ replaced by the Moore-Penrose inverse $\mathbf{A}^{+}$.

Some further related inequalities are, for example, the following—we continue to assume that $\mathbf{t}'\mathbf{t} = 1$ and that $\mathbf{A}$ is positive definite symmetric:

$$\mathbf{t}'\mathbf{A}\mathbf{t} - \frac{1}{\mathbf{t}'\mathbf{A}^{-1}\mathbf{t}} \leqslant \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2, \tag{1.7}$$

$$\frac{\mathbf{t}'\mathbf{A}^2\mathbf{t}}{(\mathbf{t}'\mathbf{A}\mathbf{t})^2} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.8}$$

$$(\mathbf{t}'\mathbf{A}^2\mathbf{t})^{1/2} - \mathbf{t}'\mathbf{A}\mathbf{t} \leqslant \frac{(\lambda_1 - \lambda_n)^2}{4(\lambda_1 + \lambda_n)}, \tag{1.9}$$

$$\mathbf{t}'\mathbf{A}^2\mathbf{t} - (\mathbf{t}'\mathbf{A}\mathbf{t})^2 \leqslant \frac{(\lambda_1 - \lambda_n)^2}{4}. \tag{1.10}$$

Equality holds in (1.7) if and only if

$$\mathbf{t}'\mathbf{A}\mathbf{t} = \lambda_1 + \lambda_n - \sqrt{\lambda_1\lambda_n} \quad \text{and} \quad \mathbf{t}'\mathbf{A}^{-1}\mathbf{t} = \frac{1}{\sqrt{\lambda_1\lambda_n}}, \tag{1.11}$$

while equality holds in (1.8)–(1.10) if and only if $\mathbf{t}'\mathbf{A}\mathbf{t}$ and $\mathbf{t}'\mathbf{A}^2\mathbf{t}$ are, respectively, equal to:

$$\text{in } (1.8) \qquad \frac{2\lambda_1\lambda_n}{\lambda_1+\lambda_n} \quad \text{and} \quad \lambda_1\lambda_n, \tag{1.12}$$

$$\text{in } (1.9) \qquad \frac{\lambda_1+\lambda_n}{4}+\frac{\lambda_1\lambda_n}{\lambda_1+\lambda_n} \quad \text{and} \quad \frac{(\lambda_1+\lambda_n)^2}{4}, \tag{1.13}$$

$$\text{in } (1.10) \qquad \frac{\lambda_1+\lambda_n}{2} \quad \text{and} \quad \frac{\lambda_1^2+\lambda_n^2}{2}. \tag{1.14}$$

Furthermore, equality holds in (1.7)–(1.10) when the vector $\mathbf{t} = \alpha\mathbf{h}_1 + \beta\mathbf{h}_n$ for certain scalar multipliers α and β where, as above, $\mathbf{h}_1$ and $\mathbf{h}_n$ are eigenvectors of $\mathbf{A}$ corresponding to λ_1 and λ_n; when the eigenvalues λ_1 and λ_n are both simple (i.e., each has multiplicity 1) then this condition is also necessary. The scalar multipliers α and β are, respectively, equal to the positive square roots of:

$$\text{for } (1.7) \qquad \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1}+\sqrt{\lambda_n}} \quad \text{and} \quad \frac{\sqrt{\lambda_n}}{\sqrt{\lambda_1}+\sqrt{\lambda_n}}, \tag{1.15}$$

$$\text{for } (1.8) \qquad \frac{\lambda_n}{\lambda_1+\lambda_n} \quad \text{and} \quad \frac{\lambda_1}{\lambda_1+\lambda_n}, \tag{1.16}$$

$$\text{for } (1.9) \qquad \frac{\lambda_1+3\lambda_n}{4(\lambda_1+\lambda_n)} \quad \text{and} \quad \frac{3\lambda_1+\lambda_n}{4(\lambda_1+\lambda_n)}, \tag{1.17}$$

$$\text{for } (1.10) \qquad \tfrac{1}{2} \quad \text{and} \quad \tfrac{1}{2}. \tag{1.18}$$

We note, therefore, that equality in (1.10) holds simultaneously with equality in the Kantorovich inequality (1.6).

The inequality (1.7) is due to Mond and Shisha (1970); cf. also Styan (1983); for (1.8), cf. Kantorovich (1948) and Greub and Rheinboldt (1959); for (1.9), cf. Mond and Shisha (1970); for (1.10), cf. Styan (1983). Mond and Pečarić (1993) provided matrix versions of (1.7), (1.8), and (1.9).

In this paper we introduce a new general matrix version of the Cauchy-Schwarz inequality, and collect together some forms of the Cauchy-Schwarz inequality that have recently appeared in the literature. We also provide matrix extensions of (1.7)–(1.10) by replacing the $n \times 1$ vector $\mathbf{t}$ with an

$n \times t$ matrix $\mathbf{T}$, allowing the symmetric matrix $\mathbf{A}$ to be nonnegative definite, and using the Löwner partial ordering.

2. A GENERALIZED MATRIX VERSION OF THE CAUCHY-SCHWARZ INEQUALITY

For a given $n \times q$ matrix $\mathbf{Y}$, we write $\mathscr{C}(\mathbf{Y})$ for the column space (range) of $\mathbf{Y}$, and $\mathbf{P}_{\mathbf{Y}}$ for the orthogonal projector onto $\mathscr{C}(\mathbf{Y})$. A generalized inverse $\mathbf{Y}^-$ of $\mathbf{Y}$ is a matrix $\mathbf{Y}^-$ that satisfies $\mathbf{YY}^-\mathbf{Y} = \mathbf{Y}$; if in addition $\mathbf{Y}^-\mathbf{YY}^- = \mathbf{Y}^-$ and both $\mathbf{YY}^-$ and $\mathbf{Y}^-\mathbf{Y}$ are symmetric, then $\mathbf{Y}^- = \mathbf{Y}^+$, the (unique) Moore-Penrose inverse of $\mathbf{Y}$.

The projector $\mathbf{P}_{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}' = \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^+\mathbf{Y}'$, since the product $\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}'$ is invariant with respect to the choice of generalized inverse $(\mathbf{Y}'\mathbf{Y})^-$ in view of the following result (Rao and Mitra, 1971, Lemma 2.2.4 and Complement 14, p. 43):

LEMMA 2.1. *Let* $\mathbf{A} \neq \mathbf{0}$ *and* $\mathbf{C} \neq \mathbf{0}$. *Then*

$$\mathbf{AB}^-\mathbf{C} = \mathbf{AB}^+\mathbf{C} \quad \textit{for all } \mathbf{B}^- \quad \Leftrightarrow \quad \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}') \textit{ and } \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B}). \tag{2.1}$$

The idempotent matrix $\mathbf{I} - \mathbf{P}_{\mathbf{Y}}$ is the orthogonal projector on the orthocomplement of $\mathscr{C}(\mathbf{Y})$ and is nonnegative definite:

$$\mathbf{I} - \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}' \geqslant_{\mathsf{L}} \mathbf{0}. \tag{2.2}$$

If $\mathbf{X}$ is an $n \times m$ matrix, then (2.2) implies (cf. Chipman, 1964, p. 1093) that

$$\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}'\mathbf{X} \leqslant_{\mathsf{L}} \mathbf{X}'\mathbf{X}, \tag{2.3}$$

which is a matrix version of (1.1). Equality in (2.3) holds if and only if $\mathbf{X}'(\mathbf{I} - \mathbf{P}_{\mathbf{Y}})\mathbf{X} = \mathbf{0}$, which is equivalent to $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{Y})$, i.e.,

$$\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}'\mathbf{X} = \mathbf{X}'\mathbf{X} \quad \Leftrightarrow \quad \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{Y}) \quad \Leftrightarrow \quad \mathbf{X} = \mathbf{YF} \tag{2.4}$$

for some $q \times m$ matrix $\mathbf{F}$. We note that when $m = q = 1$, then (2.4) reduces to (1.2), the condition for equality in the (usual) Cauchy-Schwarz inequality.

For our generalization of (2.3) below (Theorem 2.1) we introduce the following notation for powers of a matrix $\mathbf{A}$, symmetric but not necessarily nonnegative definite:

$$\mathbf{A}^{\{p\}} = \begin{cases} \mathbf{A}^{p}, & p = 1,2,\dots, & (2.5\text{a}) \\ \mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}', & p = 0, & (2.5\text{b}) \\ (\mathbf{A}^{+})^{|p|}, & p = -1,-2,\dots . & (2.5\text{c}) \end{cases}$$

If we let $\mathbf{A}$ be $n \times n$ with rank r, then we write the spectral decomposition

$$\mathbf{A} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}', \tag{2.6}$$

where $\mathbf{W}$ is an $n \times r$ matrix such that $\mathbf{W}'\mathbf{W} = \mathbf{I}_r$ and $\boldsymbol{\Lambda} = (\lambda_i)$ is an $r \times r$ diagonal matrix with the nonzero eigenvalues λ_i on its diagonal. Then

$$\mathbf{A}^{\{p\}} = \mathbf{W}\boldsymbol{\Lambda}^{p}\mathbf{W}', \qquad p = \dots, -2, -1, 0, 1, 2, \dots , \tag{2.7}$$

where $\boldsymbol{\Lambda}^{p} = (\lambda_i^{p})$, and so the nonzero eigenvalues of $\mathbf{A}^{\{p\}}$ are precisely the pth powers of the nonzero eigenvalues of $\mathbf{A}$. When $p = 0$ these eigenvalues are all equal to one; when p is negative they are the reciprocals of the nonzero eigenvalues of $\mathbf{A}$ raised to the power $|p| = -p$.

When $\mathbf{A}$ is nonnegative definite, then its nonzero eigenvalues are all positive, and by taking positive square roots of the eigenvalues in (2.7) we may define the nonnegative definite matrices

$$\mathbf{A}^{\{p\}} = \begin{cases} \mathbf{A}^{p}, & p = \tfrac{1}{2}, 1\tfrac{1}{2}, \dots, & (2.8\text{a}) \\ (\mathbf{A}^{+})^{|p|}, & p = -\tfrac{1}{2}, -1\tfrac{1}{2}, \dots . & (2.8\text{b}) \end{cases}$$

When $\mathbf{A}$ is symmetric but not nonnegative definite, then $\mathbf{A}$ has at least one negative eigenvalue and the matrix $\mathbf{A}^{1/2}$ is not real.

THEOREM 2.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *symmetric matrix with* $\mathbf{A}^{\{p\}}$ *defined by* (2.5) *and* (2.7); *let the matrices* $\mathbf{T}$ *and* $\mathbf{U}$ *be* $n \times t$ *and* $n \times u$, *respectively, and let* h *and* k *be integers (possibly negative or zero). Then*

$$\mathbf{T}'\mathbf{A}^{\{\frac{1}{2}(h+k)\}}\mathbf{U}(\mathbf{U}'\mathbf{A}^{\{k\}}\mathbf{U})^{-}\mathbf{U}'\mathbf{A}^{\{\frac{1}{2}(h+k)\}}\mathbf{T} \leqslant_{\mathrm{L}} \mathbf{T}'\mathbf{A}^{\{h\}}\mathbf{T} \tag{2.9}$$

whenever either

(a) *the matrix* $\mathbf{A}$ *is nonnegative definite, or*
(b) *the integers* h *and* k *are both even or zero.*

The following three equivalent conditions characterize equality in (2.9):

$$\mathscr{C}(\mathbf{A}^{\{\frac{1}{2}h\}}\mathbf{T}) \subset \mathscr{C}(\mathbf{A}^{\{\frac{1}{2}k\}}\mathbf{U}), \tag{2.10a}$$

$$\mathscr{C}(\mathbf{P_A T}) \subset \mathscr{C}(\mathbf{A}^{\{\frac{1}{2}(k-h)\}}\mathbf{U}), \tag{2.10b}$$

$$\mathscr{C}(\mathbf{AT}) \subset \mathscr{C}(\mathbf{A}^{\{1+\frac{1}{2}(k-h)\}}\mathbf{U}). \tag{2.10c}$$

We note that, in view of (2.1), the left-hand side of (2.9) is invariant with respect to the choice of generalized inverse. Moreover, when $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$ then (2.9) holds for all integers h, k: positive, negative, or zero. When h and k are both even or zero and $\mathbf{A}$ is symmetric but not nonnegative definite, then the matrices $\mathbf{A}^{\{\frac{1}{2}h\}}$, $\mathbf{A}^{\{\frac{1}{2}k\}}$, and $\mathbf{A}^{\{\frac{1}{2}(h+k)\}}$ are all real; when h is odd, however, the matrix $\mathbf{A}^{\{\frac{1}{2}h\}}$ will not be real.

Proof of Theorem 2.1. When either (a) the matrix $\mathbf{A}$ is nonnegative definite or (b) the integers h and k are both even or zero, the matrices $\mathbf{A}^{\{\frac{1}{2}h\}}$ and $\mathbf{A}^{\{\frac{1}{2}k\}}$ are both well defined and real. We may then substitute $\mathbf{X} = \mathbf{A}^{\{\frac{1}{2}h\}}\mathbf{T}$ and $\mathbf{Y} = \mathbf{A}^{\{\frac{1}{2}k\}}\mathbf{U}$ into (2.3) and (2.4), respectively, to yield (2.9) and (2.10a) directly. It is straightforward to show that the three conditions in (2.10) are equivalent. ∎

Many special cases of (2.9) have appeared in the literature. The case of $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$, $\mathbf{T} = \mathbf{U}$, $h = -1$, and $k = 1$ was given by Baksalary and Puntanen (1991a) as follows:

COROLLARY 2.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *nonnegative definite symmetric matrix, and let* $\mathbf{T}$ *be an* $n \times t$ *matrix. Then*

$$\mathbf{T'P_A T(T'AT)^- T'P_A T} \leqslant_{\mathsf{L}} \mathbf{T'A^+T} \tag{2.11}$$

with equality if and only if

$$\mathscr{C}(\mathbf{AT}) = \mathscr{C}(\mathbf{P_A T}). \tag{2.12}$$

Moreover, if in addition $\mathbf{A}$ *and* $\mathbf{T}$ *are such that* $\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T}$ *is idempotent*:

$$\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} = (\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T})^2, \tag{2.13}$$

then

$$(\mathbf{T}'\mathbf{A}\mathbf{T})^+ \leqslant_{\mathsf{L}} \mathbf{T}'\mathbf{A}^+\mathbf{T} \tag{2.14}$$

with equality if and only if

$$\mathbf{P}_{\mathbf{A}}\mathbf{T}\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{A}\mathbf{T}. \tag{2.15}$$

The idempotency condition (2.13) is equivalent to the condition

$$\mathbf{P}_{\mathbf{A}}\mathbf{T}(\mathbf{T}'\mathbf{P}_{\mathbf{A}})\mathbf{P}_{\mathbf{A}}\mathbf{T} = \mathbf{P}_{\mathbf{A}}\mathbf{T}, \tag{2.16}$$

since $\mathbf{P}_{\mathbf{A}} \geqslant_{\mathsf{L}} \mathbf{0}$, and so $\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T}$ is idempotent if and only if $(\mathbf{P}_{\mathbf{A}}\mathbf{T})'$ is a generalized inverse of $\mathbf{P}_{\mathbf{A}}\mathbf{T}$. Such a matrix is called a partial isometry; cf. Horn and Johnson (1994, p. 152). It follows that then $\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} = \mathbf{P}_{\mathbf{T}'\mathbf{A}}$, the orthogonal projector onto the column space $\mathscr{C}(\mathbf{T}'\mathbf{A})$.

Following Baksalary and Puntanen (1991a), we see that if $\mathbf{T}$ has full column rank, $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$, and $\mathscr{C}(\mathbf{T}) \subset \mathscr{C}(\mathbf{A})$, then $\mathbf{T}'\mathbf{A}\mathbf{T}$ is positive definite, $\mathbf{P}_{\mathbf{A}}\mathbf{T} = \mathbf{T}$, and (2.11) may be written as

$$\mathbf{T}'\mathbf{T}(\mathbf{T}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}'\mathbf{T} \leqslant_{\mathsf{L}} \mathbf{T}'\mathbf{A}^+\mathbf{T} \tag{2.17}$$

with equality if and only if

$$\mathscr{C}(\mathbf{A}\mathbf{T}) = \mathscr{C}(\mathbf{T}). \tag{2.18}$$

When $\mathscr{C}(\mathbf{T}) \subset \mathscr{C}(\mathbf{A})$, we may, in view of (2.1), replace the Moore-Penrose inverse $\mathbf{A}^+$ in (2.17) with any generalized inverse $\mathbf{A}^-$, and then (2.17) coincides with Lemma 2.1 in Gaffke and Krafft (1977), which generalizes Lemma 2c in Rao (1967).

Furthermore, if $\mathbf{A}$ is positive definite, then $\mathbf{P}_{\mathbf{A}} = \mathbf{I}_n$ and (2.16) becomes $\mathbf{T}\mathbf{T}'\mathbf{T} = \mathbf{T}$, and (2.17) leads to

$$(\mathbf{T}'\mathbf{A}\mathbf{T})^+ \leqslant_{\mathsf{L}} \mathbf{T}'\mathbf{A}^{-1}\mathbf{T}. \tag{2.19}$$

Marshall and Olkin (1990) considered the special case of (2.19) with $\mathbf{T}$ suborthogonal, i.e., $\mathbf{T}'\mathbf{T} = \mathbf{I}_t$.

Let us now partition conformably the $n \times n$ nonnegative definite symmetric matrix $\mathbf{T}$ and its Moore-Penrose inverse $\mathbf{A}^+$ as follows:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{A}^+ = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{pmatrix}, \tag{2.20}$$

with $\mathbf{A}_{11} t \times t$. Then with $\mathbf{T} = (\mathbf{I}_t : \mathbf{0})'$, Baksalary and Puntanen (1991a) showed that Corollary 2.1 implies the following:

$$\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{A}_{11} + \operatorname{rank} \mathbf{A}_{22} \quad \Rightarrow \quad \mathbf{A}_{11}^+ \leqslant_{\mathsf{L}} \mathbf{B}_{11} \tag{2.21}$$

with equality if and only if $\mathbf{A}_{12} = \mathbf{0}$; see also Baksalary and Kala (1980, Proposition 1), Chollet (1982), and Marcus (1982).

We now consider an application of (2.17) in the context of the linear statistical model. To do this, we let $\mathbf{X}$ be an $n \times m$ matrix with full column rank m, and we let $\mathbf{V}$ be an $n \times n$ positive definite symmetric matrix. Replacing $\mathbf{T}$ by $\mathbf{X}$ and $\mathbf{A}$ by $\mathbf{V}^{-1}$, we rewrite (2.17) in the form

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \leqslant_{\mathsf{L}} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{2.22}$$

This inequality has an important statistical interpretation. For this purpose, we consider the full-rank linear model

$$\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}, \tag{2.23}$$

where $\mathbf{y}$ is an $n \times 1$ observable random vector with $n \times 1$ expectation vector $\mathscr{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $n \times n$ dispersion (or covariance) matrix $\mathscr{D}(\mathbf{y}) = \mathbf{V}$. Here the $n \times m$ design (or model) matrix $\mathbf{X}$ has $\operatorname{rank} \mathbf{X} = m > 0$, and is known, while the $m \times 1$ vector $\boldsymbol{\beta}$ is unknown, and the $n \times n$ dispersion matrix $\mathbf{V}$ is positive definite symmetric and known. Then $\mathbf{Cy}$ is said to be the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ whenever the dispersion matrix of $\mathbf{Cy}$ is smallest (in the Löwner sense) among all linear unbiased estimators of $\boldsymbol{\beta}$. Since

$$\text{BLUE}\, \boldsymbol{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad \text{and} \quad \text{OLSE}\, \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{2.24}$$

the left-hand side of (2.22) is the dispersion matrix of the BLUE of $\boldsymbol{\beta}$ and the right-hand side is the dispersion matrix of the OLSE of $\boldsymbol{\beta}$, and so we have:

$$\mathscr{D}(\text{BLUE } \boldsymbol{\beta}) \leqslant_{\mathsf{L}} \mathscr{D}(\text{OLSE } \boldsymbol{\beta}). \tag{2.25}$$

We notice that in view of (2.18), equality holds in (2.25) if and only if $\mathscr{C}(\mathbf{VX}) = \mathscr{C}(\mathbf{X})$ (cf. Rao, 1967, and Zyskind, 1967); for a general discussion of the conditions for the equality of the OLSE and the BLUE and of the conditions for the equality of their dispersion matrices, see Puntanen and Styan (1989).

Furthermore, premultiplying (2.22) by $\mathbf{X}$ and postmultiplying by $\mathbf{X}'$ gives

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' \leqslant_{\mathsf{L}} \mathbf{P}_{\mathbf{X}}\mathbf{V}\mathbf{P}_{\mathbf{X}}, \tag{2.26}$$

which in statistical terms means

$$\mathscr{D}(\text{BLUE } \mathbf{X}\boldsymbol{\beta}) \leqslant_{\mathsf{L}} \mathscr{D}(\text{OLSE } \mathbf{X}\boldsymbol{\beta}). \tag{2.27}$$

The inequality (2.27) is, of course, also true when $\mathbf{X}$ has less than full rank.

Since the random vector $\mathbf{y}$ itself is an unbiased estimator for $\mathbf{X}\boldsymbol{\beta}$, we must have

$$\mathscr{D}(\text{BLUE } \mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' \leqslant_{\mathsf{L}} \mathbf{V} = \mathscr{D}(\mathbf{y}), \tag{2.28}$$

which follows from (2.9) with appropriate substitutions. As a related matter, we may mention that equality in (2.27) holds if and only if (cf. Baksalary and Puntanen, 1990)

$$\mathbf{P}_{\mathbf{X}}\mathbf{V}\mathbf{P}_{\mathbf{X}} \leqslant_{\mathsf{L}} \mathbf{V}. \tag{2.29}$$

In this context, we note that Chipman (1968, p. 120, Lemma 2.1.2; 1976, p. 562, Lemma 2.1.2) has given the following "generalized Schwarz inequality": Let $\mathbf{X}$ be an $n \times m$ matrix as above, let $\mathbf{V}$ now be an $n \times n$ nonnegative definite symmetric matrix, and let $\mathbf{X}^{\sim}$ satisfy the two conditions

$$\mathbf{X}\mathbf{X}^{\sim}\mathbf{X} = \mathbf{X} \quad \text{and} \quad \mathbf{X}\mathbf{X}^{\sim}\mathbf{V} = (\mathbf{X}\mathbf{X}^{\sim}\mathbf{V})'. \tag{2.30}$$

Then for any $n \times m$ matrix $\mathbf{F}$,

$$\mathbf{F}\mathbf{X}\mathbf{X}^{\sim}\mathbf{V}\mathbf{X}^{\sim\prime}\mathbf{X}'\mathbf{F}' \leqslant_{\mathsf{L}} \mathbf{F}\mathbf{V}\mathbf{F}', \tag{2.31}$$

with equality if and only if

$$\mathbf{FV} = \mathbf{FXX}^{\sim}\mathbf{V}; \tag{2.32}$$

cf. Chipman (1968, p. 120, Lemma 2.1.2; 1976, p. 562, Lemma 2.1.2). Rao and Mitra (1971, p. 46) refer to the transpose $\mathbf{X}^{\sim\prime}$ of $\mathbf{X}^{\sim}$, as defined by (2.30), as a minimum-$\mathbf{V}$-seminorm generalized inverse of $\mathbf{X}'$. Moreover, if $\mathbf{X}^{\sim}$ satisfies (2.30), then $\mathbf{XX}^{\sim}\mathbf{y}$ is the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, and therefore (2.31) is a parallel statement to (2.26) without any rank assumptions.

Let us now return to Theorem 2.1 and suppose that $t = u = 1$; then $\mathbf{T}$ and $\mathbf{U}$ are $n \times 1$ nonnull vectors, which we will now denote by $\mathbf{t}$ and $\mathbf{u}$, respectively. Then with $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$ and with $\mathbf{A}^{\{p\}}$ defined by (2.5) and (2.7), we obtain

$$\left(\mathbf{t}'\mathbf{A}^{\{\frac{1}{2}(h+k)\}}\mathbf{u}\right)^2 \leqslant \left(\mathbf{t}'\mathbf{A}^{\{h\}}\mathbf{t}\right)\left(\mathbf{u}'\mathbf{A}^{\{k\}}\mathbf{u}\right) \qquad \text{for} \quad h, k = \ldots, -1, 0, 1, 2, \ldots, \tag{2.33}$$

with equality if and only if [cf. (2.10c)]

$$\mathbf{At} \propto \mathbf{A}^{\{1 + \frac{1}{2}(k-h)\}}\mathbf{u}. \tag{2.34}$$

If $h = 1$ and $k = -1$, then with $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$ the inequality (2.33) becomes

$$\left(\mathbf{t}'\mathbf{P}_{\mathbf{A}}\mathbf{u}\right)^2 \leqslant \left(\mathbf{t}'\mathbf{At}\right)\left(\mathbf{u}'\mathbf{A}^{+}\mathbf{u}\right) \tag{2.35}$$

with equality if and only if

$$\mathbf{At} \propto \mathbf{P}_{\mathbf{A}}\mathbf{u}. \tag{2.36}$$

And so when $\mathbf{A} \geqslant_{\mathsf{L}} \mathbf{0}$ we have

$$\left(\mathbf{t}'\mathbf{u}\right)^2 \leqslant \left(\mathbf{t}'\mathbf{At}\right)\left(\mathbf{u}'\mathbf{A}^{-}\mathbf{u}\right) \tag{2.37}$$

for all $\mathbf{u} \in \mathscr{C}(\mathbf{A})$ [cf. (1.1) and (1.4)]; the quadratic form $\mathbf{u}'\mathbf{A}^{-}\mathbf{u}$ in (2.37) is invariant with respect to the choice of generalized inverse $\mathbf{A}^{-}$ when $\mathbf{u} \in \mathscr{C}(\mathbf{A})$ [cf. (2.1)]. For a statistical proof of (2.37), see Dey, Hande, and Tiku (1994, Theorem 2.1). Equality holds in (2.37) if and only if

$$\mathbf{At} \propto \mathbf{u}. \tag{2.38}$$

In his comment on Olkin (1992), Chaganty (1993) gave (2.37) with the Moore-Penrose inverse $\mathbf{A}^+$ instead of an arbitrary generalized inverse $\mathbf{A}^-$, and commented that he could not find a proof of this result in the literature. He pointed out that the condition $\mathbf{t} = \mathbf{A}^+\mathbf{u}$ is sufficient for equality to hold in (2.37), but gave no necessary condition. It is easy to see that $\mathbf{t} = \mathbf{A}^+\mathbf{u}$ is a special case of (2.38) when $\mathbf{u} \in \mathscr{C}(\mathbf{A})$.

In a further comment on Olkin (1992), Trenkler (1994) gave the following inequality (cf. Baksalary and Kala, 1983):

$$(\mathbf{t}'\mathbf{u})^2 \leqslant k\mathbf{t}'\mathbf{At} \tag{2.39}$$

for all $n \times 1$ vectors $\mathbf{t}$ if and only if $\mathbf{A} \geqslant_{\mathrm{L}} \mathbf{0}$, $\mathbf{u} \in \mathscr{C}(\mathbf{A})$, and $k \geqslant \mathbf{u}'\mathbf{A}^-\mathbf{u}$; cf. also Baksalary and Trenkler (1991) and Baksalary, Schipp, and Trenkler (1992), and the recent discussion between Bancroft (1994), Neudecker and Liu (1994), and Chaganty and Vaish (1994).

If we now let $\mathbf{t} = \mathbf{u} \in \mathscr{C}(\mathbf{A})$, then (2.37) simplifies further to

$$(\mathbf{t}'\mathbf{t})^2 \leqslant (\mathbf{t}'\mathbf{At})(\mathbf{t}'\mathbf{A}^-\mathbf{t}) \tag{2.40}$$

for all $\mathbf{t} \in \mathscr{C}(\mathbf{A})$ and for all choices of generalized inverse $\mathbf{A}^-$; when $\mathbf{t} \neq \mathbf{0}$, then equality holds in (2.40) if and only if $\mathbf{t}$ is an eigenvector of $\mathbf{A}$. This result is given by Dey and Gupta (1977, Lemma 2.1).

3. GENERALIZED MATRIX VERSIONS OF THE KANTOROVICH INEQUALITY

Let $\mathbf{A}$ be an $n \times n$ positive definite symmetric matrix, and let $\mathbf{T}$ be an $n \times t$ matrix such that

$$\mathbf{T}'\mathbf{T} = \mathbf{I}_t, \tag{3.1}$$

i.e., $\mathbf{T}$ is suborthogonal. For such $\mathbf{A}$ and $\mathbf{T}$, Marshall and Olkin (1990) proved the following matrix version of the Kantorovich inequality (1.6):

$$\mathbf{T}'\mathbf{A}^{-1}\mathbf{T} \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(\mathbf{T}'\mathbf{A}\mathbf{T})^{-1}, \tag{3.2}$$

and Mond and Pečarić (1993) proved the following matrix versions of (1.7), (1.8), and (1.9):

$$\mathbf{T}'\mathbf{A}\mathbf{T} - (\mathbf{T}'\mathbf{A}^{-1}\mathbf{T})^{-1} \leqslant_{\mathsf{L}} \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2 \mathbf{I}_t, \tag{3.3}$$

$$\mathbf{T}'\mathbf{A}^2\mathbf{T} \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(\mathbf{T}'\mathbf{A}\mathbf{T})^2, \tag{3.4}$$

$$(\mathbf{T}'\mathbf{A}^2\mathbf{T})^{1/2} - \mathbf{T}'\mathbf{A}\mathbf{T} \leqslant_{\mathsf{L}} \frac{(\lambda_1 - \lambda_n)^2}{4(\lambda_1 + \lambda_n)}\mathbf{I}_t. \tag{3.5}$$

Mond and Pečarić (1994) extended (3.2)–(3.5) to sums of matrices.

It is easy to see that (1.10) generalizes similarly to

$$\mathbf{T}'\mathbf{A}^2\mathbf{T} - (\mathbf{T}'\mathbf{A}\mathbf{T})^2 \leqslant_{\mathsf{L}} \tfrac{1}{4}(\lambda_1 - \lambda_n)^2\mathbf{I}_t. \tag{3.6}$$

In this section we allow $\mathbf{A}$ to be nonnegative definite, thus being possibly singular, and so generalize (3.2)–(3.6).

Baksalary and Puntanen (1991a) gave the following generalization of (3.2): Let $\mathbf{A}$ be an $n \times n$ nonnegative definite symmetric matrix of rank r with nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0$, and let $\mathbf{T}$ be an $n \times t$ matrix. Then

$$\mathbf{T}'\mathbf{A}^{+}\mathbf{T} \leqslant_{\mathsf{L}} \frac{\lambda_1 + \lambda_r}{\lambda_1\lambda_r}\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} - \frac{1}{\lambda_1\lambda_r}\mathbf{T}'\mathbf{A}\mathbf{T}. \tag{3.7}$$

If we now assume that $\mathbf{A}$ and $\mathbf{T}$ are such that

$$\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} \text{ is idempotent}, \tag{3.8}$$

i.e., $\mathbf{P}_{\mathbf{A}}\mathbf{T}$ is a partial isometry [cf. (2.13) and (2.16)], then (3.7) simplifies to

$$\mathbf{T}'\mathbf{A}^{+}\mathbf{T} \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(\mathbf{T}'\mathbf{A}\mathbf{T})^{+}. \tag{3.9}$$

Equality holds in (3.9) if and only if

$$(\mathbf{T}'\mathbf{A}\mathbf{T})^2 = \frac{\lambda_1 + \lambda_r}{2}\mathbf{T}'\mathbf{A}\mathbf{T} \quad \text{and} \quad (\mathbf{T}'\mathbf{A}^+\mathbf{T})^2 = \frac{\lambda_1 + \lambda_r}{2\lambda_1\lambda_r}\mathbf{T}'\mathbf{A}^+\mathbf{T}, \tag{3.10}$$

so that both $\mathbf{T}'\mathbf{A}\mathbf{T}$ and $\mathbf{T}'\mathbf{A}^+\mathbf{T}$ are scalar-potent. It follows that (3.10) holds if and only if either $\mathbf{A}\mathbf{T} = \mathbf{0}$ or all the nonzero eigenvalues of $\mathbf{T}'\mathbf{A}\mathbf{T}$ and $\mathbf{T}'\mathbf{A}^+\mathbf{T}$ are, respectively, equal to $(\lambda_1 + \lambda_r)/2$ and $(\lambda_1 + \lambda_r)/(2\lambda_1\lambda_r)$.

Consider now again the linear statistical model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ [cf. (2.23)], where $\mathbf{X}$ is an $n \times m$ matrix with full column rank m, and $\mathbf{V}$ is an $n \times n$ positive definite symmetric matrix. Suppose now that $\mathbf{X}$ is suborthogonal, i.e., $\mathbf{X}'\mathbf{X} = \mathbf{I}_m$. Then, in view of (2.22) and (3.9), we obtain

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \leqslant_{\mathsf{L}} \mathbf{X}'\mathbf{V}\mathbf{X} \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{3.11}$$

or equivalently

$$\mathscr{D}(\text{BLUE } \boldsymbol{\beta}) \leqslant_{\mathsf{L}} \mathscr{D}(\text{OLSE } \boldsymbol{\beta}) \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}\mathscr{D}(\text{BLUE } \boldsymbol{\beta}). \tag{3.12}$$

As noted in Section 2, equality holds in the left-hand inequality of (3.12) if and only if $\mathscr{C}(\mathbf{V}\mathbf{X}) = \mathscr{C}(\mathbf{X})$. The question of equality in the right-hand inequality of (3.12) was raised by Magness and McGuire (1962, p. 470), who conjectured that equality cannot be attained if $\mathbf{V}$ is irreducible. This conjecture was disproved, using (3.10), by Baksalary and Puntanen (1991b). We note that equality in the left-hand inequality of (3.12) may be interpreted as the OLSE being "as good as possible" or "fully efficient" with respect to the BLUE. The question of how "bad" the OLSE can be is more complicated, since there is no unique way to measure the relative goodness of the OLSE; cf. Watson (1955), Bloomfield and Watson (1975), and Puntanen (1987). Rao (1985) measured the goodness of the OLSE as the trace of the difference $\mathbf{X}'\mathbf{V}\mathbf{X} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ and, while keeping $\mathbf{V}$ fixed and letting $\mathbf{X}$ vary, found an upper bound.

We may note that corresponding to (3.11), we also have that

$$\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{A}_{11} + \operatorname{rank}\mathbf{A}_{22} \quad \Rightarrow \quad \mathbf{A}_{11}^{+} \leqslant_{\mathsf{L}} \mathbf{B}_{11} \leqslant_{\mathsf{L}} \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}\mathbf{A}_{11}^{+}; \tag{3.13}$$

cf. (2.19) and (2.21).

We generalize (3.3) in the following Theorem 3.1; generalizations of (3.4)–(3.6) appear in Theorem 3.2.

THEOREM 3.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *nonnegative definite symmetric matrix of rank* r *with nonzero eigenvalues* $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0$ *and orthogonal projector* $\mathbf{P}_{\mathbf{A}}$, *and let* $\mathbf{T}$ *be an* $n \times t$ *matrix such that* $\mathbf{P}_{\mathbf{A}}\mathbf{T}$ *is a partial isometry, i.e.,*

$$\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} \text{ is idempotent}; \tag{3.14}$$

cf. (3.8). *Then*

$$\mathbf{T}'\mathbf{A}\mathbf{T} - (\mathbf{T}'\mathbf{A}^{+}\mathbf{T})^{+} \leqslant_{\mathsf{L}} \left(\sqrt{\lambda_1} - \sqrt{\lambda_r}\right)^2 \mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T}, \tag{3.15}$$

with equality if and only if both $\mathbf{T}'\mathbf{A}\mathbf{T}$ *and* $\mathbf{T}'\mathbf{A}^{+}\mathbf{T}$ *are scalar-potent with*

$$(\mathbf{T}'\mathbf{A}\mathbf{T})^2 = \left(\lambda_1 + \lambda_r - \sqrt{\lambda_1\lambda_r}\right)\mathbf{T}'\mathbf{A}\mathbf{T} \quad \text{and} \quad (\mathbf{T}'\mathbf{A}^{+}\mathbf{T})^2 = \frac{1}{\sqrt{\lambda_1\lambda_r}}\mathbf{T}'\mathbf{A}^{+}\mathbf{T}. \tag{3.16}$$

It follows [cf. (3.10)] that equality holds in (3.15) if and only if either $\mathbf{A}\mathbf{T} = \mathbf{0}$ or all the nonzero eigenvalues of $\mathbf{T}'\mathbf{A}\mathbf{T}$ and $\mathbf{T}'\mathbf{A}^{+}\mathbf{T}$ are equal, respectively, to $\lambda_1 + \lambda_r - \sqrt{\lambda_1\lambda_r}$ and $1/\sqrt{\lambda_1\lambda_r}$.

Proof of Theorem 3.1. The key inequality that we will use in our proof [cf. Marshall and Olkin (1964, p. 509) and Styan (1983)] is

$$\lambda_i \leqslant \lambda_1 + \lambda_r - \frac{\lambda_1\lambda_r}{\lambda_i}, \qquad i = 1, 2, \ldots, r, \tag{3.17}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_i \geqslant \cdots \geqslant \lambda_r > 0$ are the eigenvalues of $\mathbf{A}$. The inequality (3.17) follows at once from

$$(\lambda_1 - \lambda_i)(\lambda_i - \lambda_r) \geqslant 0. \tag{3.18}$$

Equality holds in (3.18) if and only if equality holds in (3.17) if and only if the largest i or the smallest $n - i$ nonzero eigenvalues of $\mathbf{A}$ are all equal, i.e., either $\lambda_1 = \cdots = \lambda_i$ or $\lambda_i = \cdots = \lambda_r$, $i = 1, 2, \ldots, r$.

We write (3.17) in the matrix form

$$\boldsymbol{\Lambda} \leqslant_{\mathrm{L}} (\lambda_1 + \lambda_r)\mathbf{I}_r - (\lambda_1\lambda_r)\boldsymbol{\Lambda}^{-1}, \tag{3.19}$$

where, as before, $\boldsymbol{\Lambda}$ is an $r \times r$ diagonal matrix with the nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_i \geqslant \cdots \geqslant \lambda_r > 0$ on the diagonal. As in (2.6), we use the spectral decomposition $\mathbf{A} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}'$, where $\mathbf{W}$ is an $n \times r$ matrix such that $\mathbf{W}'\mathbf{W} = \mathbf{I}_r$. Premultiplying (3.19) by $\mathbf{T}'\mathbf{W}$ and postmultiplying by $\mathbf{W}'\mathbf{T}$, and subtracting $(\mathbf{T}'\mathbf{A}^+\mathbf{T})^+$ from both sides, we have

$$\begin{aligned}\mathbf{T}'\mathbf{A}\mathbf{T} - (\mathbf{T}'\mathbf{A}^+\mathbf{T})^+ &<_{\mathrm{L}} (\lambda_1 + \lambda_r)\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} - \lambda_1\lambda_r\mathbf{T}'\mathbf{A}^+\mathbf{T} - (\mathbf{T}'\mathbf{A}^+\mathbf{T})^+ \\ &= \left(\sqrt{\lambda_1} - \sqrt{\lambda_r}\right)^2\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T} - \mathbf{F}^2,\end{aligned} \tag{3.20}$$

where the symmetric matrix

$$\mathbf{F} = \sqrt{\lambda_1\lambda_r}\,(\mathbf{T}'\mathbf{A}^+\mathbf{T})^{1/2} - \left[(\mathbf{T}'\mathbf{A}^+\mathbf{T})^+\right]^{1/2}. \tag{3.21}$$

To confirm the equality in (3.20), we observe that

$$(\mathbf{T}'\mathbf{A}^+\mathbf{T})^{1/2}\left[(\mathbf{T}'\mathbf{A}^+\mathbf{T})^+\right]^{1/2} = \mathbf{P}_{\mathbf{T}'\mathbf{A}}, \tag{3.22}$$

the orthogonal projector onto $\mathscr{C}(\mathbf{T}'\mathbf{A})$, and from the idempotency condition (3.14) we have

$$\mathbf{P}_{\mathbf{T}'\mathbf{A}} = \mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T}. \tag{3.23}$$

The inequality (3.20) then implies (3.14), since $\mathbf{F}^2$ is nonnegative definite.

Equality holds in (3.15), therefore, if and only if equality holds throughout (3.20) and $\mathbf{F} = \mathbf{0}$; these two conditions are easily seen to be equivalent, respectively, to the two conditions in (3.16). ∎

We now generalize (3.4)–(3.6):

THEOREM 3.2. *Let the matrices* $\mathbf{A}$, $\mathbf{P_A}$, *and* $\mathbf{T}$ *be defined as in Theorem* 3.1. *Then*

$$\mathbf{T'A^2T} \leqslant_{\mathsf{L}} \frac{(\lambda_1+\lambda_r)^2}{4\lambda_1\lambda_r}(\mathbf{T'AT})^2, \tag{3.24}$$

$$(\mathbf{T'A^2T})^{1/2} - \mathbf{T'AT} \leqslant_{\mathsf{L}} \frac{(\lambda_1-\lambda_r)^2}{4(\lambda_1+\lambda_r)}\mathbf{T'P_AT}, \tag{3.25}$$

$$\mathbf{T'A^2T} - (\mathbf{T'AT})^2 \leqslant_{\mathsf{L}} \frac{(\lambda_1-\lambda_r)^2}{4}\mathbf{T'P_AT}, \tag{3.26}$$

with equality if and only if both $\mathbf{T'AT}$ *and* $\mathbf{T'A^2T}$ *are scalar-potent, and for equality—*

in (3.24):

$$(\mathbf{T'AT})^2 = \frac{2\lambda_1\lambda_r}{\lambda_1+\lambda_r}\mathbf{T'AT} \quad \textit{and} \quad (\mathbf{T'A^2T})^2 = \lambda_1\lambda_r\mathbf{T'A^2T}; \tag{3.27}$$

in (3.25):

$$(\mathbf{T'AT})^2 = \left(\frac{\lambda_1+\lambda_r}{4} + \frac{\lambda_1\lambda_r}{\lambda_1+\lambda_r}\right)\mathbf{T'AT} \quad \textit{and}$$

$$(\mathbf{T'A^2T})^2 = \frac{(\lambda_1+\lambda_r)^2}{4}\mathbf{T'A^2T}; \tag{3.28}$$

in (3.26):

$$(\mathbf{T'AT})^2 = \frac{\lambda_1+\lambda_r}{2}\mathbf{T'AT} \quad \textit{and} \quad (\mathbf{T'A^2T})^2 = \frac{\lambda_1^2+\lambda_r^2}{2}\mathbf{T'A^2T}. \tag{3.29}$$

It follows [cf. the paragraph following (3.16)] that when $\mathbf{AT} = \mathbf{0}$ equality holds throughout (3.24), (3.25), and (3.26). Then $\mathbf{AT} \neq \mathbf{0}$, however, then equality holds in (3.24), (3.25), and (3.26) if and only if all the nonzero eigenvalues of $\mathbf{T'AT}$ and $\mathbf{T'A^2T}$ are, respectively, equal to:

$$\text{for } (3.24): \qquad \frac{2\lambda_1\lambda_r}{\lambda_1+\lambda_r} \quad \text{and} \quad \lambda_1\lambda_r, \tag{3.30}$$

$$\text{for } (3.25): \qquad \frac{\lambda_1+\lambda_r}{4}+\frac{\lambda_1\lambda_r}{\lambda_1+\lambda_r} \quad \text{and} \quad \frac{(\lambda_1+\lambda_r)^2}{4}, \tag{3.31}$$

$$\text{for } (3.26): \qquad \frac{\lambda_1+\lambda_r}{2} \quad \text{and} \quad \frac{\lambda_1^2+\lambda_r^2}{2}. \tag{3.32}$$

We observe, therefore, that equality in (3.26) coincidences with equality in the generalized Kantorovich inequality (3.9); cf. (3.10), (3.26), and (3.32).

Proof of Theorem 3.2. To prove (3.24), we begin by multiplying (3.19) by $\boldsymbol{\Lambda}$, so that

$$\boldsymbol{\Lambda}^2 \leqslant_{\mathsf{L}} (\lambda_1+\lambda_r)\boldsymbol{\Lambda} - (\lambda_1\lambda_r)\mathbf{I}_r. \tag{3.33}$$

Proceeding as in the proof of Theorem 3.1, we premultiply (3.33) by $\mathbf{T'W}$ and postmultiply by $\mathbf{W'T}$ [cf. (3.19) and (3.20)] to yield

$$\mathbf{T'A^2T} \leqslant_{\mathsf{L}} (\lambda_1+\lambda_r)\mathbf{T'AT} - \lambda_1\lambda_r\mathbf{T'P_AT} \tag{3.34a}$$

$$= \frac{(\lambda_1+\lambda_r)^2}{4\lambda_1\lambda_r}(\mathbf{T'AT})^2 - \mathbf{G}_1^2, \tag{3.34b}$$

where the symmetric matrix

$$\mathbf{G}_1 = \frac{\lambda_1+\lambda_r}{2\sqrt{\lambda_1\lambda_r}}\mathbf{T'AT} - \sqrt{\lambda_1\lambda_r}\,\mathbf{T'P_AT}; \tag{3.35}$$

this follows because under the idempotency condition (3.14),

$$\mathbf{T'ATT'P_AT} = \mathbf{T'P_ATT'AT} = \mathbf{T'AT}. \tag{3.36}$$

The inequality (3.24) then follows from (3.34), since $\mathbf{G}_1^2 \geqslant_{\mathsf{L}} \mathbf{0}$. Equality holds in (3.24) if and only if $\mathbf{G}_1 = \mathbf{0}$ and equality holds throughout (3.34); these two conditions are easily seen to be equivalent, respectively, to the two conditions in (3.27). See also (3.30).

To prove (3.25) we divide (3.34a) by $\lambda_1 + \lambda_r$ and add $(\mathbf{T'A^2T})^{1/2}$ to both sides to yield

$$(\mathbf{T'A^2T})^{1/2} - \mathbf{T'AT} \leqslant_{\mathsf{L}} (\mathbf{T'A^2T})^{1/2} - \frac{\lambda_1\lambda_r}{\lambda_1 + \lambda_r}\mathbf{T'P_AT} - \frac{1}{\lambda_1 + \lambda_r}\mathbf{T'A^2T} \tag{3.37a}$$

$$= \frac{(\lambda_1 - \lambda_r)^2}{4(\lambda_1 + \lambda_r)}\mathbf{T'P_AT} - \mathbf{G}_2^2, \tag{3.37b}$$

where the symmetric matrix

$$\mathbf{G}_2 = \left(\frac{1}{\lambda_1 + \lambda_r}\mathbf{T'A^2T}\right)^{1/2} - \tfrac{1}{2}(\lambda_1 + \lambda_r)^{1/2}\mathbf{T'P_AT}; \tag{3.38}$$

this follows because under the idempotency condition (3.14),

$$(\mathbf{T'A^2T})^{1/2}\mathbf{T'P_AT} = \mathbf{T'P_AT}(\mathbf{T'A^2T})^{1/2} = (\mathbf{T'A^2T})^{1/2}; \tag{3.39}$$

cf. (3.36). The inequality (3.25) then follows from (3.37), since $\mathbf{G}_2^2 \geqslant_{\mathsf{L}} \mathbf{0}$. Equality holds in (3.25) if and only if equality holds throughout (3.37) and $\mathbf{G}_2 = \mathbf{0}$; these two conditions are easily seen to be equivalent, respectively, to the two conditions in (3.28), see also (3.31).

To complete our proof of this theorem, we now establish (3.26) and (3.29). We start by rewriting (3.37a) as

$$\mathbf{T'A^2T} \leqslant_{\mathsf{L}} (\lambda_1 + \lambda_r)\mathbf{T'AT} - \lambda_1\lambda_r\mathbf{T'P_AT}. \tag{3.40}$$

Subtracting $(\mathbf{T'AT})^2$ from both sides of (3.40) and using (3.36) gives

$$\mathbf{T'A^2T} - (\mathbf{T'AT})^2 \leqslant_{\mathsf{L}} (\lambda_1 + \lambda_r)\mathbf{T'AT} - (\mathbf{T'AT})^2 - \lambda_1\lambda_r\mathbf{T'P_AT}$$

$$= \tfrac{1}{4}(\lambda_1 - \lambda_r)^2\mathbf{T'P_AT} - \mathbf{G}_3^2, \tag{3.41}$$

where the symmetric matrix

$$\mathbf{G}_3 = \mathbf{T}'\mathbf{A}\mathbf{T} - \tfrac{1}{2}(\lambda_1 + \lambda_r)\mathbf{T}'\mathbf{P}_{\mathbf{A}}\mathbf{T}. \tag{3.42}$$

The equality (3.26) then follows from (3.41), since $\mathbf{G}_3^2 \geqslant_{\mathsf{L}} \mathbf{0}$. Equality holds in (3.26) if and only if $\mathbf{G}_3 = \mathbf{0}$ and equality holds throughout (3.41); these two conditions are easily seen to be equivalent, respectively, to the two conditions in (3.29). See also (3.32). ∎

This research was conducted while the second author was a Senior Research Fellow of the Academy of Finland. The research of the third author was supported in part by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Hans Joachim Werner for very helpful discussions while completing this paper.

REFERENCES

Baksalary, J. K. and Kala, R. 1980. Two properties of a nonnegative definite matrix, *Bull. Acad. Polon. Sci. Sér. Sci. Math.* 28:233–235.

Baksalary, J. K. and Kala, R. 1983. Partial orderings between matrices one of which is of rank one, *Bull. Acad. Polon. Sci. Sér. Sci. Math.* 31:5–7.

Baksalary, J. K. and Puntanen, S. 1990. Characterizations of the best linear unbiased estimator in the general Gauss-Markov model with the use of matrix partial orderings, *Linear Algebra Appl.* 127:363–370.

Baksalary, J. K. and Puntanen, S. 1991a. Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities, *Aequationes Math.* 41:103–110.

Baksalary, J. K. and Puntanen, S. 1991b. A counterexample to a conjecture of Magness and McGuire, in *A Spectrum of Statistical Thought, Essays in Statistical Theory, Economics and Population Genetics in Honour of Johan Fellman* (G. Rosenqvist, K. Juselius, K. Nordström, and J. Palmgren, Eds.), Publ. Swedish School of Econom. and Business Admin. 46, Helsinki, pp. 15–18.

Baksalary, J. K. Schipp, B., and Trenkler, G. 1992. Some further results on Hermitian-matrix inequalities, *Linear Algebra Appl.* 160:119–129.

Baksalary, J. K. and Trenkler, G. 1991. Nonnegative and positive definiteness of matrices modified by two matrices of rank one. *Linear Algebra Appl.* 151:169–184.

Bancroft, D. 1994. Comment on Chaganty (1993) [Letter to the Editor], *Amer. Statist.* 48:351.

Bloomfield, P. and Watson, G. S. 1975. The inefficiency of least squares, *Biometrika* 62:121–128.

Chaganty, N. R. 1993. Comment on "A matrix formulation on how deviant an observation can be" by I. Olkin (1992) [Letter to the Editor], *Amer. Statist.* 47:158.

Chaganty, N. R. and Vaish, A. K. 1994. Response to Bancroft (1994) and Neudecker and Liu (1994) [Letter to the Editor], *Amer. Statist.* 48:351–352.

Chipman, J. S. 1964. On least squares with insufficient observations, *J. Amer. Statist. Assoc.* 59:1078–1111.

Chipman, J. S. 1968. Specification problems in regression analysis, in *Proceedings of the Symposium on Theory and Applications of Generalized Inverses of Matrices* (T. L. Boullion and P. L. Odell, Eds.), Texas Tech Press, Lubbock, pp. 114–176.

Chipman, J. S. 1976. Estimation and aggregation in econometrics: An application of the theory of generalized inverses, in *Generalized Inverses and Applications* (M. Z. Nashed, Ed.), Academic, New York, pp. 549–769.

Chollet, J. 1982. On principal submatrices, *Linear and Multilinear Algebra* 11:283–285.

Dey, A. and Gupta, S. C. 1977. Singular weighing designs and estimation of total weight, *Comm. Statist. A—Theory Methods* 6:289–295.

Dey, A., Hande, S., and Tiku, M. L. 1994. Statistical proofs of some matrix results, *Linear and Multilinear Algebra* 38:109–116.

Gaffke, N. and Krafft, O. 1977. Optimum properties of Latin square designs and a matrix inequality, *Math. Operationsforsch. Statist. Ser. Statist.* 8:345–350.

Greub, W. and Rheinboldt, W. 1959. On a generalization of an inequality of L. V. Kantorovich, *Proc. Amer. Math. Soc.* 10:407–415.

Horn, R. A. and Johnson, Charles R. 1994. *Topics in Matrix Analysis*, corrected reprint ed., Cambridge U.P.

Kantorovich, V. L. 1948. Functional analysis and applied mathematics (in Russian), *Uspekhi Mat. Nauk* 3:89–185; English transl. National Bureau of Standards Report 1509, Washington, 1952.

Magness, T. A. and McGuire, J. B. 1962. Comparison of least squares and minimum variance estimates of regression parameters, *Ann. Math. Statist.* 33:462–470.

Marcus, M. 1982. A remark of the preceding paper, *Linear and Multilinear Algebra* 11:287.

Marcus, M. and Minc, H. 1992. *A Survey of Matrix Theory and Matrix Inequalities*, corrected reprint ed., Dover, New York.

Marshall, A. W. and Olkin, I. 1964. Reversal of the Lyapunov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich inequality, *J. Math. Anal. Appl.* 8:503–514.

Marshall, A. W. and Olkin, I. 1979. *Inequalities: Theory of Majorization and Its Applications*, Academic, New York.

Marshall, A. W. and Olkin, I. 1990. Matrix versions of the Cauchy and Kantorovich inequalities, *Aequationes Math.* 40:89–93.

Mond, B. and Pečarić, J. E. 1993. Matrix versions of some means inequalities, *Austral. Math. Soc. Gaz.* 20:117–120.

Mond, B. and Pečarić, J. E. 1994. A matrix version of the Ky Fan generalization of the Kantorovich inequality, *Linear and Multilinear Algebra* 36:217–221.

Mond, B. and Shisha, O. 1970. Difference and ratio inequalities in Hilbert space, in *Inequalities*, Vol. 2 (O. Shisha, ed.), Academic, New York, pp. 241–249.

Neudecker, H. and Liu, S. 1994. Comment on Chaganty (1993) [Letter to the Editor], *Amer. Statist.* 48:351.

Olkin, I. 1992. A matrix formulation on how deviant an observation can be, *Amer. Statist.* 46:205–209.

Puntanen, S. 1987. On the relative goodness of ordinary least square estimation in the general linear model, *Acta Univ. Tamper. Ser.* A, Vol. 216.

Puntanen, S. and Styan, G. P. H. 1989. The equality of the ordinary least squares estimator and the best linear unbiased estimator (With discussion by O. Kempthorne & by S. R. Searle & with "Reply" by S. Puntanen and G. P. H. Styan), *Amer. Statist.* 43:153–164.

Rao, C. R. 1967. Least squares theory using an estimated dispersion matrix and its application to measurement of signals, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (L. M. Le Cam and J. Neyman, Eds.), Univ. of California Press, Berkeley, pp. 355–372.

Rao, C. R. 1985. The inefficiency of least squares: Extensions of the Kantorovich inequality, *Linear Algebra Appl.* 70:249–255.

Rao, C. R. and Mitra, S. K. 1971. *Generalized Inverse of Matrices and Its Applications*, Wiley, New York.

Styan, G. P. H. 1983. On some inequalities associated with ordinary least squares and the Kantorovich inequality, in *Festschrift for Eino Haikala on his Seventieth Birthday*, *Acta Univ. Tamper. Ser. A* 153:158–166.

Trenkler, G. 1994. Comment on "A matrix formulation on how deviant on observation can be" by I. Olkin (1992) [Letter to the Editor], *Amer. Statist.* 48:60–61.

Wang, S.-G. and Shao, J. 1992. Constrained Kantorovich inequalities and relative efficiency of least squares, *J. Multivariate Anal.* 42:284–298.

Watson, G. S. 1955. Serial correlation in regression analysis I, *Biometrika* 42:327–341.

Zyskind, G. 1967. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models, *Ann. Math. Statist.* 38:1092–1109.

Received 17 October 1994; final manuscript accepted 16 June 1995