Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 453-507

SOME COMMENTS ON SEVERAL MATRIX INEQUALITIES WITH APPLICATIONS TO CANONICAL CORRELATIONS: HISTORICAL BACKGROUND AND RECENT DEVELOPMENTS

By S.W. DRURY
McGill University, Montréal, Canada
SHUANGZHE LIU
Australian National University, Canberra, Australia
CHANG-YU LU
East China Normal University, Shanghai, China
SIMO PUNTANEN
University of Tampere, Tampere, Finland
and
GEORGE P.H. STYAN
McGill University, Montréal, Canada

SUMMARY. We review several matrix inequalities and give some statistical applications, with special emphasis on canonical correlations; many historical and biographical remarks are also included as well as over 100 references. Our paper builds upon the recent survey by Alpargu and Styan (2000) and concentrates on recent developments. We present a new Generalized Matrix Frucht–Kantorovich inequality and show that it is "essentially equivalent" to the Generalized Matrix Wielandt inequality given by Lu (1999), extending recent results by Wang and Ip (1999). We discuss an interesting special case involving block rank additivity of a partitioned matrix and offer several characterizations. We also consider the Krasnosel′skiĭ–Kreĭn inequality and the Shisha–Mond inequality and matrix extensions due to Khatri and Rao (1981, 1982) and Rao (1985).

Paper received July 2000; revised August 2000.
MSC (2000) subject classification. Primary 15A42; secondary 62H20, 62J05.
Keywords and phrases. Antieigenvalues, block rank additivity, Bloomfield–Watson–Knott inequality, Cassels inequality, determinantal inequalities, efficiency of ordinary least squares, Frucht–Kantorovich inequality, generalized correlation coefficient, Khatri–Rao inequality, Krasnosel′skiĭ–Kreĭn inequality, Kronecker sum, Lyapunov condition, majorization of eigenvalues, measures of association, Rao inequality, Shisha–Mond inequality, trace inequalities, vector coefficient of alienation, Watson efficiency, Wielandt inequality.

Some related inequalities involving determinants and traces are also presented. We establish some new inequalities and give a proof for an upper bound for the product of canonical correlations stated by Khatri (1982) and Khatri and Rao (1982). In addition, we present a new proof of the Bloomfield–Watson–Knott inequality; the Bloomfield–Watson–Knott, Khatri–Rao and Rao inequalities are identified as essential for exciting new results on majorization of eigenvalues due to Ando (2000, 2001) and Li and Mathias (1999).

1. The Frucht–Kantorovich Inequality

The well-known inequality

$$\frac{x'Ax \cdot x'A^{-1}x}{(x'x)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \tag{1.1}$$

is usually attributed to Kantorovich (1912–1986) for the inequality he established in Kantorovich (1948, pp. 142–144), Kantorovich (1952, pp. 106–107). In (1.1) x is a real nonnull $n \times 1$ vector and A is a real $n \times n$ symmetric positive definite matrix, with λ_1 and λ_n, respectively, its (fixed) largest and smallest necessarily positive eigenvalues. Throughout this paper we assume $n \geq 2$.

As observed by Watson et al. (1997), the inequality (1.1) was established already by Frucht (1943) and so we now call (1.1) the "Frucht–Kantorovich (FK) inequality".

Frucht (1943) built on the solution to the following problem: Given positive numbers $x_1, x_2, \ldots, x_n$ such that $\sum_{i=1}^{n} x_i = k$, establish the inequalities:

$$\sum_{i=1}^{n} x_i^2 \geq \frac{k^2}{n}, \qquad \sum_{i=1}^{n} \frac{1}{x_i} \geq \frac{n^2}{k}. \tag{1.2}$$

This problem (1.2) was posed (anonymously, but presumably by Beppo Levi) in 1942 and solved by Bender (1942) and by Saleme (1942). Both inequalities in (1.2) are easily proved using the well-known Cauchy–Schwarz (CS) inequality. Moreover the "reverse" (or complementary) inequality

$$1 \leq \frac{x'Ax \cdot x'A^{-1}x}{(x'x)^2} \tag{1.3}$$

also follows at once from the CS inequality.

The upper bound in (1.1), which we will denote by μ_n, see also (1.19) below, may be written in several different ways, e.g.,

$$\mu_n = \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} = \left(\frac{(\lambda_1+\lambda_n)/2}{\sqrt{\lambda_1\lambda_n}}\right)^2, \tag{1.4}$$

the square of the ratio of the arithmetic and geometric means of λ_1 and λ_n; this is the form used in Weisstein (1999). We will call this upper bound μ_n the "Kantorovich ratio" though Barnes and Hoffman (1994) and Wolkowicz (1981) call $1-(1/\mu_n)$ the Kantorovich ratio; we will call $1-(1/\mu_n)$ the "Wielandt ratio" and write $1-(1/\mu_n)=\nu_n^2$, see (2.2) below.

The ratio of the geometric and arithmetic means of λ_1 and λ_n,

$$\frac{1}{\sqrt{\mu_n}} = \frac{\sqrt{\lambda_1\lambda_n}}{(\lambda_1+\lambda_n)/2} = \tau_1, \tag{1.5}$$

say, is called the first "antieigenvalue" of A, see Gustafson and Rao (1997, ch. 3) Applications of antieigenvalues to statistics have recently received attention, see, e.g., Khattree (2001).

Other ways of expressing μ_n include

$$\begin{aligned}\mu_n = \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} &= \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}}+\sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2 = \frac{1}{2}(\lambda_1+\lambda_n)\cdot\frac{1}{2}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_n}\right), \\ &= \left(\sqrt{\lambda_1\lambda_n}\bigg/\frac{2}{\frac{1}{\lambda_1}+\frac{1}{\lambda_n}}\right)^2 = \frac{1}{1-\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2}. \end{aligned} \tag{1.6}$$

Probably the simplest special case of (1.1) is the Schweitzer inequality

$$\frac{1}{n}(\lambda_1+\cdots+\lambda_n)\cdot\frac{1}{n}\left(\frac{1}{\lambda_1}+\cdots+\frac{1}{\lambda_n}\right) \le \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.7}$$

established by Schweitzer (1914). It follows at once from (1.1) by substituting $x=e$, the column vector of ones and $A=\text{diag}\{\lambda_i\}$. Watson et al. (1997) show that (1.7) also implies (1.1) and so these two inequalities are "essentially equivalent".

The first statistical application of the FK inequality seems to be Watson (1951, 1955), Watson and Hannan (1956), Hannan (1960, pp. 111–113), Magness and McGuire (1962) and Golub (1963). These authors studied the efficiency ϕ of the ordinary least-squares estimator in the linear model with one regressor and showed that

$$\phi = \frac{(x'x)^2}{x'Ax\cdot x'A^{-1}x} \ge \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} = \tau_1^2. \tag{1.8}$$

Here the vector x comprises the values of the regressor and A is the positive definite error dispersion matrix. It is clear that (1.8) is just the FK inequality written the other way round.

It follows at once from (1.8), and also from (1.1), that

$$\frac{x'Ax}{x'x} \leq \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\,\frac{x'x}{x'A^{-1}x}, \tag{1.9}$$

which, as we shall see, is more convenient for our matrix extensions. Indeed Marshall and Olkin (1990) generalized the FK inequality by showing that in the Löwner (partial) ordering

$$U^*AU \leq_{\mathsf{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(U^*A^{-1}U)^{-1}, \tag{1.10}$$

where the complex $n\times n$ matrix A is Hermitian positive definite with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, while the $n \times p$ complex matrix U satisfies $U^*U = I_p$. We will call (1.10) the "Marshall–Olkin Matrix FK (MOMFK) inequality". It is easy to see that A and A^{-1} in (1.10) may be interchanged and so we also find that

$$U^*A^{-1}U \leq_{\mathsf{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(U^*AU)^{-1}.$$

If we substitute $U = X(X^*X)^{-1/2}$ in (1.10), where X is any complex $n\times p$ matrix with full column rank p, then (1.10) becomes:

$$\begin{aligned}&(X^*X)^{-1/2}X^*AX(X^*X)^{-1/2}\\ &\quad\leq_{\mathsf{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*X)^{1/2}(X^*A^{-1}X)^{-1}(X^*X)^{1/2},\end{aligned}$$

or equivalently

$$X^*AX \leq_{\mathsf{L}} \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}X^*X(X^*A^{-1}X)^{-1}X^*X\,. \tag{1.11}$$

The reverse inequality

$$X^*X(X^*A^{-1}X)^{-1}X^*X \leq_{\mathsf{L}} X^*AX \tag{1.12}$$

is a special case of

$$\sum_{j=1}^k \alpha_j X_j^*X_j\Big(\sum_{j=1}^k \alpha_j X_j^*A_j^{-1}X_j\Big)^{-1}\sum_{j=1}^k \alpha_j X_j^*X_j \leq_{\mathsf{L}} \sum_{j=1}^k \alpha_j X_j^*A_jX_j, \tag{1.13}$$

apparently first established by Kiefer (1959, Lemma 3.2); see also Rao (1967, Lemma 2c), Gaffke and Krafft (1977, Lemma 2.1), and Nakamoto and Takahashi (1999, Theorem 5). In (1.13) the matrices X_j are all $n \times p$ with full column rank p and the A_j are all $n \times n$ positive definite; the scalars α_j are all positive with $\sum_{j=1}^{k} \alpha_j = 1$.

Let us partition

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}, \tag{1.14}$$

where A_{12} and A^{12} are both $p \times (n-p)$. We will use this partitioning of A and of A^{-1} throughout this paper.

If we substitute $X = (I_p : 0)'$ in (1.11) and (1.12) then together they become

$$(A^{11})^{-1} \leq_{\mathsf{L}} A_{11} \leq_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(A^{11})^{-1} \tag{1.15}$$

or equivalently

$$A_{11}^{-1} \leq_{\mathsf{L}} A^{11} \leq_{\mathsf{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} A_{11}^{-1}. \tag{1.16}$$

We now generalize the MOMFK inequality in the following theorem:

THEOREM 1. *Let the complex Hermitian $n \times n$ matrix A be nonnull and nonnegative definite with rank $r \leq n$ and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_r > 0$, and let X be any $n \times p$ complex matrix with $n \lesseqgtr p$. Then*

$$X^*AX \leq_{\mathsf{L}} \mu_r X^* P_A X (X^* A^+ X)^- X^* P_A X \tag{1.17}$$

for any generalized inverse $(X^*A^+X)^-$. Here $P_A = AA^+$ is the orthogonal projector on the column space (range) of A, with A^+ the Moore–Penrose inverse of A, and

$$\mu_r = \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} \quad \textit{for} \;\; r \geq 2 \qquad \textit{and} \quad \mu_1 = 1. \tag{1.18}$$

We will call (1.17) the "Generalized Matrix FK (GMFK) inequality". We note that the right-hand side of (1.17) is invariant for all choices of generalized inverses $(X^*A^+X)^-$, see, e.g., Rao and Mitra (1971, Lemma 2.2.4 (iii), p. 21). We will call μ_r the "generalized" Kantorovich ratio and note that the "generalized" antieigenvalue $\tau_1 = 1/\sqrt{\mu_r}$.

Baksalary and Puntanen (1991) showed that

$$Z^*AZ \leq_{\mathsf{L}} \mu_r (Z^*A^+Z)^+$$

when the complex $n \times p$ matrix Z is such that Z^*P_AZ is idempotent.

The reverse inequality

$$X^*P_AX(X^*A^+X)^-X^*P_AX \leq_{\mathsf{L}} X^*AX$$

is due to Baksalary and Puntanen (1991); see also Pečarić et al. (1996, Corollary 2.1).

Proof of Theorem 1. To prove the GMFK inequality, we write $A = Q\Lambda Q^*$, where Q is $n \times r$ such that $Q^*Q = I_r$ and $\Lambda = \{\lambda_i\}$ is $r \times r$ diagonal, and so (1.17) becomes

$$X^*Q\Lambda Q^*X \leq_{\mathsf{L}} \mu_r X^*QQ^*X(X^*Q\Lambda^{-1}Q^*X)^-X^*QQ^*X, \tag{1.19}$$

since $P_A = QQ^*$. We now substitute in (1.19) the normalized full-rank decomposition $Q^*X = KL^*$, where K and L have full column rank $s = \operatorname{rank}(Q^*X)$ and K is normalized so that $K^*K = I_s$, to obtain

$$LK^*\Lambda KL^* \leq_{\mathsf{L}} \mu_r LK^*KL^*(LK^*\Lambda^{-1}KL^*)^-LK^*KL^* \tag{1.20}$$

and hence $K^*\Lambda K \leq_{\mathsf{L}} \mu_r L^*(LK^*\Lambda^{-1}KL^*)^-L$ since L has full column rank and $K^*K = I_s$. But $L^*(LK^*\Lambda^{-1}KL^*)^-L = (K^*\Lambda^{-1}K)^{-1}$ since the $s \times s$ matrix $E = L^*(LK^*\Lambda^{-1}KL^*)^-LK^*\Lambda^{-1}K$ is idempotent with rank $s = \operatorname{rank}(LK^*)$ and so $E = I_s$. Hence

$$K^*\Lambda K \leq_{\mathsf{L}} \mu_r (K^*\Lambda^{-1}K)^{-1},$$

which is the MOMFK inequality, and our proof of (1.17) is complete. □

2. The Wielandt Inequality

The inequality (1.17) allows us to offer a quick proof of the Generalized Matrix Wielandt (GMW) inequality, extending a recent result by Wang and Ip (1999). It seems that the Wielandt inequality, (2.2) below, was introduced by Bauer and Householder (1960), who showed that for any two nonnull real vectors x and y and real positive definite matrix A, the inequalities

$$\frac{(x'y)^2}{x'x \cdot y'y} \leq \cos^2\phi \quad \text{and} \quad 0 \leq \phi \leq \frac{\pi}{2} \quad \Longrightarrow \quad \frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \cos^2\theta, \tag{2.1}$$

where $\cot^2(\theta/2) = \psi \cot^2(\phi/2)$ and the condition number $\psi = \lambda_1/\lambda_n$ with, as earlier, $\lambda_1 \geq \cdots \geq \lambda_n$ the necessarily positive eigenvalues of A.

When $\phi = \pi/2$ the vectors x and y must be orthogonal: $x'y = 0$ and then, since now $\psi = \cot^2(\theta/2)$,

$$\frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \left(\frac{\psi - 1}{\psi + 1}\right)^2 = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 = \nu_n^2, \tag{2.2}$$

say. We will refer to (2.2) as the "Wielandt inequality"[1], since Bauer and Householder (1960) credit (2.2) to Helmut Wielandt "and also private communication" [with Wielandt]. We have found it difficult to deduce (2.2) from the results in Wielandt (1953). We will call the upper bound in (2.2) the "Wielandt ratio" though Barnes and Hoffman (1994) and Wolkowicz (1981) refer to it as the "Kantorovich ratio", though it is different from (1.4) above. For the geometric meaning of the Wielandt inequality, see Gustafson (1999).

It is clear that both (2.1) and (2.2) strengthen the CS inequality since

$$\frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \cos^2\theta \leq 1 \quad \text{and} \quad \frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \leq 1.$$

The first appearance of (2.2) in a statistical context seems to be by Eaton (1976). Let the random vector h have dispersion matrix A; then the maximum of the squared correlation

$$\max_{x,y:\, x'y=0} \operatorname{corr}^2(x'h, y'h) = \max_{x,y:\, x'y=0} \frac{(x'Ay)^2}{x'Ax \cdot y'Ay} = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2,$$

see (5.35) in §5 below.

In his seminal book Householder (1975, p. 83), seems to be the first person to observe that when

$$y = A^{-1}x - \left(\frac{x'A^{-1}x}{x'x}\right)x = \left(I - \frac{xx'}{x'x}\right)A^{-1}x \tag{2.3}$$

then the Wielandt inequality (2.2) becomes the FK inequality.

Very recently, Bhatia (1999), Bhatia and Davis (2000) and Zhang (2001) have shown that the FK inequality also implies the Wielandt inequality and so these two inequalities are essentially equivalent; see also Bhatia and Davis (2001).

[1]This is not the only inequality so called: we found 11 references in *Mathematical Reviews* to the "Wielandt inequality" but only two of these are about (2.2).

We have already seen that the FK inequality is essentially equivalent to the Schweitzer inequality. Watson et al. (1997) showed that the FK inequality is also essentially equivalent to the Pólya–Szegő inequality, the Greub–Rheinboldt inequality, the Cassels inequality (Cassels, 1951, 1955) and the Krasnosel′skiĭ–Kreĭn (KK) inequality (Krasnosel′skiĭ and Kreĭn, 1952).

Pólya and Szegő (1925) showed that

$$\frac{\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2}{(\sum_{i=1}^n a_i b_i)^2} \le \frac{(ab+AB)^2}{4abAB}, \tag{2.4}$$

where $0 < a \le a_i \le A, 0 < b \le b_i \le B \quad (i = 1, \ldots, n)$. Greub and Rheinboldt (1959) obtained this "weighted" version of the Pólya–Szegő inequality (2.4):

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \le \frac{(ab+AB)^2}{4abAB},$$

where the a_i and b_i (and a, b, A, B) are as in (2.4) and the $w_i > 0$ $(i = 1, \ldots, n)$.

The Cassels inequality may be expressed as:

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \le \frac{(m+M)^2}{4mM},$$

where $a_i > 0,\ b_i > 0$ and $w_i > 0$ $(i = 1, \ldots, n)$, with

$$m = \min_i \frac{a_i}{b_i} \quad \text{and} \quad M = \max_i \frac{a_i}{b_i}.$$

The KK inequality, see also (6.5) in §6 below, is

$$\frac{w'A^2w \cdot w'w}{(w'Aw)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{2.5}$$

which follows at once from the FK inequality by substituting $x = A^{1/2}w$. The KK inequality (2.5) may also be written as

$$\frac{w'Aw}{\sqrt{w'A^2w \cdot w'w}} \ge \frac{\sqrt{\lambda_1\lambda_n}}{(\lambda_1 + \lambda_n)/2} = \tau_1, \tag{2.6}$$

the first antieigenvalue as in (1.5). It follows at once from (2.6) that

$$\tau_1 = \min_w \operatorname{corr}(w'Au, w'u),$$

i.e., τ_1 is the minimum correlation between $w'Au$ and $w'u$ over all nonrandom vectors w, where the random vector u has dispersion matrix I; see also Khattree (2001).

We will establish that the FK inequality and the Wielandt inequality are essentially equivalent by showing that our GMFK inequality is equivalent to the "GMW inequality" given by Lu (1999), extending results of Wang and Ip (1999).

2.1 *The generalized matrix Wielandt inequality and the generalized matrix Frucht–Kantorovich inequality.* Wang and Ip (1999) generalized the Wielandt inequality as follows: Let A be an $n \times n$ complex positive definite Hermitian matrix, and let X and Y be complex $n \times p$ and $n \times q$ matrices, respectively. If $X^*Y = 0$, then for all generalized inverses $(Y^*AY)^-$

$$X^*AY(Y^*AY)^-Y^*AX \leq_{\mathsf{L}} \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 X^*AX \tag{2.7}$$

in the Löwner partial ordering, where λ_1 and λ_n are the largest and smallest, necessarily real and positive, eigenvalues of A.

A very interesting and important special case of (2.7) occurs when $n = p + q$ and

$$X = \begin{pmatrix} I_p \\ 0 \end{pmatrix} \qquad \text{and} \qquad Y = \begin{pmatrix} 0 \\ I_{n-p} \end{pmatrix}. \tag{2.8}$$

If we partition A as in (1.14), then

$$X^*AX = A_{11}, \;\; X^*AY = A_{12}, \;\; Y^*AX = A_{21}, \;\; Y^*AY = A_{22}, \tag{2.9}$$

and (2.7) becomes

$$A_{12}A_{22}^{-1}A_{21} \leq_{\mathsf{L}} \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 A_{11}.$$

Lu (1999) extended the inequality (2.7) to

$$X^*AY(Y^*AY)^-Y^*AX \leq_{\mathsf{L}} \nu_r^2 X^*AX, \tag{2.10}$$

where

$$\nu_r = \frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r} \quad \text{for } r \geq 2 \qquad \text{and} \quad \nu_1 = 0. \tag{2.11}$$

The $n \times n$ matrix A is now nonnegative definite with rank $r \leq n$ and $X^* P_A Y = 0$; when A is positive definite, then the projector $P_A = AA^+ = I$ and $X^* P_A Y = X^* Y = 0$. In (2.11) λ_r is the smallest nonzero (positive) eigenvalue of A. We call (2.10) the GMW inequality. It follows at once from (2.11) and (1.19) that the "generalized Wielandt ratio"

$$\left(\frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r}\right)^2 = \nu_r^2 = 1 - \frac{1}{\mu_r} \quad \text{for } r \geq 1, \tag{2.12}$$

where μ_r is our "generalized Kantorovich ratio".

To prove the GMW inequality, we will use

LEMMA 1. *Let A be an $n \times n$ complex nonnull nonnegative definite matrix and let the complex matrices X and Y be $n \times p$ and $n \times q$, respectively. If*

$$X^* P_A Y = 0, \tag{2.13}$$

where $P_A = AA^+$ is the orthogonal projector on the column space of A and A^+ is the Moore–Penrose inverse of A, then

$$A - P_A X (X^* A^+ X)^- X^* P_A \geq_{\mathsf{L}} AY(Y^* AY)^- Y^* A, \tag{2.14}$$

and hence

$$X^* AX - X^* P_A X (X^* A^+ X)^- X^* P_A X \geq_{\mathsf{L}} X^* AY(Y^* AY)^- Y^* AX \tag{2.15}$$

for any choices of generalized inverses $(X^* A^+ X)^-$ and $(Y^* AY)^-$.

Equality holds in (2.14) *if and only if*

$$\operatorname{rank}(A) = \operatorname{rank}(AX) + \operatorname{rank}(AY) \tag{2.16}$$

and then equality holds in (2.15).

PROOF. It suffices to notice that

$$E = P_A - (A^+)^{1/2} X (X^* A^+ X)^- X^* (A^+)^{1/2} - A^{1/2} Y (Y^* AY)^- Y^* A^{1/2} \tag{2.17}$$

is Hermitian idempotent and hence nonnegative definite. To see that $E = E^2$ we note that all the matrices in (2.17) are idempotent and the two products of the last two matrices are both equal to 0. □

The result that equality holds in (2.14) if and only if (2.16) holds also follows at once from the rank equality

$$\operatorname{rank}(A - P_A X(X^* A^+ X)^- X^* P_A - AY(Y^* AY)^- Y^* A) \\ = \operatorname{rank}(A) - \operatorname{rank}(AX) - \operatorname{rank}(AY),$$

due to Tian (2000). We observe that the orthogonality condition $X^* P_A Y = 0$ in (2.13) implies that

$$\operatorname{rank}(A) \geq \operatorname{rank}(AX) + \operatorname{rank}(AY) \tag{2.18}$$

from the Frobenius rank inequality, see e.g., Rao and Rao (1998, p. 134). And so if equality holds, as in (2.16), then the column space of the matrix $P_A Y$ coincides with the null (column) space of $X^* P_A$.

When the $n \times (p+q)$ matrix $(X : Y)$ has full row rank n and $X^* Y = 0$ so that the column spaces of X and Y are orthogonal and together span the "whole space" $\mathbb{C}_n$, then, as we will prove in Lemma 2 below, we have

$$\operatorname{rank}(X^* P_A Y) = \operatorname{rank}(AX) + \operatorname{rank}(AY) - \operatorname{rank}(A) \tag{2.19}$$

and so then we have $X^* P_A Y = 0$ if and only if rank is "additive" in the sense that $\operatorname{rank}(A) = \operatorname{rank}(AX) + \operatorname{rank}(AY)$, i. e., (2.13) $\Leftrightarrow$ (2.16). To prove (2.19) (Puntanen et al., 2001, show this in a slightly more general setting), we use the following lemma due to Puntanen (1985, p. 12; 1987, Th. 3.4.1, p. 34); see also Baksalary and Styan (1993) and Tian and Styan (2001):

LEMMA 2. *Let the complex matrices F and G be $n \times p$ and $n \times q$, respectively, and let $P_F = FF^+$ and $P_G = GG^+$ denote the corresponding orthogonal projectors, with $Q_F = I_n - P_F$ and $Q_G = I_n - P_G$. Then*

$$\operatorname{rank}(P_F P_G Q_F) = \operatorname{rank}(P_F P_G) + \operatorname{rank}(P_F : P_G) - \operatorname{rank}(F) - \operatorname{rank}(G) \tag{2.20}$$

$$= \operatorname{rank}(P_F P_G) + \operatorname{rank}(Q_F P_G) - \operatorname{rank}(G) \tag{2.21}$$

$$= \operatorname{rank}(P_G P_F) + \operatorname{rank}(Q_G P_F) - \operatorname{rank}(F) = \operatorname{rank}(P_G P_F Q_G). \tag{2.22}$$

If we put $F = X$ and $G = A$ then the left-hand side of (2.20) becomes the left-hand side of (2.19), i. e., $\operatorname{rank}(X^* P_A Y)$, since with $(X : Y)$ of full row rank we have $Q_X = P_Y$, and (2.21) becomes the right-hand side of (2.19), since $\operatorname{rank}(P_Z P_A) = \operatorname{rank}((P_Z P_A)^*) = \operatorname{rank}(P_A P_Z) = \operatorname{rank}(AZ)$ with $Z = X$ or Y. We notice that in Lemma 2 $\operatorname{rank}(P_F P_G Q_F) = \operatorname{rank}(P_G P_F Q_G)$ and so we can interchange F and G.

PROOF OF LEMMA 2. We will make repeated use of the formula

$$\operatorname{rank}(K : L) = \operatorname{rank}(K) + \operatorname{rank}(Q_K L) = \operatorname{rank}(L : K), \tag{2.23}$$

established by Marsaglia and Styan (1974a, Th. 5, p. 274). We have

$$\begin{aligned}
\operatorname{rank}(P_F P_G Q_F) &= \operatorname{rank}((P_F P_G Q_F)^*) = \operatorname{rank}(Q_F P_G P_F) && (2.24)\\
&= \operatorname{rank}(P_F : P_G P_F) - \operatorname{rank}(F) && (2.25)\\
&= \operatorname{rank}(Q_G P_F : P_G P_F) - \operatorname{rank}(F) && (2.26)\\
&= \operatorname{rank}(Q_G P_F) + \operatorname{rank}(P_G P_F) - \operatorname{rank}(F) && (2.27)\\
&= \operatorname{rank}(P_F : P_G) - \operatorname{rank}(G) + \operatorname{rank}(P_G P_F) - \operatorname{rank}(F),
\end{aligned}$$

which proves (2.20). To go from (2.25) to (2.26), we have used the elementary property that $\operatorname{rank}(P_F : P_G P_F) = \operatorname{rank}(P_F - P_G P_F : P_G P_F) = \operatorname{rank}(Q_G P_F : P_G P_F)$, while to go from (2.26) to (2.27), we use the virtual disjointness of the column spaces of $Q_G P_F$ and $P_G P_F$. Otherwise we have used only (2.23), which we may also use to deduce (2.21) from the right-hand side of (2.20); since the right-hand side of (2.20) is "symmetric" in F and G, the equalities (2.22) follow at once, using also $\operatorname{rank}(P_F P_G) = \operatorname{rank}((P_F P_G)^*) = \operatorname{rank}(P_G P_F)$. □

We now have the following theorem:

THEOREM 2. *The GMFK inequality*

$$X^*AX \leq_{\mathsf{L}} \mu_r X^* P_A X (X^* A^+ X)^- X^* P_A X, \tag{2.28}$$

and the GMW inequality

$$X^*AY(Y^*AY)^- Y^*AX \leq_{\mathsf{L}} \nu_r^2 X^*AX \qquad \textit{with} \quad X^* P_A Y = 0, \tag{2.29}$$

are essentially equivalent.

PROOF. To prove that the GMFK inequality implies (2.29), we see from (2.15) and (2.28) that

$$\begin{aligned}
X^*AY(Y^*AY)^- Y^*AX &\leq_{\mathsf{L}} X^*AX - X^* P_A X (X^* A^+ X)^- X^* P_A X \\
&\leq_{\mathsf{L}} X^*AX - (1/\mu_r) X^*AX = \nu_r^2 X^*AX, \quad (2.30)
\end{aligned}$$

since $\nu_r^2 = 1 - (1/\mu_r)$ and our proof that (2.28) $\Rightarrow$ (2.29) is complete.

To prove that (2.29) also implies the GMFK inequality, and so they are essentially equivalent, we choose Y so that rank additivity (i.e., equality) holds in (2.14). One such choice of Y is the orthogonal projector on the null space of X^*P_A

$$Y_0 = P_A - P_AX(X^*P_AX)^-X^*P_A. \tag{2.31}$$

Clearly $X^*P_AY_0 = 0$ and so

$$\begin{aligned}\nu_r^2 X^*AX &\geq_{\mathsf{L}} X^*AY_0(Y_0^*AY_0)^-Y_0^*AX \\ &= X^*AX - X^*P_AX(X^*A^+X)^-X^*P_AX\,;\end{aligned}$$

hence $X^*P_AX(X^*A^+X)^-X^*P_AX \geq_{\mathsf{L}} (1/\mu_r)X^*AX$ and our proof is complete. □

Another choice of Y is

$$Y_1 = Y_0A^+X = (P_A - P_{P_AX})A^+X = (P_A - P_AX(X^*P_AX)^-X^*P_A)A^+X. \tag{2.32}$$

When $X = x$ and $Y = y$ are vectors and A is positive definite, then (2.32) becomes (2.3), which Householder (1975, p.83) used to show that the Wielandt inequality implies the FK inequality. With this choice of $Y = Y_1$ we do not usually have rank additivity (i.e., equality) in (2.14), but the GMW inequality does still become the GMFK inequality. To prove this with Y_1, however, seems to be more complicated than with Y_0. We write

$$F = X^*AX, \quad G = X^*A^+X, \quad H = X^*P_AX$$

and note that $HH^-F = F = FH^-H$and $HH^-G = G = GH^-H$. We wish to prove that $K = W$, where $K = X^*AX - X^*P_AX(X^*A^+X)^-X^*P_AX = F - HG^-H$ and $W = X^*AY_1(Y_1^*AY_1)^-Y_1^*AX$. It clearly suffices to show that $F - W = HG^-H$. After some algebraic manipulation, we find that $W = (H - FH^-G)(GH^-FH^-G - G)^-(H - GH^-F)$ and so $GH^-WH^-G = GH^-FH^-G = G$. Hence $G(H^-(F - W)H^-)G = G$ and thus $H^-(F - W)H^- = G^-$. Since $HH^-F = F$ and $HH^-W = W$ we have $HG^-H = H(H^-(F - W)H^-)H = F - W$, and so our proof that the GMW inequality with the choice $Y = Y_1$ implies the GMFK inequality is complete. □

2.2 *Block rank additivity.* When X and Y are as in (2.8) and A is partitioned as in (1.14), then

$$X^*AX = A_{11}, \quad X^*AY = A_{12}, \quad Y^*AX = A_{21}, \quad Y^*AY = A_{22}, \tag{2.33}$$

see also (2.9), and the rank additivity condition (2.16) becomes the "block rank additivity" condition

$$\mathrm{rank}(A) = \mathrm{rank}(A_{11}) + \mathrm{rank}(A_{22}). \tag{2.34}$$

The GMW inequality now becomes

$$A_{12}A_{22}^{-}A_{21} \leq_{\mathsf{L}} \nu_r^2 A_{11}, \tag{2.35}$$

which holds provided the upper right $p \times (n-p)$ submatrix

$$(AA^{+})_{12} = X^{*}AA^{+}Y = X^{*}P_{A}Y = 0. \tag{2.36}$$

Marsaglia and Styan (1974b) showed that the orthogonality condition (2.36) holds if the block rank additivity condition (2.34) holds. The GMFK inequality now becomes

$$A_{11} \leq_{\mathsf{L}} \mu_r \left((P_A)_{11} \left((A^{+})_{11} \right)^{-} (P_A)_{11} \right), \tag{2.37}$$

where $(\cdot)_{11}$ denotes the upper left $p \times p$ submatrix. We believe the inequality (2.37) to be new.

The GMW inequality, together with (2.37), leads to the following theorem:

THEOREM 3. *Let A be an $n \times n$ complex nonnull nonnegative definite matrix with rank r partitioned as follows:*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{12} is $p \times (n-p)$, and let $P_A = AA^{+}$. Let the generalized Schur complement

$$A_{11.2} = A_{11} - A_{12}A_{22}^{-}A_{21} \tag{2.38}$$

and let

$$\nu_r = \frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r} \quad \textit{for } r \geq 2 \qquad \textit{and} \quad \nu_1 = 0, \qquad \textit{with} \quad \mu_r = \frac{1}{1-\nu_r^2};\ r \geq 1,$$

where $\lambda_1 \geq \cdots \geq \lambda_r$ are the nonzero, necessarily positive, r largest eigenvalues of A.

Then the following nine conditions are equivalent:

(a) $$\operatorname{rank}(A) = \operatorname{rank}(A_{11}) + \operatorname{rank}(A_{22})$$

(b) $$\operatorname{rank}(A_{11}) = \operatorname{rank}(A_{11} - A_{12}A_{22}^{-}A_{21})$$

(c) $$(P_A)_{12} = (AA^{+})_{12} = 0$$

(d) $$A_{12}A_{22}^{-}A_{21} \leq_{\mathsf{L}} \nu_r^2 A_{11}$$

(e) $$A_{11} \leq_{\mathsf{L}} \mu_r(A_{11} - A_{12}A_{22}^{-}A_{21})$$

(f) $$(P_A)_{11}((A^{+})_{11})^{-}(P_A)_{11} = A_{11} - A_{12}A_{22}^{-}A_{21}$$

(g) $$A^{+} = \begin{pmatrix} A_{11}^{+} + A_{11}^{+}A_{12}A_{22.1}^{+}A_{21}A_{11}^{+} & -A_{11}^{+}A_{12}A_{22.1}^{+} \\ -A_{22.1}^{+}A_{21}A_{11}^{+} & A_{22.1}^{+} \end{pmatrix}$$

(h) $$A^{+} = \begin{pmatrix} A_{11.2}^{+} & -A_{11.2}^{+}A_{12}A_{22}^{+} \\ -A_{22}^{+}A_{21}A_{11.2}^{+} & A_{22}^{+} + A_{22}^{+}A_{21}A_{11.2}^{+}A_{12}A_{22}^{+} \end{pmatrix}$$

(i) $$\operatorname{rank}(A^{+}) = \operatorname{rank}((A^{+})_{11}) + \operatorname{rank}((A^{+})_{22}).$$

PROOF. That (a) $\Leftrightarrow$ (b) is well known from rank properties of the Schur complement and that (a) $\Leftrightarrow$ (c) follows from Lemma 2. Marsaglia and Styan (1974b) showed that (a) $\Leftrightarrow$ (g) and (a) $\Leftrightarrow$ (h) and that (a) $\Rightarrow$ (f). The GMW inequality shows that (c) $\Rightarrow$ (d) and that (d) $\Leftrightarrow$ (e) follows at once from $\mu_r = 1 - \nu_r^2$. The GMFK inequality (2.37) shows that (f) $\Rightarrow$ (e). Since A and A^{+} are interchangeable in (c), it follows from Lemma 2 that (c) $\Leftrightarrow$ (i). [We are most grateful to Jürgen Groß for drawing our attention to the fact that block rank additivity of A and A^{+} are equivalent.]

To complete the proof it suffices to prove that (e)$\Rightarrow$(b). To do this, we note that when (e) holds then

$$A_{11} \geq_{\mathsf{L}} A_{11} - A_{12}A_{22}^{-}A_{21} \geq_{\mathsf{L}} \mu_r A_{11}$$

and so

$$\operatorname{rank}(A_{11}) \geq \operatorname{rank}(A_{11} - A_{12}A_{22}^{-}A_{21}) \geq \operatorname{rank}(\mu_r A_{11}) = \operatorname{rank}(A_{11}),$$

since $\mu_r > 0$. Thus $\operatorname{rank}(A_{11}) = \operatorname{rank}(A_{11} - A_{12}A_{22}^{-}A_{21})$ which is (b), and our proof is complete. □

Building on a result of Marsaglia and Styan (1974b), Baksalary and Kala (1980) proved that when the matrix A is nonnegative definite and the block rank additivity condition

$$\operatorname{rank}(A) = \operatorname{rank}(A_{11}) + \operatorname{rank}(A_{22}) \tag{2.39}$$

holds, as in (a) both in Theorem 3 and in Corollary 1, then

$$(A^+)_{11} \geq_{\mathsf{L}} (A_{11})^+ \quad \text{and} \quad (A^+)_{22} \geq_{\mathsf{L}} (A_{22})^+ \tag{2.40}$$

where $(A^+)_{11}$ is the top-left and $(A^+)_{22}$ is the bottom-right block of the Moore–Penrose inverse A^+. (An explicit formula for the complete partitioned matrix A^+ is obtained by Groß, 2000.)

Baksalary and Kala (1980) conjectured that (2.40) also implies the block rank additivity condition (2.39); this conjecture has, however, been proven false by Virtanen (2000) with the 4×4 matrix

$$A = \begin{pmatrix} 122 & 59 & 139 & 202 \\ 59 & 30 & 60 & 89 \\ 139 & 60 & 194 & 273 \\ 202 & 89 & 273 & 386 \end{pmatrix} = A_0,$$

say. It is easy to see that A_0 has rank 3, since the sum of rows one and three equals the sum of rows two and four; moreover A_0 is nonnegative definite, and thus positive semi-definite singular. We partition A_0 so that its top-left principal submatrix $A_{11}^{(0)}$ is 2×2, and so $p = n - p = m = 2 = h = r_1 = r_2$. Thus block rank additivity does not hold since $r_1 + r_2 = 4 > 3 = r$. Then, to one decimal place

$$(A_0^+)_{11} - (A_{11}^{(0)})^+ = \begin{pmatrix} 30.6 & -56.5 \\ -56.5 & 104.4 \end{pmatrix} \;\&\; (A_0^+)_{22} - (A_{22}^{(0)})^+ = \begin{pmatrix} 90.5 & -58.9 \\ -58.9 & 38.3 \end{pmatrix}$$

are both positive definite and so (2.40) holds for $A = A_0$.

Though (2.40) therefore does not imply (2.39), it would be interesting to find a "nice" additional condition under which (2.40) $\Rightarrow$ (2.39).

When the matrix A is positive definite, then block rank additivity (a) holds and Theorem 3 becomes the following corollary:

COROLLARY 1. *Let A be an $n \times n$ complex positive definite matrix partitioned as follows:*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{12} is $p \times (n-p)$ and let the Schur complements $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Then the following hold:

(a) $$\operatorname{rank}(A) = \operatorname{rank}(A_{11}) + \operatorname{rank}(A_{22})$$

(b) $$\operatorname{rank}(A_{11}) = \operatorname{rank}(A_{11} - A_{12}A_{22}^{-1}A_{21})$$

(c) $$A_{12}A_{22}^{-1}A_{21} \leq_{\mathsf{L}} \nu_n^2 A_{11}$$

(d) $$A_{11} \leq_{\mathsf{L}} \mu_n(A_{11} - A_{12}A_{22}^{-1}A_{21})$$

(e) $$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix}$$

(f) $$A^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

(g) $$\operatorname{rank}(A^{-1}) = \operatorname{rank}((A^{-1})_{11}) + \operatorname{rank}((A^{-1})_{22}).$$

3. The Shisha–Mond and Rao Inequalities

An inequality which does not seem to be essentially equivalent to the FK inequality is

$$\frac{x'Ax}{x'x} - \frac{x'x}{x'A^{-1}x} \leq \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2, \tag{3.1}$$

established by Shisha and Mond (1967), and so we will call (3.1) the "Shisha–Mond (SM) inequality". Rao (1985) extended this by showing that

$$\begin{aligned} \operatorname{tr}\Big((X'X)^{-1/2}X'AX(X'X)^{-1/2} - (X'X)^{1/2}(X'A^{-1}X)^{-1}(X'X)^{1/2}\Big) \\ = \operatorname{tr}\Big(X'AX(X'X)^{-1} - X'X(X'A^{-1}X)^{-1}\Big) \\ \leq \sum_{i=1}^{m}\left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2, \end{aligned} \tag{3.2}$$

where the $n \times p$ matrix X has full column rank p and $m = \min(p, n-p)$, while, as earlier, $\lambda_1 \geq \cdots \geq \lambda_n$ are the necessarily positive eigenvalues of the $n \times n$ positive definite matrix A. When $X = (I_p : 0)'$ and A and A^{-1} are partitioned as in (1.14), then (3.2) simplifies to

$$\operatorname{tr}(A_{11} - (A^{11})^{-1}) = \operatorname{tr}(A_{12}A_{22}^{-1}A_{21}) \leq \sum_{i=1}^{m} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2. \tag{3.3}$$

Indeed (3.3) also implies (3.2) and so they are equivalent. [We are most grateful to Chi-Kwong Li for drawing this to our attention.] To see this, we write (3.3) for an $n \times n$ positive definite matrix B with eigenvalues $\beta_1 \geq \cdots \geq \beta_n$,

$$\operatorname{tr}(B_{11} - (B^{11})^{-1}) \leq \sum_{i=1}^{m} \left(\sqrt{\beta_i} - \sqrt{\beta_{n-i+1}}\right)^2, \tag{3.4}$$

with B_{11} and B^{11} the top-left submatrices, respectively, of B and of B^{-1}. We now let $B = L'AL$ with $L'L = I_n$; then A and B have identical eigenvalues. We write $L = (U : V)$ and choose U as the $n \times p$ matrix $X(X'X)^{-1/2}$. Then $B_{11} = U'AU = (X'X)^{-1/2}X'AX(X'X)^{-1/2}$, and $B^{-1} = L'A^{-1}L$ so that $B^{11} = U'A^{-1}U = (X'X)^{-1/2}X'A^{-1}X(X'X)^{-1/2}$. Substituting in (3.4) yields (3.2).

In an early preprint of the paper by Liu and King (2002), it was claimed that the determinant

$$|A_{11} - (A^{11})^{-1}| \leq \prod_{i=1}^{m} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2, \tag{3.5}$$

but this inequality does not hold in general. As observed by Virtanen (2000), there are many $n \times n$ positive definite matrices for which n is even, $\lambda_{n/2} = \lambda_{(n/2)+1}$ and the left-hand side of (3.5) is positive; but the right-hand side of (3.5) is always 0 when $\lambda_{n/2} = \lambda_{(n/2)+1}$. The same idea was used by Bartmann and Bloomfield (1981, p.71) to show that the inequality (5.37) below, given by Khatri (1978), does not hold in general.

One way to construct matrices with $\lambda_{n/2} = \lambda_{(n/2)+1}$ is with the Kronecker sum (see, e.g., Friedman, 1961; Horn and Johnson, 1994, p. 268):

$$A = \begin{pmatrix} E & I_{n/2} \\ I_{n/2} & E \end{pmatrix} = I_2 \otimes E + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes I_{n/2} \tag{3.6}$$

which has eigenvalues those of $E \pm 1$. When $n = 4$ and E has eigenvalues γ and δ, then (3.6) has eigenvalues $\gamma+1, \gamma-1, \delta+1, \delta-1$, and $\gamma - 1 = \delta + 1 \Leftrightarrow \delta = \gamma - 2$. Our Kronecker sum (3.6) then has eigenvalues $\gamma+1, \gamma-1, \gamma-1, \gamma-3$ and so is positive definite provided $\gamma > 3$. When $\gamma = 4$ then

$$E = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \tag{3.7}$$

and (3.6) has eigenvalues $5, 3, 3, 1$. When A is the Kronecker sum (3.6) with E defined by (3.7) and $X = (I_2 : 0)'$, then the left-hand side of (3.5) equals

$$|A_{11} - (A^{11})^{-1}| = |A_{12}A_{22}^{-1}A_{21}| = \frac{1}{|E|} = \frac{1}{8} > 0; \tag{3.8}$$

here A^{11} is the top-left submatrix of A^{-1}.

We now establish the following inequality, which we believe to be new.

PROPOSITION 1. *Let the real $n \times p$ matrix X have full column rank p and let $m = \min(p, n-p)$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the necessarily positive eigenvalues of the $n \times n$ positive definite matrix A, with A and A^{-1} partitioned as in* (1.14). *Then*

$$\frac{|X'AX - X'X(X'A^{-1}X)^{-1}X'X|}{|X'X|} \leq \max_{\sigma} \prod_{i=1}^{m} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-\sigma(i)+1}}\right)^2, \tag{3.9}$$

or equivalently

$$|A_{11} - (A^{11})^{-1}| = |A_{12}A_{22}^{-1}A_{21}| \leq \max_{\sigma} \prod_{i=1}^{m} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-\sigma(i)+1}}\right)^2. \tag{3.10}$$

In (3.9) *and* (3.10) *the maximum is over all permutations σ of* $\{1, 2, \dots, m\}$.

The proof of Proposition 1 is given in §8.3 of the Appendix; for a different proof/derivation of a special case, see Liu and King (2002). [To show that (3.9) and (3.10) are equivalent, we use the same argument that we used to show that (3.2) and (3.3) are equivalent.]

Liski et al. (1992) extended (3.2) by showing that

$$\operatorname{tr}(P_X A Q_X (Q_X A Q_X)^+ Q_X A P_X) \leq \sum_{i=1}^{s} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2, \tag{3.11}$$

where $s = \min(\operatorname{rank}(X), \operatorname{rank}(Q_X A))$; here again $P_X = AA^+$ and $Q_X = I - P_X$. When $P_X P_A Q_X = 0$ then (see Baksalary et al., 1990),

$$P_X A Q_X (Q_X A Q_X)^+ Q_X A P_X = P_X A P_X - P_X (P_X A^+ P_X)^+ P_X.$$

When A is nonsingular (positive definite) and X has full column rank then

$$\operatorname{tr}(P_X A P_X - P_X (P_X A^+ P_X)^+ P_X) = \operatorname{tr}(X'AX(X'X)^{-1} - X'X(X'A^{-1}X)^{-1}),$$

which is the left-hand side of (3.2). It follows at once that the left-hand sides of (3.2) and (3.11) coincide with $s = \min(p, n-p)$, since the $n \times p$ matrix X now has full column rank p.

Ando (2001) showed that for a complex $n \times n$ Hermitian positive definite matrix A the following majorization of eigenvalues holds:

$$\left\{\mathrm{ch}_i\Big(U^*AU - (U^*A^{-1}U)^{-1}\Big)\right\}_{i=1}^{m} \prec_{\mathsf{w}} \left\{\left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2\right\}_{i=1}^{m}, \quad (3.12)$$

where the $n \times p$ complex matrix U satisfies $U^*U = I_p$ and $m = \min(p, n-p)$. Here ch_i denotes the ith largest eigenvalue and $\prec_{\mathsf{w}}$ weak majorization in the sense that, see Marshall and Olkin (1979), given two sequences of m positive numbers

$$\{\alpha_i\} \text{ with } \alpha_1 \geq \cdots \geq \alpha_m \quad \text{and} \quad \{\beta_i\} \text{ with } \beta_1 \geq \cdots \geq \beta_m, \quad (3.13)$$

then $\{\alpha_i\}$ is said to be "weakly majorized" by $\{\beta_i\}$ whenever

$$\sum_{i=1}^{h} \alpha_i \leq \sum_{i=1}^{h} \beta_i \qquad h = 1, \ldots, m, \quad (3.14)$$

and we will then write

$$\{\alpha_i\}_{i=1}^{m} \prec_{\mathsf{w}} \{\beta_i\}_{i=1}^{m}. \quad (3.15)$$

If in addition $\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \beta_i$, then the adjective "weakly" is dropped and $\{\alpha_i\}$ is said to be "majorized" by $\{\beta_i\}$.

It follows at once from (3.12) that

$$\max_{U} \mathrm{ch}_{\max}\Big(U^*AU - (U^*A^{-1}U)^{-1}\Big) \leq \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2, \quad (3.16)$$

where the maximum is over all $n \times m$ matrices U such that $U^*U = I_m$ with $m \leq n/2$. When $m = 1$ and U is the real $n \times 1$ vector $u = x/\sqrt{x'x}$ then the left-hand side of (3.16) reduces to the maximum of the left-hand side of the SM inequality over all $n \times 1$ vectors x.

When $U = (I_p : 0)'$, then (3.12) becomes

$$\left\{\mathrm{ch}_i\Big(A_{12}A_{22}^{-1}A_{21}\Big)\right\}_{i=1}^{m} \prec_{\mathsf{w}} \left\{\left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right)^2\right\}_{i=1}^{m}, \quad (3.17)$$

since $A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$. This suggests that

$$\left\{\mathrm{ch}_i^{1/2}\left(A_{12}A_{22}^{-1}A_{21}\right)\right\}_{i=1}^{m} \prec_{\mathsf{w}} \left\{\sqrt{\lambda_i}-\sqrt{\lambda_{n-i+1}}\right\}_{i=1}^{m} \tag{3.18}$$

may hold as conjectured by Ando (2000). We have not been able to prove (3.18), but note that clearly (3.18) implies (3.17).

Ando (2001) presents his results in terms of C^*-algebras and calls U^*AU the "compression" $\Phi_{\mathrm{c}}(A)$ of the "unital positive map" $\Phi(A)$ from $n \times n$ matrices to $p \times p$ matrices. The map Φ between C^*-algebras is said to be unital positive if it is unit-preserving and positivity-preserving, respectively. Ando (2001) gives (3.2) in the form

$$\mathrm{tr}\left(\Phi_{\mathrm{c}}(A)-\{\Phi_{\mathrm{c}}(A^{-1})\}^{-1}\right) \le \sum_{i=1}^{m}\left(\sqrt{\lambda_i}-\sqrt{\lambda_{n-i+1}}\right)^2, \tag{3.19}$$

with $m = \min(p, n-p)$, and calls (3.19) the "Rao inequality".

Pečarić et al. (1996) showed that

$$X^*AX-(X^*A^+X)^+ \le_{\mathsf{L}} \left(\sqrt{\lambda_1}-\sqrt{\lambda_r}\right)^2 X^*P_AX,$$

where A is Hermitian nonnegative definite with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_r > 0$ and X is any complex $n \times p$ matrix. As earlier $P_A = AA^+$ is the orthogonal projector on the column space of A.

4. The Bloomfield–Watson Trace Inequality, a New Set of Inequalities, and Bounds for the Spread and Variance

4.1 *The Bloomfield–Watson trace inequality.* An inequality of the same type as those we have already discussed but which holds for all real symmetric (complex Hermitian) matrices A, not necessarily nonnegative definite (or positive definite), was established by Bloomfield and Watson (1975):

$$\mathrm{tr}(U'A^2U-(U'AU)^2) \le \frac{1}{4}\sum_{i=1}^{m}(\lambda_i-\lambda_{n-i+1})^2. \tag{4.1}$$

We will call (4.1) the "Bloomfield–Watson (BW) trace inequality".

In (4.1) the real matrix U is $n \times p$ with $U'U = I_p$ and $m = \min(p, n-p)$, and where, as before, $\lambda_1 \ge \cdots \ge \lambda_n$ are the ordered eigenvalues of the real symmetric matrix A, which need not be nonnegative (or positive) definite here, since both sides of (4.1) remain unchanged if we replace A by $A + kI$, where k is a scalar (positive or negative). Indeed the matrix difference is invariant:

$$U'A^2U - (U'AU)^2 = U'(A+kI)^2U - (U'(A+kI)U)^2. \tag{4.2}$$

When $U = (I_p : 0)'$, and A is partitioned as in (1.14), then (4.1) becomes

$$\operatorname{tr}(A_{12}A_{21}) \le \frac{1}{4}\sum_{i=1}^{m}(\lambda_i - \lambda_{n-i+1})^2. \tag{4.3}$$

It follows that (4.3) also implies (4.1) using the same argument that we used to show that (3.2) and (3.3) are equivalent.

Very recently Ando (2001) established that

$$\left\{\operatorname{ch}_i\Big(U'A^2U - (U'AU)^2\Big)\right\}_{i=1}^{m} \prec_{\mathsf{w}} \left\{\frac{1}{4}\Big(\lambda_i - \lambda_{n-i+1}\Big)^2\right\}_{i=1}^{m}, \tag{4.4}$$

from which (4.1) follows at once. The inequalities (4.4) may also be written as

$$\left\{\operatorname{sg}_i^2(A_{12})\right\}_{i=1}^{m} \prec_{\mathsf{w}} \left\{\frac{1}{4}\Big(\lambda_i - \lambda_{n-i+1}\Big)^2\right\}_{i=1}^{m}. \tag{4.5}$$

When $A_{11} = I_p$ and $A_{22} = I_{n-p}$ then equality holds throughout (4.5). To see this, let $\operatorname{sg}_i(A_{12}) = \kappa_i$; $i = 1, \ldots, h = \operatorname{rank}(A_{12})$. Then

$$\lambda_i = 1 + \kappa_i; \quad \lambda_{n-i+1} = 1 - \kappa_i; \qquad i = 1, \ldots, h = \operatorname{rank}(A_{12}), \tag{4.6}$$

and $\lambda_{h+1}, \ldots, \lambda_{n-h} = 1$. Thus

$$\operatorname{sg}_i(A_{12}) = \kappa_i = \tfrac{1}{2}(\lambda_i - \lambda_{n-i+1}); \quad i = 1, \ldots, m = \min(p, n-p), \tag{4.7}$$

and equality holds throughout (4.5). We note that when A is nonnegative definite then A is a correlation matrix and the κ_i are canonical correlations; but (4.6) and (4.7) hold even when the symmetric matrix A is not nonnegative definite and so not a correlation matrix and the κ_i are not canonical correlations.

4.2 *A new set of inequalities related to the BW trace inequality.* In an early preprint of the paper by Liu and King (2002), it was claimed that

$$|U'A^2U - (U'AU)^2| \le \prod_{i=1}^{m}\Big(\tfrac{1}{4}(\lambda_i - \lambda_{n-i+1})^2\Big) \tag{4.8}$$

but this does not hold in general. Consider again our Kronecker sum matrix (3.6)

$$A = \begin{pmatrix} E & I \\ I & E \end{pmatrix} = \begin{pmatrix} 4 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix} \qquad \text{and} \qquad U = \begin{pmatrix} I_2 \\ 0 \end{pmatrix}.$$

Then the right-hand side of (4.8) is 0 since A now has eigenvalues $5, 3, 3$ and 1, while the left-hand side is $|A_{12}A_{21}| = |I|^2 = 1 > 0$.

We will now establish a set of inequalities, which we believe to be new:

PROPOSITION 2. *Let the $n \times n$ real matrix A be symmetric but not necessarily nonnegative definite. Then*

$$\prod_{j=1}^{k} \mathrm{sg}_j(A_{12}) \leq \max_{\sigma \in S_k} \prod_{j=1}^{k} \left(\tfrac{1}{2}(\lambda_j - \lambda_{n-\sigma(j)+1}) \right) \quad \text{for } k = 1, 2, \ldots, m, \tag{4.9}$$

where A_{12} is the top right $p \times (n-p)$ submatrix of A, $m = \min(p, n-p)$, and the maximum is over all the permutations $\sigma(j)$ in the symmetric group S_k, for $j = 1, 2, \ldots, k$ and $k = 1, 2, \ldots, m$.

Furthermore

$$|A_{12}A_{21}| \leq \max_{\sigma \in S_k} \prod_{i=1}^{m} \left(\tfrac{1}{4}(\lambda_j - \lambda_{n-\sigma(j)+1})^2 \right), \tag{4.10}$$

or equivalently

$$|U'A^2U - (U'AU)^2| \leq \max_{\sigma \in S_k} \prod_{i=1}^{m} \left(\tfrac{1}{4}(\lambda_j - \lambda_{n-\sigma(j)+1})^2 \right), \tag{4.11}$$

where the real matrix U is $n \times p$ with $U'U = I_p$.

The proof of (4.9) is given in §8.1 of the Appendix; for a different proof/derivation of a special case, see Liu and King (2002). [The special case of (4.9) with $k = m$ is (4.10); the equivalence of (4.10) and (4.11) follows easily with the same argument that we used above to show that (3.2) and (3.3) are equivalent.]

4.3 *Bounds for the spread and variance.* The important special case of (4.1) when $m = 1$ seems only to have been "discovered" almost ten years later — by Styan (1983) and Khatri (1984); see also Jia (1996), Jiang (1998):

$$\frac{x'A^2x}{x'x} - \left(\frac{x'Ax}{x'x} \right)^2 \leq \frac{1}{4}(\lambda_1 - \lambda_n)^2. \tag{4.12}$$

The left-hand side of (4.12) yields a set of lower bounds with different choices of x for the "spread" $\lambda_1 - \lambda_n$ of the matrix A. For example, if $x = Pe$, where P is an orthogonal matrix of eigenvectors of A and $e = (1, 1, \dots, 1)'$, then (4.12) becomes

$$\frac{1}{n}\mathrm{tr}(A^2) - \frac{1}{n^2}(\mathrm{tr}A)^2 \leq \frac{1}{4}(\lambda_1 - \lambda_n)^2, \tag{4.13}$$

see Wolkowicz and Styan (1980). The special case of the majorization inequality (4.5) with $m = 1$ is

$$\max_i \sum_{j \neq i} a_{ij}^2 \leq \frac{1}{4}(\lambda_1 - \lambda_n)^2 \tag{4.14}$$

with $A = \{a_{ij}\}$. The inequality (4.14) was strengthened by Jiang and Zhan (1997), who showed that

$$\max_\alpha \frac{1}{|\alpha|} \sum_{i \in \alpha, j \notin \alpha} |a_{ij}|^2 \leq \frac{1}{4}(\lambda_1 - \lambda_n)^2, \tag{4.15}$$

where $\emptyset \neq \alpha \subset \{1, 2, \dots, n\}$ and the cardinality $|\alpha| \leq n/2$. As observed by Jiang and Zhan (1997), the inequality (4.14) is a special case of (4.15); see also Tu (1984).

We may express (4.14) in the form

$$\max_i (a_{ii}^{(2)} - a_{ii}^2) \leq \frac{1}{4}(\lambda_1 - \lambda_n)^2, \tag{4.16}$$

since the left-hand sides of (4.14) and (4.16) are clearly equal; here $a_{ii}^{(2)}$ is the ith diagonal element of A^2. Moreover, (4.16) follows immediately from (4.12) with $x = e_i$, where e_i is the $n \times 1$ vector with 1 in the ith position and 0 elsewhere.

When $x = e$, $A = \mathrm{diag}\{x_i\}$, $\lambda_1 = x_{\max} = \max x_i$ and $\lambda_n = x_{\min} = \min x_i$ in (4.12) then it provides an upper bound for the variance s^2 (with divisor n) of the n numbers $x_1, \dots, x_n$ with mean $\bar{x} = \sum x_i / n$

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \leq \frac{1}{4}(x_{\max} - x_{\min})^2 \tag{4.17}$$

due to Popoviciu (1935); the inequality (4.17) was rediscovered in a statistical context by Nair (1947, p. 113; 1948) and so we will call (4.17) the Popoviciu–Nair inequality. See also, in chronological order, Thomson (1955), Rayner (1975, 1981), Page and Murty (1982, 1983), Jensen (1999), Shiffler and Lackritz (1993), and the series of papers in *Teaching Statistics* by Shiffler and Harsha (1980), MacLeod and Henderson (1984), Boyd (1985), Farnum and Suich (1987) and Priddis (1987).

A tighter upper bound for the variance is

$$s^2 \leq (x_{\max} - \bar{x})(\bar{x} - x_{\min}) \tag{4.18}$$

due to Muilwijk (1966); see also Murthy and Sethi (1965, Appendix 1), Barnett et al. (1999), Barnett and Dragomir (1999), and Bhatia and Davis (2000). Barnes (1995, §2) gives (4.18) stating that it was derived by Barnes and Hoffman (1981), but we could not find (4.18) there. For further discussion and related results see Alpargu and Styan (2000), Jensen (1999) and Seaman and Odell (1988).That

$$(x_{\max} - \bar{x})(\bar{x} - x_{\min}) \leq \frac{1}{4}(x_{\max} - x_{\min})^2$$

follows at once from the arithmetic mean-geometric mean inequality and so (4.18) is a tighter bound for the variance than (4.17).

5. The Bloomfield–Watson–Knott Inequality and Canonical Correlations

A well-known determinantal extension of the FK inequality is the "Bloomfield–Watson–Knott (BWK) inequality":

$$\frac{|X'X|^2}{|X'AX| \cdot |X'A^{-1}X|} \geq \prod_{i=1}^{m} \frac{4\lambda_i \lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}, \tag{5.1}$$

where the $n \times p$ matrix X has full column rank p with $m = \min(p, n-p)$, and where $\lambda_1 \geq \cdots \geq \lambda_n$ denote the necessarily positive eigenvalues of the $n \times n$ positive definite matrix A.

If we substitute $X = (I_p : 0)'$ in the inequality (5.1) then it becomes

$$|I - A_{11}^{-1} A_{12} A_{22}^{-1} A_{21}| \geq \prod_{i=1}^{m} \frac{4\lambda_i \lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}, \tag{5.2}$$

where A is partitioned as in (1.14). It is clear that (5.2) also implies (5.1).

The inequality (5.1) was originally conjectured in 1955 by James Durbin (see Watson, 1955, p. 331), but first established (for $p > 1$) only twenty years later by Bloomfield and Watson (1975) and Knott (1975). Further proofs of (5.1) were provided by Khatri and Rao (1981, 1982) and by Yang (1990); Alpargu (1996, § 4) and Alpargu et al. (1998) present a proof of the BWK inequality based closely on the proof given by Bloomfield and Watson (1975) but which avoids using Lagrange multipliers; Alpargu et al. (1998) also show that the proof (using the Poincaré eigenvalue separation theorem) by Yang (1990) is incomplete. Ando (2001) uses the BWK inequality to prove the majorization inequality (5.18) below from which, clearly, the BWK inequality follows at once. In §8.2 of the Appendix we present a new proof of the BWK inequality.

The interest of these authors centered on the efficiency of the ordinary least squares estimator in the usual Gauss–Markov linear model:

$$y = X\beta + \varepsilon, \quad \mathrm{E}(\varepsilon) = 0, \quad \mathrm{D}(\varepsilon) = A, \tag{5.3}$$

where y is an $n \times 1$ vector of observations, β is a $p \times 1$ vector of unknown parameters, X is an $n \times p$ known design matrix of rank $k < n$, ε is an $n \times 1$ vector of random errors and the $n \times n$ dispersion (variance-covariance) matrix A is positive definite.

The Best Linear Unbiased Estimator (BLUE) of β is the generalized least squares estimator $\tilde{\beta} = (X'A^{-1}X)^{-1}X'A^{-1}y$, which has dispersion matrix

$$\mathrm{D}(\tilde{\beta}) = (X'A^{-1}X)^{-1},$$

while the Ordinary Least Squares Estimator (OLSE) $\hat{\beta} = (X'X)^{-1}X'y$ has dispersion matrix

$$\mathrm{D}(\hat{\beta}) = (X'X)^{-1}X'AX(X'X)^{-1},$$

and so the matrix

$$(X'X)^{-1}X'AX(X'X)^{-1} - (X'A^{-1}X)^{-1} \geq_{\mathsf{L}} 0, \tag{5.4}$$

as in (1.12). To "measure" how far away the matrix difference (5.4) is from the null matrix or equivalently to see how "bad" the OLSE could be with respect to the BLUE, we consider the "Watson efficiency" (Watson, 1951, §3.3; Watson, 1955, p. 330) ϕ of the OLSE $\hat{\beta}$ with respect to the BLUE $\tilde{\beta}$ as the ratio of their generalized variances defined as the determinants of the corresponding dispersion matrices:

$$\phi = \mathrm{eff}(\hat{\beta}) = \frac{|\mathrm{D}(\tilde{\beta})|}{|\mathrm{D}(\hat{\beta})|} = \frac{|(X'A^{-1}X)^{-1}|}{|(X'X)^{-1}X'AX(X'X)^{-1}|}$$

$$= \frac{|X'X|^2}{|X'AX| \cdot |X'A^{-1}X|} \geq \prod_{i=1}^{m} \frac{4\lambda_i \lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}, \quad (5.5)$$

the BWK inequality. Here $m = \min(p, n-p)$ and the $n \times p$ matrix X has full column rank $p < n$.

5.1 *Canonical correlations and the inefficiency of ordinary least squares.* Another measure of how "bad" the OLSE could be with respect to the BLUE is provided by the set of canonical correlations κ_i between the OLSE $X\hat{\beta} = Hy$ and the associated residual vector $r = y - X\hat{\beta} = My$, see Bloomfield and Watson (1975), Bartmann and Bloomfield (1981), Puntanen (1987, Th. 3.4.1, p. 34). Here the hat matrix $H = P_X$ and $M = I - H$; we now suppose that X does not necessarily have full column rank. We also now allow the possibility that the dispersion matrix A may be singular. Then, as shown by Khatri (1976), Seshadri and Styan (1980), and Rao (1981), the κ_i^2 are the nonzero, necessarily positive, eigenvalues of $(HAH)^{-}HAM(MAM)^{-}MAH$, and these are invariant for all choices of generalized inverses $(HAH)^{-}$ and $(MAM)^{-}$. We here assume that $HAM \neq 0$.

Let us now consider the following function of the canonical correlations κ_i

$$\phi_1 = \prod_{i=1}^{h} (1 - \kappa_i^2), \quad (5.6)$$

where $h = \mathrm{rank}(HAM) \geq 1$. Clearly, we must assume that no canonical correlation κ_i is equal to 1 for if a $\kappa_i = 1$, then $\phi_1 = 0$; hence, following Puntanen (1987, Th. 3.4.1, p. 34), we may (and do) characterize this condition with

$$HP_AM = 0; \quad (5.7)$$

see also the orthogonality condition in (2.29). From Lemma 2, we see that (5.7) is equivalent to $\mathrm{rank}(A) = \mathrm{rank}(AH) + \mathrm{rank}(AM)$.

We have the following theorem.

THEOREM 4. *Let the $n \times 1$ random vector u have dispersion matrix A with rank $r \leq n$ and let $\lambda_1 \geq \cdots \geq \lambda_r$ be the r nonzero, necessarily positive, eigenvalues of A. Let the real matrices X and Y be $n \times p$ and $n \times t$, respectively. Let $r_1 = \mathrm{rank}(AX), r_2 = \mathrm{rank}(AY)$ and $h = \mathrm{rank}(X'AY) \geq 1$.*

Then the nonzero canonical correlations between $X'u$ and $Y'u$ are

$$\kappa_i = \mathrm{ch}_i^{1/2}((X'AX)^- X'AY(Y'AY)^- Y'AX) \text{ for } i = 1, \dots, h = \mathrm{rank}(X'AY).$$

If $X'P_AY = 0$, where $P_A = AA^+$ is the orthogonal projector on the column space of A, then for any choices of generalized inverses $(X'AX)^-$ and $(Y'AY)^-$

$$|I-(X'AX)^- X'AY(Y'AY)^- Y'AX| = \prod_{i=1}^{h}(1-\kappa_i^2) \geq \prod_{i=1}^{\min(r_1,r_2)} \frac{4\lambda_i\lambda_{r-i+1}}{(\lambda_i+\lambda_{r-i+1})^2}. \tag{5.8}$$

PROOF. Let the $n \times n$ matrix A have rank r and write $A = TDT'$, where T is $n \times r$ and D is the $r \times r$ diagonal matrix of the r nonzero, necessarily positive, eigenvalues of A. Let

$$r_1 = \mathrm{rank}(AX) = \mathrm{rank}(T'X) \quad \text{and} \quad r_2 = \mathrm{rank}(AY) = \mathrm{rank}(T'Y) \tag{5.9}$$

and let $T'X = G_1F_1'$ and $T'Y = G_2F_2'$ be normalized full-rank decompositions with G_j being $r \times r_j$ $(j = 1, 2)$ and $F_j'F_j = I_{r_j}$ $(j = 1, 2)$. Then

$$\begin{aligned}
\phi_1 &= |I - (X'AX)^- X'AY(Y'AY)^- Y'AX| && (5.10)\\
&= |I - (G_1'DG_1)^{-1}G_1'DG_2(G_2'DG_2)^{-1}G_2'DG_1| && (5.11)\\
&= \frac{|G_1'DG_1 - G_1'DG_2(G_2'DG_2)^{-1}G_2'DG_1|}{|G_1'DG_1|} && (5.12)\\
&\geq \frac{|G_1'G_1(G_1'D^{-1}G_1)^{-1}G_1'G_1|}{|G_1'DG_1|} && (5.13)\\
&= \frac{|G_1'G_1|^2}{|G_1'DG_1| \cdot |G_1'D^{-1}G_1|} \geq \prod_{i=1}^{r_1} \frac{4\lambda_i\lambda_{r-i+1}}{(\lambda_i + \lambda_{r-i+1})^2}. && (5.14)
\end{aligned}$$

To go from (5.10) to (5.11), we note that $|I - BC| = |I - CB|$ for any two conformable matrices B and C, and when F has full column rank and Q is nonsingular, then $F'(FQF')^- F = Q^{-1}$. To establish the inequality in (5.13) we note that when (5.7) holds, then we may use (2.15) in Lemma 1.

We may reverse the roles of G_1 and G_2 in the inequality string (5.10)–(5.13) and hence see that

$$\phi_1 \geq \frac{|G_1'G_1|^2}{|G_1'DG_1| \cdot |G_1'D^{-1}G_1|} \geq \prod_{i=1}^{\min(r_1,r_2)} \frac{4\lambda_i\lambda_{r-i+1}}{(\lambda_i+\lambda_{r-i+1})^2}, \tag{5.15}$$

and our proof is complete. □

When the $n \times (p+t)$ matrix $(X : Y)$ has full row rank n, then from Lemma 2 we see that

$$r = \text{rank}(A) = \text{rank}(AX) + \text{rank}(AY) = r_1 + r_2 \tag{5.16}$$

and hence equality holds in (5.13). When $X = P_X = H$ and $Y = I - P_X = M$, then (5.16) holds and the Watson efficiency

$$\phi = \phi_1 = |I - (H'AH)^- H'AM(MAM)^- MAH| = \prod_{i=1}^{h}(1-\kappa_i^2), \tag{5.17}$$

see also (5.5), where the κ_i are now the canonical correlations between the "fitted values" Hy and the "residuals" My in the linear model (5.3) and h=rank$(HAM)\geq 1$; see Puntanen (1985, 1987).

Very recently Ando (2001, Theorem 3.1a) proved that for a Hermitian positive definite matrix A and compression $\Phi_c(A)$

$$\left\{\log \text{ch}_i\Big((\Phi_c(A^{-1}))^{1/2}\Phi_c(A)(\Phi_c(A^{-1}))^{1/2}\Big)\right\}_{i=1}^m \prec_w \left\{\log \frac{(\lambda_i+\lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}\right\}_{i=1}^m \tag{5.18}$$

or equivalently

$$\prod_{i=1}^k \text{ch}_i\Big((\Phi_c(A^{-1}))^{1/2}\Phi_c(A)(\Phi_c(A^{-1}))^{1/2}\Big) \leq \prod_{i=1}^k \frac{(\lambda_i+\lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}},\ k=1,...,m \tag{5.19}$$

extending the result by Bartmann and Bloomfield (1981, p.70) which shows that

$$\prod_{i=1}^k (1-\kappa_i^2) \geq \prod_{i=1}^k \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i+\lambda_{n-i+1})^2}; \qquad k = 1, ..., m. \tag{5.20}$$

The special case of (5.20) with $k = m$ is the BWK inequality in the form (5.2) with the canonical correlations $\kappa_i = \mathrm{ch}_i^{1/2}(A_{11}^{-1}A_{12}A_{22}^{-1}A_{21})$, $i = 1, ..., m$. The special case of (5.20) with $k = 1$ is the Wielandt inequality (2.2) above, and (5.35) below with $r = n$.

The special case of (5.19) with $k = m$ is this form of the BWK inequality:

$$\left|(\Phi_c(A^{-1}))^{1/2}\Phi_c(A)(\Phi_c(A^{-1}))^{1/2}\right| \leq \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{5.21}$$

We note that the left-hand side of (5.21) may also be written as $|\Phi_c(A)| \cdot |\Phi_c(A^{-1})|$. When $\Phi_c(A)$ is the real matrix

$$\Phi_c(A) = (X'X)^{-1/2}X'AX(X'X)^{-1/2} \tag{5.22}$$

then (5.21) reduces to (5.1).

The inequalities (5.20) may also be written in the form

$$\prod_{i=1}^{k}(1-\kappa_i^2) \geq \prod_{i=1}^{k}\left(1 - \left(\frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}}\right)^2\right); \qquad k = 1, ..., m, \tag{5.23}$$

From this Bartlett and Styan (1980) suggest the inequalities

$$\prod_{i=1}^{k}(1-\kappa_i) \geq \prod_{i=1}^{k}\left(1 - \frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}}\right) = \prod_{i=1}^{k}\left(\frac{2\lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}}\right); \; k = 1, ..., m. \tag{5.24}$$

They generated 7000 matrices A with $n = 4$ and $m = p = n - p = 2$ and found no counter-example to (5.24). Of course (5.24) holds for $k = 1$, which is the Wielandt inequality (5.35) below.

While we have not been able to prove the Bartlett–Styan conjecture[2] (5.24) we do note that (5.24) holds for all 4×4 matrices of the Kronecker sum type

$$A = \begin{pmatrix} E & I_{n/2} \\ I_{n/2} & E \end{pmatrix} = I_2 \otimes E + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes I_{n/2}\,, \tag{5.25}$$

see also (3.6) above. To see this we let γ and δ denote the eigenvalues of E and assume, without loss of generality, that $\gamma \geq \delta$. Then A has eigenvalues

[2] *Note added in proof.* Established by Drury (2002).

$\gamma+1, \gamma-1, \delta+1, \delta-1$ and so A is positive definite if and only if $\delta > 1$. Clearly $\lambda_1 = \gamma+1$ and $\lambda_4 = \delta - 1$ so that $\lambda_1 + \lambda_4 = \gamma + \delta = \lambda_2 + \lambda_3$. The canonical correlations are $\kappa_1 = 1/\delta$ and $\kappa_2 = 1/\gamma$. Substituting (5.25) in (5.24) with $k = m = 2$ yields

$$\left(1 - \frac{1}{\delta}\right)\left(1 - \frac{1}{\gamma}\right) \geq \frac{2(\delta-1)}{\gamma+\delta}\frac{2\lambda_3}{\gamma+\delta}, \tag{5.26}$$

which simplifies to

$$(\gamma+\delta)^2(\gamma-1) \geq 4\gamma\delta\lambda_3. \tag{5.27}$$

Since $(\gamma+\delta)^2 \geq 4\gamma\delta$ and $\gamma - 1 \geq \lambda_3$ we see that (5.27) holds and so (5.24) is corroborated with A as given by (5.25) and $m = p = n - p = 2$.

5.2 *The vector correlation coefficient or the largest squared canonical correlation.* When the $n \times n$ nonnegative definite matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

is the dispersion matrix of the $n \times 1$ random vector

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where x_1 is $p \times 1$, x_2 is $(n-p) \times 1$ and A_{12} is $p \times (n-p)$, then the canonical correlations $\kappa_1 \geq \cdots \geq \kappa_h$ between x_1 and x_2 are the square roots of the eigenvalues of $A_{11}^{-}A_{12}A_{22}^{-}A_{21}$. We may consider as our first measure of association ρ_1 between x_1 and x_2 the largest canonical correlation squared

$$\rho_1 = \kappa_1^2 \leq 1 - \prod_{i=1}^{h}(1-\kappa_i^2) = 1 - \phi_1,$$

see also (5.6); here $h = \operatorname{rank}(A_{12}) \geq 1$. We again assume the condition (5.7) that no κ_i be equal to 1 which here is equivalent to block rank additivity

$$\operatorname{rank}(A) = \operatorname{rank}(A_{11}) + \operatorname{rank}(A_{22}), \tag{5.28}$$

see our Theorem 3 above. We find, therefore, from Theorem 4 that then

$$\rho_1 \leq 1 - \prod_{i=1}^{\min(r_1,r_2)} \frac{4\lambda_i\lambda_{r-i+1}}{(\lambda_i + \lambda_{r-i+1})^2}, \tag{5.29}$$

where $r_1 = \operatorname{rank}(A_{11})$ and $r_2 = \operatorname{rank}(A_{22})$.

Hotelling (1936) and Anderson (1958, p. 344; 1984, p. 397) call ρ_1 the "vector correlation coefficient" and $\phi_1 = 1 - \rho_1$ the "vector coefficient of alienation" and observe that when A_{11} is nonsingular (positive definite) then

$$\rho_1 = 1 - \phi_1 = 1 - \frac{|A_{11} - A_{12}A_{22}^{-}A_{21}|}{|A_{11}|}\,. \tag{5.30}$$

5.3 *Product of the squared canonical correlations and the geometric mean.* Another scalar measure of association between x_1 and x_2 is

$$\rho_2 = \prod_{i=1}^{h} \kappa_i^2, \tag{5.31}$$

which Lin (1987) finds to be "less attractive" than ρ_1. Wang and Ip (1999) call ρ_2 the "Hotelling correlation coefficient". Of course when $h = 1$, then $\rho_2 = \rho_1 = \kappa_1^2$.

When A_{11} is nonsingular (positive definite), then it is easy to see that (5.31) may be expressed as

$$\rho_2 = \prod_{i=1}^{h} \kappa_i^2 = \frac{|A_{12}A_{22}^{-}A_{21}|}{|A_{11}|} = \frac{|A_{12}|^2}{|A_{11}| \cdot |A_{22}|}, \tag{5.32}$$

when n is even and A_{11} and A_{22} are both $n/2 \times n/2$ nonsingular (positive definite), see Anderson (1984, p. 396). We will refer to this latter situation as "full-rank equal split".

As an alternate measure of association we may also use the geometric mean of the squared canonical correlations as considered by Khatri (1978, 1982)

$$\rho_{\text{gm}} = \rho_2^{1/m} = \left(\prod_{i=1}^{m} \kappa_i^2 \right)^{1/m}. \tag{5.33}$$

From Theorem 3(d) we recall that with block rank additivity (5.28)

$$A_{12}A_{22}^{-}A_{21} \leq_{\mathsf{L}} \nu_r^2 A_{11} \tag{5.34}$$

and so we see that

$$\kappa_1 = \max_{1 \leq i \leq h} \kappa_i \leq \nu_r = \frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r} \quad \text{for } i = 1, \ldots, h, \tag{5.35}$$

with $r = r_1 + r_2$, where $r = \operatorname{rank}(A)$, $r_1 = \operatorname{rank}(A_{11})$ and $r_2 = \operatorname{rank}(A_{22})$. Eaton (1976) and Wang and Ip (1999) proved (5.35) with A nonsingular (positive definite). From (5.35) we see that the measure of association

$$\rho_2 = \prod_{i=1}^{h} \kappa_i^2 \leq (\nu_r^2)^h \leq \left(\frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r}\right)^{2h} \quad \text{with } r = r_1 + r_2. \tag{5.36}$$

Khatri (1978, Theorem 1(i)) claimed that in the full rank situation with $r = \operatorname{rank}(A) = n$, and with $m = \min(p, n-p)$ as before,

$$\rho_2 = \prod_{i=1}^{h} \kappa_i^2 \leq \prod_{i=1}^{m} \left(\frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}}\right)^2. \tag{5.37}$$

But this inequality (5.37) does not hold in general, as observed by Bartlett and Styan (1980) and by Bartmann and Bloomfield (1981, p.71), and as acknowledged by Khatri (1982). Indeed, if we choose A as the Kronecker sum matrix (3.6)

$$A_1 = \begin{pmatrix} 4 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix},$$

then $r = n = 4$; the eigenvalues are $5, 3, 3$ and 1 and so the right-hand side of (5.37) is 0. We partition A so that the top-left submatrix A_{11} is 2×2, and thus $h = 2 = p = n - p = m = r_1 = r_2$; hence $\kappa_1 = 1/2$ and $\kappa_2 = 1/4$ and the left-hand side of (5.37) is $\kappa_1^2 \kappa_2^2 = 1/64 > 0$.

We now establish a tight upper bound for the measure of association $\rho_2 = \prod_{i=1}^{h} \kappa_i^2$, see also (5.31), and hence for the product of canonical correlations, when the dispersion matrix A is positive definite and the matrix A_{12} has full rank $h = m = \min(p, n-p)$. The main inequality, (5.39) below, appears[3] in Khatri (1982) and Khatri and Rao (1982, p. 92).

[3]The result (5.39) was discovered and proved by Drury in August 2000. In a personal communication in December 2000, however, C.R. Rao pointed out (to Puntanen and Styan) that in 1981 he had given the result (5.39) *with proof* to C.G. Khatri (1931–1989); in the correction note by Khatri (1982), however, it is only noted that "some corrections are necessary in the proof" of (5.37) given earlier by Khatri (1978) and that "this can be done". We believe that no proof of (5.39) has been published to date.

THEOREM 5. *Let the $n \times n$ positive definite matrix*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{5.38}$$

with necessarily positive eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, be the dispersion matrix of the $n \times 1$ random vector

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where x_1 is $p \times 1$, x_2 is $(n-p) \times 1$ and A_{12} is $p \times (n-p)$, and let $m = \min(p, n-p)$. Suppose that A_{12} has rank m and let $\kappa_1 \geq \cdots \geq \kappa_m$ denote the canonical correlations between x_1 and x_2. Then

$$\rho_2 = \prod_{i=1}^{m} \kappa_i^2 \leq \max_{\alpha,\beta} \prod_{i=1}^{m} \left(\frac{\lambda_{\alpha(i)} - \lambda_{\beta(i)}}{\lambda_{\alpha(i)} + \lambda_{\beta(i)}} \right)^2 = \prod_{i=1}^{m} \left(\frac{\lambda_i - \lambda_{n-m+i}}{\lambda_i + \lambda_{n-m+i}} \right)^2. \tag{5.39}$$

The maximum in (5.39) is taken over all possible partial matchings $\{\alpha, \beta\}$ of $\{1, 2, \ldots, n\}$ into m pairs, the i-th pair being denoted $\{\alpha(i), \beta(i)\}$.

Hence when the dispersion matrix A is positive definite:

$$\prod_{i=1}^{m} \left(\frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}} \right)^2 \leq \prod_{i=1}^{m} \left(\frac{\lambda_i - \lambda_{n-m+i}}{\lambda_i + \lambda_{n-m+i}} \right)^2 \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right)^n. \tag{5.40}$$

The first ratio in (5.40) is the upper bound for ρ_2 conjectured in (5.37), the middle ratio is our new upper bound (5.39) and the last ratio is the upper bound constructed from the upper bound for the largest canonical correlation, see also (5.35) and (5.36). When $n = 4$, $m = 2 = p = n - p$, then (5.40) becomes

$$\left(\frac{\lambda_1 - \lambda_4}{\lambda_1 + \lambda_4} \right)^2 \left(\frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \right)^2 \leq \left(\frac{\lambda_1 - \lambda_3}{\lambda_1 + \lambda_3} \right)^2 \left(\frac{\lambda_2 - \lambda_4}{\lambda_2 + \lambda_4} \right)^2 \leq \left(\frac{\lambda_1 - \lambda_4}{\lambda_1 + \lambda_4} \right)^4. \tag{5.41}$$

We find the left-hand inequalities in (5.40) and (5.41) to be surprising!

When A is our Kronecker sum matrix (3.6), then our new upper bound (5.39), i.e., the middle ratio in (5.40) and (5.41):

$$\prod_{i=1}^{n/2} \left(\frac{\lambda_i - \lambda_{p+i}}{\lambda_i - \lambda_{p+i}} \right)^2 = \left(\frac{\lambda_1 - \lambda_3}{\lambda_1 + \lambda_3} \right)^2 \left(\frac{\lambda_2 - \lambda_4}{\lambda_2 + \lambda_4} \right)^2 = \frac{1}{64} = \kappa_1^2 \kappa_2^2 = \rho_2,$$

and so equality is attained (5.39). The last ratio in (5.41)

$$\left(\frac{\lambda_1 - \lambda_4}{\lambda_1 + \lambda_4} \right)^4 = \frac{16}{81}$$

is much higher than $1/64$ and so in this example, therefore, inequality (5.39) is a considerable improvement over the inequality based on the upper bound for the first canonical correlation, see also (5.35) and (5.36).

Before entering into the proof of Theorem 5, we point out that it will be used as a model for all the proofs in Appendix A. The key elements of the proof that will vary according to context are the function f, the Lyapunov condition (5.44) and the matrix X which it involves, the commutation pair (5.46), the matrices B and Z, the interval J and the mapping φ.

PROOF OF THEOREM 5. In order to prove (5.39) we will use calculus-type techniques. We will assume, without loss of generality, that $n-p \leq p$ and so $m = \min(p, n-p) = n-p$. We will write $q = n-p$. We let $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, which we regard as fixed and let V be an $n \times n$ orthogonal matrix so that $A = V'\Lambda V$ is a generic positive definite matrix with the given eigenvalues. We partition A as before. Then the determinant function

$$f(V) = |A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}| = \frac{|A_{21}A_{11}^{-1}A_{12}|}{|A_{22}|}, \tag{5.42}$$

is infinitely differentiable on the (compact) orthogonal group $O(n)$ and must, therefore, attain its supremum. Let us denote by V an element of $O(n)$ at which f takes a local maximum value. We will use the following lemma.

LEMMA 3. *Let the real $n \times n$ positive definite matrix A be partitioned as in* (5.38) *with A_{12} being $p \times (n-p)$ and let*

$$X = A_{12} - A_{12}(A_{21}A_{11}^{-1}A_{12})^{-1}A_{22}. \tag{5.43}$$

Then at any local maximum point of the determinant function $f(V)$ defined in (5.42) *the following identity must hold*

$$A_{11}X = XA_{22}\,. \tag{5.44}$$

PROOF OF LEMMA 3. We perturb V by $V(t) = \exp(tS)V$ where S is a $n \times n$ skew-symmetric matrix. Further, let us partition V as $V = (V_1 : V_2)$, where V_1 and V_2 are $n \times p$ and $n \times q$ matrices respectively. Then, clearly $A_{jk} = V_j'\Lambda V_k$. The corresponding relationships hold for the perturbed matrices, so that

$$A_{jk}(t) = V_j'(t)\Lambda V_k(t)$$

where $V_j(t) = V_j + tSV_j + \cdots$ where the $\cdots$ will always denote terms of order $O(t^2)$. Note that A_{11} and A_{22} are necessarily invertible. We have

$$A_{12}(t) = V_1'(t)\Lambda V_2(t) = V_1'(\Lambda + t[\Lambda, S] + \cdots)V_2 = A_{12} + tV_1'KV_2 + \cdots\,,$$

where $K = [\Lambda, S]$. Similarly

$$A_{11}(t)^{-1} = A_{11}^{-1} - tA_{11}^{-1}V_1'KV_1A_{11}^{-1} + \cdots$$

There are analogous formulae for $A_{22}(t)^{-1}$ and $A_{21}(t)$. Combining these yields

$$\begin{aligned} &A_{22}(t)^{-1}A_{21}(t)A_{11}(t)^{-1}A_{12}(t) = \\ &\quad A_{22}^{-1}A_{21}A_{11}^{-1}A_{12} - tA_{22}^{-1}V_2'KV_2A_{22}^{-1}A_{21}A_{11}^{-1}A_{12} + tA_{22}^{-1}V_2'KV_1A_{11}^{-1}A_{12} \\ &\quad -tA_{22}^{-1}A_{21}A_{11}^{-1}V_1'KV_1A_{11}^{-1}A_{12} + tA_{22}^{-1}A_{21}A_{11}^{-1}V_1'KV_2 + \cdots \end{aligned}$$

Let us denote $C = A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}$. Then $|C|$ is always nonnegative and at a local maximum point, it cannot vanish. Thus C is invertible. We use the well-known formula

$$|C + tQ + \cdots| = |C| + t|C|\operatorname{tr}(C^{-1}Q) + \cdots$$

to obtain an expansion of $|A_{22}(t)^{-1}A_{21}(t)A_{11}(t)^{-1}A_{12}(t))|$. At a critical point of f we will have $\operatorname{tr}(C^{-1}Q) = 0$. In our case, this yields

$$\begin{aligned} &\operatorname{tr}(-C^{-1}A_{22}^{-1}V_2'KV_2A_{22}^{-1}A_{21}A_{11}^{-1}A_{12} + C^{-1}A_{22}^{-1}V_2'KV_1A_{11}^{-1}A_{12} \\ &\quad -C^{-1}A_{22}^{-1}A_{21}A_{11}^{-1}V_1'KV_1A_{11}^{-1}A_{12} + C^{-1}A_{22}^{-1}A_{21}A_{11}^{-1}V_1'KV_2) = 0. \end{aligned}$$

If we use the cyclical property of the trace and rearrange the terms, then

$$\begin{aligned} &\operatorname{tr}(S[\Lambda, -V_2A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}C^{-1}A_{22}^{-1}V_2' + V_1A_{11}^{-1}A_{12}C^{-1}A_{22}^{-1}V_2' \\ &\quad -V_1A_{11}^{-1}A_{12}C^{-1}A_{22}^{-1}A_{21}A_{11}^{-1}V_1' + V_2C^{-1}A_{22}^{-1}A_{21}A_{11}^{-1}V_1']) = 0. \quad (5.45) \end{aligned}$$

The second term in the commutator can be written as

$$-V_2A_{22}^{-1}V_2' + V_1A_{11}^{-1}A_{12}YV_2' - V_1A_{11}^{-1}A_{12}YA_{21}A_{11}^{-1}V_1' + V_2YA_{21}A_{11}^{-1}V_1',$$

where $Y = (A_{21}A_{11}^{-1}A_{12})^{-1}$, which reveals that it is symmetric. The commutator in (5.45) is skew symmetric, so since its trace with every skew symmetric matrix S vanishes, we deduce that it must vanish identically

$$[\Lambda,\ V_1A_{11}^{-1}A_{12}YA_{21}A_{11}^{-1}V_1' + V_2A_{22}^{-1}V_2' - V_2A_{12}^{-1}V_1' - V_1A_{21}^{-1}V_2'] = 0.$$

But we can express this as the fact that

$$(V_1 : V_2)\begin{pmatrix} A_{11}^{-1}A_{12}YA_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}Y \\ -YA_{21}A_{11}^{-1} & A_{22}^{-1} \end{pmatrix}\begin{pmatrix} V_1' \\ V_2' \end{pmatrix}$$

commutes with Λ, or equivalently, that the matrices

$$\begin{pmatrix} A_{11}^{-1}A_{12}YA_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}Y \\ -YA_{21}A_{11}^{-1} & A_{22}^{-1} \end{pmatrix} \text{ and } \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{5.46}$$

commute. When we compare the products of the above matrices, we find that the diagonal blocks are independent of the order of multiplication. So, finally this commutation condition boils down to $-A_{12}Y + A_{12}A_{22}^{-1} = A_{11}^{-1}A_{12} - A_{11}^{-1}A_{12}YA_{22}$ which is equivalent to (5.44). □

PROOF OF THE LEFT-HAND PART OF (5.39). Since A_{11} and A_{22} are themselves positive definite, they can be diagonalized by the Spectral Theorem. So, applying a similarity by a block diagonal orthogonal matrix to A, we can assume without loss of generality that A_{11} and A_{22} are diagonal matrices with strictly positive diagonal elements which we denote $\mu_1, \ldots, \mu_p$ and $\nu_1, \ldots, \nu_q$, respectively. So, from (5.44) we obtain the fact that either $\mu_j = \nu_k$ or $x_{jk} = 0$.

We split up the pattern of X into components. Two elements of the pattern are connected if there is a path with horizontal or vertical segments going from one element to the other while remaining in the pattern. Clearly, distinct components are both row and column disjoint, and if row j and row k are involved in the same component, then $\mu_j = \mu_k$ and similarly for the columns. We now apply the singular value decomposition on each (in general rectangular) block corresponding to each component. The key point is that the block diagonal orthogonal similarity which replaces A_{11} by $P'A_{11}P$, A_{12} by $P'A_{12}Q$, A_{21} by $Q'A_{21}P$, and A_{22} by $Q'A_{22}Q$ transforms X into $P'XQ$. Thus, diagonalizing X on each rectangular block does not change A_{11} or A_{22}. After a further similarity by a block permutation matrix, (which does in general permute the diagonal elements of A_{11} and A_{22}) we can assume that the only possible non-zero entries of X are the entries x_{jj} for $j = 1, \ldots, q$. For each such j, there are two cases. Either $x_{jj} \neq 0$, in which case $\mu_j = \nu_j$, or $x_{jj} = 0$, in which case we may have $\mu_j \neq \nu_j$.

We would now like to deduce that the matrix A_{12} itself is diagonal. For this, let $B = A_{11}^{-1/2}A_{12}A_{22}^{-1/2}$. Let $Z = B - B(B'B)^{-1}$. Then $X = A_{11}^{1/2}ZA_{22}^{1/2}$. So Z is rectangular diagonal since X is. Consider the singular value decomposition $B = P'\Delta Q$ of B where Δ is the $p \times q$ diagonal matrix with $b_1, b_2, \ldots, b_q$ on the main diagonal and P and Q are orthogonal. Now it is easy to see that

$$\begin{pmatrix} I & B \\ B' & I \end{pmatrix}$$

is strictly positive definite. Thus B is a strict contraction. Also we observe that the b_j are non-zero at a local maximum point. Thus $b_j \in J$ for $j = 1, 2, \ldots, q$ where J denotes the open interval $(0, 1)$.

Let $\varphi(t) = t - t^{-1}$ for $t \in J$. Now we have $Z = B - B(B'B)^{-1} = P'\Delta_1 Q$ is diagonal where Δ_1 is the $p \times q$ diagonal matrix with $\varphi(b_1), \varphi(b_2), \ldots, \varphi(b_q)$ on the main diagonal. But Z is already rectangular diagonal, so we obtain information from the uniqueness assertion of the singular value decomposition. The key point here is that the transformation φ is 1-1 on J, so that two diagonal entries of Δ_1 are equal only if the corresponding entries of Δ are equal. Thus Q and the first q columns of P are block diagonal with the blocks Q_ℓ and P_ℓ respectively corresponding to blocks of constancy of $\varphi(b_j)$. It follows that $B = P'\Delta Q$ is also diagonal.

Now that we know that A_{12} is diagonal, we can write (after simultaneously rearranging rows and columns) A as a block diagonal matrix with q diagonal blocks of size 2×2 and $n - 2q$ blocks of size 1×1. However, our result is easy in the 2×2 case and we therefore find that the left-hand part of (5.39) holds. □

We are now ready to prove the right-hand part of (5.39).

Proof of the right-hand part of (5.39). It remains to show that of all possible partial matchings, the one that matches j with $p + j$ for $j = 1, 2, \ldots, q$ is the one that achieves the maximum value. For simplicity in this section, we will assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$. If the result is established in this special case, then it will suffice to use perturbations to obtain the general case. In a partial matchings $\{\alpha, \beta\}$ of $\{1, 2, \ldots, n\}$ into q pairs, the ith pair being denoted $\{\alpha(i), \beta(i)\}$, we can always assume that $\alpha(i) < \beta(i)$ for $i = 1, 2, \ldots, q$. We claim first that a partial matching that has $\max_{i=1}^q \alpha_i > \min_{i=1}^q \beta_i$ cannot be maximal. Such a matching must match two elements n_1 and n_2 and match two elements n_3 and n_4, where $n_1 > n_2 > n_3 > n_4$. So, we have, without loss of generality, $\lambda_{n_1} > \lambda_{n_2} > \lambda_{n_3} > \lambda_{n_4} > 0$, then it is clear that

$$\frac{\lambda_{n_1} - \lambda_{n_2}}{\lambda_{n_1} + \lambda_{n_2}} \cdot \frac{\lambda_{n_3} - \lambda_{n_4}}{\lambda_{n_3} + \lambda_{n_4}} < \frac{\lambda_{n_1} - \lambda_{n_3}}{\lambda_{n_1} + \lambda_{n_3}} \cdot \frac{\lambda_{n_2} - \lambda_{n_4}}{\lambda_{n_2} + \lambda_{n_4}}$$

since both

$$\frac{\lambda_{n_2}}{\lambda_{n_1}} > \frac{\lambda_{n_3}}{\lambda_{n_1}} \quad \text{and} \quad \frac{\lambda_{n_4}}{\lambda_{n_3}} > \frac{\lambda_{n_4}}{\lambda_{n_2}}.$$

Secondly, we observe that the α_i must fill out $\{1, 2, \ldots q\}$ and the β_i must fill out $p + 1, p + 2, \ldots, n$. For example, if $\alpha_i > q$, there is necessarily an integer k with $1 \le k \le q$ which is not an α_j. We will then do better to replace α_i with k.

This means that the upper bound in (5.39) is of the form

$$\max_{\sigma \in S_q} \prod_{j=1}^{q} \left(\frac{\lambda_j - \lambda_{p+\sigma(j)}}{\lambda_j + \lambda_{p+\sigma(j)}} \right)^2 ,$$

for some permutation σ. We now show that a maximal σ cannot invert any pairs. For if so, then we have $n_1 < n_2 < n_3 < n_4$ with n_1 matched to n_4 and n_2 matched to n_3. We claim that

$$\frac{\lambda_{n_1} - \lambda_{n_4}}{\lambda_{n_1} + \lambda_{n_4}} \cdot \frac{\lambda_{n_2} - \lambda_{n_3}}{\lambda_{n_2} + \lambda_{n_3}} < \frac{\lambda_{n_1} - \lambda_{n_3}}{\lambda_{n_1} + \lambda_{n_3}} \cdot \frac{\lambda_{n_2} - \lambda_{n_4}}{\lambda_{n_2} + \lambda_{n_4}} .$$

Indeed, an unpleasant calculation reveals

$$\begin{aligned} &(\lambda_{n_1} + \lambda_{n_4})(\lambda_{n_2} + \lambda_{n_3})(\lambda_{n_1} - \lambda_{n_3})(\lambda_{n_2} - \lambda_{n_4}) \\ &\qquad -(\lambda_{n_1} - \lambda_{n_4})(\lambda_{n_2} - \lambda_{n_3})(\lambda_{n_1} + \lambda_{n_3})(\lambda_{n_2} + \lambda_{n_4}) \\ &\quad = 2(\lambda_{n_1} - \lambda_{n_2})(\lambda_{n_3} - \lambda_{n_4})\gamma > 0, \end{aligned}$$

where

$$\gamma = (\lambda_{n_1} - \lambda_{n_2})\lambda_{n_2} + (\lambda_{n_2} - \lambda_{n_3})(\lambda_{n_2} + \lambda_{n_3}) + (\lambda_{n_3} - \lambda_{n_4})(\lambda_{n_3} + 2\lambda_{n_4}) + 2\lambda_{n_4}^2 > 0.$$

It follows that the maximum is attained at the identity permutation, and our result follows. □

5.4 *Sum of the squared canonical correlations and the arithmetic mean.* Khatri (1978, 1982), Zhang (1978) and Lin (1987) considered the arithmetic mean of the squared canonical correlations

$$\rho_{\mathrm{am}} = \frac{1}{h} \sum_{i=1}^{h} \kappa_i^2 = \frac{1}{h} \mathrm{tr}(A_{11}^{-} A_{12} A_{22}^{-} A_{21}). \tag{5.47}$$

Khatri (1978) claimed that in the full rank situation with $r = \mathrm{rank}(A) = n$ and $h = m = \min(p, n-p)$ then the arithmetic mean

$$\rho_{\mathrm{am}} = \frac{1}{m} \sum_{i=1}^{m} \kappa_i^2 \le \frac{1}{m} \sum_{i=1}^{m} \left(\frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}} \right)^2 , \tag{5.48}$$

see also (5.37). But as noted by Bartlett and Styan (1980) and Khatri (1982), the inequality (5.48) does not hold in general. Indeed let A be the Kronecker sum

$$A = \begin{pmatrix} 25 & 0 & 8 & 0 \\ 0 & 9 & 0 & 8 \\ 8 & 0 & 25 & 0 \\ 0 & 8 & 0 & 9 \end{pmatrix} = I_2 \otimes \begin{pmatrix} 25 & 0 \\ 0 & 9 \end{pmatrix} + \begin{pmatrix} 0 & 8 \\ 8 & 0 \end{pmatrix} \otimes I_2 ,$$

see also (3.6) above. This Kronecker sum has eigenvalues $25+8 = 33$, $25-8 = 17$, $9+8 = 17$ and $9-8 = 1$ and so is positive definite, with $r = n = 4$. We partition A so that the top-left submatrix A_{11} is 2×2, and thus $h = 2 = p = n - p = m = r_1 = r_2$. We again, therefore, have a full-rank equal split, see also (5.32). The left-hand side of (5.48) is $\rho_{\mathrm{am}} = 22592/50625 \simeq 0.4463$, while the right-hand side equals $128/289 \simeq 0.4429 < 0.4463$.

We conjecture that in the full-rank situation the following inequality holds:

$$\rho_{\mathrm{am}} = \frac{1}{m}\sum_{i=1}^{m}\kappa_i^2 = \frac{1}{m}\mathrm{tr}(A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}) \le \frac{1}{m}\max_{\alpha,\beta}\sum_{i=1}^{m}\left(\frac{\lambda_{\alpha(i)} - \lambda_{\beta(i)}}{\lambda_{\alpha(i)} + \lambda_{\beta(i)}}\right)^2 . \tag{5.49}$$

As in (5.39) the maximum here is taken over all possible partial matchings $\{\alpha, \beta\}$ of $\{1, 2, \ldots, n\}$ into m pairs, the ith pair being denoted $\{\alpha(i), \beta(i)\}$. It follows as in the proof of (5.39) that the optimal matching must match $\{1, 2, \ldots, m\}$ with $\{n-m+1, n-m+2, \ldots, n\}$. Thus, the right-hand side of (5.49)

$$\frac{1}{m}\max_{\alpha,\beta}\sum_{i=1}^{m}\left(\frac{\lambda_{\alpha(i)} - \lambda_{\beta(i)}}{\lambda_{\alpha(i)} + \lambda_{\beta(i)}}\right)^2 = \frac{1}{m}\max_{\sigma\in S_m}\sum_{i=1}^{m}\left(\frac{\lambda_i - \lambda_{n+\sigma(i)-m}}{\lambda_i + \lambda_{n+\sigma(i)-m}}\right)^2 \tag{5.50}$$

The maximum is over the set S_m of all permutations σ of $\{1, 2, \ldots, m\}$. We note that (5.49) was given in a preprint of Khatri (1982) but does not appear in the published version. It is not given by Bartlett and Styan (1980), by Bartmann and Bloomfield (1981), or by Khatri and Rao (1982).

While we have not been able to establish the inequality in (5.49) we note that the idea of the proof of (5.39) very nearly works in this context. The analogue of (5.44) is $A_{11}^{-1}X = XA_{22}^{-1}$, where $X = A_{12} - A_{12}A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}$. Our sticking point is that the mapping $\varphi(t) = t - t^3$ is not 1–1 on $J = [0, 1)$.

When $A_{11} = I_p$ and $A_{22} = I_{n-p}$, then A is a correlation matrix, and block rank additivity implies that A is positive definite and so $r = n$. Then, cf. (4.6) and (4.7) above, $\lambda_i = 1 + \kappa_i$ and $\lambda_{n-i+1} = 1 - \kappa_i$ $(i = 1, \ldots, h)$, and $\lambda_{h+1} = \cdots = \lambda_{n-h} = 1$. Hence

$$\frac{\lambda_i - \lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}} = \kappa_i; \qquad i = 1, \ldots, h, \tag{5.51}$$

and we have equality in (5.39) and in (5.49) but not necessarily in (5.36) or in (5.54).

5.5 *The harmonic mean of the squared canonical correlations and the smallest squared canonical correlation.* Khatri (1978, 1982), Zhang (1978) and Lin (1987) also considered as measures of association the harmonic mean of the squared canonical correlations

$$\rho_{\text{hm}} = 1 \Big/ \frac{1}{h}\sum_{i=1}^{h} 1/\kappa_i^2 \tag{5.52}$$

and the smallest squared canonical correlation $\rho_{\min} = \kappa_h^2 = \min_{1\le i\le h}\kappa_i^2$. From the harmonic-geometric-arithmetic mean inequality:

$$\rho_{\min} \le \rho_{\text{hm}} \le \rho_{\text{gm}} \le \rho_{\text{am}} \le \rho_{\max} \tag{5.53}$$

with $\rho_{\text{hm}} = \rho_{\text{gm}} \Leftrightarrow \rho_{\text{gm}} = \rho_{\text{am}}$ if and only if $\kappa_1 = \cdots = \kappa_h$.

When $h = 1$, then all of these five measures of association are equal to $\kappa_1^2 = \rho_1 = \rho_2$, as defined above.

It follows at once from (5.53) and (5.35) that with $r = r_1 + r_2$ we have

$$\rho_{\min} \le \rho_{\text{hm}} \le \rho_{\text{gm}} \le \rho_{\text{am}} = \frac{1}{h}\sum_{i=1}^{h}\kappa_i^2 \le \left(\frac{\lambda_1 - \lambda_r}{\lambda_1 + \lambda_r}\right)^2 . \tag{5.54}$$

Khatri (1982) and Khatri and Rao (1982) stated, without proof, that also in the full rank situation with $h = m = \min(p, n-p)$ and $r = \text{rank}(A) = n$, the harmonic mean ρ_{hm} satisfies the inequality[4]

$$\frac{1}{\rho_{\text{hm}}} = \frac{1}{m}\sum_{i=1}^{h}\frac{1}{\kappa_i^2} \ \ge\ \frac{1}{m}\min_{\sigma\in S_m}\sum_{i=1}^{m}\left(\frac{\lambda_i + \lambda_{n+\sigma(i)-m}}{\lambda_i - \lambda_{n+\sigma(i)-m}}\right)^2 , \tag{5.55}$$

where the minimum is over the set S_m of all permutations σ of $\{1, 2, \ldots, m\}$.

6. The Khatri–Rao and Krasnosel′skiĭ–Kreĭn Inequalities

Khatri and Rao (1981, 1982) generalized the BWK inequality in several directions. Let T and U be $n \times p$ matrices such that $T'T = U'U = I_p$ and let

[4] *Note added in proof.* Established by Drury (2002).

K be an $n \times n$ nonsingular matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_n > 0$. Then with $m = \min(p, n-p)$

$$|T'KU| \cdot |U'K^{-1}T| \leq \prod_{i=1}^{m} \frac{(\sigma_i + \sigma_{n-i+1})^2}{4\sigma_i \sigma_{n-i+1}}. \tag{6.1}$$

When $p = 1$ then $m = 1$ and (6.1) reduces to the inequality:

$$\frac{t'Ku \cdot u'K^{-1}t}{t't \cdot u'u} \leq \frac{(\sigma_1 + \sigma_n)^2}{4\sigma_1 \sigma_n},$$

obtained by Strang (1960); here t and u are $n \times 1$ vectors.

Moreover, Khatri and Rao (1981, 1982) showed that

$$\operatorname{tr}(T'KU \cdot U'K^{-1}T) \leq \sum_{i=1}^{m} \frac{(\sigma_i + \sigma_{n-i+1})^2}{4\sigma_i \sigma_{n-i+1}} + p - m. \tag{6.2}$$

Now let B and C be symmetric $n \times n$ nonsingular matrices such that $BC^{-1} = C^{-1}B$ and is positive definite, let X be an $n \times p$ matrix of rank p, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the, necessarily positive, eigenvalues of BC^{-1}. Then Khatri and Rao (1982) showed that

$$\frac{|X'B^2X| \cdot |X'C^2X|}{|X'BCX|^2} \leq \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}} \tag{6.3}$$

and

$$\operatorname{tr}\Big(X'B^2X(X'BCX)^{-1}X'C^2X(X'BCX)^{-1}\Big) \leq \sum_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}} + p - m.$$

If we let $C = I$ and $B = A$ in (6.3), then (6.3) becomes

$$\frac{|X'A^2X| \cdot |X'X|}{|X'AX|^2} \leq \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}} \tag{6.4}$$

and when $p = 1$ and $X = w$ an $n \times 1$ vector, then (6.4) becomes

$$\frac{w'A^2w \cdot w'w}{(w'Aw)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}, \tag{6.5}$$

the KK inequality, see also (2.5) above.

When $X = (I_p : 0)'$ then (6.4) may be written as

$$|I + A_{21}A_{11}^{-2}A_{12}| \le \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{6.6}$$

which Ando (2001) extends by proving that

$$\{\mathrm{ch}_i(A_{21}A_{11}^{-2}A_{12})\}_{i=1}^{m} \prec_{\mathrm{w}} \left\{\frac{(\lambda_i - \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}\right\}_{i=1}^{m}; \tag{6.7}$$

this suggests that

$$\{\mathrm{ch}_i^{1/2}(A_{21}A_{11}^{-2}A_{12})\}_{i=1}^{m} \prec_{\mathrm{w}} \left\{\frac{\lambda_i - \lambda_{n-i+1}}{2\sqrt{\lambda_i\lambda_{n-i+1}}}\right\}_{i=1}^{m}. \tag{6.8}$$

We have not been able to prove (6.8), but note that clearly (6.8) implies (6.7).

Ando (2001) presents (6.4) in the form:

$$|\{\Phi_{\mathrm{c}}(A)\}^{-1}\Phi_{\mathrm{c}}(A^2)\{\Phi_{\mathrm{c}}(A)\}^{-1}| = \frac{|\Phi_{\mathrm{c}}(A^2)|}{|\Phi_{\mathrm{c}}(A)|^2} \le \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{6.9}$$

which he calls the “Khatri–Rao (KR) inequality”.

Liu (1997) posed the following inequality as a problem, solved by Bebiano, da Providencia and Li (1998) and by Liu (1998):

$$1 + \frac{\mathrm{tr}(A_{12}A_{21})}{\mathrm{tr}(A_{11}^2)} = \frac{\mathrm{tr}(A_{11}^2 + A_{12}A_{21})}{\mathrm{tr}(A_{11}^2)} \le \frac{\sum_{i=1}^{m}(\lambda_i + \lambda_{n-i+1})^2}{4\sum_{i=1}^{m}\lambda_i\lambda_{n-i+1}},$$

or equivalently

$$\frac{\mathrm{tr}(U'A^2U)}{\mathrm{tr}((U'AU)^2)} \le \frac{\sum_{i=1}^{m}(\lambda_i + \lambda_{n-i+1})^2}{4\sum_{i=1}^{m}\lambda_i\lambda_{n-i+1}}, \tag{6.10}$$

where the $n \times n$ real matrix A is positive definite, and the $n \times p$ real matrix U satisfies $U'U = I_p$ with $m = \min(p, n-p)$.

Moreover Ando (2001) has shown that

$$\Big\{\log \mathrm{ch}_i\Big(\{\Phi_{\mathrm{c}}(A)\}^{-1}\Phi_{\mathrm{c}}(A^2)\{\Phi_{\mathrm{c}}(A)\}^{-1}\Big)\Big\}_{i=1}^{m} \prec_{\mathrm{w}} \left\{\log\left(\frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}\right)\right\}_{i=1}^{m} \tag{6.11}$$

from which (6.9) follows at once; Ando (2001), however, uses the KR inequality (6.9) to prove (6.11).

7. Summary of Open Problems

We summarize here the various open problems[5] presented in this paper. All these problems involve an $n \times n$ postive definite matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{7.1}$$

where A_{12} is $p \times (n-p)$. We write $m = \min(p, n-p)$. The necessarily positive eigenvalues of A are denoted $\lambda_1 \geq \cdots \geq \lambda_n$ and the associated canonical correlations are $\kappa_1 \geq \cdots \geq \kappa_h$, where $\kappa_i^2 = \mathrm{ch}_i(A_{11}^{-1}A_{12}A_{22}^{-1}A_{21})$; $i = 1, \ldots, h = \mathrm{rank}(A_{12}) \geq 1$, with ch_i the ith largest eigenvalue.

Problem 7.1 Upper Bound for the Arithmetic Mean of Squared Canonical Correlations.

Establish the following upper bound, see (5.49) and (5.50) above, for the arithmetic mean ρ_{am} of squared canonical correlations

$$\rho_{\mathrm{am}} = \frac{1}{m}\sum_{i=1}^{m} \kappa_i^2 = \frac{1}{m}\mathrm{tr}(A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}) \leq \frac{1}{m}\max_{\sigma \in S_m}\sum_{i=1}^{m}\left(\frac{\lambda_i - \lambda_{n+\sigma(i)-m}}{\lambda_i + \lambda_{n+\sigma(i)-m}}\right)^2, \tag{7.2}$$

where the maximum is over the set S_m of all permutations σ of $\{1, 2, \ldots, m\}$.

Problem 7.2. Lower Bound for the Reciprocal of the Harmonic Mean of Squared Canonical Correlations.

Prove the following lower bound, (5.55) above, for the reciprocal of the harmonic mean ρ_{hm} of squared canonical correlations

$$\frac{1}{\rho_{\mathrm{hm}}} = \frac{1}{m}\sum_{i=1}^{h}\frac{1}{\kappa_i^2} \geq \frac{1}{m}\min_{\sigma \in S_m}\sum_{i=1}^{m}\left(\frac{\lambda_i + \lambda_{n+\sigma(i)-m}}{\lambda_i - \lambda_{n+\sigma(i)-m}}\right)^2, \tag{7.3}$$

where the minimum is over the set S_m of all permutations σ of $\{1, 2, \ldots, m\}$.

Problem 7.3. Two Possible Sets of Majorization Inequalities.

Prove or disprove the following two sets of majorization inequalities, (3.18) and (6.8) above, proposed by Ando (2000),

$$\left\{\mathrm{ch}_i^{1/2}\left(A_{12}A_{22}^{-1}A_{21}\right)\right\}_{i=1}^{m} \prec_{\mathrm{w}} \left\{\sqrt{\lambda_i} - \sqrt{\lambda_{n-i+1}}\right\}_{i=1}^{m}, \tag{7.4}$$

$$\{\mathrm{ch}_i^{1/2}(A_{12}A_{22}^{-2}A_{21})\}_{i=1}^{m} \prec_{\mathrm{w}} \left\{\frac{\lambda_i - \lambda_{n-i+1}}{2\sqrt{\lambda_i\lambda_{n-i+1}}}\right\}_{i=1}^{m}. \tag{7.5}$$

[5]*Note added in proof.* Problems 7.2 and 7.4 are solved in Drury (2002).

PROBLEM 7.4. Multiplicative Majorization-type Inequalities Associated with Canonical Correlations.

Prove that the following multiplicative inequalities, (5.24) above, hold:

$$\prod_{i=1}^{k}(1-\kappa_i) \geq \prod_{i=1}^{k}\left(1-\frac{\lambda_i-\lambda_{n-i+1}}{\lambda_i+\lambda_{n-i+1}}\right) = \prod_{i=1}^{k}\left(\frac{2\lambda_{n-i+1}}{\lambda_i+\lambda_{n-i+1}}\right); \quad k=1,...,m. \tag{7.6}$$

We have established (7.6) for certain special cases.

PROBLEM 7.5. Block Rank Additivity and Löwner Ordering.

Find a "nice" additional condition so that together with the Löwner ordering

$$(A^+)_{11} \geq_{\mathsf{L}} (A_{11})^+ \quad \text{and} \quad (A^+)_{22} \geq_{\mathsf{L}} (A_{22})^+, \tag{7.7}$$

as in (2.40), the two conditions together imply block rank additivity as in (2.39),

$$\text{rank}(A) = \text{rank}(A_{11}) + \text{rank}(A_{22}). \tag{7.8}$$

8. Appendix: Three Proofs

All the proofs in this Appendix are modelled on our proof of Theorem 5. We leave most of the tedious details to the reader.

8.1 *Proof of Proposition 2.* We start by establishing (4.11). In our proof we will assume, without loss of generality, that $n-p \leq p$ and so $m = \min(p, n-p) = n-p$. We apply the method of Theorem 5 with $f(V) = |A_{21}A_{12}|$ and find that the matrix entering the commutation condition (5.46) is

$$\begin{pmatrix} 0 & A_{21}^+ \\ A_{12}^+ & 0 \end{pmatrix},$$

the matrix X entering the Lyapunov condition (5.44) is A_{12}^+. The matrix B whose singular values we wish to estimate is A_{12}. The matrix Z is again A_{12}^+ in this case, the interval $J = (0, \infty)$ and $\varphi(t) = t^{-1}$. The method of Theorem 5 shows that we can assume without loss of generality that A_{11}, A_{22} are diagonal and A_{12} is rectangular diagonal. Thus, as before the problem is reduced to the case $n = 2, p = 1, m = 1$, where it is easily established by direct calculation.

To reduce the set of matchings that need to be considered, we will assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$. If the result is established in this special case, then it will suffice to use perturbations to obtain the general case. In a partial matching $\{\alpha, \beta\}$ of $\{1, 2, \ldots, n\}$ into m pairs, the ith pair being denoted $\{\alpha(i), \beta(i)\}$, we can always assume that $\alpha(i) < \beta(i)$ for $i = 1, 2, \ldots, m$. We claim first that a partial matching that has $\max_{i=1}^m \alpha_i > \min_{i=1}^m \beta_i$ cannot be maximal. Such a matching must match two elements n_1 and n_2 and match two elements n_3 and n_4, where $n_1 > n_2 > n_3 > n_4$. So, we have, without loss of generality, $\lambda_{n_1} > \lambda_{n_2} > \lambda_{n_3} > \lambda_{n_4} > 0$, then it is clear that

$$(\lambda_{n_1} - \lambda_{n_2})^2(\lambda_{n_3} - \lambda_{n_4})^2 < (\lambda_{n_1} - \lambda_{n_3})^2(\lambda_{n_2} - \lambda_{n_4})^2$$

since both $(\lambda_{n_1} - \lambda_{n_3})^2 > (\lambda_{n_1} - \lambda_{n_2})^2$ and $(\lambda_{n_2} - \lambda_{n_4})^2 > (\lambda_{n_3} - \lambda_{n_4})^2$.

Secondly, we observe that the α_i must fill out $\{1, 2, \ldots m\}$ and the β_i must fill out $p+1, p+2, \ldots, n$. For example, if $\alpha_i > m$, there is necessarily an integer k with $1 \le k \le m$ which is not an α_j. We will then do better to replace α_i with k.

This means that the upper bound in (4.11) is of the form

$$\max_{\sigma \in S_m} \prod_{j=1}^{m} (\lambda_j - \lambda_{n+1-\sigma(j)})^2,$$

for some permutation $\sigma \in S_m$ as required.

To establish (4.9), we start by applying a similarity by a block-diagonal orthogonal matrix to reduce to the case where A_{12} is rectangular diagonal with its singular values on the diagonal in *increasing* order. We now repartition the matrix so that the upper left block is $(n-k) \times (n-k)$ and the lower right block is $k \times k$. We then apply (4.11) with $p = n - k$ and $m = k$.

8.2 *A new proof of the BWK inequality (5.1).* The first step is to recognize that we can restate (5.1) in the form

$$|A_{11}||(A^{-1})_{11}| \le \prod_{i=1}^{m} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}}. \tag{8.1}$$

We have $(A^{-1})_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ and take $f(V) = |I - A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}|$. A rather unpleasant calculation shows that the matrix entering the analogue of the commutation condition (5.46) is

$$\begin{pmatrix} C^{-1} - A_{11}^{-1} & -C^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}C^{-1} & A_{22}^{-1}A_{21}C^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

where C now denotes $A_{11} - A_{12}A_{22}^{-1}A_{21}$. The analogue of the Lyapunov condition eventually simplifies right down to

$$A_{11}A_{12} = A_{12}A_{22},$$

Thus we can arrange that the blocks A_{12} are rectangular diagonal and that A_{11} and A_{22} are diagonal. It is clear from an examination of the 2×2 case that

$$|A_{11}||(A^{-1})_{11}| \leq \max_{\alpha,\beta} \prod_{i=1}^{m} \frac{(\lambda_{\alpha(i)} + \lambda_{\beta(i)})^2}{4\lambda_{\alpha(i)}\lambda_{\beta(i)}},$$

where the maximum is over all possible partial matchings $\{\alpha, \beta\}$ of $\{1, 2, \dots, n\}$ into m pairs. We leave the reader to check that the maximum is actually taken as indicated by the right-hand side of (8.1). □

8.3 *Proof of Proposition 1.* We restate (3.9) in the form

$$|A_{11} - (A^{-1})_{11}^{-1}| \leq \max_{\sigma} \prod_{i=1}^{m} \left(\sqrt{\lambda_i} - \sqrt{\lambda_{n-\sigma(i)+1}}\right)^2.$$

Furthermore, we have $A_{11} - (A^{-1})_{11}^{-1} = A_{12}A_{22}^{-1}A_{21}$. We take $f(V) = |A_{12}A_{22}^{-1}A_{21}|$ and, contrary to our usual convention, we are assuming that $m = n - p \geq p$ or else the determinant will vanish and the result will be trivial. Let us denote $C = A_{12}A_{22}^{-1}A_{21}$, then the matrix occurring in the analogue of the commutation condition (5.46) is

$$\begin{pmatrix} 0 & C^{-1}A_{12}A_{22}^{-1} \\ A_{22}^{-1}A_{21}C^{-1} & -A_{22}^{-1}A_{21}C^{-1}A_{12}A_{22}^{-1} \end{pmatrix}.$$

The analogue of the Lyapunov condition becomes

$$A_{11}C^{-1}A_{12}A_{22}^{-1} - A_{12}A_{22}^{-1} = C^{-1}A_{12} \tag{8.2}$$

We multiply on the right by A_{21} (losing information in the process) to obtain $A_{11} - C = C^{-1}A_{12}A_{21}$ or equivalently

$$CA_{11} = C^2 + A_{12}A_{21}\,. \tag{8.3}$$

But, since C, A_{11} and the right-hand side of (8.3) are symmetric, we find that C and A_{11} commute. This means that (8.2) can be rewritten as

$$CA_{11} = A_{11}C, \qquad (A_{11} - C)A_{12} = A_{12}A_{22}\,.$$

We can diagonalize A_{11}, C and A_{22} simultaneously. We note that A_{11} and C transform identically under a similarity by a block diagonal orthogonal matrix. Following the proof of (5.39), we apply such a similarity which renders A_{12} rectangular diagonal and $A_{11} - C$ and A_{22} diagonal. But then $C = A_{12}A_{22}^{-1}A_{21}$ is again diagonal and hence so is A_{11}. Thus, at a critical point, we can assume that A_{12} is rectangular diagonal and that A_{11} and A_{22} are diagonal. We leave the other details of the proof to the reader. □

Acknowledgements. We are most grateful to C. Radhakrishna Rao, and to Gülhan Alpargu, T. Ando, Jerzy K. Baksalary, Cynthia L. Bartlett, Rajendra Bhatia, Ka Lok Chu, Chandler Davis, Gene H. Golub, Jürgen Groß, Allan Gut, Tõnu Kollo, Chi-Kwong Li, Jorma Kaarlo Merikoski, Yongge Tian, Ari Virtanen, and Fuzhen Zhang for helpful discussions. Special thanks go to T. Ando, Rajendra Bhatia, Ravindra Khattree, Tuulikki Mäkeläinen, Jorma Kaarlo Merikoski, Götz Trenkler and Fuzhen Zhang for providing us with copies of various papers. The research of the first and fifth authors was supported in part by the Natural Sciences and Engineering Research Council of Canada. This paper was completed while the fifth author was on sabbatical leave in the Department of Mathematics, Statistics and Philosophy, University of Tampere, Tampere, Finland, May–December 2000; support by the Academy of Finland is most gratefully acknowledged.

Note added in proof. A solution to Problem 7.2, as well as a proof of the Bartlett–Styan conjecture (Problem 7.4), are given by Drury (2002). The paper by Bartlett and Styan (1980) is reprinted, with slight editorial modifications, as Appendix B in §9 of the earlier version of this paper by Drury et al. (2000), which also includes some biographical/anecdotal information about mathematicians and about 50 references that are not given below.

References

Alpargu, G. (1996). *The Kantorovich Inequality, With Some Extensions and With Some Statistical Applications.* M.Sc. thesis, Dept. of Mathematics and Statistics, McGill University, Montréal.

Alpargu, G., Drury, S.W. and Styan, G.P.H. (1998). Some remarks on the Bloomfield–Watson–Knott Inequality and on some other inequalities related to the Kantorovich Inequality. In *Proceedings of the Conference in Honor of S.R. Searle, August 9–10, 1996*, Biometrics Unit, Cornell University, Ithaca, New York, 125-143.

Alpargu, G. and Styan, G.P.H. (2000). Some comments and a bibliography on the Frucht–Kantorovich and Wielandt inequalities. In *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker* (R.D.H. Heijmans, D.S.G. Pollock and A. Satorra, eds.), Kluwer, Dordrecht, 1-38.

ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis.* Second Edition. Wiley, New York. (First Edition: 1958.)

ANDO, T. (2000). From determinant and trace inequalities to majorization of eigenvalues. Unpublished lecture notes, Faculty of Economics, Hokusei Gakuen University, Sapporo, Japan, 6 pp.

ANDO, T. (2001). Bloomfield–Watson–Knott type inequalities for eigenvalues. *Taiwanese J. Math.* **5**, 443-469.

BAKSALARY, J.K. AND BAKSALARY, O.M. (2001). Two rank equalities associated with blocks of an orthogonal projector: Solution 25-4.1 to Problem 25-4 (proposed by Tian, 2000). *Image: Bull. Int. Linear Algebra Soc.* **26**, 6–7.

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 PUNTANEN, S. (1991). Generalized matrix versions of the Cauchy–Schwarz and Kantorovich inequalities. *Aequationes Math.* **41**, 103-110.

BAKSALARY, J.K., PUNTANEN, S. and STYAN, G.P.H. (1990). A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model. *Sankhyā Ser. A* **52**, 279-296.

BAKSALARY, J.K. and STYAN, G.P.H. (1993). Around a formula for the rank of a matrix product with some statistical applications. In *Graphs, Matrices, and Designs: Festschrift in Honor of Norman J. Pullman on his Sixtieth Birthday* (R.S. Rees, ed.), Lecture Notes in Pure and Applied Mathematics, vol. **139**, Marcel Dekker, New York, 1-18.

BARNES, E.R. (1995). An inequality for probability moments with applications. *SIAM J. Discrete Math.* **8**, 347-358.

BARNES, E.R. and HOFFMAN, A.J. (1981). On bounds for eigenvalues of real symmetric matrices. *Linear Algebra Appl.* **40**, 217-223.

BARNES, E.R. and HOFFMAN, A.J. (1994). Bounds for the spectrum of normal matrices. *Linear Algebra Appl.* **201**, 79-90.

BARNETT, N.S., CERONE, P., DRAGOMIR, S.S. and ROUMELIOTIS, J. (1999). Some inequalities for the dispersion of a random variable whose pdf is defined on a finite interval. *RGMIA Research Report Collection* **2**, 1005-1025.

BARNETT, N.S. and DRAGOMIR, S.S. (1999). Some elementary inequalities for the expectation and variance of a random variable whose pdf is defined on a finite interval. *RGMIA Research Report Collection* **2**, 1067-1074.

BARTLETT, C.L. AND STYAN, G.P.H. (1980). On some inequalities between the eigenvalues of a covariance matrix and the associated canonical correlations. Unpublished research report, Dept. of Mathematics, McGill University, Montréal, 7 pp., 4 August 1980. [Reprinted, with some slight editorial modifications, as Appendix B in §9 of Drury et al. (2000). Abstract #80t-131: *The IMS Bulletin* **9**, 294 (1980).]

BARTMANN, F.C. and BLOOMFIELD, P. (1981). Inefficiency and correlation. *Biometrika* **68**, 67-71.

BAUER, F.L. and HOUSEHOLDER, A.S. (1960). Some inequalities involving the Euclidean condition of a matrix. *Numer. Math.* **2**, 308-311.

BEBIANO, N., DA PROVIDENCIA, J. and LI, C.-K. (1998). Bounds for a ratio of matrix traces: Solution 18-4.1 to Problem 18-4 (proposed by Liu, 1997). *Image: Bull. Int. Linear Algebra Soc.* **20**, 23-25.

BENDER, A.H. (1942). Siendo $x_1, x_2, x_3, \ldots x_n$ números positivos con la condición $\sum x_i = k$, demostrar las desigualdades $\sum_{i=1}^n x_i^2 \geq k^2/n, \quad \sum_{i=1}^n 1/x_i \geq n^2/k$: Problema N° 21—1° Solución. *Mathematicæ Notæ–Boletin del Instituto de Matemática "Beppo Levi" (Rosario)* **2**, 195-197. (English Translation: Appendix A of Watson et al., 1997.)

BHATIA, R. (1999). The matrix version of the inequalities of Kantorovich and Wielandt. Unpublished research report, Indian Statistical Institute–Delhi Centre, 2 pp.

BHATIA, R. and DAVIS, C. (2000). A better bound on the variance. *Amer. Math. Monthly* **107**, 353-357.

BHATIA, R. and DAVIS, C. (2001). A key inequality for functions of matrices. *Linear Algebra Appl.* **323**, 1-5.

BLOOMFIELD, P. and WATSON, G.S. (1975). The inefficiency of least squares. *Biometrika* **62**, 121-128.

BOYD, A.V. (1985). The standard deviation and absolute deviations from the mean. *Teaching Statist.* **7**, 78-82.

CASSELS, J. W. S. (1951). Appendix 1: A proof of the inequality (3.3.2). In Watson (1951, pp. 138-139). (Preprint of Cassels, 1955.)

CASSELS, J. W. S. (1955). Appendix: A proof of the inequality (3.2). In Watson (1955, pp. 340-341).

DRURY, S.W. (2002). The canonical correlations of a 2×2 block matrix with given eigenvalues. *Linear Algebra Appl.* **354**, 103-117.

DRURY, S.W., LIU, S, LU, C.-Y., PUNTANEN, S. and STYAN, G.P.H. (2000). Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments. Report A332 (December 2000), Dept. of Mathematics, Statistics and Philosophy, University of Tampere, Tampere, Finland, 63 pp.

EATON, M.L. (1976). A maximization problem and its applications to canonical correlation. *J. Multivariate Anal.* **6**, 422-425.

FARNUM, N.R. and SUICH, R.C. (1987). Bounds on the sample standard deviation. *Teaching Statist.* **9**, 51-56.

FRIEDMAN, B. (1961). Eigenvalues of composite matrices. *Proc. Cambridge Philos. Soc.* **57**, 37-49.

FRUCHT, R. (1943). Sobre algunas desigualdades: Observación relativa a la solución del Problema N° 21, indicada por el Ing. Ernesto M. Saleme (1942) (with an untitled appendix by Beppo Levi). *Mathematicæ Notæ–Boletin del Instituto de Matemática "Beppo Levi" (Rosario)* **3**, 41–46. (English Translation: Appendix A of Watson et al., 1997.)

GAFFKE, N. and KRAFFT, O. (1977). Optimum properties of Latin square designs and a matrix inequality. *Math. Operationsforsch. Statist. Ser. Statist.* **8**, 345-350.

GOLUB, G.H. (1963). Comparison of the variance of minimum variance and weighted least squares regression coefficients. *Ann. Math. Statist.* **34**, 984-991.

GREUB, W. and RHEINBOLDT, W. (1959). On a generalization of an inequality of L. V. Kantorovich. *Proc. Amer. Math. Soc.* **10**, 407–415.

GROß, J. (2000). The Moore–Penrose inverse of a partitioned nonnegative definite matrix. *Linear Algebra Appl.* **321**, 113-121.

GUSTAFSON, K.E. (1999). The geometrical meaning of the Kantorovich–Wielandt inequalities. *Linear Algebra Appl.* **296**, 143-151.

GUSTAFSON, K.E. and RAO, D.K.M. (1997). *Numerical Range: The Field of Values of Linear Operators and Matrices.* Springer-Verlag, New York.

HANNAN, E.J. (1960). *Time Series Analysis.* Methuen, London. (See pp. 111-113.)

HORN, R.A. and JOHNSON, C.R. (1994). *Topics in Matrix Analysis.* Corrected Reprint Edition. Cambridge University Press.

HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321-377.

HOUSEHOLDER, A.S. (1975). *The Theory of Matrices in Numerical Analysis.* Second Edition (reprint). Dover, New York.

JENSEN, S.T. (1999). *The Laguerre–Samuelson Inequality with Extensions and Applications in Statistics and Matrix Theory.* M.Sc. thesis, Dept. of Mathematics and Statistics, McGill University, Montréal.

JIA, Z. (1996). An extension of Styan's inequality (in Chinese). *Chinese J. Engrg. Math.* **13** (1), 122-126. (See also Jiang, 1998.)

JIANG, E. and ZHAN, X. (1997). Lower bounds for the spread of a Hermitian matrix. *Linear Algebra Appl.* **256**, 153-163.

JIANG, Y.Q. (1998). An extension of Styan's inequality: Comment on Jia (1996) (in Chinese). *J. Xuzhou Normal Univ. Natural Sci. Edition* **16**, 4-5.

KANTOROVICH, L.V. (1948). Funkcional'nyi analiz i prikladnaya matematika (in Russian). *Uspekhi Mat. Nauk Nov. Ser.* **3** (6/28), 89–185. (See pp. 142-144. English Translation: Kantorovich, 1952.)

KANTOROVICH, L.V. (1952). *Functional Analysis and Applied Mathematics.* Translated from Russian (Kantorovich, 1948) by C.D. Benster and edited by G.E. Forsythe. NBS Report no. 1509 (1101-10-5100), U.S. Dept. of Commerce, National Bureau of Standards, Los Angeles. (See pp. 106-109.)

KHATRI, C.G. (1976). A note on multiple and canonical correlation for a singular covariance matrix. *Psychometrika* **41**, 465-470.

KHATRI, C.G. (1978). Some optimization problems with applications to canonical correlations and sphericity tests. *J. Multivariate Anal.* **8**, 453–467. (Corrected by Khatri, 1982.)

KHATRI, C.G. (1982). Erratum (to Khatri, 1978). *J. Multivariate Anal.* **12**, 612.

KHATRI, C.G. (1984). An inequality concerning the difference of variances of two different estimates. *Math. Today* **2**, 29-32.

KHATRI, C.G. and RAO, C.R. (1981). Some extensions of the Kantorovich inequality and statistical applications. *J. Multivariate Anal.* **11**, 498-505.

KHATRI, C.G. and RAO, C.R. (1982). Some generalizations of Kantorovich inequality. *Sankhyā Ser. A* **44**, 91-102.

KHATTREE, R. (2001). On the calculation of antieigenvalues and antieigenvectors. *J. Interdiscip. Math.* **4**, 195-199.

KIEFER, J. (1959). Optimum experimental designs. (With discussion.) *J.R. Stat. Soc. Ser. B* **21**, 272–319. [Reprinted in *Jack Carl Kiefer Collected Papers III: Design of Experiments* (L.D. Brown, I. Olkin, J. Sacks and H.P. Wynn, eds.), Springer-Verlag, New York, 54–101 (1985).]

KNOTT, M. (1975). On the minimum efficiency of least squares. *Biometrika* **62**, 129-132.

KRASNOSEL'SKIĬ, M.A. and KREĬN, S.G. (1952). An iteration process with minimal residuals (in Russian). *Mat. Sbornik Nov. Ser.* **31** (73), 315–334. (See pp. 323-325.)

LI, C.-K. and MATHIAS, R. (1999). Inequalities on the singular values of an off-diagonal block of a Hermitian matrix. *J. Inequal. Appl.* **3**, 137-142.

LIN, P. (1987). Measures of association between vectors. *Comm. Statist. Theory Methods* **16**, 321-338.

LISKI, E.P., PUNTANEN, S. and WANG, S. (1992). Bounds for the trace of the difference of the covariance matrices of the OLSE and BLUE. *Linear Algebra Appl.* **176**, 121-130.

LIU, S. (1997). Bounds for a ratio of matrix traces: Problem 18-4. *Image: Bull. Int. Linear Algebra Soc.* **18**, 32 and **19**, 27. (Solved by Bebiano et al., 1998 and by Liu, 1998.)

LIU, S. (1998). Bounds for a ratio of matrix traces: Solution 18-4.2 to Problem 18-4 (proposed by Liu, 1997). *Image: Bull. Int. Linear Algebra Soc.* **20**, 25-26.

LIU, S. and KING, M.L. (2002). Two Kantorovich-type inequalities and efficiency comparisons between the OLSE and BLUE. *J. Inequal. Appl.* **7**, 169-177.

LU, C.-Y. (1999). Generalized matrix versions of the Wielandt inequality with some statistical applications. Unpublished research report, Dept. of Mathematics, Northeast Normal University, Changchun (Jilin), China, 8 pp.

MACLEOD, A.J. and HENDERSON, G.R. (1984). Bounds for the sample standard deviation. *Teaching Statist.* **6**, 72-76.

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. [Acknowledgement of Priority: **34** (1963), 352.]

MARSAGLIA, G. and STYAN, G.P.H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra* **2**, 269-292.

MARSAGLIA, G. and STYAN, G.P.H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā Ser. A* **36**, 437-442.

MARSHALL, A.W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

MARSHALL, A.W. and OLKIN, I. (1990). Matrix versions of the Cauchy and Kantorovich inequalities. *Aequationes Math.* **40**, 89-93.

MUILWIJK, J. (1966). Note on a theorem of M.N. Murthy and V.K. Sethi. *Sankhyā Ser. B* **28**, 183.

MURTHY, M.N. and SETHI, V.K. (1965). Self-weighting design at tabulation stage. *Sankhyā Ser. B* **27**, 201-210.

NAIR, K.R. (1947). *Studies in Design and Analysis of Experiments.* Ph.D. thesis, University College, London. (Part V published as Nair, 1948.)

NAIR, K. R. (1948). Certain symmetrical properties of unbiased estimates of variance and covariance. *J. Indian Soc. Agricultural Statist.* **1**, 162–172. (Original version is Part V of Nair, 1947.)

NAKAMOTO, R. and TAKAHASHI, S. (1999). Generalizations of an inequality of Marcus. *Math. Japon.* **50**, 35-39.

PAGE, W. and MURTY, V.N. (1982). Nearness relations among measures of central tendency and dispersion: Part 1. *Two-Year College Math. J.* **13**, 315-327.

PAGE, W. and MURTY, V.N. (1983). Nearness relations among measures of central tendency and dispersion: Part 2. *Two-Year College Math. J.* **14**, 8-17.

PEČARIĆ, J.E., PUNTANEN, S. and STYAN, G.P.H. (1996). Some further matrix extensions of the Cauchy–Schwarz and Kantorovich inequalities, with some statistical applications. *Linear Algebra Appl.* **237/238**, 455-476.

PÓLYA, G. and SZEGŐ, G. (1925). *Aufgaben und Lehrsätze aus der Analysis, Band I: Reihen, Integralrechnung, Funktionentheorie.* Julius Springer, Berlin. (See #92-93, pp. 57, 213-214; English Translation: Pólya and Szegő, 1998.)

PÓLYA, G. and SZEGŐ, G. (1998). *Problems and Theorems in Analysis, Volume I: Series, Integral Calculus, Theory of Functions.* Classics in Mathematics, Springer-Verlag, Berlin. (Corrected reprint of the revised translation from the German (Pólya and Szegő, 1995) by Dorothee Aeppli of the 1978 fourth German edition: Grundlehren der Mathematischen Wissenschaften, vol. **193**. See #92-93, pp. 71-72, 253-255.)

POPOVICIU, T. (1935). Sur les équations algébriques ayant toutes leurs racines réeles. *Mathematica (Cluj)* **9**, 129-145.

PRIDDIS, M.J.M. (1987). An upper bound for the standard deviation. *Teaching Statist.* **9**, 78-80.

PUNTANEN, S. (1985). Properties of the canonical correlations between the least squares fitted values and the residuals. In *Proc. First International Tampere Seminar on Linear Statistical Models and their Applications: University of Tampere, Tampere, Finland, August 30th to September 2nd, 1983* (T. Pukkila and S. Puntanen, eds.), Dept. of Mathematical Sciences/Statistics, University of Tampere, 269-284.

PUNTANEN, S. (1987). On the relative goodness of ordinary least squares and on some associated canonical correlations. In *On the Relative Goodness of Ordinary Least Squares Estimation in the General Linear Model* by S. Puntanen [Ph.D. thesis, Dept. of Mathematical Sciences, University of Tampere], Acta Universitatis Tamperensis Series A, University of Tampere, vol. **216**, Paper [1], 77 pp.

PUNTANEN, S., STYAN, G.P.H. and TIAN, Y. (2001). Two rank equalities associated with blocks of an orthogonal projector: Solution 25-4.3 to Problem 25-4 (proposed by Tian, 2000). *Image: Bull. Int. Linear Algebra Soc.* **26**, 8-9.

RAO, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals, in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966*, vol. **1**, L.M. Le Cam and J. Neyman, eds., Univ. of California Press, Berkeley, 355-372.

RAO, C.R. (1981). A lemma on g-inverse of a matrix and computation of correlation coefficients in the singular case. *Comm. Statist. Theory Methods* **10**, 1-10.

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.

RAO, C.R. and RAO, M.B. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics.* World Scientific, Singapore.

RAYNER, J.C.W. (1975). Variance bounds. *Sankhyā Ser. B* **37**, 135-138.

RAYNER, J.C.W. (1981). Bounds for the sample standard deviation. *The New Zealand Statistician* **16**, 42-43.

Saleme, E.M. (1942). Siendo $x_1, x_2, x_3, ..., x_n$ números positivos con la condición $\sum x_i = k$, demostrar las desigualdades $\sum_{i=1}^{n} x_i^2 \geq k^2/n, \quad \sum_{i=1}^{n} 1/x_i \geq n^2/k$: Problema N° 21—2° Solución y generalización. *Mathematicæ Notæ–Boletin del Instituto de Matemática "Beppo Levi" (Rosario)* **2**, 197-199 (1942). (English Translation: Appendix A of Watson et al., 1997.)

Schweitzer, P. (1914). Egy egyenlőtlenség az aritmetikai középértékről [An inequality about the arithmetic mean]. *Matematikai és Physikai Lapok* **23**, 257-261. (English Translation: Appendix A of Alpargu, 1996, and Appendix B of Watson et al., 1997.)

Seaman, J.W., Jr. and Odell, P.L. (1988). Variance, upper bounds. In *Encyclopedia of Statistical Sciences, Volume 9: Strata Chart to Zyskind–Martin Models* (S. Kotz, N.L. Johnson and C.B. Read, eds.), Wiley, New York, 480-484.

Seshadri, V. and Styan, G.P.H. (1980). Canonical correlations, rank additivity and characterizations of multivariate normality. In *Analytic Function Methods in Probability Theory: Proc. Colloquium on the Methods of Complex Analysis in the Theory of Probability and Statistics held at the Kossuth L. University, Debrecen, Hungary, August 29–September 2, 1977*, Colloquia Mathematica Societatis János Bolyai, vol. **21**, János Bolyai, Budapest and North-Holland, Amsterdam, 331-344.

Shiffler, R.E. and Harsha, P.D. (1980). Upper and lower bounds for the sample standard deviation. *Teaching Statist.* **2**, 84-86.

Shiffler, R.E. and Lackritz, J.R. (1993). Bounds for the standard deviation for grouped data. *Internat. J. Math. Ed. Sci. Tech.* **24**, 359-363.

Shisha, O. and Mond, B. (1967). Bounds on differences of means. In *Inequalities: Proceedings of a Symposium held at Wright–Patterson Air Force Base, Ohio, August 19–27, 1965* (O. Shisha, ed.), Academic Press, New York, 293-308.

Strang, W.G. (1960). On the Kantorovich inequality. *Proc. Amer. Math. Soc.* **11**, 468.

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 Universitatis Tamperensis Series A, University of Tampere, vol. **153**, 158-166.

Thomson, G.W. (1955). Bounds for the ratio of range to standard deviation. *Biometrika* **42**, 268-269.

Tian, Y. (2000). Two rank equalities associated with blocks of an orthogonal projector: Problem 25-4. *Image: Bull. Int. Linear Algebra Soc.* **25**, 16. (Solved by Baksalary and Baksalary, 2001; by Puntanen et al., 2001; and by Werner, 2001.)

Tian, Y. and Styan, G.P.H. (2001). Rank equalities for idempotent and involutory matrices. *Linear Algebra Appl.* **335**, 101-117.

Tu, B. (1984). On the spread of a matrix (in Chinese). *J. Fudan Univ. Nat. Sci.* **23**, 435-442.

Virtanen, A. (2000). Personal communication.

Wang, S. and Ip, W.C. (1999). A matrix version of the Wielandt inequality and its applications to statistics. *Linear Algebra Appl.* **296**, 171-181.

Watson, G.S. (1951). *Serial Correlation in Regression Analysis.* Ph.D. thesis, Dept. of Experimental Statistics, North Carolina State College, Raleigh.

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

WATSON, G.S., ALPARGU, G. and STYAN, G.P.H. (1997). Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor. *Linear Algebra Appl.* **264**, 13-53. (Includes English translations of Bender, 1942, Frucht, 1943 and Saleme, 1942.)

WATSON, G.S. and HANNAN, E.J. (1956). Serial correlation in regression analysis, II. *Biometrika* **42**, 436-448.

WEISSTEIN, E.W. (1999). *CRC Concise Encyclopedia of Mathematics.* Chapman & Hall/CRC, Boca Raton, Florida.

WERNER, H.J. (2001). Two rank equalities associated with blocks of an orthogonal projector: Solution 25-4.2 to Problem 25-4 (proposed by Tian 2000). *Image: Bull. Int. Linear Algebra Soc.* **26**, 7.

WIELANDT, H. (1953). Inclusion theorems for eigenvalues. In *Simultaneous Linear Equations and the Determination of Eigenvalues* (L.J. Paige and O. Taussky, eds.), National Bureau of Standards Applied Mathematics Series vol. **29**, U.S. Government Printing Office, Washington, D.C., 75-78. [Reprinted, with commentaries by C. Davis and by H. Schneider, and an editor's note, In *Helmut Wielandt: Mathematische Werke/Mathematical Works–Volume 2: Linear Algebra and Analysis* (B. Huppert and H. Schneider, eds.), Walter de Gruyter, Berlin, 156-162 (1996).]

WOLKOWICZ, H. (1981). Bounds for the Kantorovich ratio. Unpublished research report, Dept. of Mathematics, University of Alberta, Edmonton, 15 pp.

WOLKOWICZ, H. and STYAN, G.P.H. (1980). Bounds for eigenvalues using traces. *Linear Algebra Appl.* **29**, 471-506.

YANG, H. (1990). A brief proof on the generalized variance bound of the relative efficiency in statistics. *Comm. Statist. Theory Methods* **19**, 4587-4590.

ZHANG, F. (2001). Equivalence of the Wielandt inequality and the Kantorovich inequality. *Linear and Multilinear Algebra* **48**, 275-279.

ZHANG, Y.-T. (1978). Generalized correlation coefficients and their applications (in Chinese). *Acta Math. Appl. Sinica* **1**, 312-320.

S.W. DRURY
DEPARTMENT OF MATHEMATICS
AND STATISTICS, MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTRÉAL, QUÉBEC, CANADA H3A 2K6

SHUANGZHE LIU
SCHOOL OF MATHEMATICAL SCIENCES
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200, AUSTRALIA
E-mail: lius@maths.anu.edu.au

CHANG-YU LU
DEPARTMENT OF STATISTICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI 200062, CHINA
E-mail: cylu@nenu.edu.cn

SIMO PUNTANEN
DEPARTMENT OF MATHEMATICS,
STATISTICS AND PHILOSOPHY
FIN-33014
UNIVERSITY OF TAMPERE, FINLAND
E-mail: sjp@uta.fi

GEORGE P.H. STYAN
DEPARTMENT OF MATHEMATICS
AND STATISTICS
MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTRÉAL, QUÉBEC
CANADA H3A 2K6
E-mail: styan@math.mcgill.ca