Matrix Inequalities and their Applications to Statistics

by

Rakesh Rajan Singh Bisen

A Thesis

in

The Department

of

Mathematics & Statistics

Presented in Partial Fulfillment of the Requirements for

the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

June, 1992

ISBN 0-315-80960-4

Abstract

Matrix Inequalities and their Applications to Statistics

by

Rakesh Rajan Singh Bisen

The thesis consists of a survey of matrix inequalities and their applications to statistics. Chapter one provides some definitions which are used in the sequel. In chapter two a matrix generalisation of Minkowski, Cauchy-Schwartz inequalities along with the inequalities concerning positive definite, Hermitian matrices, λ and P matrices. The third chapter includes a variety of results on the localisation of the characteristic roots of a matrix and inequalities thereof.

This thesis concerns further with the application of some of the inequalities in statistics.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. T. D. Dwivedi for his guidance, valuable insight, thorough understanding and generous support during the preparation of this thesis.

I am also grateful to Prof. Y. P. Chaubey for giving valuable suggestions which have helped me to improve the presentation of this thesis.

My heartiest thanks go to Dr. T. N. Srivastava and Prof. Z. Khalil for their deep concern and continuous encouragement at various stages of this programme.

I would further like to thank Mrs. Bonnie Campbell, Mrs. Cynthia Parkinson, Miss Kathryn Nye and Miss Nancy Gravenor for their valuable suggestions and help during typing the manuscript. I would also like to thank my friends Christos Grecos, Pankaj Kamthan and Satya Prakash for their critical remarks towards the improvement of the thesis.

Last but not the least, I am indebted to my parents for their patience, understanding and encouragement during my studies.

(Rakesh Rajan Singh Bisen)

Special Notations

$\Re$	the set of all real numbers.
$\mathbb{C}$	the set of all complex numbers.
O	zero matrix.
I	identity matrix.
A^T	transpose of A (A^t or A').
A^*	conjugate transpose of A.
$\bar{A}$	conjugate of A.
A^{-1}	inverse of A.
A^+	generalized inverse.
A_{ij}	cofactor of a_{ij}.
Adj A	adjoint of A.
$A \oplus B$	direct sum.
$A \otimes B$	Kronecker product.
Tr (A)	trace of the matrix A (or tr A).
r (A)	rank of the matrix A.
λ	eigenvalue, or parameter in $(A - \lambda I)X = B$.
λ_i, X_i	eigenvalues and corresponding eigenvectors of matrix, where $i = 1, \ldots, n$.
$W^\perp$	subspace of vectors orthogonal to W.
$\mathrm{diag}(a_1,\ldots,a_n)$	the diagonal matrix with $a_1,\ldots,a_n$ down the main diagonal.

TABLE OF CONTENTS

CHAPTER 1

Introduction

Matrix theory started with the work of Hamilton, Cayley and Sylvester. The fact, the term "Matrix" was coined by J. J. Sylvester in 1850. Even though its one of the recent mathematical development, it has become a great tool of research workers in almost every discipline; for example in Quantum Theory, Classical Mechanics, Aeronautical, Mechanical, Electrical Engineering, Linear Programming, Economics, Environmental Studies, Psychology and Statistics. Thus the knowledge of matrix analysis has become must for researchers in almost every branch of knowledge.

The modern research in matrix analysis blends an extraordinary variety of mathematical tools; this is one of the features that makes matrix analysis exciting, as well as adds to the vitality of mathematics by linking many mathematical subjects to applications. The rest of this chapter includes the preliminaries for the development of the subject matter included in the thesis.

Preliminaries

Basic Definitions

Singular Value: If A is a general $m \times n$ matrix, μ is a nonzero number, and U, V are vectors such that $A\,U = \mu\, V$, $A^H\, V = \mu\, U$, then μ is called a *singular value* of A and U,V are called a pair of singular vectors corresponding to A.

Involutory: A square matrix A such that $A^2 = I$ is called *involutory*.

Normal Matrix: A matrix A is said to be *normal* if and only if $A^H A = A A^H$. Simple examples of normal matrices are unitary, hermitian, and skew-hermitian matrices, and diagonal matrices with arbitrary elements.

Definition: A point $x \in R$ is called a *contact point* of a set $M \subset R$ if every neighborhood of x contains at least one point of M. The set of all contact points of a set M is denoted by [M] and is called the *closure* of M. Obviously $M \subset [M]$, since every point of M is a contact point of M.

Matrix Norms.

Definition: The *matrix norm* of a square matrix A is a non-negative number denoted by $\| A \|$, associated with A, such that

(a) $\| A \| > 0$ for $A \neq 0$, $\| A \| = 0 \Rightarrow A = 0$

(b) $\| k A \| = k \| A \|$ for any scalar k.

(c) $\| A + B \| \leq \| A \| + \| B \|$

(d) $\| A B \| \leq \| A \| \| B \|$

A matrix norm is said to be compatible with a vector norm $\| X \|$ if

(e) $\| A X \| \leq \| A \| \| X \|$

We now show how to derive, from a given vector norm, a matrix norm that is compatible with the vector norm. Since we must have $\| A X \| \leq \| A \| \| X \|$, we define

$$\| A \| = \sup_{X \neq 0} \frac{\| A X \|}{\| X \|}$$

Where sup denotes least upper bound for all $X \neq 0$. If we introduce $Z = \dfrac{X}{\| X \|}$,

we see that $\| Z \| = 1$ and $\dfrac{\| A X \|}{\| X \|} = \| A Z \|$

$$\|A\| = \sup_{\|Z\|=1} \|AZ\| \qquad (1)$$

Definition: A matrix norm constructed by means of (1) is said to be the natural norm associated with the vector norm. We also say that the matrix norm is subordinate to, or induced by the vector norm.

Spectral matrix: Let A be an $n \times n$ matrix that has n linearly independent eigenvectors X_1, $X_2, \ldots, X_n$ which correspond to the eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_n$. Define

$$M = [\, X_1, X_2, \ldots, X_n \,] \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \lambda_n \end{bmatrix}$$

Here M is called a *modal matrix*, and D is called a *spectral matrix* for A.

Kronecker Product: If $A = [\, a_{ij} \,] \in M_{m,n}$ and $B = [\, b_{ij} \,] \in M_{p,q}$ are given, then their *Kronecker product* $A \otimes B$ (also called the *tensor product*) is the mp-by-nq matrix $[a_{ij}B]$.

Hadamard Product: The *Hadamard product* of two matrices $A = [\, a_{ij} \,]$ and $B = [\, b_{ij} \,]$ with the same dimensions (not necessarily square) with entries in a given ring is the entry-wise product $A \circ B \equiv [\, a_{ij}b_{ij} \,]$, which has the same dimensions as A and B. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{ and } B = \begin{bmatrix} i & \pi & 0 \\ e & 0 & 1 \end{bmatrix} \text{ then } A \circ B = \begin{bmatrix} i & 2\pi & 0 \\ 4e & 0 & 6 \end{bmatrix}.$$

If A and B are matrices with the same dimensions but not square, then the ordinary matrix product is not defined, but $A \circ B$ is defined; in the square case both products are defined.

Stable Matrix: A matrix A is called *stable* if the real parts of all its eigenvalues are less than zero.

Lyapunov Equation: In the study of stability of local equilibria for a systems of ordinary differential equations, one is naturally led to consider matrices whose eigenvalues all lie in the half-plane; a square complex matrix is said to be positive stable if all its eigenvalues have positive real part. A positive definite matrix is positive stable, of course, but there is a more subtle connection between the positive definite matrices is positive stable matrices. A celebrated theorem of Lyapunov says that a given matrix $A \in M_n$ is positive stable if and only if there is a positive definite $G \in M_n$ such that the Hermitian matrix $(GA) + (GA)^* = GA + A^*G$ is positive definite.

Diagonal Dominant: A square matrix A is said to be *diagonal dominant*, if n×n matrix A having this property,

$$|a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}|, \qquad i = 1, 2, \ldots, n.$$

Spectral Radius: The *spectral radius* of an n×n (complex) matrix A having eigenvalues $\lambda_1, \ldots, \lambda_n$ is defined by

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

Companion Matrices: The nth degree polynomial

$$a(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n$$

an n×n *companion matrix* (the n×n matrix having 1's along the diagonal immediately above the principal diagonal, having last row $|-a_n, -a_{n-1}, \ldots, a_2, a_1|$, and zero everywhere else.)

$$C = \begin{bmatrix} 0 & & & I_{n-1} \\ -a_n & \ldots & -a_2 & -a_1 \end{bmatrix},$$

for which

$$\det(\lambda I_n - C) = a(\lambda).$$

Block Companion Matrices: A second , and quite different , type of primeness associated with two polynomial matrices involves their determinants. Suppose that the $n \times n$ matrix

$$B(\lambda) = B_0\lambda^M + B_1\lambda^{M-1} + \cdots + B_{M-1}\lambda + B_M,$$

where each B_i is a constant $n \times n$ matrix $B(\lambda)$ having degree M is *regular*, so that we can set $B_0 = I_n$. By analogy with the companion matrix associated with a scalar polynomial we define the $nM \times nM$ *block companion matrix*

$$C_B = \begin{bmatrix} 0 & I_n & & & \\ 0 & & I_n & & 0 \\ & & & \ddots & \\ & & & & I_n \\ -B_M & -B_{M-1} & \cdots & -B_2 & -B_1 \end{bmatrix},$$

which has the fundamental property

$$\det(\lambda I_{nM} - C_B) = \det B(\lambda).$$

When $n = 1$, C_B reduces to the companion matrix C.

CHAPTER 2

Matrix Inequalities

2.1 Minkowski inequality and M-Matrices.

The well known Minkowski's inequality (Magnus and Neudecker [75]) in its most simple form is given as follows:

$$\left[(x_1 + y_1)^p + (x_2 + y_2)^p \right]^{1/p} \le (x_1^p + x_2^p)^{1/p} + (y_1^p + y_2^p)^{1/p} \tag{2.1.1}$$

for every non-negative x_1, x_2, y_1, y_2, and $p > 1$. It is noted that (2.1.1) can be extended in the following two directions. We have

$$\left(\sum_{i=1}^{m} (x_i + y_i)^p \right)^{1/p} \le \left(\sum_{i=1}^{m} x_i^p \right)^{1/p} + \left(\sum_{i=1}^{m} y_i^p \right)^{1/p} \tag{2.1.2}$$

for every $x_i \ge 0$, $y_i \ge 0$ and $p > 1$; and also

$$\left[\left(\sum_{j=1}^{n} x_j \right)^p + \left(\sum_{j=1}^{n} y_j \right)^p \right]^{1/p} \le \sum_{j=1}^{n} (x_j^p + y_j^p)^{1/p} \tag{2.1.3}$$

for every $x_j \ge 0$, $y_j \ge 0$ and $p > 1$. Notice that if in (2.1.3) we replace x_j by $x_j^{1/p}$, y_j by $y_j^{1/p}$ and then p by 1/p, we obtain

$$\left(\sum_{j=1}^{n} (x_j + y_j)^p \right)^{1/p} \ge \left(\sum_{j=1}^{n} x_j^p \right)^{1/p} + \left(\sum_{j=1}^{n} y_j^p \right)^{1/p} \tag{2.1.4}$$

for every $x_j \geq 0$, $y_j \geq 0$ and $0 < p < 1$.

It is not difficult to see that all these cases are included in the following inequality.

Let $X = (x_{ij})$ be a non-negative m×n matrix (that is, $x_{ij} \geq 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$), and let $p > 1$. Then

$$\left[\sum_{i=1}^{m}\left(\sum_{j=1}^{n} x_{ij}\right)^{p}\right]^{1/p} \leq \sum_{j=1}^{n}\left(\sum_{i=1}^{m} x_{ij}^{p}\right)^{1/p} \qquad (2.1.5)$$

with equality if and only if $r(X) = 1$.

The extension of the above matrices follows as given below. Let A and B be two positive semidefinite matrices of the same order, then for $p > 1$,we have

$$[\operatorname{tr}(A+B)^{p}]^{1/p} \leq (\operatorname{tr} A^{p})^{1/p} + (\operatorname{tr} B^{p})^{1/p}. \qquad (2.1.6)$$

with equality if and only if $A = \mu B$ where $\mu(>0)$ is a real number.

The proof of this is straight forward (Magnus and Neudecker [75]).

We now state the Minkowski's determinant theorem [84].

Theorem 2.1.1 : If A and B are each n×n positive definite (symmetric) matrices, then

$$|A|^{1/n} + |B|^{1/n} \leq |A+B|^{1/n} \qquad (2.1.7)$$

where |.| denotes determinant. An analogue of the above result for the M–matrix can be obtained. First we define an M-matrix.

Definition 2.1.1 : An M-matrix is a square matrix whose off-diagonal entries are nonpositive and which has all positive principal minors.

Using the M-matrices an analogue of the Minkowski's determinant theorem can be proved as follows.

Theorem 2.1.2 [79] **:** Suppose each of A and B is an $n \times n$ M-matrix with the property that there is a vector $u > 0$ such that $Au > 0$ and $Bu > 0$. Then

$$|A| + |B| \leq |A+B|$$

and

$$|A^{-1}| + |B^{-1}| \leq |A^{-1} + B^{-1}|.$$

Proof : Since $Au > 0$, there is a positive diagonal matrix D such that AD is row diagonally dominant. In fact D can be chosen as $D = \mathrm{diag}(u_1, \ldots, u_n)$ where u_i are co-ordinates of u. Similarly, for the same diagonal matrix D, we have that BD is row diagonally dominant, and also (A+B)D is row diagonally dominant. Haynsworth [50] proved that if A and B are diagonally dominant $n \times n$ matrices with positive diagonal, then

$$|A| + |B| \leq |A+B|$$

Hence we have $|(A+B)D| \geq |AD| + |BD|$, and the first determinantal inequality holds. Since $|A^{-1}| > 0$ and $|B^{-1}| > 0$, then we have

$$|A^{-1}|\,|A+B|\,|B^{-1}| \geq |A^{-1}|\,|\{|A| + |B|\}|\,|B^{-1}|,$$

and the second inequality follows.

Observation about the above theorem. If

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

where B is the transpose of A, then

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow u_1 - u_2 > 0, \; u_2 > 0.$$

and

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow u_1 > 0, \quad -u_1 + u_2 > 0.$$

Thus $u_1 - u_2 > 0$ and $u_2 - u_1 > 0$ is a contradiction.

Therefore it is obvious that strong row diagonal dominance does not necessarily imply (2.1.7).

2.2 Comparison of H-matrices with M-matrices.

Recently multicolinearity in the design matrix in linear models have been talked about in the literature to a great extent. Several methods have been designed in numerical analysis to determine the inverse of such matrices which could provide meaningful least square estimate. These methods depend upon the condition number of the matrix in the question. H-matrices produce an equality which helps us in estimating the condition numbers and constructing the error bounds for the solution of equations. In order to proceed we first provide some definitions.

Definition 2.2.1 (Comparison matrix) : (Ostrowski [90]) Associates with every matrix A = (a_{ik}) the comparison matrix

$$\langle A \rangle = (\in_{ik} |a_{ik}|), \qquad \in_{ik} = \begin{cases} 1 & \text{if } i = k \\ -1 & \text{otherwise} \end{cases} \tag{2.2.1}$$

$\langle A \rangle$ is obtained from A if we replace the diagonal entries by their absolute values, and the off-diagonal entries by their negative absolute values. The matrix A is an M-matrix if $\langle A \rangle$ = A and if there is a vector $u > 0$ with $Au > 0$.

Definition 2.2.2 : The matrix A is called an H-matrix if there is a vector $u > 0$ with

$\langle A\rangle u > 0$, where $\langle A\rangle$ is the comparison matrix corresponding to A.

An old theorem of Ostrowski states that the absolute value of the inverse of an H-matrix is, componentwise, bounded by the inverse of a related M-matrix.

Ostrowski [90] showed that for H-matrices

$$|A^{-1}| \leq \langle A\rangle^{-1} \tag{2.2.2}$$

Equation (2.2.2) is relevant in the context of estimating matrix condition numbers and constructing error bounds for solutions of equations.

Theorem 2.2.1 [86] **:** Let A be an H-matrix and

$$\Delta := I - \mathrm{Diag}(A)^{-1}A , \tag{2.2.3}$$

$$\Omega := \left(I - |\Delta|\right)^{-1} \left(|\Delta| - \Delta\right) \tag{2.2.4}$$

Then $\Omega \geq 0$ and

$$|A^{-1}| \leq \langle A\rangle^{-1} \leq (1 + \Omega)\,|A^{-1}| \tag{2.2.5}$$

Corollary 2.2.1 : Under the hypothesis of Theorem 2.2.1,

$$\| \, |A^{-1}|c \, \| \leq \| \langle A\rangle^{-1}c \, \| \leq (1 + \omega) \| \, |A^{-1}|c \, \| \tag{2.2.6}$$

for all $c > 0$ and every monotone matrix norm with $\| I \| = 1$ and $\| \Delta \| < 1$;

here

$$\omega = \frac{\| \, |\Delta| - \Delta \, \|}{1 - \| \Delta \|}. \tag{2.2.7}$$

Theorem 2.2.2 : Let A be a nonsingular $n\times n$ matrix possessing a decomposition $A = LR$ into the product of two triangular matrices. Then

$$|A^{-1}| \leq \langle R\rangle^{-1}\langle L\rangle^{-1}. \tag{2.2.8}$$

Moreover, if A is an H-matrix then

$$| A^{-1} | \leq \langle R \rangle^{-1} \langle L \rangle^{-1} \leq \langle A \rangle^{-1}. \qquad (2.2.9)$$

The proofs are given in Neumaier [86] .

2.3 Variations on Cauchy-Schwartz inequality.

Let V be an n-dimensional inner product space, and let A, B, P, Q be linear on V. What relations must exist among these operators so that the inequality (Marcus [77])

$$(Av, u)(Bu, v) \leq (Pu, u)(Qv, v) \qquad (2.3.1)$$

holds for all u and v in V. If V is Euclidean (i.e the underlying field is $\mathbb{R}$), the situation can be chaotic. For example, take $A = I$, $B = -I$, P skew-symmetric, Q arbitrary then (2.3.1) holds.

Theorem 2.3.1: Let V a unitary space, dim $V \geq 3$, and assume that A, B, P, Q are nonsingular. Then (2.3.1) holds for all u and v iff

(i) $P = \alpha H$, $Q = \beta K$, $\alpha\beta = \epsilon = \pm 1$, H and K are definite Hermitian and

(ii) $A^* = \lambda B$, λ real, so that (1) reads

$$\lambda \, | (Bu,v) |^2 \leq \epsilon \, (Hu,u)(Kv,v) \qquad (2.3.2)$$

and

(iii) If $\epsilon = 1$, $\lambda > 0$, then H and K have the same sign (i.e both positive definite or both negative definite) and

$$\lambda_{max} (P^{-1}AQ^{-1}B) \leq 1 \qquad (2.3.3)$$

or

(iv) If $\epsilon = 1$, $\lambda < 0$, then H and K have the same sign; or

(v) If $\epsilon = -1$, $\lambda > 0$, then H and K have opposite signs

and

$$\lambda_{max}(P^{-1}AQ^{-1}B) \le 1. \tag{2.3.4}$$

(vi) If $\epsilon = -1$, $\lambda < 0$, then H and K have opposite signs.

(λ_{max} is the maximum eigenvalue of the indicated operator.)

It should be noted that for $n = 2$ the inequality (2.3.2) can hold with both H and K indefinite. Simply take

$$B = I_2, \quad H = K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \lambda = \epsilon = -1$$

The inner product in V induces an inner product in the second tensor space $V \otimes V$ that satisfies

$$(x \otimes y,\ u \otimes v) = (x,u)(y,v) \tag{2.3.5}$$

with respect to inner product (2.3.5) the interchange operator $\sigma : V \otimes V \to V \otimes V$

$$\sigma x \otimes y = y \otimes x$$

is obviously Hermitian and satisfies $\sigma^2 = I$. It is convenient to interpret (2.3.1) in terms of mappings on $V \otimes V$. Let

$$\mathcal{L} = P \otimes Q - (A \otimes B)\sigma, \tag{2.3.6}$$

and note that

$$(\mathcal{L}u \otimes v,\ u \otimes v) = (Pu,u)(Qv,v) - (Av,u)(Bu,v). \tag{2.3.7}$$

Let $\mathcal{D}$ denote the set of decomposable elements in $V \otimes V$, i.e. elements of the form $u \otimes v$, where u & v in V. Then we see from (2.3.6) and (2.3.7) that (2.3.1) holds iff

$$(\mathcal{L}z, z) \ge 0 \text{ for all } z \in \mathcal{D}. \tag{2.3.8}$$

We remark that the condition (2.3.8) is not equivalent to $\mathcal{L} \ge 0$ (i.e $\mathcal{L}$ positive semidefinite). Let $S_1 = I + \sigma$, $S_2 = I - \sigma$ on span of dimension 2 and compute that

$$(S_1 - S_2 v_1 \otimes v_2,\ v_1 \otimes v_2) = \text{per } A - \det A,$$

where A is a 2×2 matrix

$$A = \begin{bmatrix} (v_1, v_1) & (v_1, v_2) \\ (v_2, v_1) & (v_2, v_2) \end{bmatrix}$$

An inequality of Schur [99] states that per A ≥ det A. However

$$(S_1 - S_2)S_2 v_1 \otimes v_2 = -2S_2 v_1 \otimes v_2.$$

Lemma 2.3.1 : If $\mathcal{L}$ is any linear operator on $V \otimes V$, then $\mathcal{L} = 0$ iff $(\mathcal{L}z,z) = 0$ for all $z \in \mathcal{D}$.

Lemma 2.3.2 : (a) $P \otimes Q$ is Hermitian iff $P = \alpha H$, $Q = \beta K$, where H and K are Hermitian and $|\alpha| = |\beta| = 1$, $\alpha\beta = \epsilon = \pm 1$.

(b) $A \otimes B\sigma$ is Hermitian iff $A^* = \lambda B$, $B^* = \mu A$, λ and μ are real and $\lambda\mu = 1$.

Lemma 2.3.3 : The inequality (2.3.1) holds iff

$$(\mathcal{L}z,z) \geq 0, \qquad \text{all } z = u \otimes v \in \mathcal{D} \tag{2.3.9}$$

for the operator

$$\mathcal{L} = \epsilon H \otimes K - \lambda(B^* \otimes B)\sigma, \quad \lambda \in \mathbb{R}, \quad \epsilon = \pm 1.$$

Note that (2.3.11) is precisely the same as

$$\epsilon (Hu,u)(Kv,v) \geq \lambda\, |(Bu,v)|^2, \qquad u,v \in V \tag{2.3.10}$$

Lemma 2.3.4 : If $\epsilon = 1$, $\lambda > 0$, then (2.3.1) holds iff H and K are definite of the same sign and

$$\lambda_{max}(P^{-1}AQ^{-1}B) \leq 1.$$

Lemma 2.3.5 : If $\epsilon = -1$, $\lambda > 0$, then (2.3.1) holds iff H and K are definite of opposite signs and

$$\lambda_{max}(P^{-1}AQ^{-1}B) \leq 1 \tag{2.3.11}$$

Lemma 2.3.6 : Let $n \geq 3$; assume that B, H, K are nonsingular and that H and K are Hermitian. Then

$$|(Bu, v)|^2 \geq (Hu,u)\,(Kv,v), \qquad u,v \in V \tag{2.3.12}$$

iff H and K are of opposite signs.

Proof of Lemmas : See [77] .

The Theorem 2.3.1 is now simply a combination of Lemmas 2, 4, 5, and 6. We remark that replacing A by –A and P by –P in the inequality

$$(Av, u)\,(Bu, v) \geq (Pu, u)\,(Qv, v) \tag{2.3.13}$$

reduces (2.3.13) to (2.3.1) and the theorem can be applied to the operators –A, B, -P, Q.

2.4 λ roots of an λ-matrix.

It is well known that the eigenvalues play an important role in the study of matrices. So in the following, we give the relationship of the eigenvalues of M-matrices.

$$\text{Let } A(\lambda) = I\,\lambda^n + A_1\,\lambda^{n-1} + A_2\,\lambda^{n-2} + \ldots + A_{n-1}\,\lambda + A_n,$$

be the λ-matrix where the coefficients A_i are s×s complex matrices,

$$A_i \in M_{s,s}(\mathbb{C}),\ (i = 1,2,\ldots,n).$$

The latent roots λ of A(λ) are the zeros of det A(λ). Denoting by C the block companion matrix of A(λ), forming a matrix C^HC; letting

$D = \operatorname{diag}(\,\|A_{n-1}\|_i\, I,\ \|A_{n-2}\|_i\, I, \ldots, \|A_1\|_i\, I,\ I\,)$, where $\|.\|_i$, $(i = 1, \infty)$ is a scalar norm;

For $B = (\beta_{ij}) \in M_{s,s}(\mathbb{C})$, we let

$$\|B\|_1 := \operatorname*{Max}_{j=1,2,\ldots,s} \left\{ \sum_{i=1}^{s} |\beta_{ij}| \right\}, \quad \|B\|_\infty := \operatorname*{Max}_{i=1,2,\ldots,s} \left\{ \sum_{j=1}^{s} |\beta_{ij}| \right\}$$

and take the scalar norm $\|.\|_i$ of each block of C^HC– so obtaining a matricial norm Φ_i, $(i = 1, \infty)$ as follows (Vitoria [105])

$$\Phi_i(D^{-1}C^HCD) \le \left(\begin{array}{c|c} I & \begin{matrix}1\\ \cdot\\ \cdot\\ \cdot\\ 1\end{matrix} \\ \hline \|A_{n-1}^H\|_i \cdot \|A_{n-1}\|_i \ldots \|A_1^H\|_i \|A_1\|_i : & \sum_{l=1}^{n} \|A_l^H\|_i \|A_l\|_i \end{array}\right) \quad (*)$$

$$(i = 1, \infty).$$

Given matrix M we have the following relations.

$\rho(M) \le \rho(\Phi_i(M))$ where ρ is spectral radius, $\Phi_i(M)$ is matricial norm. We specialize (*) in two way

(i) By taking a (scalar) norm, we get the following inequality

$$|\lambda| \le \left(1 + \sum_{l=1}^{n} \|A_l^H\|_i \|A_l\|_i\right)^{1/2}, \qquad (i = 1, \infty). \qquad (2.4.1)$$

(ii) By calculation of the eigenvalues, one obtains

$$|\lambda| \le \left(\frac{1 + \sum_{l=1}^{n} \|A_l^H\|_i \|A_l\|_i + \sqrt{\left(1 + \sum_{l=1}^{n} \|A_l^H\|_i \|A_l\|_i\right)^2 - 4\|A_n^H\|_i \|A_n\|_i}}{2}\right)^{1/2} \qquad (2.4.2)$$

$$(i = 1, \infty)$$

which gives a sharper bound, but is less simple.

A particular case is given with respect to the complex polynomial

$$p(z) = z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n$$

We have respectively, from (2.4.1) and (2.4.2):

$$|z| \leq \left(1 + \sum_{i=1}^{n} |a_i|^2\right)^{1/2}$$

which is an inequality by Vitoria [105] and

$$|z| \leq \left(\frac{1 + \sum_{i=1}^{n} |a_i|^2 + \sqrt{\left(1 + \sum_{i=1}^{n} |a_i|^2\right)^2 - 4|a_n|^2}}{2} \right)^{1/2}$$

an inequality by Parodi [91].

2.5 Determinantal inequality involving the Moore-Penrose inverse.

The inverse of a matrix is defined when the matrix is square and non-singular. For many purposes it is useful to generalize the concept of invertibility to singular matrices and, indeed, to non-square matrices. One such generalization that is particularly useful because of its uniqueness is the Moore-Penrose inverse.

Let A be a rectangular matrix of complex numbers whose m rows of $A \in C^{m \times n}$ are partioned into r arbitrary blocks:

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_r \end{bmatrix}, \qquad A_i \in C^{k_i \times n}, \qquad i = 1, \ldots, r, \quad k_1 + \cdots + k_r = m.$$

Definition: The Moore-Penrose inverse of a matrix A_i, $i = 1, \ldots, r$ is the unique matrix A_i^+ satisfying the relations

$$A_iA_i^+A_i = A_i, \quad A_i^+A_iA_i^+ = A_i^+, \quad (A_iA_i^+)^* = A_iA_i^+, \quad (A_i^+A_i)^* = A_i^+A_i,$$

where * denotes the conjugate transpose. These are used to form the matrix $B = (A_1^+, \ldots ,A_r^+) \in C^{n\times m}$.

Clearly, the determinant of the m×m matrix AB will be zero unless A has full row rank. Therefore, from now on, it will assumed that rank(A) = m ≤ n. Thus

$$A_i^+ = A_i^*(A_iA_i^*)^{-1}, \qquad i = 1, \ldots, r;$$

hence

$$B = (A_1^*, \ldots ,A_r^*)\begin{pmatrix} (A_1A_1^*) & & 0 \\ & \cdot & \\ 0 & \cdot & (A_rA_r^*)^{-1} \end{pmatrix} = A^*D,$$

where D is m×m. This leads to

$$\det(AB) = \det(AA^*)\det(D) = \frac{\det(AA^*)}{\prod_{i=1}^{r} \det(A_iA_i^*)} \tag{2.5.1}$$

equal to a real positive number. We will show that det(AB) ≤ 1.

Observe that

$$AB = \begin{pmatrix} I & A_1A_2^+ & \ldots & A_1A_r^+ \\ A_2A_1^+ & I & \ldots & A_2A_r^+ \\ \vdots & \vdots & & \vdots \\ A_rA_1^+ & A_rA_2^+ & \ldots & I \end{pmatrix},$$

and consequence of the formula

$$\det\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} = \det(M_1)\det(M_4 - M_3M_1^{-1}M_2)\ ,$$

M_1 nonsingular, is the relation

$$\det(AB) = \det\left(\begin{bmatrix} A_2 \\ \vdots \\ A_r \end{bmatrix}(I - A_1^+A_1)(A_2^+, \ldots, A_r^+)\right)$$

$$= \det(\tilde{A}G\tilde{A}^*)\det(\tilde{D})\ , \qquad (2.5.2)$$

where

$$\tilde{A} = \begin{bmatrix} A_2 \\ \vdots \\ A_r \end{bmatrix}, \qquad G = I - A_1^+A_1, \qquad \text{and} \qquad \tilde{D} = \begin{pmatrix} (A_2A_2^*)^{-1} & 0 \\ 0 & (A_rA_r^*)^{-1} \end{pmatrix}$$

Since G and $A_1^+A_1$ are both Hermitian and idempotent,

$$(\tilde{A}G)(\tilde{A}G)^* = \tilde{A}G\tilde{A}^* = \tilde{A}\tilde{A}^* - (\tilde{A}A_1^+A_1)(\tilde{A}A_1^+A_1)^*. \qquad (2.5.3)$$

The two matrices on the right can be simultaneously reduced to the diagonal form. Indeed, let V be the unitary matrices which reduces the Hermitian positive definite matrix $\tilde{A}\tilde{A}^*$ to diag$(\mu_1, \ldots, \mu_{m-k_1})$. Since all μ_i are real and positive, let

$$E = \mathrm{diag}\,(\mu_1^{-1/2}, \ldots, \mu_{m-k_1}^{-1/2}) = E^*,$$

then (2.5.3) becomes

$$(EV\tilde{A}G)(EV\tilde{A}G)^* = I - (EV\tilde{A}A_1^+A_1)(EV\tilde{A}A_1^+A_1)^*$$

Now let U be the unitary matrix which reduces the last matrix on the right to diag($\sigma_1, \ldots$ $,\sigma_{m-k_1}$), $\sigma_i \geq 0$. Then

$$(UEV\tilde{A}G)(UEV\tilde{A}G)^* = \mathrm{diag}\,(1-\sigma_1, \ldots, 1-\sigma_{m-k_1}).$$

But, from (2.5.1) and (2.5.2), $\tilde{A}G\tilde{A}^*$ is nonsingular, so that the Hermitian matrix on the left is positive definite. Thus $0 \leq \sigma_i < 1$, $i = 1, \ldots, m-k_1$, and taking determinants on each side yields,

$$\det(\tilde{A}G\tilde{A}) = \det(E^{-2}) \cdot \prod_{i=1}^{m-k_1} (1 - \sigma_i)$$

$$= \prod_{i=1}^{m-k_1} \mu_i (1 - \sigma_i) \leq \prod_{i=1}^{m-k_1} \mu_i = \det(\tilde{A}\tilde{A}^*).$$

In view of (2.5.2), we have

$$\det(AB) \leq \det\left(\begin{bmatrix} A_2 \\ \vdots \\ A_r \end{bmatrix} (A_2^+, \ldots, A_r^+)\right)$$

Replacing the same procedure, we find, eventually (Lavoie [70]) that

$$\det(AB) \leq \det(A_r A_r^+) = 1.$$

This completes the proof.

2.6 Inequality of Goldberg and Straus.

Let $c = \{c_j\}$ and $z = \{z_j\}$ be sequences of complex numbers for $j = 1,2,\dots,m$ with $m \geq 2$. Leaving c fixed but permuting z, we can form various sums of the products, such as

$$P = c_1 z_1 + c_2 z_2 + \dots + c_m z_m$$

$$Q = c_1 z_m + c_2 z_{m-1} + \dots + c_m z_1$$

$$R = c_1 z_m + c_2 z_2 + \dots + c_{m-1} z_{m-1} + c_m z_1$$

It is desired to find at least one expression like this which is large. More specifically, one would like to find a positive constant C such that

$$\max_\pi \left| \sum_{i=1}^{m} c_i z_{\pi(i)} \right| \geq C \max_i |z_i|, \qquad (2.6.1)$$

where π denotes a permutation of $(1, 2, \dots, m)$, the sequence c is fixed and z is arbitrary.

One cannot have $C > 0$ in (2.6.1) if the sum of all the c_j is 0, since then we could take $z_j = 1$. If all the c_j are equal, the inequality fails again since we could choose z_j to be any numbers whose sum if 0. Hence the constant C in (2.6.1) ought to involve at least the two parameters

$$s = |c_1 + c_2 + \dots + c_m| \quad \text{and} \quad d = \max_{i,j} |c_i - c_j| \qquad (2.6.2)$$

Theorem 2.6.1 : If $sd \neq 0$ inequality (2.6.1) holds with $C = sd/(2s + d)$

For the proof see Goldberg and Straus [43].

Goldberg has asked whether the constant C given by Theorem 2.6.1 is sharp. The answer of this question is given in the following theorem.

Theorem 2.6.2 : If $sd \neq 0$, the best value of the constant C in (2.6.1) satisfies

$$\frac{sd}{2s+d-2s/m} \leq C \leq \min\left\{s, \frac{sd}{2s+d-2s/m-2d/m}\right\}$$

and the inequality on the right becomes an equality when c and z are real.

This shows that, while the Goldberg-Straus constant is not optimum for any m, it is the best that can be chosen independently of m even if the sequences are real. We shall see that Theorem 2.6.2 remains valid when $c_j \in V$, $z_j \in \mathbf{C}$ or $c_j \in \mathbf{C}$, $z_j \in V$, where V is an arbitrary vector space over the complex field $\mathbf{C}$. The fact that the optimum constant (independent of m) is the same for this case as for the real case is another surprising aspect of Theorem 2.6.1.

Proof of the left hand inequality.

Let the z_j be so numbered that $|z_m| = \max |z_j|$, and so that

$$|z_m - z_1| = \max_j |z_m - z_j| = t\,|z_m|, \tag{2.6.3}$$

where t is defined by this equation. We suppose the c_j numbered so that $d = |c_m - c_1|$. Subtracting the above expressions P, R we then get

$$2\max(|P|, |R|) \geq |P - R| \geq dt\,|z_m|.$$

This shows that the constant C always satisfies $C \geq dt/2$.

Now let us make a cyclic permutation of z_j to get $P_j = c_1 z_{1+j} + c_2 z_{2+j} + \cdots + c_m z_{m+j}$.

(For a notational convenience, $z_{m+1} = z_i$.) Then

$$|P_1 + P_2 + \ldots + P_m| = s\,|z_1 + z_2 + \ldots + z_m|$$

$$= s\,|mz_m + (z_1 - z_m) + (z_2 - z_m) + \ldots + (z_{m-1} - z_m)|$$

$$\geq s\,|z_m|\,[m - (m-1)t],$$

where the last expression follows from (2.6.3). Since the largest $|P_j|$ is at least equal to the average, it follows that

$$\max_j |P_j| \geq s\,|z_m|\,[1-(1-1/m)t]$$

and hence $C \geq s - st + st/m$. Thus

$$C \geq \max(dt/2,\ s - st + st/m)$$

no matter what value $t \geq 0$ may have. The choice of t giving the poorest value of C is that for which the two expressions are equal. This completes the proof.

Proof of the right hand inequality : c_i and z_j are real. Since c_i can be multiplied by -1, we assume $s > 0$; and since the desired inequality is homogeneous, we let $\max |z_j| = 1$. Multiplying z_j by -1, if necessary we see that there is no loss of generality in assuming

$$c_1 \leq c_2 \leq \ldots \leq c_m, \qquad -1 \leq z_1 \leq z_2 \leq \ldots \leq z_m = 1, \tag{2.6.4}$$

as well as

$$c_1 + c_2 + \ \ldots\ + c_m = s > 0, \quad c_m - c_1 = d \tag{2.6.5}$$

We refer to (2.6.4) & (2.6.5) as the constrains.

With P and Q as above, the value of the sum for any permutation satisfies

$$Q \leq \sum c_i z_{\pi(i)} \leq P$$

(The general case follows from the case $m = 2$.)

Hence the constant C for the given choice of c and z satisfies

$$C(c,z) = \max(P, -Q).$$

The problem is choice of c and z, subject to the constraints, in such a way that the expression $C(c,z)$ is minimized.

Lemma 2.6.1: Let a_j and b_j be increasing sequences of real numbers for $j = 1,2,\ldots,n$ and let

$$p = a_1b_1 + a_2b_2 + \cdots + a_nb_n, \quad q = a_1b_n + a_2b_{n-1} + \cdots + a_nb_1.$$

Then if all a_j are replaced by their arithmetic mean, the value of p is not increased, and value of q is not diminished.

Let p_j denote the sum obtained by cyclic permutation of the a_j, add the inequalities $q \le p_j \le p$, and divided by n.

According to the Lemma, we can replace the values $z_1, z_2, \ldots, z_{m-1}$ by their mean which we denote by t. This change preserves the constraints and it does not increase P or diminish Q, as will now be shown. We have

$$P = (c_1z_1 + \cdots + c_{m-1}z_{m-1}) + c_mz_m \ge t(c_1 + \cdots + c_{m-1}) + c_mz_m$$

$$Q = c_1z_m + (c_2z_{m-1} + \cdots + c_mz_1) \le c_1z_m + t(c_2 + \cdots + c_m),$$

where we have applied the Lemma, with $n = m - 1$. In the first case, $a_j = z_j$, $b_j = c_j$ in the second case $a_j = z_j$, $b_j = c_{j+1}$.

In view of constrains,

$$P(t) = c_m + t(s - c_m), \qquad Q(t) = P(t) - d(1-t)$$

Evidently P and Q are linear and $P(1) = Q(1) = s$. The constant associated with these sequences

$$C(t) = \max[P(t), -Q(t)]$$

and we want to choose t, $-1 \le t \le 1$, so that this is minimum.

Case 1: $s + d > ms$. In this the constant on the right in Theorem 2.6.2 is s. Since $c_j \le c_m$, we have $s + d \le mc_m$; hence $ms \le mc_m$, in other words, $s < c_m$. This shows that P(t) is decreasing. The value of C(t) is least when t = 1, and in that case C(t) = s. This gives the desired result.

Case 2: $s + d < ms$, $P(-1) + Q(-1) > 0$. If $c_m \geq s$ we can reason as in case 1 and get the value $C = P(1) = s$. Since this is the larger than the constant in Theorem 2.6.2, we consider sequence with $c_m < s$. In this case $P(t)$ is increasing and, since $P(-1) + Q(-1) > 0$, it is readily checked that the t is minimum $C(t)$ is $t = -1$. The value of minimum is $C(-1) = P(-1) = 2c_m - s$.

Case 3: $s + d < ms$, $P(-1) + Q(-1) < 0$. In this case the value of t for minimum $C(t)$ satisfies

$$P(t) + Q(t) = 0, \qquad -1 \leq t \leq 1.$$

$$C(t) = P(t) = -Q(t) = \frac{sd}{(2s + d - 2c_m)} . \tag{2.6.6}$$

Evidently $C(t)$ is minimum when c_m is least, and since

$$c_2 + c_3 + \cdots + c_{m-1} + 2c_m = d + s,$$

c_m is minimized by taking $c_j = c_m$ for $j \geq 2$. The value c_1 is then determined by $c_m - c_1 = d$. Thus the critical sequence are

$$c = \{c_1, c_m, \ldots, c_m\}, \qquad z = \{t, t, \ldots, t, 1\}$$

with $c_m - c_1 = d$, $c_1 + (m-1)c_m = s$.

Thus $mc_m = d + s$ and (2.6.6) reduces to expression in Theorem 2.6.2. The hypothesis $P(-1) + Q(-1) \leq 0$ is equivalent to $d + s \geq 2c_m$ and it holds since $d + s = mc_m$ and $m \geq 2$. This completes the proof.

2.7 Lower bounds for singular values of matrix sums.

It is pointed out by the Lemma 2.7.1 that the result on lower bounds for the absolute singular values of sum of matrices given in Marshall and Olkin [80. p.243 & p.246] is not valid. So in the following, a new result is established on such lower bounds.

Let A and B be n×n complex matrices. The absolute singular values of A, $\sigma_i(A)$, are defined as $\lambda_i^{1/2}(AA^*)$ and the real singular values of A, $\rho_i(A)$, are defined as $\lambda_i(A+A^*/2)$. The singular values are ordered so that they are nonincreasing; e.g. $\sigma_1(A) \geq \ldots \geq \sigma_n(A)$. The result on lower bound for the absolute singular values of $C = A + B$ in terms of $\sigma_i(A)$ and $\sigma_i(B)$ proved in Marshall and Olkin [80, p. 243 and 246] is not valid by a counterexample. A result on lower bounds for $\sigma_i(C)$ is derived in terms of the real singular values, $\rho_i(A)$ and $\rho_i(B)$.

Lemma 2.7.1 : If A and B are n×n complex matrices, then

$$\sum_{i=1}^{k} \sigma_i(A + B) \geq \sum_{i=1}^{k} \sigma_i(A) + \sum_{i=1}^{k} \sigma_{n-i+1}(B) \tag{2.7.1}$$

for k = 1, . . ,n.

For a counterexample to the above Lemma, let A = I and B = −I where I is the identity matrix. This leads to obviously false inequalities. Apparently, the following crucial step used in 'proving' (2.7.1).

$$\min R \operatorname{tr} UBV^* = \sum_{i=1}^{k} \sigma_{n-i+1}(B)$$

where minimum is over k×n unitary matrices U and V, is not valid.

It might be added in passing that the bounds on the eigenvalues of HA, where H is a Hermitian positive definite matrix and A is Hermitian matrix, given in [80, p.510] do not hold in general, unless A is also positive definite. This can easily seen by setting A = −H which leads to obviously invalid inequalities.

Theorem 2.7.1: For n×n complex matrices A, B and C = A + B,

$$\sum_{s=1}^{m} \sigma_{i_s + j_s + m - s - n}(C) \geq \sum_{s=1}^{m} \rho_{i_s}(A) + \sum_{s=1}^{m} \rho_{j_s}(B) \tag{2.7.2}$$

where

$$1 \leq i_1 < \ldots < i_m \leq n$$

$$1 \leq j_1 < \ldots < j_m \leq n \qquad i_1 + j_1 \geq n + 2 - m \tag{2.7.3}$$

Proof : For any n×n complex matrix C, it is shown in [32, p. 112] that

$$\sigma_k(C) \geq \rho_k(C) \tag{2.7.4}$$

for k = 1, . . . ,n. With C = A + B in (4) we have,

$$\sigma_{k_s}(C) \geq \lambda_{k_s}\left(\frac{A + A^*}{2} + \frac{B + B^*}{2}\right) \tag{2.7.5}$$

where

$$1 \leq k_s = i_s + j_s + m - s - n \leq n \tag{2.7.6.}$$

For Hermitian matrices G and H, with (2.7.3) and (2.7.3) holding, it is shown in [102, p. 369] that

$$\sum_{s=1}^{m} \lambda_{k_s}(G + H) \geq \sum_{s=1}^{m} \lambda_{i_s} + \sum_{s=1}^{m} \lambda_{j_s}(H) \tag{2.7.7}$$

Summing (2.7.5) over s = 1 to m and using (2.7.7) in the right hand side of the resultant inequality, yields (2.7.2).

2.8 Intervals of P-matrices and related matrices.

We consider the set of the real n×n matrices the usual partial ordering is defined entrywise. We call an interval of matrices with respect to this partial ordering a matrix interval. We consider matrix intervals which contain only P-matrices or related matrices. A matrix is a P-matrix if all of its principal minors are positive.

All matrices and vectors are assumed to be real and of order n (matrices are square). If A is a matrix, its entries are denoted by a_{ij}. For matrices $\underline{A} = (\underline{a}_{ij})$, $\overline{A} = (\overline{a}_{ij})$ satisfying $\underline{A} \leq \overline{A}$ the matrix interval $[A] = [\underline{A}, \overline{A}]$ is defined by $[\underline{A}, \overline{A}] := \{A \mid \underline{A} \leq A \leq \overline{A}\}$. Vector intervals are defined in an analogous way.

With matrix interval [A] we associate the three subsets,

(i) $V_1[A]$, the set of all matrices A for which $a_{ij} \in \{\underline{a}_{ij}, \overline{a}_{ij}\}$, i,j=1(1)n;

(ii) $V_2[A]$, the set of all matrices $A \in V_1[A]$ for which for i = 1(1)n, $a_{ii} = \underline{a}_{ii}$, and if $a_{ij} = \underline{a}_{ij}$ (respectively, $\overline{a}_{ij}$) then $a_{ji} = \underline{a}_{ji}$ (respectively, $\overline{a}_{ji}$), j = i + 1(1) n; and

(iii) $V_3[A]$, the set of all matrices $A \in V_1[A]$ which can be transformed by simultaneous interchange of rows and columns to a matrix partitioned as

$$\begin{bmatrix} B_1 & C_1 \\ C_2 & B_2 \end{bmatrix},$$

where B_1, B_2 are square and the entries of B_i and C_i are entries of $\underline{A}$ and $\overline{A}$, respectively, i = 1, 2.

Then we can readily confirm that the cardinalities of $V_1[A]$, $V_2[A]$ and $V_3[A]$ are at most 2^{n^2}, $2^{n(n-1)/2}$, and 2^{n-1} respectively. We write V_1, V_2 and V_3 when the matrix interval in the question can be inferred from the context.

We shall consider the following classes of matrices;

P : of all matrices whose principal minors are positive, the P matrices;

A : of all matrices A for which a positive definite diagonal matrix D exists such that $AD + DA^T$ is positive definite, the diagonally stable matrices;

Theorem 2.8.1 [18] : Let [A] be a matrix interval.Then

(i) $[A] \subset P$ iff $V_3 \subset P$,

(ii) $[A] \subset A$ iff $V_2 \subset P$.

Proof : (i) We make use of the following characterization of P-matrices [64, p.98]: $A \in P$ iff for every signature matrix (i.e. diagonal matrix with diagonal elements ± 1) S, there exists a vector $x > 0$ such that $SASx > 0$.

Let $V_3 \subset P$ and $A \in [A]$, let S be any signature matrix, and set $J := \{ i \mid s_{ii} = 1 \}$. Define the matrix B by

$$b_{ij} := \begin{cases} \underline{a}_{ij} & \text{if } j \in J \\ \bar{a}_{ij} & \text{if } j \notin J \end{cases} \qquad \text{for } i \in J,$$

$$b_{ij} := \begin{cases} \bar{a}_{ij} & \text{if } j \in J \\ \underline{a}_{ij} & \text{if } j \notin J \end{cases} \qquad \text{for } i \notin J.$$

Then $B \in V_3$, and by the above characterization a vector x exists such that $SBSx > 0$.

Since

$$\sum_{j \in J} a_{ij} x_j - \sum_{j \notin J} a_{ij} x_j \geq \sum_{j \in J} \underline{a}_{ij} x_j - \sum_{j \notin J} \bar{a}_{ij} x_j \qquad \text{for } i \in J,$$

$$\sum_{j \notin J} a_{ij} x_j - \sum_{j \in J} a_{ij} x_j \geq \sum_{j \notin J} \underline{a}_{ij} x_j - \sum_{j \in J} \bar{a}_{ij} x_j \qquad \text{for } i \notin J$$

we have $SASx \geq SBSx$. Hence $A \in P$.

(ii) Let $V_2 \subset A$ and $A \in [A]$; Let C be any nonzero positive semidefinite matrix. Then for all i $\displaystyle\sum_{k=1}^{n} c_{ik} a_{ki} \geq \sum_{c_{ik} \geq 0} c_{ik} \underline{a}_{ki} + \sum_{c_{ik} < 0} c_{ik} \bar{a}_{ki}$.

Define the matrix B by

$$b_{ij} := \begin{cases} \underline{a}_{ij} & \text{if } c_{ij} \geq 0, \\ \bar{a}_{ij} & \text{if } c_{ij} < 0 \end{cases} \qquad i, j = 1(1)n;$$

then $B \in V_2$, and the i^{th} entry of the diagonal of CA is bounded below by the i^{th} entry of the diagonal of CB. The matrix CB has a positive diagonal entry by the theorem of Barker, Berman and Plemmons [11], which completes the proof.

2.9 Hermitian matrix inequalities and a conjecture.

In many ways Hermitian matrices resemble real numbers. Indeed, all eigenvalues of a Hermitian matrix are real and the matrix is diagonalizable.This similitude may lead an unwary mind to wrong conclusions.This is true in the study of inequalities involving Hermitian matrices.

In the sequel we use capital letters A, B, . . ,X, etc, to denote $n \times n$ Hermitian matrices where n is some integer greater than 1; $A = A^*$ where A^* denotes the conjugate of the transpose of A. We use u and v to denote complex column vectors in C^n furnished with the usual inner product (u,v). We define

1. $A \geq (>)\ 0$ if all eigenvalues of A are nonnegative (positive) or equivalently if $(u, Au) \geq (>)\ 0$ for all nonzero vectors $u \in C^n$.

2. $A \geq (>)\ B$ if $A - B \geq (>)\ 0$,or equivalently if $(u, Au) \geq (>)\ (u,Bu)$ for all nonzero vectors $u \in C^n$.

This ordering is only a partial one. Thus it is not true that if A is not greater than or equal to B then A must be smaller than B. Another source of trouble is the fact that matrix multiplication is not compatible with the ordering (unless all matrices involved are mutually commutative). Most inequalities involving multiplications of real numbers cease to hold when the real numbers are replaced by Hermitian matrices. For example:

1. It is not true that $A \geq 0$ and $B \geq 0$ imply that $AB + BA \geq 0$. Simply take

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

2. It is not true that $A \geq B \geq 0$ implies that $A^2 \geq B^2$. Take

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

To illustrate the delicacy of Hermitian inequalities further,

3. Let A and B be nonnegative Hermitian such that $A + B > 0$,and let X be an $n \times n$ Hermitian matrix that satisfies the inequality

$$(A+B)X + X(A+B) \geq AB + BA.$$

In the scalar case ($n = 1$), it is obvious that the real number X must be positive. However that is not always true for $n \geq 2$. A counterexample is

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Then $(A+B)X + X(A+B) \geq AB + BA$ but $X \ngeq 0$.

Some general results are provided below.

Theorem 2.9.1 : Let A, B, C, D be $n \times n$ Hermitian matrices. Suppose that A commutes with C and B commutes with D. If $A \geq B \geq 0$ and $C \geq D \geq 0$, then for any positive r and s such that $r + s \leq 1$ we have $A^r C^s \geq B^r D^s$.

Theorem 2.9.2 : If $A \geq B \geq 0$, then $A^{1/2} \geq B^{1/2}$.

Theorem 2.9.3 : If $A \geq B \geq 0$,then $A^r \geq B^r$ for all $r \in [0,1]$.

Lemma 2.9.1 : Suppose C and P are square matrices of the same size with $C > 0$. The equation $CX + XP = P$ has a unique solution X. Moreover if $P \geq 0$, then so is X.

Lemma 2.9.2 : Suppose $C \geq 0$. Then $A \geq B$ implies that $CAC \geq CBC$. If furthermore $C > 0$, then the converse also holds.

Lemma 2.9.3 : If $A \geq B \geq 0$, then $B^{-1} \geq A^{-1}$.

Proof of the theorems and Lemmas can be found in Chan and Kwong [22].

A Conjecture (Chan and Kwong [22]) : Taking square roots of course restores the inequality that is destroyed after squaring. It is interesting to ask if the following sequence of operations will preserve the inequality;

1. Squaring the sides of the inequality $A \geq B \geq 0$ to get (A^2, B^2).
2. Multiplying both on the right and on the left by some $C > 0$ to get (CA^2C, CB^2C).
3. Taking square roots .

The question is whether

$$(CA^2C)^{1/2} \geq (CB^2C)^{1/2}.$$

Take $A = \begin{pmatrix} 5 & 2 \\ 2 & 101 \end{pmatrix}$, $\quad B = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}$, $\quad$ and $\quad C = \begin{pmatrix} 1 & 0 \\ 0 & 0.2 \end{pmatrix}$

$A - B \geq 0.$

$$E = (CA^2C)^{1/2} - (CB^2C)^{1/2} = \begin{pmatrix} 4.116681.. & 1.679158... \\ 1.679158.. & 0.134061... \end{pmatrix}.$$

The eigenvalues of E are -0.479. . . and 4.730.Thus

$$(CA^2C)^{1/2} \ngeq (CB^2C)^{1/2}.$$

The inequality persists with the special choice C = B or A.

If $A \geq B \geq 0$, then

$$(BA^2B)^{1/2} \geq B^2 \tag{2.9.1}$$

and $$A^2 \geq (AB^2A)^{1/2} \tag{2.9.2}$$

Second inequality is a consequence of the first one. We may assume that without loss of generality that $B > 0$ so that B^{-1} exists. By hypothesis, $B^{-1} \geq A^{-1} > 0$. The first inequality then gives $(A^{-1}B^{-2}A^{-1})^{1/2} \geq A^{-2}$. Taking inverse now gives (2.9.2).

Now $$(AB^2A)^{1/2} \geq B^2 \tag{2.9.3}$$

also follows from the hypothesis of the conjecture. If this were true, then by transitivity, it would follow that $A^2 \geq B^2$, which we know is false. Repeated use of (2.9.1) shows that

$$(B^3A^2B^3)^{1/2} \geq B^4,$$

and more generally

$$(B^{m-1}A^2B^{m-1})^{1/2} \geq B^m \tag{2.9.4}$$

for $m = 2^k$; where k = 1, 2, . . . We conclude that (2.9.4) is true in two dimensional case but it is not true in general.

2.10 A counterexample to a conjecture regarding a Hermitian matrix inequality.

Let A, B and C be nonnegative Hermitian matrices such that $A \leq C$, $B \leq C$. Chan and Kwong [22] posed a question, wheather it is true that $(A^2 + B^2)^{1/2} \leq \sqrt{2}\ C$?

Now take

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and $\quad C = A + B = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$

A, B and C satisfy the hypothesis of the problem. And

$$D = \sqrt{2}\, C - (A^2 + B^2)^{1/2}$$

The eig values of D are -0.154214. . . and 1.5989. . .

Consequently

$$(A^2 + B^2)^{1/2} \not\leq \sqrt{2}\, C.$$

2.11 Inequality concerning minors of a semidefinite matrix.

Let $S = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$ be a Hermitian nxn matrix of complex numbers, where A, B and C are its submatrices of size mxm, m by (n-m), and (n-m) by (n-m), respectively ($0 < m < n$). Here M* denotes the complex conjugate of the transpose of a matrix M.

Theorem 2.11.1: If S is semidefinite (either positive or negative) then

$$|\det S| \leq |\det A \cdot \det C| \qquad (2.11.1)$$

For definite S equality in (2.11.1) holds iff B = 0.

Inequality (2.11.1) was proved originally by E. Beckenbach and R. Bellman [13].

An alternative proof of (2.11.1) which is purely algebraic is given by Kimelfeld [66].

Remark 2.11.1 : The example of $S = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ with m = 2 shows that the word "definite" in the last assertion of the theorem cannot be replaced by "semidefinite".

Using the theorem for m = 1, and proceeding inductively, one derives the following well-known corollary.

Corollary 2.11.1: If S is semidefinite, then $|\det S|$ is less than or equal to absolute value of the product of all elements of the main diagonal of S. For definite S the equality occurs iff S is diagonal.

2.12 Bergstrom's Inequality.

For all positive definite matrices A and B the inequality

$$\frac{\det(A+B)}{\det(A_n+B_n)} \geq \frac{\det A}{\det A_n} + \frac{\det B}{\det B_n}$$

is valid.

Proof: For all $x, z \in R^n$ the inequality

$$(Ax, x)^{1/2} \cdot (A^{-1}z, z)^{1/2} \geq (x, z) \qquad (2.12.1)$$

Equality is attained iff the vectors z and Ax are proportional.

By virtue of the inequality (2.12.1) one has

$$(A^{-1}z, z)^{-1} = \min_{x \neq 0} \frac{(Ax, x)}{(x, z)^2}$$

setting $z = (0, \ldots, 0, 1)$ one gets the equality [14]

$$\frac{\det A}{\det A_n} = \min_{x_n = 1} (Ax, x)$$

from which there follows, by the criteria

"The pointwise supremum $f(x) = \sup_\alpha f_\alpha(x)$of an arbitrary collection of convex functions $(f_\alpha(x), M)$ is convex on the set $\{x \in M \mid f(x) < +\infty\}$", the concavity of the function

$$\frac{\det A}{\det A_n}$$

on the convex cone of positive definite matrices.

2.13 Generalization of $A \circ A^{-1} \geq I$.

A matrix inequality is obtained, in an elementary way, for the Schur product of two positive definite matrices. If $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times n$ matrices, let $A \circ B$ denote their Schur (or Hadamard) product; thus $A \circ B = (a_{ij} b_{ij})$. If A is Hermitian positive definite (semidefinite) matrix, we write $A > 0$ ($A \geq 0$). It is a basic result of Schur [15, p.95] that if $A \geq 0$, $B \geq 0$, then $A \circ B \geq 0$. If $A \geq 0$, $B \geq 0$, then $A \geq B$ will mean that $A - B$ is positive semidefinite. The determinant, the complex conjugate, and transpose of the matrix A are denoted by $|A|$, A, and A^T respectively.

Theorem 2.13.1 [10] **:** If $A > 0$, then $A \circ A^{-1} \geq I$.

Alternative proofs as well as generalizations of Theorem 2.13.1 have been given by Johnson [57] and Ando [4]. A result which is much more general than Theorem 2.13.1. is provided in 2.13.2. First we require the following lemma.

Lemma 2.13.1 : Let $A > 0$ be an $n \times n$ matrix, and suppose

$$A = \begin{pmatrix} B & C \\ C^* & D \end{pmatrix}$$

where B is a square matrix. Then

(i) $B \geq CD^{-1}C^*$.

(ii) $$\begin{pmatrix} CD^{-1}C^* & C \\ C^* & D \end{pmatrix} \geq 0.$$

Proof (i) :

$$B - CD^{-1}C^* = \begin{pmatrix} I & CD^{-1} \end{pmatrix} \begin{pmatrix} B & C \\ C^* & D \end{pmatrix} \begin{pmatrix} I \\ D^{-1}C^* \end{pmatrix} \geq 0.$$

(ii):

$$\begin{pmatrix} CD^{-1}C^* & C \\ C^* & D \end{pmatrix} = \begin{pmatrix} CD^{-1/2} \\ D^{1/2} \end{pmatrix} \begin{pmatrix} D^{-1/2}C^* & D^{1/2} \end{pmatrix} \geq 0.$$

The following is the main result.

Theorem 2.13.2 : Let A, B be positive definite n×n matrices, and suppose A, A^{-1}, B, B^{-1} are conformally partioned as follows :

$$A = \begin{pmatrix} X & Y \\ Y^* & Z \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} \tilde{X} & \tilde{Y} \\ \tilde{Y}^* & \tilde{Z} \end{pmatrix},$$

$$B = \begin{pmatrix} U & V \\ V^* & W \end{pmatrix}, \qquad B^{-1} = \begin{pmatrix} \tilde{U} & \tilde{V} \\ \tilde{V}^* & \tilde{W} \end{pmatrix}.$$

Then

$$A \circ B \geq \begin{pmatrix} X \circ \tilde{U}^{-1} & 0 \\ 0 & \tilde{Z}^{-1} \circ W \end{pmatrix}.$$

In particular,

$$A \circ A^{-1} \geq \begin{pmatrix} X \circ X^{-1} & 0 \\ 0 & \tilde{Z} \circ \tilde{Z}^{-1} \end{pmatrix},$$

Proof : It is well known that

$$Z - \tilde{Z}^{-1} = Y^*X^{-1}Y \qquad \text{and} \qquad U - \tilde{U}^{-1} = VW^{-1}V^*$$

Thus, by (ii) of Lemma 2.13.1,

$$\begin{pmatrix} X & Y \\ Y^* & Z-\tilde{Z}^{-1} \end{pmatrix} \geq 0 \quad \text{and} \quad \begin{pmatrix} U-\tilde{U}^{-1} & V \\ V^* & W \end{pmatrix} \geq 0,$$

and hence

$$\begin{pmatrix} X & Y \\ Y^* & Z-\tilde{Z}^{-1} \end{pmatrix} \circ \begin{pmatrix} U-\tilde{U}^{-1} & V \\ V^* & W \end{pmatrix} \geq 0.$$

A trivial simplication of the above inequality gives the first part of the theorem. To prove the second part, set $B = A^{-1}$, and the proof is complete.

2.14 The relationship between Hadamard and conventional multiplication for positive definite matrices.

If A is positive definite Hermitian, then $A \cdot A^{-1} \geq I$ in the positive semidefinite ordering. The new result of Johnson and Elsner [59] is a converse to this inequality : under certain weak regularity assumptions about a function F on the positive definite matrices, $A \cdot F(A) \geq AF(A)$ for all positive definite A iff F(A) is a positive multiple of A^{-1}. In addition to inequality $A \cdot A^{-1} \geq I$ and furthermore $\lambda_{min}(A \cdot B) \geq \lambda_{min}(AB^T)$, for A, B positive definite Hermitian. We also show that $\lambda_{min}(A \cdot B) \leq \lambda_{min}(AB)$ and note that $\lambda_{min}(AB)$ and $\lambda_{min}(AB^T)$ can be quite different from A, B positive definite Hermitian.

Let P_n denote the set of nxn positive definite Hermitian matrices, while $\overline{P}_n$ denote its closure, If we take P_n and $\overline{P}_n$ to be partially ordered via

$$B \geq A \qquad \text{if and only if} \qquad B - A \in \overline{P}_n$$

$$B > A \qquad \text{if and only if} \qquad B - A \in P_n \cdot$$

the positive semidefinite and positive definite orderings, respectively.

The Hadamard (or entrywise) product of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ of the same dimensions is denoted and defined by [56]

$$A \cdot B = (a_{ij} b_{ij})$$

while conventional matrix multiplication is indicated, as usual, by juxtaposition.

It is often attributed to Schur and has been shown that P_n (as well as $\overline{P}_n$) is closed under the Hadamard product (see [56]).

Theorem 2.14.1 : If A, B $\in$ P_n (respectively $\overline{P}_n$), then $A \cdot B \in P_n$ (respectively $\overline{P}_n$).

It is easy to observe that if $A \in P_n$ and $B \in \overline{P}_n$, then $A \cdot B \in P_n$, unless B has a diagonal entry equal to 0. Of course P_n is not closed under conventional multiplication, but A, B $\in$ P_n does imply that the eigenvalues of AB are positive real numbers and thus that AB $\in$ P_n if A and B commute (For details see [53]).

It was first noted by Fiedler [33, 34] that

Theorem 2.14.2a : If $A \in P_n$, then

$$A \cdot A^{-1T} \geq I$$

This inequality cannot be strict; for letting $e = (1,1,\ldots,1)^T$, one has $(A \cdot (A^{-1})^T)e = e$, so that 1 is an eigenvalue of $A \cdot (A^{-1})^T$.

It was noted by Johnson [57] that

Theorem 2.14.2b : If $A \in P_n$, then

$$A \cdot A^{-1} \geq I.$$

It is possible for this inequality to be strict when A is complex. For example, let

$$A = \begin{bmatrix} 3 & 1-i & -i \\ 1+i & 2 & 1 \\ i & 1 & 1 \end{bmatrix}.$$

Of course Theorem 2.14.2.a and 2.14.2.b coincide when A is real. The maps A $\rightarrow$

$A \cdot A^{-1}$ and $A \to A \cdot (A^{-1})^T$ are quite interesting and have been given in Johnson and Elsner [58].

For nxn matrix A , all of whose eigenvalues are real, denote the (algebraically) smallest of these eigenvalues by λ_{min} (A) and the largest by λ_{max} (A).

Theorem 2.14.3a : For A, B $\in P_n$, we have

$$\lambda_{min} (A \cdot B) \geq \lambda_{min} (AB^T).$$

Theorem 2.14.3b : For A, B $\in P_n$, we have

$$\lambda_{min} (A \cdot B) \geq \lambda_{min} (AB).$$

Theorem 2.14.3a and 2.14.3b coincide if A and B are real (in fact, if A or B is real). But if A and B are complex, the eigenvalues of AB and AB^T can be quite different, and λ_{min} (AB) and λ_{min} (AB^T) can differ. For example, if

$$A = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & -2i \\ 2i & 2 \end{bmatrix}$$

then

$$\sigma(AB) = \{2 \pm \sqrt{2}\}, \ \sigma(AB^T) = \{6 \pm \sqrt{34}\} \text{ and } \sigma(A \cdot B) = \{4 \pm \sqrt{8}\};$$

and we have $\lambda_{min} (A \cdot B) = 4 - \sqrt{8} > \lambda_{min} (AB) = 2 - \sqrt{2} > \lambda_{min} (AB^T) = 6 - \sqrt{34}$.

The inequality of the Theorem 2.14.2b may be rewritten as

$$A^{-1} \geq AA^{-1},$$

so that the Hadamard multiplication dominates the conventional multiplication when the two multiplicands are functionally related (namely by the inversion function). Of course the left hand side is positive definite by Theorem 2.14.1, but that it should dominate the usual product is remarkable. The implication of the converse is that inversion is essentially unique in this regard.

We say that F: $P_n \rightarrow P_n$ is an ordinary function on P_n if, for $A \in P_n$ with unitary diagonalization

$$A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\, U,$$

$F(A) = U^* \operatorname{diag}\big(f_1(\lambda_1, \ldots, \lambda_n), \ldots, f_n(\lambda_1, \ldots, \lambda_n)\big)\, U$ for some given functions f_i: $R_n^+ \rightarrow R_n^+$, $i = 1, \ldots, n$. Polynomials with positive coefficients, inversion, and exponentiation are examples of ordinary functions on P_n, but in each of these cases f_i depends only upon λ_i for all f_i.

Theorem 2.14.4 : Let F be an ordinary function on P_n. Then

$$A \cdot F(A) \geq AF(A) \qquad \text{for all } A \in P_n$$

iff for each $A \in P_n$, F(A) is a positive scalar multiple of A^{-1}.

Proof : See Johnson and Elsner [59].

2.15 Inequality for the second immanant.

For n×n complex matrix we use the notation $A \geq 0$ to denote that A is positive semidefinite. Suppose that λ is any (irreducible, complex) character of a subgroup G of S_n, the nth symmetric group. The generalized matrix function corresponding to λ is defined on n×n complex matrices by

$$d_\lambda(A) = \sum_{\sigma \in G} \lambda(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}$$

and is referred to as an Immanant when $G = S_n$. The characters S_n are indexed by partitions of n via the theory of Young Diagrams, and when λ corresponds to a partition of the (k,1,. . .,1) we call the corresponding character a single hook character, and use the notation λ_k.

For example, if we let fix(σ) denote the number of fixed points $\sigma \in S_n$, then

$$\lambda_n \ (\sigma) = 1, \text{ the principal character,}$$

$$\lambda_{n-1}(\sigma) = \text{fix}\,(\sigma) - 1,$$

$$\lambda_2(\sigma) = \epsilon(\sigma) \qquad (\text{fix}(\sigma) - 1),$$

$$\lambda_1 \ (\sigma) = \epsilon(\sigma), \text{ the alternating character.}$$

The immanants d_n and d_1 are therefore the familiar permanent and determinant functions.

The Hadamard Determinant Theorem states that

$$\prod_{i=1}^{n} a_{ii} \geq \det(A), \qquad \text{all} \qquad A \geq 0.$$

Schur [99] proved that for any (irreducible, complex) character of a subgroup of S_n

$$\frac{d_\lambda(A)}{\lambda(\text{id})} \geq \det(A), \qquad \text{all} \qquad A \geq 0,$$

which reduces to Hadamard's Theorem by letting $G = \{\text{id}\}$ and $\lambda = 1$.

Marcus' [76] proved the following permanental analogue of Hadamard's Theorem:

$$\text{per}(A) \geq \prod_{i=1}^{n} a_{ii}, \qquad \text{all} \qquad A \geq 0.$$

Marcus' result suggests : Is it true for any (irreducible, complex) character of a subgroup G of S_n that

$$\text{per}(A) \geq \frac{d_\lambda(A)}{\lambda(\text{id})}, \qquad \text{for all} \qquad A \geq 0\,?$$

Lieb [72], showed that if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \geq 0, \qquad A_{11} \text{ square,}$$

then

$$\operatorname{per}(A) \geq \operatorname{per}(A_{11}) \operatorname{per}(A_{22}).$$

Pate [92], has shown that if $A = [A_{ij}] \geq 0$ where A_{ij} is m×m , all i, j = 1,. . .,k, then

$$\operatorname{per}(A) \geq \operatorname{per}[\operatorname{per}(A_{ij})],$$

except perhaps for finitely many values of m for any fixed k. It is also easy to see from $\operatorname{per}(A) \geq \det(A) \geq 0$ that

$$\operatorname{per}(A) \geq \operatorname{alt}(A) = \sum_{\sigma \text{ even}} \prod_{i=1}^{n} a_{i\sigma(i)}.$$

and Merris and Watkins [81] have shown that

$$\operatorname{per}(A) \geq \frac{d_{n-1}(A)}{\lambda_{n-1}(\mathrm{id})}, \qquad \text{all} \qquad A \geq 0,$$

and

$$\operatorname{per}(A) \geq \frac{d_{n-2}(A)}{\lambda_{n-2}(\mathrm{id})}, \qquad \text{all} \qquad A \geq 0.$$

Theorem 2.15.1 : $\operatorname{per}(A) \geq \dfrac{d_2(A)}{\lambda_2(\mathrm{id})}, \qquad \text{all} \qquad A \geq 0.$

Furthermore, equality holds iff A is diagonal or has a zero row. Since $\lambda_2(\mathrm{id}) = n-1$, this result can also be stated as

$$(n-1)\operatorname{per}(A) \geq d_2(A), \qquad \text{all} \qquad A \geq 0,$$

with equality iff A is diagonal or has a zero row.

The proof of the theorem can be found in Grone [46].

2.16 Fischer inequality for the second immanant.

Let M_n be the space of nxn complex matrices. Let $H_n \subset M_n$ be the cone of positive semidefinite Hermitian matrices, and $C_n \subset H_n$ the compact convex subset consisting of those matrices all of whose main diagonal entries equal 1 (e.g. the correlation matrices). Let d_2 be the generalized matrix function (or immanant) afforded by the symmetric group S_n and irreducible degree n-1 character corresponding to the partition (2,1,1, . . ,1).

Suppose $A \in H_n$ is partioned into blocks,

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

If $n \geq 4$, then $d_2(\mathrm{diag}(A_{11}, A_{22})) \geq d_2(A)$. This follows from the fact that d_2 is a Schur-concave function of the spectrum for $A \in C_n$.

Let G be a subgroup of the symmetric group, S_n, and let X be an irreducible character of G. The generalized matrix function (GMF) afforded by G and X is defined by

$$d(A) = \sum_{\sigma \in G} X(\sigma) \prod_{t=1}^{n} a_{t\sigma(t)}, \qquad (2.16.1)$$

Where $A = (a_{ij}) \in M_n$. If $G = S_n$, then (following Littlewood) d is an immanant, and if G = {id}, then d is Hadamard function :

$$h(A) = \prod_{t=1}^{n} a_{tt}.$$

In case X(id) > 1, the normalized GMF, $\bar{d}(A) = \dfrac{d(A)}{X(id)}$

An explicit formula for X is $X(\sigma) = \epsilon(\sigma)\,(f(\sigma) - 1)$, where ϵ is the alternating or signum character and $f(\sigma)$ is the number of fixed points of s. [Note that X(id) = n-1]. The immanant d_2 is referred to as the second immanant and satisfies [73, 81]

$$d_2(A) = \sum_{t=1}^{n} a_{tt} \det A(t) \; - \det A, \tag{2.16.2}$$

Where $A(t) \in M_{n-1}$ is the principal submatrix of A obtained by deleting i^{th} row and column.

Theorem 2.16.1 : If $n \geq 4$, and $d_2(A)$ is a Schur-concave function of the spectrum of A on C_n. Furthermore, d_2 is strictly Schur-concave on the matrices in C_n of rank at least $n-1$.

Theorem 2.16.2 : Suppose $A \in H_n$ is partioned as

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

where A_{11} and A_{22} are square. Let

$$\tilde{A} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}.$$

In $n \geq 4$, then $d_2(\tilde{A}) \geq d_2(A)$ with equality iff $A = \tilde{A}$, A has a zero row (and column), or rank $\tilde{A} < n-1$.

Corollary 2.16.1 : If $A \in H_n$, $n \geq 4$, then $\bar{d}_2(A) \leq h(A)$ with equality iff A is either diagonal or has a zero row (and column).

Proof : See Grone and Merris [47].

2.17 A note on the analogue of Oppenheim's inequality for permanents.

All matrices are nxn Hermitian matrices which are positive definite or semipositive definite. John Chollet [27] asked if there was a permanental analogue for the Hadamard product to Oppenheim's inequality for determinants : i.e. it is true that

$$\text{per}\,(A \circ B) = \text{per}\, A\ \text{per}\, B?$$

He showed that if suffices to prove only

$$\text{per}\,(A \circ \overline{A}) \le (\text{per}\, A)^2, \tag{2.17.1}$$

where $\overline{A}$ is complex conjugate of A.

Suppose A is positive semidefinite (or definite) Hermitian 2×2 matrix, say

$$A = \begin{pmatrix} d_1 & x \\ \overline{x} & d_2 \end{pmatrix}.$$

Then $\quad \text{per}\, A = d_1 d_2 + \overline{x}\, x$

and $\quad \text{per}(A \circ \overline{A}) = d_1^2 d_2^2 + (x\overline{x}\,)^2.$

In a positive semidefinite matrix $d_1, d_2 \ge 0$, then clearly $(\text{per}\, A)^2 \ge \text{per}(A \circ \overline{A})$, since $x\overline{x} = |\,x\,|^2 \ge 0.$

Suppose

$$A_1 = \begin{pmatrix} d_1 & x_1 & y_1 \\ \overline{x}_1 & d_2 & z_1 \\ \overline{y}_1 & \overline{z}_1 & d_3 \end{pmatrix}$$

Note that

$$\text{per}\, A_1 = d_1 d_2 d_3\, \text{per}\begin{pmatrix} 1 & x & y \\ \overline{x} & 1 & z \\ \overline{y} & \overline{z} & 1 \end{pmatrix} = d_1 d_2 d_3\, \text{per}\, A\,,$$

where A is again positive semidefinite Hermitian, and if $\text{per}\,(A \circ \overline{A}) \le (\text{per}\, A)^2$, then $\text{per}(A_1 \circ \overline{A}_1) \le (\text{per}\, A_1)^2$. Thus we consider matrices A having ones down the main diagonal

$$A \circ \overline{A} = \begin{pmatrix} 1 & x\overline{x} & y\overline{y} \\ x\overline{x} & 1 & z\overline{z} \\ y\overline{y} & z\overline{z} & 1 \end{pmatrix}$$

We have

$$\text{per } A = 1 + x\overline{y}z + \overline{x}y\overline{z} + |x|^2 + |y|^2 + |z|^2$$

and

$$(\text{per } A)^2 = 1 + \underline{x^2\overline{y}^2z^2 + \overline{x}^2y^2\overline{z}^2} + |x|^4 + |y|^4 + |z|^4$$

$$+ 2(x\overline{y}z + \overline{x}y\overline{z})\left(1 + |x|^2 + |y|^2 + |z|^2\right) + 2\underline{|x|^2|y|^2|z|^2}$$

$$+ 2\left(|x|^2 + |y|^2 + |z|^2 + |x|^2|y|^2 + |x|^2|z|^2 + |y|^2|z|^2\right).$$

Interchange the underlined terms and write

$$(x\overline{y}z)^2 + (\overline{x}y\overline{z})^2 = (x\overline{y}z + \overline{x}y\overline{z})^2 - 2|x|^2|y|^2|z|^2$$

to get,

$$(\text{per } A)^2 = \text{per}(A \circ \overline{A}) + (x\overline{y}z + \overline{x}y\overline{z})^2 + 2\Big[\left(|x|^2|y|^2 - |x|^2|y|^2|z|^2\right)$$

$$+ \left\{(x\overline{y}z + \overline{x}y\overline{z})\left(1 + |x|^2 + |y|^2 + |z|^2\right) + |x|^2 + |y|^2 + |z|^2\right\}$$

$$+ |x|^2|z|^2 + |y|^2|z|^2\Big].$$

The principal minors of A are nonnegative, so $|z|^2 \leq 1$, from which it follows that

$$|x|^2|y|^2 - |x|^2|y|^2|z|^2 \geq 0.$$

If $x\overline{y}z + \overline{x}y\overline{z} = 2\,\text{Re}(x\overline{y}z) \geq 0$,then all terms in $(\text{per } A)^2 - \text{per}(A \circ \overline{A})$ are nonnegative and the result follows.

Suppose that $\text{Re}(x\overline{y}z) < 0$. Note $|\text{Re}(x\overline{y}z)| \leq |x||y||z|$. To bound

$|x||y||z|$ apply the inequality of the geometric mean and arithmetic mean to get

$$|x|^2|y|^2|z|^2 \le \left(\frac{|x|^2+|y|^2+|z|^2}{3}\right)^3.$$

Expanding det $A \ge 0$,

$$1 > 1 + 2\,\mathrm{Re}(x\bar{y}z) \ge |x|^2 + |y|^2 + |z|^2,$$

so $$|x||y||z| \le \frac{1}{\sqrt{27}}\left(|x|^2+|y|^2+|z|^2\right)^{3/2}$$

$$< \frac{1}{5}\left(|x|^2+|y|^2+|z|^2\right).$$

Hence

$$\left|(x\bar{y}z+\bar{x}y\bar{z})\left(1+|x|^2+|y|^2+|z|^2\right)\right| \le 2|x||y||z|\left(1+|x|^2+|y|^2+|z|^2\right)$$

$$\le \frac{2}{5}\left(|x|^2+|y|^2+|z|^2\right) \qquad (2.17.2)$$

$$(\text{Since } |x|^2+|y|^2+|z|^2 < 1)$$

$$< |x|^2+|y|^2+|z|^2.$$

It follows that the term

$$\left\{(x\bar{y}z+\bar{x}y\bar{z})\left(1+|x|^2+|y|^2+|z|^2\right)+|x|^2+|y|^2+|z|^2\right\}$$

is positive in the expression for $(\text{per } A)^2 - \text{per}(A \circ \bar{A})$. Since all remaining terms are nonnegative,

$$(\text{per } A)^2 \ge \text{per}(A \circ \bar{A}).$$

2.18 An inequality for sum of elements of matrix power.

Let su denotes the sum of elements, then inequality $3 \text{ su } A^3 \ge \text{su } A \text{ su } A^2$, is known to hold for any symmetric nonnegative 3×3 matrix A. This section presents some sharpened version of this inequality.

We study real matrices only. It is known [74] that

$$n \text{ su } A^3 \geq \text{su } A \text{ su } A^2 \qquad (2.18.1)$$

for any symmetric (elementwise) nonnegative $n \times n$ matrix A if (and only if) $n \leq 3$, with equality iff A or A^2 has equal row sums.

The case $n = 1$ is trivial and $n = 2$ is easy to verify.

We now give a sample proof for $n = 3$.

Write

$$A = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \geq 0,$$

and denote

$$f(A) = 3 \text{ su } A^3 - \text{su } A \text{ su } A^2,$$

$$r_1 = a + b + c, \quad r_2 = b + d + e, \quad r_3 = c + e + f, \quad R = (r_1, r_2, r_3)^T, \quad E = (1, 1, 1)^T,$$

Then (see Kankaanpaa [60] for details)

$$\begin{aligned} f(A) \quad &= (E^TE)\,(E^TA^3E) - (E^TAE)\,(E^TA^2E) \\ &= E^T(EE^TA - AEE^T)A^2E = E^T(ER^T - RE^T)AR = \dots. \\ &= (r_1 - r_2)^2(r_1 + r_2 - 3b) + (r_1 - r_3)^2(r_1 + r_3 - 3c) + (r_2 - r_3)^2(r_2 + r_3 - 3e) = \dots. \\ &= (r_1 - r_2)^2(2a + d + 2e + f) + (r_2 - r_3)^2(a + 2b + d + 2f) \\ &\qquad\qquad + 2(r_1 - r_2)(r_2 - r_3)(a + b + e + f - c). \end{aligned}$$

assuming, without loss of generality, $r_1 \geq r_2 \geq r_3$, we have

$$\begin{aligned} f(A) \quad &\geq (a + d + f)(r_1 - r_2)^2 + (a + d + f)(r_2 - r_3)^2 + 2(a - d + f)\,(r_1 - r_2)\,(r_2 - r_3) \\ &\geq (a + d + f)(r_1 - r_2)^2 + (a + d + f)(r_2 - r_3)^2 - 2(a + d + f)(r_1 - r_2)(r_2 - r_3) \end{aligned}$$

$$= (a + d + f)\,[(r_1 - r_2) - (r_2 - r_3)]^2 = \mathrm{tr}\,A\,(r_1 + r_3 - 2r_2)^2.$$

Thus

$$3\,\mathrm{su}\,A^3 - \mathrm{su}\,A\,\mathrm{su}\,A^2 \ \ge\ \mathrm{tr}\,A\,(\mathrm{su}\,A - 3r_{(2)})^2 \qquad (2.18.2)$$

for any symmetric nonnegative 3×3 matrix A. Here $r_{(2)}$ denotes the second largest row sum.

An obvious way to sharpen (2.18.1) is by applying it with A replaced by $A - \alpha I$, $\alpha \le \min_i a_{ii}$. Then, assuming $n \le 3$ and denoting

$$f(A) \quad = n\,\mathrm{su}\,A^3 - \mathrm{su}\,A\,\mathrm{Su}\,A^2, \qquad s_k = \mathrm{su}\,A^k,$$

we obtain by (2.18.1)

$$0 \le f(A - \alpha I) = n\,(s_3 - 3\alpha s_2 + 3\alpha^2 s_1 - \alpha^3 n) - (s_1 - \alpha n)(s_2 - 2\alpha s_1 + \alpha^2 n)$$

$$= f(A) - 2\alpha(ns_2 - s_1^2).$$

Thus $f(A) \ge 2\alpha(ns_2 - s_1^2)$, and since $ns_2 - s_1^2 \ge 0$, the right hand side is best for $\alpha = \min_i a_{ii}$, and we have (Kankaanpaa, Merikoski and Virtanen [61])

$$n\,\mathrm{su}\,A^3 - \mathrm{su}\,A\,\mathrm{su}\,A^2 \ge 2\min_i a_{ii}\,[\,n\,\mathrm{su}\,A^2 - (\mathrm{su}\,A)^2\,].$$

2.19 A note on the variation of permanents.

For any two nxn complex matrices A, B the inequality

$$|\,\mathrm{per}(A) - \mathrm{per}(B)\,| \ \le\ n\,\|A - B\|\max(\|A\|, \|B\|)^{n-1}$$

holds, if $\|\cdot\|$ is either the row sum or the column sum norm. It is conjectured that this result holds for any operator norm.

R. Bhatia [17] has proved that for any two n×n matrices A, B the inequality

$$|\,\mathrm{per}(A) - \mathrm{per}(B)\,| \ \le\ n\,\|A - B\|_2\max(\|A\|_2, \|B\|_2)^{n-1} \qquad (2.19.1)$$

holds. Here $\|\cdot\|_2$ denotes the spectral norm.

An analogous result for the row-sum and the column-sum norm can also proved.

Recall that

$$\| A \|_\infty = \max_i \sum_k | a_{ik} | ,$$

$$\| A \|_1 = \max_k \sum_i | a_{ik} |$$

are the operator norms corresponding to the vector norms

$$\| x \|_1 = \sum_{i=1}^{n} | x_i | \qquad \text{and} \qquad \| x \|_\infty = \max_i | x_i |$$

respectively, where $A = (a_{ik})$, $x = (x_1, \ldots, x_n)^T$. We show

$$| \operatorname{per}(A) - \operatorname{per}(B) | \le n \| A - B \|_p \max (\| A \|_p, \| B \|_p)^{n-1}, \quad p = 1, \infty. \qquad (2.19.2)$$

As $\| A \|_1 = \| A^T \|_\infty$, it suffices to prove the case $p = 1$. We make use of the result in [83, p.113]

$$| \operatorname{per} A | \le \| a_1 \|_1 \| a_2 \|_1 \ldots \| a_n \|_1,$$

where $A = (a_1, \ldots, a_n)$ and a_i denotes the ith column of A. If $B = (b_1, \ldots, b_n)$, define

$$A_k = (a_1, a_2, \ldots, a_k, b_{k+1}, \ldots, b_n), \qquad k = 1, \ldots, n-1,$$

$A_o = B$ and $A_n = A$. Then

$$| \operatorname{per}(A_i) - \operatorname{per}(A_{i-1}) | = | \operatorname{per}(a_1, a_2, \ldots, a_{i-1}, a_i - b_i, b_{i+1}, \ldots, b_n) |$$

$$\le \| a_i - b_i \|_1 \prod_{j<i} \| a_j \|_1 \prod_{j>i} \| b_j \|_1 .$$

Hence $| \operatorname{per}(A) - \operatorname{per}(B) | \le \sum_{i=1}^{n} | \operatorname{per}(A_i) - \operatorname{per}(A_{i-1}) |$

$$\le n \max_i \| a_i - b_i \|_1 \max_i \left(\prod_{j<i} \| a_j \|_1 \prod_{j>i} \| b_j \|_1 \right)$$

$$\leq n \| A - B \|_1 \max\left(\| A \|_1, \| B \|_1 \right)^{n-1}.$$

This establishes (2.19.2).

Conjecture of Elsner [29]: If $\| \ \|$ denotes any operator norm for $n \times n$ matrices, then

$$| \operatorname{per}(A) - \operatorname{per}(B) | \leq n \| A - B \| \max (\| A \|, \| B \|)^{n-1}.$$

2.20 Matrix trace inequality.

Theorem 2.20.1 [110] : If A and B are two $n \times n$ positive definite matrices, then

(I) $\operatorname{tr}(AB) > 0$ and

(II) $\sqrt{\operatorname{tr}(AB)} < \dfrac{(\operatorname{tr} A + \operatorname{tr} B)}{2}$.

Proof : Let P be an orthogonal matrix such that

$$P'AP = \operatorname{diag}(k_1, \ldots, k_n) = J.$$

Then

$$\operatorname{tr}(AB) = \operatorname{tr}(P'ABP) = \operatorname{tr}(P'APP'BP) = \operatorname{tr}(JC) \tag{2.20.1}$$

where $C = P'BP$ is still a positive definite matrix.

Now we have

$$JC = \operatorname{diag}(k_1, \ldots, k_n) \cdot (c_{ij})$$

$$= \begin{pmatrix} k_1c_{11} & & & * \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ * & & & k_nc_{nn} \end{pmatrix}$$

and $\quad \operatorname{tr}(JC) = k_1c_{11} + \ldots + k_nc_{nn} > 0. \qquad (2.20.2)$

On the other hand, we can compute

$$(\mathrm{tr}J + \mathrm{tr}C)^2 - 4\mathrm{tr}(JC) = (k_1 - c_{11})^2 + \ldots + (k_n - c_{nn})^2 + \text{Positive terms} > 0. \quad (2.20.3)$$

from (2.20.1) and (2.20.2) we get (I). From (2.20.1) and (2.20.3) we get

$$\mathrm{tr}(AB) = \mathrm{tr}(JC) < \frac{(\mathrm{tr}\,J + \mathrm{tr}\,C)^2}{4} = \left(\frac{\mathrm{tr}\,A + \mathrm{tr}\,B}{2}\right)^2.$$

From the proof we see that the same is true for Hermitian matrices.

CHAPTER 3

Inequalities for Eigenvalues

The matrix eigenvalue problem arises in a wide variety of areas in the physical and social sciences as well as in engineering, most typically, for example, in the stability analysis of physical systems that are modeled by linear systems of equations, differential equations, and so on.

3.1 The largest and the smallest characteristic roots of a positive definite matrix.

Let A be a positive definite matrix. It follows then that all the characteristic roots are real and positive. Furthermore, we know that the characteristic roots of the powers of A are the corresponding powers of the characteristic roots of A. We also know that the trace, the sum of the elements along the main diagonal, is the sum of the characteristic roots.

Consider the sequence $A, A^2, \ldots, A^n, \ldots$. Let us define

$$u_n = \operatorname{tr}(A^n). \tag{3.1.1}$$

Then for the largest eigenvalue λ_1,

$$\frac{u_{n+1}}{u_n} \le \lambda_1 \le u_n^{1/n}. \tag{3.1.2}$$

A simple application of Cauchy-Schwarz inequality [75, 16] shows that the lower bound is monotone increasing. A simple result from the theory of inequalities shows that the upper bound is monotone decreasing.

Let us now examine the arithmetic. It takes the same labor to multiply two matrices together as to square a matrix. Hence, we consider the sequence $A, A^2, \ldots, A^{2n}, \ldots$. Here, each matrix is the square of the preceding. Let us now define the sequence

$$v_n = \operatorname{tr}(A^{2n}). \tag{3.1.3}$$

Then, we have the inequalities

$$\left(\frac{v_{n+1}}{v_n}\right) \leq \lambda_1 \leq v_n^{1/2^n} . \tag{3.1.4}$$

As before, the upper bound is monotone decreasing. A simple application of Holder's inequality shows that the lower bound is monotone increasing.

We see that we have to take 2^nth roots. This can be done in several ways. We can use logarithms. Or, we can take repeated square roots. Here, we can use the simple recoverance relation of Hiro and obtain arbitrary accuracy. Even if we have a computer with a limited number of significant figures, a simple use of algebra can overcome this.

In the following we will discuss the accuracy of the estimates.

Let λ_1 denote the largest characteristic value, and λ_2 denote the next largest. The accuracy of the bounds depends upon the ratio λ_1 to λ_2. We shall obtain estimates for this ratio by obtaining estimates for $\lambda_1\lambda_2$.

Consider the elementary sum $\sum_{i \neq j} \lambda_i \lambda_j$. We can obtain an expression for this directly from the matrix. However, it is easier to proceed as follows. We have the elementary identity :

$$2\sum_{i\neq j}\lambda_i\lambda_j = \left(\sum_i \lambda_i\right)^2 - \sum_i \lambda_i^2 . \qquad (3.1.5)$$

This is equal to

$$(\mathrm{tr}\, A)^2 - \mathrm{tr}\,(A^2). \qquad (3.1.6)$$

We can now proceed as above and obtain similar bounds.

The same method may be employed to obtain the bound for the smaller characteristic root using the inverse matrix.

3.2 Inequality between the diagonal elements and the eigenvalues of an oscillating matrix.

A real matrix A is totally nonnegative (totally positive) if all the minors of A are nonnegative (positive); A is oscillating if it is totally nonnegative and if some power of it is totally positive. Gantmacher and Krein [38] have shown that in particular case each totally positive matrix is an oscillating matrix.

(i) If $A = (\alpha_{ij})$ is an n×n totally nonnegative matrix, then A is an oscillating matrix iff $\det A \neq 0$ and $\alpha_{i,i+1}, \alpha_{i+1,i} > 0$, $i = 1(1)n\text{-}1$.

(ii) If A is an oscillating matrix, then any principal submatrix formed from consecutive rows and columns of A is an oscillating matrix.

(iii) If A is an oscillating matrix, then the eigenvalues of A are positive and simple, and strictly interlace those of the two principal submatrices of A of order n-1 obtained by deleting the last row and column or the first row and column.

The eigenvalues and the diagonal elements of a matrix A arranged in decreasing order will be denoted by $\lambda_i(A)$ and $\delta_i(A)$, respectively . Trace of A is denoted by trA.

Theorem 3.2.1 : Let $n \geq 2$ and $A = (\alpha_{ij})$ be an n by n oscillating matrix. Then

$$\sum_{i=1}^{k} \delta_i(A) < \sum_{i=1}^{k} \lambda_i(A), \qquad k=1(1)n-1 \tag{3.2.1}$$

The theorem can be easily proved by induction. From (3.2.1) it follows that, in particular,

$$\delta_1(A) < \lambda_1(A), \ \delta_n(A) > \lambda_n(A).$$

The following corollary follows from the above theorem.

Corollary 3.2.1 : If $n \geq 2$ and A and B are $n \times n$ oscillating matrices, then

$$\mathrm{tr}AB > \sum_{i=1}^{n} \lambda_i(A)\, \lambda_{n-i+1}(B) \tag{3.2.2}$$

Proof : See Garloff [39].

3.3 Eigenvalue bounds for algebraic Riccati and Lyapunov equations.

Consider the algebraic Riccati equation

$$A^*K + KA - KRK = Q \tag{3.3.1}$$

and the Lyapunov matrix equation

$$A^*K + KA = -Q \tag{3.3.2}$$

with A being a stable matrix, K a Hermitian positive definite matrix, and Q and R being Hermitian positive semidefinite matrices. It is well known [21] that (3.3.1) arises in the fields of optimal control and filtering theory and (3.3.2) in the study of stability of time-invariant linear system.

We now proceed to obtain information on the stability and settling time of the linear time-invariant system,

$$\dot{x}(t) = A\, x(t), \qquad x(t_0) = x_0.$$

All matrices are n×n, $\lambda_i(X)$ denote the eigenvalues of a matrix X, Re $\lambda_i(X)$ denote the real part of $\lambda_i(X)$, and Tr(X) denotes the trace of X. The eigenvalue of a matrix X are ordered such that the real parts are nonincreasing, i.e.

$$\operatorname{Re}\lambda_1(X) \geq \operatorname{Re}\lambda_2(X) \geq \ldots \geq \operatorname{Re}\lambda_n(X).$$

Lemma 3.3.1 [2] **:** For Hermitian matrices V and W , with $1 \leq i, j \leq n$

$$\lambda_{i+j-1}(V+W) \leq \lambda_j(V) + \lambda_i(W)\ ; \ i+j \leq n+1 \tag{3.3.3}$$

and $$\lambda_{i+j-n}(V+W) \geq \lambda_j(V) + \lambda_i(W)\ ; \ i+j \geq n+1\ . \tag{3.3.4}$$

Lemma 3.3.2 [2], [31] **:** For Hermitian nonnegative definite matrices X and Y , with $1 \leq i, j \leq n$.

$$\lambda_{i+j-1}(XY) \leq \lambda_j(X)\,\lambda_i(Y)\ ; \qquad i+j \leq n+1 \tag{3.3.5}$$

$$\lambda_{i+j-n}(XY) \geq \lambda_j(X)\,\lambda_i(Y)\ ; \qquad i+j \geq n+1 \tag{3.3.6}$$

Theorem 3.3.1 : For the algebraic Riccati equation (3.3.1), with m=1,2,....,n.

$$\sum_{i=1}^{m} 2\operatorname{Re}\lambda_i(A) \leq \sum_{i=1}^{m} \lambda_i(RK - K^{-1}Q) \tag{3.3.7}$$

$$-\sum_{i=1}^{m} 2\operatorname{Re}\lambda_{n-i+1}(A) \leq -\sum_{i=1}^{m} \lambda_{n-i+1}(RK - K^{-1}Q) \tag{3.3.8}$$

where the equality holds for m = n.

This result with R = 0 was originally obtained by Wimmer [109] for the Lyapunov equation (3.3.2) using the following result due to Fan [30].

Lemma 3.3.3 [30] **:** If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Z and $\alpha_1, \ldots, \alpha_n$ are eigenvalues of $\frac{(Z+Z^*)}{2}$, then for m = 1,2,. . .,n

$$\sum_{i=1}^{m} \operatorname{Re} \lambda_i(Z) \le \sum_{i=1}^{m} \alpha_i \qquad (3.3.9)$$

with the equality being true for m = n.

For proofs of (3.3.7) and (3.3.8) see Wimmer [109].

Proof of Theorem 3.3.1 : Since K is a Hermitian positive definite matrix, it has a unique square root matrix, $K^{1/2}$,that is also Hermitian positive definite [93]. Now define the matrices

$$B = K^{1/2} A K^{-1/2}$$

$$D = K^{-1/2} (KRK - Q) K^{-1/2} \qquad (3.3.10)$$

so that

$$B + B^* = K^{-1/2}(KA+A^*K)K^{-1/2}$$

or using (1)

$$B + B^* = D \qquad (3.3.11)$$

from Lemma 3.3.3 to (3.3.11) and the facts that the eigenvalues of B are the same as those of A and the eigenvalues of D are the same as those of

$$K^{-1/2} D K^{1/2} = RK - K^{-1}Q$$

yields (3.3.7) , similarly (3.3.8) follows by writing (3.3.11) as

$$-(B + B^*) = -D$$

and noting that $\lambda_i(-x) = -\lambda_{n-i+1}(x)$. This completes the proof.

Several bounds on the extremal eigenvalues of K can be obtained via Theorem 3.3.1. To simplify the notation, define

$$q_m = \sum_{j=1}^{m} \lambda_j(Q)$$

$$q_{-m} = \sum_{j=1}^{m} \lambda_{n-j+1}(Q)$$

$$a_m = -\sum_{j=1}^{m} \mathrm{Re}\lambda_j(A)$$

$$a_{-m} = -\sum_{j=1}^{m} \mathrm{Re}\lambda_{n-j+1}(A)$$

so that $q_n = q_{-n} = \mathrm{Tr}(Q)$ and $a_n = a_{-n} = -\mathrm{Tr}(A)$

Lemma 3.3.4 : The maximum eigenvalue of algebraic Riccati equation (3.3.1) satisfies the inequalities,

$$\lambda_1(K) \geq \frac{q_{-m}}{a_m + \left[a_m^2 + m\lambda_1(R)\, q_{-m}\right]^{1/2}} \qquad \text{for } m = 1,2,\ldots,n.$$

Lemma 3.3.5 : The minimum eigenvalue of K of(3.3.1) satisfies the inequalities

$$\lambda_n(K) \geq \frac{q_m}{a_{-m} + \left[a_{-m}^2 + m\lambda_n(R)\, q_m\right]^{1/2}} \qquad \text{for } m = 1,2,\ldots,n.$$

For proofs of Lemma 3.3.4 and Lemma 3.3.5 (see Karanam [63]) .

3.4 The product of complementary principal minors of a positive definite matrix.

An upper bound is given for the product of complementary principal minors of a positive definite matrix in terms of its eigenvalues.

Let H be an n×n positive definite matrix (i.e. H has real entries, $H^T=H$ and all the eigenvalues of H are positive). Let the eigenvalues of H be $\alpha_1, \alpha_2, \ldots, \alpha_n$ where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Let r be an integer with $1 \leq r \leq n-1$, and H be partitioned in the form

$$H = \begin{bmatrix} H_1 & H_2 \\ H_2^T & H_3 \end{bmatrix}$$

where H_3 is r×r.

A lower bound for the value of the product $\det H_1 . \det H_3$ by E. Fischer [36] is

$$\det H_1 . \det H_3 \geq \det H,$$

or equivalently, in terms of the eigenvalues of H,

$$\det H_1 . \det H_3 \geq \alpha_1 . \alpha_2 , \ldots , \alpha_n .$$

This note gives an upper bound for $\det H_1 . \det H_3$ in terms of the eigenvalues of H.

Lemma 3.4.1 : Let A be a nonsingular n×n matrix. Let r be an integer with $1 \leq r \leq n-1$. Let A and A^{-1} be partitioned

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix},$$

where A_4 and B_4 are both r×r. Suppose that A_1 is nonsingular. It then follows that

$$\det A_1 = \det A . \det B_4 . \qquad (3.4.1)$$

Proof : See Aitken [1, p.99].

Theorem 3.4.1 : Let H be a positive definite n×n matrix with eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_n$, where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Let r be an integer with $1 \leq r \leq n-1$. Let H be partitioned

$$H = \begin{bmatrix} H_1 & H_2 \\ H_2^T & H_3 \end{bmatrix}$$

where H_3 is r by r. Let q = min(r, n-r). Then

$$\det H_1 . \det H_3 \le \alpha_{q+1} . \alpha_{q+2} \cdot \cdot . \alpha_{n-q} \prod_{k=1}^{q} \left(\frac{\alpha_k + \alpha_{n-k+1}}{2} \right)^2 . \qquad (3.4.2)$$

Proof : See Murphy [85].

3.5 Diagonal elements and eigenvalues of a symmetric matrix.

Let Λ be a real symmetric matrix of order m with eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_m$.

A. Horn [51] if $d_1 \le d_2 \le \ldots \le d_m$ are real numbers such that

$$\sum_{i=1}^{k} d_i \ge \sum_{i=1}^{k} \lambda_i \ , \qquad k = 1,2, \ldots, m-1$$

and $$\sum_{i=1}^{m} d_i = \sum_{i=1}^{m} \lambda_i \qquad (3.5.1)$$

Then there exists an orthogonal matrix P of order m such that the diagonal elements of $P'\Lambda P$ are $d_1, d_2, \ldots, d_m$.

We may assume, without loss of generality, that the matrix Λ is diagonal (with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_m$ arranged in ascending order of magnitude), and proof follows by induction. For m = 2 condition (3.5.1) becomes $\lambda_1 \le d_1 \le d_2 \le \lambda_2$ and $d_2 = \lambda_1 + \lambda_2 - d_1$. If $\lambda_1 = \lambda_2$, the theorem is trivial. Otherwise, if $\lambda_1 < \lambda_2$, write

$$P = (\lambda_2 - \lambda_1)^{-1/2} \begin{bmatrix} \sqrt{(\lambda_2 - \lambda_1)} & -\sqrt{(d_1 - \lambda_1)} \\ \sqrt{(d_1 - \lambda_1)} & \sqrt{(\lambda_2 - d_2)} \end{bmatrix}$$

P is orthogonal and P'ΛP has d_1 and d_2 as diagonal elements. Hence the theorem holds for $m = 2$.

Now suppose the theorem holds for $m \geq 2$ and let $L = \mathrm{diag}(\lambda_1, \ldots, \lambda_{m-1})$ be diagonal matrix of order m+1 with diagonal elements $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{m+1}$. Since condition (3.5.1) implies $\lambda_1 \leq d_1 \leq d_{m+1} \leq \lambda_{m+1}$, we can find the least integer $j > 1$ such that $\lambda_{j-1} \leq d_1 \leq \lambda_j$. There is always a permutation matrix P_1 such that

$$P'_1 \Lambda P_1 = \mathrm{diag}(\lambda_1, \lambda_j, \lambda_2, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_{m+1})$$

For submatrix $\mathrm{diag}(\lambda_1, \lambda_j)$ of order 2, since

$$\lambda_1 \leq \min\{d_1, \lambda_1 + \lambda_j - d_1\} \leq \max\{d_1, \lambda_1 + \lambda_j - d_1\} \leq \lambda_j,$$

there exists an orthogonal matrix Q of order 2 such that the matrix $Q'\mathrm{diag}(\lambda_1, \lambda_j)Q$ has diagonal elements d_1 and $\lambda_1 + \lambda_j - d_1$, in view for the case m=2. Consider the orthogonal matrix (of order m+1)

$$P_2 = \begin{bmatrix} Q & O' \\ O & I \end{bmatrix},$$

where I is identity matrix of order $m-1$ and O is an approximate zero matrix. Then

$$P'_2(P'_1 \Lambda P_1)P_2 = \begin{bmatrix} d_1 & b' \\ b & \Lambda_1 \end{bmatrix}$$

where $\Lambda_1 = \mathrm{diag}(\lambda_1 + \lambda_j - d_1, \lambda_2, \ldots \lambda_{j-1}, \lambda_{j+1}, \ldots \lambda_{m+1})$ and b is an appropriate column vector. The m diagonal elements of Λ_1 and m numbers $d_2, \ldots, d_{m+1}$ satisfy condition (3.5.1). In fact, as $\lambda_i \leq d_1$ for all $i = 2, \ldots, j-1$, we have

$$\sum_{i=2}^{k} d_i \geq (k-1)d_1 \geq \sum_{i=2}^{k} \lambda_i, \qquad k = 2, \ldots, j-1 \qquad (3.5.2)$$

for $k = j, \ldots, m+1$, we have

$$\sum_{i=2}^{k} d_i = \sum_{i=1}^{k} d_i - d_1 \geq \sum_{i=1}^{k} \lambda_i - d_1 = (\lambda_1 + \lambda_j - d_1) + \sum_{i=2}^{j-1} \lambda_i + \sum_{i=j+1}^{k} \lambda_i \ . \quad (3.5.3)$$

The right-hand sides of (3.5.2) and (3.5.3) are not less than the sum of the smallest k-1 elements of Λ_1. Under the inductive assumption, there exists therefore an orthogonal matrix Q_1 of order m such that the diagonal elements of $Q_1' \Lambda Q_1$ are precisely $d_2, \ldots d_{m+1}$.

We now set

$$P_3 = \begin{bmatrix} I & O' \\ O & Q_1 \end{bmatrix}.$$

Thus $P = P_1 P_2 P_3$ is the required orthogonal matrix (see [23]).

3.6 A note on the Hadamard Product of Matrices.

The smallest eigenvalue of the Hadamard Product $A \times B$ of two positive definite Hermitian matrices is bounded from below by the smallest eigenvalue of AB^T.

The Hadamard (or Schur) product of two matrices $A = (a_{ik})$, $B = (b_{ik})$ of the same dimensions is the matrix $A * B = (a_{ik} b_{ik})$. If $C = (c_{ik})$ is a square complex matrix, g(C) the spectral norm (the matrix norm generated by the Euclidean vector norm), i.e. $\sqrt{\lambda}$, where λ is the maximum eigenvalue of CC^* or C^*C; N(C) the Frobenius norm of C i.e.

$$N(C) = \Big(\sum_{i,k} |c_{ik}|^2 \Big)^{1/2} = (\mathrm{tr}CC^*)^{1/2}$$

If C has all real eigenvalues, m(C) will mean the smallest eigenvalue of C. The set of all complex n×n matrices will be denoted by M_n.

Theorem 3.6.1 [35]: If $A \in M_n$, $B \in M_n$ are both positive definite Hermitian, then

$$m(A*B) \geq m(AB^T);$$

equality is attained iff AB^T is a multiple of I.

Proof : We shall need two well-known results of Householder [54] and Schur [98], formulated as following Lemmas.

Lemma 3.6.1 : For any matrices $A \in M_n$, $B \in M_n$,

$$N(AB) \leq g(A)\, N(B)$$

Corollary 3.6.1 : For any invertible $A \in M_n$ and $B \in M_n$,

$$N(AB) \geq [g(A^{-1})]^{-1} N(B).$$

It is easily seen that if B is nonnegative, then equality is attained in each of these inequalities iff A is a multiple of unitary matrix.

Lemma 3.6.2 : For any diagonal $X \in M_n$ and any invertible $S \in M_n$,

$$N(S^{-1}XS) \geq N(X).$$

To return to the proof, let $x = (x_1, \ldots, x_n)^T$, $X = \mathrm{diag}\{x_i\}$.

Then, for A, B positive definite,

$$\begin{aligned} m(A*B) &= \min \left\{ \sum_{i,k=1}^{n} a_{ik}\bar{x}_k b_{ik} x_i \,;\quad x^*x=1 \right\} \\ &= \min \{ \mathrm{tr}(AX^*B^TX);\ N(X) = 1 \} \\ &= \min \{ N^2[(B^T)^{1/2}XA^{1/2}];\ N(X) = 1 \} \\ &= \min \{ N^2[(B^T)^{1/2}A^{1/2}.\, A^{-1/2}XA^{1/2}];\ N(X) = 1 \} \\ &\geq \min \{ [g(\,[(B^T)^{1/2}A^{1/2}]^{-1})]^{-2}\, N^2(A^{-1/2}XA^{1/2});\ N(X)=1 \} \\ &\geq [g([(B^T)^{1/2}A^{1/2}]^{-1})]^{-2} = m(AB^T), \end{aligned}$$

Since

$$g([(B^T)^{1/2} A^{1/2}]^{-1}) = [\, m\{(B^T)^{1/2} A^{1/2} A^{1/2} (B^T)^{1/2}\}]^{-1/2}$$

$$= [m(AB^T)]^{-1/2}.$$

Let equality be attained. Then $(B^T)^{1/2} A^{1/2}$ is a multiple of a unitary matrix, i.e. AB^T is a multiple of identity matrix, and

$$m\{A*(A^T\}^{-1}) \geq 1,$$

since $A*(A^T) - I$ is positive semidefinite singular (Fiedler [33]).

3.7 Perturbation theorems on matrix eigenvalues.

The perturbation theorems on matrix eigenvalues which are concerned with localization of eigenvalues, i.e., to produce region in the complex plane in which eigenvalues of a given matrix lie.

The vector and matrix norms are taken as usual q-norm (or Z_q-norm), $1 \leq q \leq \infty$. For each n = 1,2,. . . . ., let E^n denote a real or complex vector space of column vectors of dimension n, $x = (x_1, x_2, \ldots, x_n)^T$. The q-norm on E^n is defined as follows:

$$x = (x_1, \ldots, x_n)^T \in E^n \Rightarrow$$

$$\| x \|_q = (\, | x_1 |^q + \ldots + | x_n |^q)^{1/q}, \quad 1 \leq q < \infty$$

$$\| x \| = \max_i \, | x_i |, \; q = \infty \tag{3.7.1}$$

Let B be any nxp real or complex matrix. Let $\|B\|_{q,q'}$ denote the norm of B as a linear transformation from E^p to E^n, where E^n is given the q-norm and E^p is given the q'-norm i.e.

$$\| B \|_{q,q'} = \max \{ \, \|By\|_q / \|y\|_{q'} : y \neq 0, y \in E^p\} \tag{3.7.2}$$

$\| B \|_{q,q'}$ will be denoted by simply $\| B \|_q$ if q = q'. From the definition of the matrix norm $\|\cdot\|_{q,q'}$, we have

$$\|By\|_q \le \|B\|_{q,q'} \cdot \|y\|_{q'}, \quad y \in E^p \tag{3.7.3}$$

We now consider some examples.

Example 1[101, p.179] **:** Let B be any n×p matrix. Then

$$\|B\|_1 = \max_{1\le j\le p} \sum_{i=1}^{n} |b_{ij}| \text{ , matrix column-sum norm;}$$

$$\|B\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |b_{ij}| \text{, matrix row-sum norm;}$$

$$\|B\|_{\infty,1} = \max_{1\le i\le n,\ 1\le j\le p} |b_{ij}| \text{ , infinity norm.}$$

Example 2[89, p.52] **:** By diag$\{d_1, \ldots, d_n\}$ we denote the diagonal matrix with diagonal elements $d_1, \ldots, d_n$. Then

$$\|\mathrm{diag}\{d_1, \ldots, d_n\}\|_{q,q'} = \max_i |d_i| \quad , \quad 1 \le q' \le q \le \infty \tag{3.7.4}$$

In particular $\| I \|_{q,q'} = 1$, $1\le q'\le q\le\infty$. Inequality (3.7.4) is generally false if $q' > q$. For example $\| I \|_{1,\infty} = n$, where I is the n-th order identity matrix.

Fundamental Inequality : Let A, X and B be n×n, n×p and p×p matrices, respectively, where $p \le n$. Let β be an eigenvalue of B but not of A. Then for $1 \le q, q' \le \infty$,

$$\min_{y\neq 0} \left(\|Xy\|_q / \|y\|_{q'} \right) \le \|(A-\beta I)^{-1} (AX - XB)\|_{q,q'} \tag{3.7.5}$$

Proof: Let $Bv = \beta v$, $v \neq 0$. Since β is not an eigenvalue of A, $(A - \beta I)^{-1}$ exists and we compute

$$\begin{aligned}\|Xv\|_q &= \|(A-\beta I)^{-1} (A - \beta I) Xv\|_q \\ &= \|(A-\beta I)^{-1} (AX - XB)\, v\|_q \\ &\le \|(A-\beta I)^{-1} (AX - XB)\|_{q,q'} \cdot \|v\|_{q'} \qquad \text{by (3.7.3).}\end{aligned}$$

From this (3.7.5) follows.

3.8 Simple estimates for singular values of a matrix.

For eigenvalues of a square matrix $A = (a_{ij})$ there is widely used Gerschgorin theorem (Stewart [101]).

Theorem 3.8.1 : Let $A = (a_{ij}) \in C^{n \times n}$. Then each eigenvalue of A lies in one of the disks in the complex plane

$$D_i = \left\{ \lambda : |\lambda - a_{ii}| \le r_i := \sum_{j=1,\ j \ne i}^{n} |a_{ij}| \right\}, \qquad i = 1, \ldots, n \tag{3.8.1}$$

Furthermore, if k disks continue a connected region but are disconnected from the other n-k disks, then exactly k eigenvalues lie in the region.

For singular values [88, p.446] of a rectangular matrix A, we can apply Gerschgorin Theorem to A*A to get estimates. However, there are two disadvantages : (I) it is little complicated to use the elements of A*A; (II) the smallest singular value will be very badly conditioned in this process [44]. In many cases, we cannot use this process to give a nonzero lower bound for the smallest singular value.

The estimation Theorem 3.8.2 uses only the elements of A itself. For a square matrix $A = (a_{ij})$, this Theorem 3.8.2 simply uses n real intervals

$$B_i := m \left\{ \left[\left(|a_{ii}| - s_i \right)_+ ,\ |a_{ii}| + s_i \right] \right\}, \qquad i = 1, \ldots, n \tag{3.8.2}$$

to replace the n disks in Theorem 3.8.1, where

$$s_i := \max\left(\sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}|,\ \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ji}| \right), \qquad i = 1, \ldots, n \tag{3.8.3}$$

a for a real number a, we denote $a_+ := \max(0, a)$

this theorem gives a sharper bound for the smallest singular values of A than the Gerschgorin Theorem applied to A*A.

Suppose $A = (a_{ij}) \in C^{m\times n}$, write

$$r_i := \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \quad c_i := \sum_{\substack{j=1 \\ j\neq i}}^{m} |a_{ji}|, \quad s_i := \max\{r_i, c_i\}, \quad a_i := |a_{ii}| \quad (3.8.4)$$

for i = 1,2, . . ,min(m,n). For $m \neq n$, define

$$s = \begin{cases} \max\limits_{n+1\le i \le m} \left(\sum\limits_{j=1}^{n} |a_{ij}|\right) & \text{for } m > n, \\ \max\limits_{m+1\le i \le n} \left(\sum\limits_{j=1}^{m} |a_{ji}|\right) & \text{for } m < n. \end{cases}$$

We consider the theorem for $m \geq n$. For $m < n$, the result is similar.

Theorem 3.8.2 : With the above notation, each singular value of A lies in one of the real intervals

$$B_i = [\,(a_i - s_i)_+,\ a_i + s_i],\qquad i = 1, \ldots, n$$

$$B_{n+1} = [0,s] \qquad (3.8.5)$$

If $m = n$ or if $m > n$ and $a_i \geq s_i + s$, $i = 1, \ldots, n$ then B_{n+1} is not needed in the above statement. Furthermore, every component interval of the union of B_i, $i = 1,2, \ldots, n+1$(n for m = n), contains exactly k singular values if it contains k intervals $B_1, \ldots, B_n$.

Proof : See Qi [96].

3.9 Perron-Frobenius eigenvector for nonnegative integral matrices whose largest eigenvalue is integral.

If A is an integral nonnegative irreducible $k \times k$ matrix whose largest nonnegative eigenvalue is an integer n, then the right eigenspace for n is spanned by a positive vector with integer components.

Lemma 3.9.1 : If A is a $k \times k$ nonnegative real matrix with largest nonnegative eigenvalue λ_0, then

$$|\det(\lambda I - A)| \le \lambda^k - \lambda_0^k \quad (\lambda_0 \le \lambda)$$

Proof : The proof of lemma proceeds by induction on k. For $k = 1$ is clear. Since A is nonnegative, the $(k\text{-}1) \times (k\text{-}1)$ matrix $A(i|i)$ obtained by deleting the i^{th} row and i^{th} column from A has largest nonnegative eigenvalue $\lambda_i \le \lambda_0$.

$$|\mathrm{adj}(\lambda I - A)_{ii}| = |\det(\lambda I - A(i|i)| \le \lambda^{k-1} - \lambda_i^{k-1} \le \lambda^{k-1} \qquad (\lambda_i \le \lambda)$$

By the definition of det, we have

$$\frac{d}{d\lambda}\det(\lambda I - A) = \sum_{i=1}^{k} \mathrm{adj}(\lambda I - A)_{ii}\,.$$

so $$\left|\frac{d}{d\lambda}\det(\lambda I - A)\right| \le \sum_{i=1}^{k} |\mathrm{adj}(\lambda I - A)_{ii}| \le k\lambda^{k-1} \qquad (\lambda_0 \le \lambda)$$

Hence

$$|\det(\lambda I - A)| \le \int_{\lambda_0}^{\lambda} k\lambda^{k-1}\, d\lambda = \lambda^k - \lambda_0^k \qquad (\lambda_0 \le \lambda)\,.$$

In extreme case of equality. If $\lambda_0 = 0$, then A is nilpotent and the characteristic polynomial of A is λ^n, so $\det(\lambda I - A) = \lambda^k = \lambda^k - \lambda_0^k$. Otherwise, we have

Lemma 3.9.2 : If A is $k \times k$ nonnegative real matrix with largest nonnegative eigenvalue $\lambda_0 > 0$, and there is a $\bar{\lambda} > \lambda_0$ with

$$\det(\lambda I - A) = \bar{\lambda}^k - \lambda_0^k\,,$$

then $A = \lambda D^{-1} P D\,,$

where D is a diagonal, nonnegative matrix,and P is a cyclic permutation matrix. Moreover, A can be expressed uniquely, up to a scalar multiples of D.

Proof : See Ashley [5].

3.10 The distance between two permanental roots of a matrix.

The permanent spread of a complex square matrix A is defined to be the greatest distance between two roots of the equation $\operatorname{per}(zI - A) = 0$.

Let A be a complex matrix, , $n \geq 2$, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its characteristic roots. By $\omega_1, \ldots, \omega_n$ we denote the roots of equation

$$\operatorname{per}(zI - A) = 0, \tag{3.10.1}$$

where per is the permanent function, z a complex variable , and I the identity matrix. The numbers $\omega_1, \ldots, \omega_n$ will be referred to as the permanental roots of A.

The symbols

$$s_d(A) = \max_{i,j} | \lambda_i - \lambda_j | \tag{3.10.2}$$

and

$$s_p(A) = \max_{i,j} | \omega_i - \omega_j | \tag{3.10.3}$$

are the determinental spread of A and the permanental spread of A respectively. And the numbers $s_{d,R}(A)$, $s_{d,I}(A)$ by replacing λ_i in (3.10.2) with Re λ_i, Im λ_i, respectively, and numbers $s_{p,R}(A)$, $s_{p,I}(A)$ by replacing ω_i in (3.10.3) with Re ω_i, Im ω_i, respectively.

Two nxn matrices A and B will be called PD-similar if there exists an nxn permutation matrix P and a nonsingular nxn diagonal matrix such that

$$B = D^{-1}P^{-1}APD \tag{3.10.4}$$

or

$$B = D^{-1}P^{-1}A^TPD \tag{3.10.5}$$

The transformations of A given by (3.10.4) and (3.10.5) are special cases of those linear mappings which preserve the permanent of matrix ([78] ,[20]).

Lemma 3.10.1: If A and B are are PD-similar matrices, then

$$\mathrm{per}(zI - A) = \mathrm{per}(zI - A) \tag{3.10.6}$$

Lemma 3.10.2 : If A is a semitriangular square matrix, then

$$s_p(A) = s_d(\tilde{A}) \tag{3.10.7}$$

$$s_{p,R}(A) = s_{d,R}(\tilde{A}) \tag{3.10.8}$$

$$s_{p,I}(A) = s_{d,I}(\tilde{A}) \tag{3.10.9}$$

Proposition 3.10.1 : If A is an n by n matrix, , $n \geq 2$ then

$$s_{d,I}(A) \leq \{\|A\|^2 - \mathrm{Re}[\mathrm{tr}(A^2)] - (2/n)[\mathrm{Im}(\mathrm{tr}A)]^2\}^{1/2} \tag{3.10.10}$$

Equality occurs in (3.10.13) if and only if A is normal and the imaginary parts of the characteristic roots of A are segment bisecting points.

Proposition 3.10.2 : If $A = (a_{ij})$ is a normal square matrix then

$$s_{d,I}(A) \geq \max_{i \neq j} |a_{ij} - \bar{a}_{ji}| \tag{3.10.11}$$

Proof : See Krauter [68].

3.11 Bounds for the real eigenvalues of a cascade matrix.

Bounds are derived for the real eigenvalues of a special matrix. Consider the real $n \times n$ matrix

$$A_n = \begin{bmatrix} a & -b & & & & & \\ 0 & a & -b & & & O & \\ -c & 0 & a & -b & & & \\ & \cdot & \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \cdot & & \\ & & & -c & 0 & a & -b \\ & O & & & -c & 0 & a \end{bmatrix}$$

where b, c > 0. Such matrices arise in the design of two-up one down cascades for isotope separation. Geldenhuys and Sippel [40] derived bounds independent of n for the real eigenvalues of A_n.

Write $\quad A_n = aI - B_n$ and let $\Delta_n(\lambda) = \det(B_n - \lambda I)$.

then $$\Delta_n(\lambda) = -\lambda\Delta_{n-1}(\lambda) + b^2c\Delta_{n-3}(\lambda), \qquad n \geq 3, \tag{3.11.1}$$

$$\Delta_0(\lambda) = 1, \quad \Delta_1(\lambda) = -\lambda, \qquad D_2(\lambda) = \lambda^2. \tag{3.11.2}$$

considering only real values of λ.

By using a generating function for $\Delta_n(\lambda)$

$$\Delta_n(\lambda) = \sum_{k=0}^{[n/3]} \binom{n-2k}{k} (b^2c)^k \, (-\lambda)^{n-3k}, \tag{3.11.3}$$

where [n/3] denotes the integral part of n/3.

For $n \geq 3$, B_n is an irreducible nonnegative matrix. Then it follows from the Perron-Frobenius Theorem that B_n has a unique maximum real eigenvalue γ_n. From Verga [103]

$$\gamma_n < \gamma_{n+1} \qquad \text{for } n = 2,3,\ldots \tag{3.11.4}$$

The difference equation (3.11.1) has the characteristic equation

$$f(r) = r^3 + \lambda r^2 - b^2c = 0$$

The function values at the turning points of f(r) are $f(0) = -b^2c$ and $f(-2\lambda/3) = 4\lambda^3/(27 - b^2c)$. The only real values of λ that will make the second of these function values zero is

$$\lambda = \tau = 3\left(\frac{b^2c}{4}\right)^{1/3}.$$

This implies that $f(-2\lambda/3) > 0$ when $\lambda > \tau$.

Theorem 3.11.1 : All real eigenvalues of B_n are less than t.

Proof : τ is not an eigenvalue of B_n, since if $\lambda = \tau$, we find from the solution of (3.11.1) with initial conditions (3.11.2) that

$$\Delta_n(\tau) = (1/9)\,[1 + (-2)^n\,(8+6n)]\,(b^2c/4)^{n/3}$$

which can not be zero.

For $\lambda > \tau$, if x,y and z are the three distinct zeros of f(r),we find from the initial conditions (3.11.2) that the constants c_1, c_2, c_3 appearing in the solution $\Delta_n(\lambda) = c_1x^n + c_2y^n + c_3z^n$ have the values

$$c_1 = \frac{(\lambda+y)(\lambda+z}{(x-y)(x-z)}$$

$$c_2 = \frac{(\lambda+x)(\lambda+z)}{(y-x)(y-z)}$$

$$c_3 = \frac{(\lambda+x)(\lambda+y)}{(z-x)(z-y)},$$

without loss of generality we assume $x < y < 0 < z$. Note that $f(-\lambda) = -b^2c$, so that $-\lambda < x$. Then

$$(\lambda + x)(\lambda + y)(x-y)z^n < 0,$$

and for n even and positive

$$(\lambda + y)(\lambda + z)(y - z)x^n + (\lambda + x)(\lambda + z)(z - x)y^n < 0$$

These two inequalities show that, for n an even, positive integer, it is impossible for λ to be an eigenvalue of B_n, that is, to satisfy

$$(\lambda + y)(\lambda + z)(y - z)x^n + (\lambda + x)(\lambda + z)(z - x)y^n + (\lambda + x)(\lambda + y)(x - y)z^n = 0.$$

Thus, the only real eigenvalue of B_1 is 0 and $0 < \tau$. We have $\gamma_{2n} < \tau$ for n = 1,2, . . . , and from (3.11.4) we know that $\gamma_n < \gamma_{2n+1}$ for n = 2,3,. . .

3.12 The monotonicity theorem, Cauchy's Interlace Theorem, and the Courant-Fischer Theorem.

The simple dimensional identity

$$\dim(S_1 \cap S_2) = \dim S_1 + \dim S_2 - \dim(S_1+S_2) \qquad (3.12.1)$$

where S_1 and S_2 are subspaces of a finite dimensional vector space. The following basic facts used in the subsequent proofs (a) the eigenvalues of a Hermitian matrix are real and the corresponding eigenvector maybe taken to be orthonormal (b) letting $\alpha_1 \leq \ldots \leq \alpha_k$ denote a subset of eigenvalues of a Hermitian matrix A and letting $u_1, \ldots, u_k$ denote orthonormal set of corresponding eigenvectors , we have $\alpha_1 \leq x^H A x \leq \alpha_k$ for any x in the span of $u_1, \ldots, u_k$ where $x^H x = 1$ (the symbol "H" denotes conjugate transpose).

The Monotonicity Theorem [55]**:** Let A and B be Hermitian and let A + B = C. Let the eigenvalues of A, B and C be $\alpha_1 \leq \ldots \leq \alpha_n$, $\beta_1 \leq \ldots \leq \beta_n$ and $\gamma_1 \leq \ldots \leq \gamma_n$ respectively. Then

(1) $\alpha_j + \beta_{i-j+1} \leq \gamma_i \qquad (i \geq j)$

(2) $\gamma_i \leq \alpha_j + \beta_{i-j+n} \qquad (i \leq j)$

(3) $\alpha_i + \beta_1 \leq \gamma_i \leq \alpha_i + \beta_n$

Proof : Let $\quad Au_i = \alpha_i u_i, \quad Bv_i = \beta_i v_i, \quad Cw_i = \gamma_i w_i$

$$u_i^H u_j = v_i^H v_j = w_i^H w_j = \delta_{ij}, \quad i,j=1,2,\ldots,n.$$

consider first the case $i \geq j$ and let

$S_1 = \text{span}\{u_j, \ldots, u_n\}, \qquad \dim S_1 = n-j+1;$

$S_2 = \text{span}\{v_{i-j+1}, \ldots, v_n\}, \qquad \dim S_2 = n-i+j;$

$S_3 = \text{span}\{w_1, \ldots, w_i\}, \qquad \dim S_3 = i$

Then (3.12.1) gives

$$\dim(S_1 \cap S_2 \cap S_3) \geq \dim S_1 + \dim S_2 + \dim S_3 - 2n = 1$$

This assures the existence of an $x \in S_1 \cap S_2 \cap S_3$ such that $x^H x = 1$. For this x we have

$$\alpha_j + \beta_{i-j+1} \leq x^H A x + x^H B x = x^H C x \leq \gamma_i,$$

proving (1). Application of (1) to (-A) + (-B) = C proves (2) . Setting i = j in (1) and (2) gives (3).

Cauchy Interlace Theorem [55]: Let

$$A = \begin{bmatrix} B & C \\ C^H & D \end{bmatrix}$$

be nxn Hermitian matrix, where B has size mxm. Let eigenvalues of A and B be $\alpha_1 \leq \ldots \leq \alpha_n$ and $\beta_1 \leq \ldots \leq \beta_m$ respectively. Then

$$\alpha_k \leq \beta_k \leq \alpha_{k+n-m} \qquad k = 1, \ldots, m.$$

Proof : Let $Au_i = \alpha_i u_i, \quad u_i^H u_j = \delta_{ij}, \qquad i,j = 1, \ldots, n,$

$$Bv_i = \beta_i v_i, \quad v_i^H v_j = \delta_{ij}, \qquad i,j = 1, \ldots, m,$$

$$w_i = \begin{bmatrix} v_i \\ 0 \end{bmatrix} \qquad i = 1, \ldots, m.$$

Let $1 \leq k \leq m$ and let

$$S_1 = \mathrm{span}\{u_k, \ldots, u_n\}, \qquad \dim S_1 = n-k+1$$

$$S_2 = \mathrm{span}\{w_1, \ldots, w_k\}, \qquad \dim S_2 = k .$$

Again by (3.12.1), the existence of an $x \in S_1 \cap S_2$ such that $x^H x = 1$ is assured and we have

$$\alpha_k \leq x^H A x \leq \beta_k$$

Application of this inequality to $-A$ gives $\beta_k \leq \alpha_{k+n-m}$.

The Courant-Fischer Theorem (Minimax Characterization) : Let A be Hermitian and let $\alpha_1 \leq \ldots \leq \alpha_n$ be eigenvalues of A. Then for $k = 1, \ldots, n$.

$$\alpha_k = \min_{S^k} \max \{ v^H A v : v \in S^k, v^H v = 1 \}$$

$$= \max_{S^{k-1}} \min \{ v^H A v : v \perp S^{k-1}, v^H v = 1 \},$$

where S^k denotes an arbitrary k-dimensional subspace of complex n-vectors.

Proof : Let $Au_i = \alpha_i u_i$, $\quad u_i^H u_j = \delta_{ij}$, $\quad j=1, \ldots, n$.

Let

$$S_1 = \text{span}\{u_k, \ldots, u_n\} \text{ and } S_2 = S^k \text{ (any k-dimensional subspace)}$$

Then (3.12.1) guarantees the existence of an $x \in S_1 \cap S^k$, $x^H x = 1$, giving $x^H A x \geq \alpha_k$.

on the other hand ,for any $u \in \text{span}\{u_1, \ldots, u_k\}$, a k-dimensional subspace, we have $u^H A u \leq \alpha_k$ and $u_k^H A u_k = \alpha_k$ proving the first equality of the theorem.

To prove the second equality, choose

$$S_1 = \text{span}\{u_1, \ldots, u_k\}, \qquad S_2 = (S^{k-1})^{\perp},$$

and proceed with the same argument as above.

3.13 Matrices with some extremal properties.

Let A be an nxn complex matrix. The eigenvalues of Hermetian matrices $(A+A^*)/2$ and $(A-A^*)/2$ are called the real singular values and imaginary singular values of A respectively.

We adopt the following notation:

$\mathbf{C}_{nxn}$	: algebra of all $n \times n$ complex matrices.
$U_n(\mathbf{C})$	: group of all $n \times n$ unitary matrices.
$\mathbf{C}^n$	: linear space of complex row vector$(x_1, \ldots, x_n)$.
S_n	: symmetric group of degree n.

A^t : transpose of A.

A^* : conjugate transpose of A.

tr A : trace of A.

$A_1 \oplus A_2$: direct sum of the square matrices A_1 and A_2.

$\mathrm{diag}(\alpha_1, \ldots, \alpha_n)$: diagonal matrices with diagonal entries $\alpha_1, \ldots, \alpha_n$.

We always assume that a matrix $A \in \mathbf{C}_{nxn}$ has singular values $\alpha_1 \geq \ldots \geq \alpha_n$; real singular values $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$; imaginary singular values $\mu_1 \geq \ldots \geq \mu_n$; eigenvalues $\alpha_1, \ldots, \alpha_n$ with $|\alpha_1| \geq \ldots \geq |\alpha_n|$.

We now discuss in detail the singular value and eigenvalues of matrices.

Let $A \in \mathbf{C}_{nxn}$. Weyl [42, (p. 35-36)] has shown that

$$\left|\prod_{j=1}^{k} \alpha_j\right| \leq \prod_{j=1}^{k} a_j \qquad (k = 1, \ldots, n-1), \tag{3.13.1}$$

$$\left|\prod_{j=1}^{n} \alpha_j\right| \leq \prod_{j=1}^{n} a_j . \tag{3.13.2}$$

all the inequalities (3.13.1) become equalities simultaneously if and only if A is normal. Horn [52] proved the converse theorem, namely if $x_1 \geq \ldots \geq x_n \geq 0$ and $y_1 \geq \ldots \geq y_n \geq 0$ satisfy

$$\prod_{j=1}^{k} y_j \leq \prod_{j=1}^{k} x_j \qquad (k = 1, \ldots, n-1) \qquad \text{and} \quad \prod_{j=1}^{n} y_j = \prod_{j=1}^{n} x_j ,$$

then there exists $A \in \mathbf{C}_{nxn}$ such that $|\alpha_j| = y_j$ and $a_j = x_j$ $(j=1, \ldots, n)$. Now consider the condition on the matrix A for which one of the inequalities in (3.13.1) becomes equality.

Lemma 3.13.1 : Let $A = (\alpha_{ij}) \in \mathbf{C}_{n\mathrm{x}n}$. If $A_1 = (\alpha_{ij})_{1\leq i,j\leq k}$ has singular values $a_1, \ldots, a_k$ then $A = A_1 \oplus A_2$, where $A_2 = (\alpha_{ij})_{k<i,j\leq n}$.

Proof : Let A satisfy the hypothesis of the lemma. Then

$$\sum_{1\leq j\leq k} a_j^2 = \sum_{1\leq i,j\leq k} |\alpha_{ij}|^2 \leq \sum_{\substack{1\leq i\leq n \\ 1\leq j\leq k}} |\alpha_{ij}|^2 \leq \sum_{1\leq j\leq k} a_j^2$$

(the last inequality holds because the sum of the first k diagonal elements of A^*A is not greater than the sum of its k largest eigenvalues [80, p.218]).

Hence $\alpha_{ij} = 0$ for those (i, j) pairs with $1\leq j\leq k < i \leq n$ or $1\leq i\leq k < j\leq n$. It follows that $A = A_1 \oplus A_2$.

Theorem 3.13.1 : Let $A \in \mathbf{C}_{n\mathrm{x}n}$ and $1\leq k <n$. Then $|\prod_{j=1}^{k} \alpha_j| = \prod_{j=1}^{k} a_j$ iff

(i) A is of rank less than k,or

(ii) A is unitarily similar to $A_1 \oplus A_2$,where $A_1 \in \mathbf{C}_{k\mathrm{x}k}$ has singular values $a_1, \ldots, a_k$.

Proof: If $|\prod_{j=1}^{k} \alpha_j| = \prod_{j=1}^{k} a_j = 0$,then the rank of A is less than k. Suppose $|\prod_{j=1}^{k} \alpha_j| = \prod_{j=1}^{k} a_j > 0$. By the Schur Triangularization Lemma (see [71]), A is unitarily similar to $A' = (\alpha_{ij})$ in the lower triangular form with diagonal entries $\alpha_1, \ldots, \alpha_n$. Let $a'_1 \geq \ldots \geq a'_k$ be singular values of $(\alpha_{ij})_{1\leq i,j\leq k}$. We have $a'_j \leq a_j$ $(j = 1, \ldots, k)$. As $\prod_{j=1}^{k} a'_j = |\prod_{j=1}^{k} \alpha_j| = \prod_{j=1}^{k} a_j > 0$, it follows that $a'_j = a_j$ $(j=1, \ldots, k)$. By Lemma 3.13.1, $A' = A_1 \oplus A_2$ as required.

If the rank of A is less than k, than $\left|\prod_{j=1}^{k} \alpha_j\right| = \prod_{j=1}^{k} a_j = 0$. Suppose A is unitarily similar to $A_1 \oplus A_2$, where $A_1 \in \mathbf{C}_{k\times k}$ has singular values $a_1, \ldots, a_k$. Let $\alpha_{i_1}, \ldots, \alpha_{i_k}$ $(1 \le i_1 < \ldots < i_k \le n)$ be the eigenvalues of A_1. Then

$$\prod_{j=1}^{k} a_j = |\det A_1| = \left|\prod_{j=1}^{k} \alpha_{i_j}\right| \le \left|\prod_{j=1}^{k} \alpha_j\right|.$$

Corollary 3.13.1 : Let $A \in \mathbf{C}_{n\times n}$ be nonnegative and $1 \le k < n$. Then $\left|\prod_{j=k+1}^{n} \alpha_j\right| = \prod_{j=k+1}^{n} a_j$ iff condition (ii) of Theorem 3.13.1 holds.

Proof : By the fact that $\left|\prod_{j=k+1}^{n} \alpha_j\right| = \prod_{j=k+1}^{n} a_j$ iff A is nonsingular matrix .

Apart from (3.13.1) and (3.13.2) there is another set of inequalities [80, p.232] related to the eigenvalues and singular values of a matrix A in $\mathbf{C}_{n\times n}$ namely,

$$\sum_{j=1}^{k} |\alpha_j| \le \sum_{j=1}^{k} a_j \qquad (k = 1, \ldots, n). \tag{3.13.3}$$

In the case that an equality holds, we have

Theorem 3.13.2 : Let $A \in \mathbf{C}_{n\times n}$ and $1 \le k \le n$. Then $\sum_{j=1}^{k} |\alpha_j| \le \sum_{j=1}^{k} a_j$ iff $|\alpha_j| = a_j$ $(j=1, \ldots, k)$ and A is unitarily similar to $\mathrm{diag}(\alpha_1, \ldots, \alpha_k) \oplus B$.

Proof : Suppose A is unitarily similar to $A' = (\alpha_{ij})$ in the lower triangular form with

Theorem 3.13.2 : Let $A \in \mathbf{C}_{nxn}$ and $1 \le k \le n$. Then $\sum_{j=1}^{k} |\alpha_j| \le \sum_{j=1}^{k} a_j$ iff $|\alpha_j| = a_j$ ($j=1, \ldots, k$) and A is unitarily similar to $\mathrm{diag}(\alpha_1, \ldots, \alpha_k) \oplus B$.

Proof : Suppose A is unitarily similar to $A' = (\alpha_{ij})$ in the lower triangular form with diagonal entries $\alpha_1, \ldots, \alpha_n$. By Gogberg and Kerin [42, p.39-41], if $\sum_{j=1}^{k} |\alpha_j| \le \sum_{j=1}^{k} a_j$ then $|a_j| = a_j$ ($j=1, \ldots, k$)

Moreover, since

$$\sum_{j=1}^{k} a_j^2 = \sum_{j=1}^{k} |\alpha_j|^2 \le \sum_{\substack{1\le i \le n \\ 1\le j \le k}} |\alpha_{ij}|^2 \le \sum_{j=1}^{k} a_j^2 ,$$

we have $\alpha_{ij} = 0$ for all (i,j) pairs with and $1 \le j \le k$ and $j \le i$. Hence $A' = \mathrm{diag}(\alpha_1, \ldots, \alpha_k) \oplus B$.

CHAPTER 4

Applications of Matrix Inequalities to Statistics

The theory of the linear model and multivariate analysis include important areas in statistics, such as design of experiments, correlation, analysis of variance, regression, least squares, components of variance-areas that comprise a large segment of the matrix theory and application of statistics. Matrices originated in mathematics more than a century ago, but their broad adaptation in science is relatively recent, prompted by the widespread acceptance of statistical analysis of data, and of computers to do the analysis; both statistics and computing rely heavily on matrix algebra.

4.1 Generalized Hadamard's inequalities and their applications to Statistics.

Let $A = (a_{ij})$ be p.d. (positive definite real symmetric) N×N matrix. Then the most famous inequality is due to Hadamard (see [75]):

$$|A| \leq \prod_{i=1}^{N} a_{ii} \,, \tag{4.1.1}$$

where the equality holds iff $a_{ij} = 0$ $(i \neq j)$.

Let us partition A as follows

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{pmatrix}, \text{ where } A_{ii}\ (1 \le i \le k) \text{ is the square matrix.}$$

Then the Fischer's inequality is

$$|A| \le \prod_{i=1}^{k} |A_{ii}| , \tag{4.1.2}$$

where the equality holds iff $A_{ij} = 0$ $(i \neq j)$.

A number of inequalities in multivariate statistical analysis can be derived by the inequality (4.1.2).

Example 1: Let $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}$ (p×p, p.d.) be the covarience matrix of a random vector $X' = (X_1, X_2, \ldots ,X_p)$. Then the square of multiple correlation coefficient between X_1 and $(X_2, \ldots ,X_p)$ is given by

$$\rho^2_{1,(2,\ldots,p)} = 1 - \frac{\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{vmatrix}}{\sigma_{11} \, |\Sigma_{22}|} . \tag{4.1.3}$$

Thus the inequality (4.1.2) shows that $0 \le \rho^2_{1,(2,\ldots,p)} < 1$ and $\rho^2_{1,(2,\ldots,p)} = 0$ iff $\sigma_{12} = 0$.

Example 2: Let p-dimensional random vector X be distributed according to $N_p(\mu, \Sigma)$. We partition Σ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} \cdots & \Sigma_{2k} \\ \vdots & \vdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} \cdots & \Sigma_{kk} \end{pmatrix},$$ where Σ_{ii} $(1 \leq i \leq k)$ is the square matrix of order $p_i \times p_i$.

Let S be the matrix of the sum of cross-products from p–variate sample $(X_1, X_2, \ldots, X_p)$, $S = \sum_{\alpha=1}^{n} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'$, which is partioned correspondingly as $S = (S_{ij})$, where S_{ii} $(1 \leq i \leq k)$ is the square matrix of order p_i.

Then the likelihood ratio statistic λ for testing the hypothesis H: $\Sigma_{ij} = 0$ $(i \neq j)$(see Anderson [3], Ch.9) given the multivariate normal population, becomes

$$\lambda^{2/n} = \frac{|S|}{\prod_{i=1}^{k} S_{ii}} \tag{4.1.4}$$

The inequality (4.1.2) implies that $0 < \lambda \leq 1$, and it is interesting to note $\lambda = 1$ iff $S_{ij} = 0$, $i \neq j$.

There is a refinement of Fischer's inequality (4.1.2) due to Faguet (Smirnov [100], p.70):

$$|A| \leq \frac{\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix}}{|A_{22}|} \tag{4.1.5}$$

where the equality holds iff $A_{13} - A_{12} A_{22}^{-1} A_{23} = 0$.

For square matrices A and D, we have

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{cases} |A|\,|D - CA^{-1}B|, & \text{when A is nonsingular} \\ |D|\,|A - BD^{-1}C|, & \text{when D is nonsingular} \end{cases} \tag{4.1.6}$$

Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ be p.d., where A_{11} and A_{22} are square matrices. Then it follows that A^{-1}, $A_{11.2}$ and $A_{22.1}$ are p.d. and

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11.2}^{-1} & , -A_{11.2}^{-1} A_{12} A_{22}^{-1} \\ -A_{22}^{-1} A_{21} A_{11.2}^{-1}, & A_{22}^{-1} + A_{22}^{-1} A_{21} A_{11.2}^{-1} A_{12} A_{22}^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22.1}^{-1} A_{21} A_{11}^{-1}, & -A_{11} A_{12} A_{22\,1}^{1} \\ -A_{22.1}^{-1} A_{21} A_{11}^{-1} & , A_{22.1}^{-1} \end{pmatrix}, \tag{4.1.7}$$

where $A_{ij.k} = A_{ij} - A_{ik} A_{kk}^{-1} A_{kj}$.

Lemma 4.1.1 : Let A and B are p. d. matrices of order N.

(i) If $x'Ax \geq x'Bx$ for any x then $|A| \geq |B|$.

(ii) In addition to the condition in (i), if $x_0' A x_0 > x_0' B x_0$ for some x_0, then $|A| > |B|$.

Remark : This lemma is still valid for Hermite positive definite matrices because

$$\int_{\{\binom{u}{v} \mid x' A x \leq 1\}} du\, dv = \frac{V_{2N}}{|A|},$$ where $x = u + \sqrt{-1}\, v$ (u and v are real N-vectors) and $V_{2N} = \dfrac{\pi^N}{\Gamma(N+1)}$.

Proof : It is easily seen that the condition of (i) and (ii) yield the following (i)' and (ii)', respectively.

(i)' $\{ x \mid x' A x \leq 1 \} \subset \{ x \mid x' B x \leq 1 \}$.

(ii)' $\{x \mid x'Ax \le 1\} \subset \{x \mid x'Bx \le 1\}$ and the difference set $\{x \mid x'Bx \le 1\} - \{x \mid x'Ax \le 1\}$ has positive Lebesgue measure. Considering the definite integrals

$$\int_{\{x \mid x'Ax \le 1\}} dx = \frac{V_N}{|A|^{1/2}}, \qquad \int_{\{x \mid x'Bx \le 1\}} dx = \frac{V_N}{|B|^{1/2}},$$

where $V_N = \dfrac{\pi^{N/2}}{\Gamma(\frac{N}{2}+1)}$ is the volume of N-dimensional unit hypersphere, we have the conclusions of the theorem.

Proof of the Fischer's Inequality (2) : For k = 2, we can obtain (4.1.2) for any k inductively. The inequality (4.1.2) is equivalent to $|A_{22} - A_{21}A_{11}^{-1}A_{12}| \le |A_{22}|$, since

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}| \quad \text{and} \quad |A_{11}| > 0.$$

Considering Lemma (i), (ii) and positive definiteness of A_{11}^{-1} and $A_{22} - A_{21}A_{11}^{-1}A_{12}$, we have (4.1.2).

Proof of Faguet's Inequality (4.1.5) is given in [87].

4.2 Inefficiency and correlation.

The inequality for the generalized efficiency of least squares estimates relative to best linear unbiased estimator (Bloomfield & Watson [19]; Knott [67]) has found applications in other areas. A consequence was used by Venables [104] in testing for sphericity of a Gaussian distribution, and by Khatri [65] in proving extremal results for canonical correlations.

The common feature of all these results is the distinguished position occupied by certain subspaces. Suppose that U is an n-dimensional random vector with dispersion matrix Γ, and the Γ has eigenvalues and eigenvectors $\lambda_1 \geq \ldots \geq \lambda_n$ and $\gamma_1, \ldots, \gamma_n$, respectively. Let M be the subspace spanned by $\gamma_1+\gamma_n, \gamma_2+\gamma_{n-1}, \ldots, \gamma_k+\gamma_{n-k+1}$, for some $k \leq \frac{1}{2}n$. There is an associated subspace M' spanned by $\gamma_1-\gamma_n, \gamma_2-\gamma_{n-1}, \ldots$ Note that because of the arbitrariness of signs of eigenvectors, there are in fact 2^k such pairs (M, M').

These subspaces appear in the work of Bloomfield & Watson [19] and Knott [67]. There the aim was to find, in the linear model $y = X\beta+u$, regression subspaces for which the performance of the least squares estimators would be the worst when compared with best linear unbiased estimators. The worst situations are when the columns of X span one such subspace M.

Khatri [65] also considered these subspaces, in deriving extermal results for certain functions of canonical correlations. The first reference to such problem is Venables [104].

We show that the occurrence of the same subspaces in the two different contexts of regression and of correlation is no coincidence. We show first that the relative efficiency of least squares is a function of the correlation between the fitted values and the residuals.

We consider the linear model $y = X\beta+u$. We observe y, an $n\times 1$ vector, and X, an $n\times k$ matrix. The parameter vector β, to be estimated, is $k\times 1$ and u is a $n\times 1$ error vector with $E(u) = 0$ and $E(uu^T) = \Gamma$. We assume that rank $(X) = k$ and rank $(\Gamma) = n$. We also assume that $n \geq 2k$.

The least squares estimator $\tilde{\beta}$ and the best linear unbiased estimator $\hat{\beta}$ are given by $\tilde{\beta} = (X^TX)^{-1} X^Ty$ and $\hat{\beta} = (X^T\Gamma^{-1}X)^{-1} X^T\Gamma^{-1}y$, respectively. Their variances are

$$\mathrm{var}(\tilde{\beta}) = (X^T X)^{-1} X^T \Gamma X (X^T X)^{-1}, \qquad \mathrm{var}(\hat{\beta}) = (X^T \Gamma^{-1} X)^{-1}.$$

There is no loss of generality in supposing that $X^T X = I_k$. In this framework, it is well known that $\tilde{\beta}$ can equal $\hat{\beta}$ almost surely, even when Γ is not a multiple of the identity matrix when $n \to \infty$. The equality holds whenever the columns of X span the same subspace as the columns of ΓX (Watson [106] & [107]), or equivalently, if and only if the fitted values $\tilde{y} = X\tilde{\beta}$ and the residuals $y - \tilde{y}$ are uncorrelated.

To see this consider N, an $n \times (n-k)$ matrix such that $(X: N)^T (X: N) = I_n$. Using a relation for the inverses of partitioned matrices, given, for example, by ([97], p. 67), we find that

$$(X^T \Gamma^{-1} X)^{-1} = X^T \Gamma X - X^T \Gamma N (N^T \Gamma N)^{-1} N^T \Gamma X.$$

The nonsingularity of $N^T \Gamma N$ implies that $(X^T \Gamma^{-1} X)^{-1}$ and $X^T \Gamma X$ are equal, or equivalently, that $\tilde{\beta}$ and $\hat{\beta}$ are equal, almost surely, if and only if $X^T \Gamma X$ is null. Other equivalent conditions are given by Haberman [48].

The only coordinate-free measures of efficiency are functions of the eigenvalues of $(X^T \Gamma^{-1} X)^{-1}$ with respect to $X^T \Gamma X$. However, these are the eigenvalues of

$$I_k - (X^T \Gamma X)^{-\frac{1}{2}} X^T \Gamma N (N^T \Gamma N)^{-1} N^T \Gamma X (X^T \Gamma X)^{-\frac{1}{2}},$$

and are thus of the form $(1 - \rho_i^2)$, where ρ_i is the ith canonical correlation between the least squares fitted values $XX^T y$ and the residuals $NN^T y$.

Thus all coordinate-free measures of the efficiency of least squares estimates, relative to the best linear unbiased estimates, are functions only of the canonical correlations between the least squares fitted values and residuals.

Let $Z = (z_1,\ldots,z_n)$ be a set of random variables such that $E(Z) = 0$ and $E(ZZ^T) = \Gamma$. Let S and T be two n×k matrices, both with rank k, such that $S^TT = 0$. Clearly, we must have $k \le \frac{1}{2}n$. Let $\rho_1^2 \ge \ldots \ge \rho_k^2$ be the canonical correlation between S^TZ and T^TZ. Then we have the following.

Lemma 4.2.1 : The product

$$\prod_{i=1}^{k} (1 - \rho_i^2)$$

is a minimum when the columns of S span M and the columns of T span M'.

Theorem 4.2.1 : If Φ is a monotone increasing convex function, then

$$\sum_{i=1}^{k} \Phi\left\{-\log\,(1 - \rho_i^2)\right\}$$

is maximized when the subspace spanned by the columns of S is M and the subspace spanned by the columns of T is M'. Clearly, the roles of M and M' may be reversed, and in fact all of the possible pairs (M, M') give the same maximized value.

Proof: We start by showing that the products

$$\prod_{i=1}^{j} (1 - \rho_i^2) \qquad (j = 1, \ldots, k)$$

are all minimal when

$$\prod_{i=1}^{k} (1 - \rho_i^2)$$

is minimal. However, this follows from the following observations:

(i) that the lower bound for such products is

$$\prod_{i=1}^{j} \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}$$

since if it were not we could produce a counterexample to the lemma, with k replaced by j; and

(ii) that the product attains this bound for the subspace specified in this theorem. We have effectively shown that the sums

$$\sum_{i=1}^{j} \log(1-\rho_i^2) \qquad (j = 1, \ldots, k)$$

are minimal. and hence our theorem reduces to the Theorem A.2 of Marshall & Olkin ([80], p. 116).

4.3 A note on the matrix ordering of special C-matrices.

In statistics, the consideration of inequalities of the form $\Delta_t - tt' \leq \Delta_r - rr'$, where $t = (t_1, \ldots, t_n)'$ is a positive stochastic vector in $\mathbb{R}^n$ ($t_i > 0$ and $\sum t_i = 1$), and r is a positive stochastic vector in $\mathbb{R}^n$ and Δ_t and Δ_r are diagonal matrices with t and r on their diagonals. The ordering $\leq$ denotes the Loewner matrix ordering (Marshall & Olkin [80], Ch. 16E).

The matrix $\Delta_t - tt'$ appears as a special case of a C-matrix (Baksalary and Pukelsheim [7]) in experimental design theory, i.e., as the information matrix for the treatment contrasts of an experimental design, with treatment replication vector t [95, Theorem 4(a)] and [28, Lemma 2]. Also $\Delta_t - tt'$ is the dispersion matrix of a multinomial distribution with cell probability vector t.

Theorem 4.3.1 : Suppose t and r are positive stochastic vectors of dimension n such that $t \neq r$. Then

$$\Delta_t - tt' \le \Delta_r - rr' \tag{4.3.1}$$

if and only if there exists some subscript i such that

$$t_i > r_i\,, \tag{4.3.2}$$

$$t_j > r_j\,, \qquad \text{for all } j \ne i \tag{4.3.3}$$

$$\frac{r_i t_i}{t_i - r_i} \ge \sum_{j \ne i} \frac{r_j t_j}{r_j - t_j} \tag{4.4.4}$$

Theorem 4.3.2 : Suppose t is a positive stochastic vector of dimension n. Then there exists some positive stochastic vector $r \ne t$ such that the inequality (4.3.1) holds if and only if there exists some subscript i such that $t_i > \frac{1}{2}$.

Lemma 4.3.1 : Suppose D is a positive definite nxn matrix, b is a nonzero vector in $\mathbb{R}^n$, and α is a positive scalar. Then

$$D \ge \alpha bb' \Leftrightarrow \frac{1}{\alpha} \ge b' D^{-1} b.$$

Proof : For the first part, premultiplying with $b'D^{-1}$ and post multiplying with its transpose yields $b' D^{-1} b \ge \alpha (b' D^{-1} b)^2$, i.e. $\frac{1}{\alpha} \ge b \ge b' D^{-1} b$. For the converse part, the Cauchy-Schwarz inequality leads to

$$\alpha (b' x)^2 = \alpha (b' D^{-1/2} D^{1/2} x)^2 \le \alpha (b' D^{-1} b) (x' D x)$$

for every vector x in $\mathbb{R}^n$, Hence $\frac{1}{\alpha} \ge b' D^{-1} b$ implies $D \ge \alpha bb'$.

Proof of Theorem 4.3.1: Since $t \ne r$ and $\sum t_i = 1 = \sum r_i$, there must exist some subscript i such that $t_i > r_i$. Without loss of generality we may take i = 1, and thus assume $t_1 > r_1$. Define the matrix $K_n = I_n - 1_n 1_n' / n$, where 1_n is the n-dimensional vector with all elements unity.

As all components of t are assumed to be positive, we have

$$(\Delta_t - tt')\, K_n\, \Delta_t^{-1}\, K_n \;=\; K_n.$$

Hence $K_n\, \Delta_t^{-1}\, K_n$ is seen to be the Moore-Penrose inverse of $\Delta_t - tt'$, and rank $(\Delta_t - tt')$ = $n - 1$. It now follows from Theorem 4.3.1 in [82] that the inequality (4.3.1) is equivalent to the converse ordering

$$K_n\, \Delta_t^{-1}\, K_n \;\geq\; K_n\, \Delta_r^{-1}\, K_n$$

among the Moore-Penrose inverse. Premultiplying with $(-1_{n-1} : I_{n-1})$ and postmultiplying with its transpose leads to another equivalent form of (4.3.1)

$$D \;\geq\; \alpha_1\, 1_{n-1}\, 1'_{n-1}\,, \tag{4.3.5}$$

Where D is the (n-1)×(n-1) diagonal matrix with for $\alpha_j \;=\; \frac{1}{t_j} - \frac{1}{r_j}$, $j \geq 2$, on its diagonal, and $\alpha_1 \;=\; \frac{1}{t_1} - \frac{1}{r_1}$. By assumption a1 > 0, and the forces D to be positive definite. Thus (4.3.5) entails (4.3.2); and the Lemma implies.

$$\frac{1}{\alpha_1} \geq \sum_{j>1} \frac{1}{\alpha_j}\,,$$

i.e. (4.3.4). Conversely (4.3.2), (4.3.3), (4.3.4) and the Lemma establish (4.3.5).

Proof of Theorem 4.3.2 : The inequalities $t_i - t_i^2 \;\leq\; r_i - r_i^2$, obtained from (4.3.1), and $t_i > r_i$, given in (4.3.2), can hold simultaneously only if $t_i > 1/2$ and $r_i \;\in\;]1 - t_i\, ; t_i)$. This establishes the direct part. For converse part, choose some $r_i \;\in\;]1 - t_i\, ; t_i)$, and for $j \neq i$ define

$$r_j \;=\; \frac{1 - r_i}{1 - t_i}\, t_j \;=\; t_j \;+\; \frac{t_i - r_i}{1 - t_i}\, t_j\,.$$

Then $r = (r_1,\ldots, r_n)'$ is a positive stochastic vector satisfying (4.3.2) and (4.3.3). Because of

$$\sum_{j \neq i} \frac{r_j t_j}{r_j - t_j} = \frac{1 - r_i}{t_i - r_i}(1 - t_i) \leq \frac{r_i t_i}{t_i - r_i},$$

it also fulfils (4.3.4). The inequality (4.3.1) now follows from Theorem 4.3.1.

4.4 Schur- convexity for A-optimal designs.

Consider a linear regression model

$$y = X\beta + \varepsilon,$$

where y is an $m \times 1$ vector of observations, X is an $m \times n$ matrix to be called the design matrix, β is an $n \times 1$ vector of unknown parameters, and ε is an $m \times 1$ vector of random variables with mean the $m \times 1$ zero vector and known covariance matrix Λ. Assume that m $\geq$ n and denote the eigenvalues of Λ is ascending order of magnitude by

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots \leq \lambda_m .$$

For a given design matrix X of rank n, an unbiased estimate of the parameter β based on the observation y is the ordinary least squares estimate

$$(X' X)^{-1} X'y,$$

whose convariance matrix is given by

$$(X' X)^{-1} X'\Lambda X (X' X)^{-1}. \tag{4.4.1}$$

One of the design problems is to choose X from a given experimental region such that the trace of the matrix in (4.4.1) is minimal. This is a problem in the A-optimal designs of regression experiments, and the experimental region under consideration is taken to be the set H of all $m \times n$ real matrices of rank n whose ith column has a Euclidean norm not exceeding c_i, i = 1, . . . ,n, where the c_i are given positive numbers.

Chan [26] shown that for any matrix X in H the trace of the matrix in (4.4.1) has a lower bound of

$$\left(\sum_{i=1}^{n} c_i^2\right)^{-1} \left(\sum_{i=1}^{n} \lambda_i^{1/2}\right)^2, \tag{4.4.2}$$

and that when $\lambda_1 > 0$ and the c_i, $i = 1, \ldots, n$, are arranged in ascending order of magnitude, a necessary and sufficient condition for the existence of an X in H to attain the lower bound is that

$$\left(\sum_{i=1}^{n} c_i^2\right)^{-1} \sum_{i=1}^{k} c_i^2 \geq \left(\sum_{i=1}^{n} \lambda_i^{1/2}\right)^{-1} \sum_{i=1}^{k} \lambda_i^{1/2}, \qquad k = 1, \ldots, n-1. \tag{4.4.3}$$

Lemma 1.4.1 : Suppose that $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and $0 < b_1 \leq b_2 \leq \ldots \leq b_n$ are each arranged in ascending order of magnitude. Then the condition

$$b_1^{-1} \lambda_1^{1/2} \geq b_2^{-1} \lambda_2^{1/2} \geq \cdots \geq b_n^{-1} \lambda_n^{1/2} \tag{4.4.4}$$

is equivalent to the condition that for any $z_1, \ldots, z_n$ such that

$$0 < z_1 \leq z_2 \leq \ldots \leq z_n,$$

$$\sum_{i=1}^{k} z_i \leq \sum_{i=1}^{k} b_i, \qquad k = 1, \ldots, n-1, \tag{4.4.5}$$

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} b_i,$$

there is the inequality

$$\sum_{i=1}^{n} z_i^{-1} \lambda_i \geq \sum_{i=1}^{n} b_i^{-1} \lambda_i .$$

The proof of the Lemma can be found in Chan [24].

Theorem 4.4.1 : Suppose that the positive numbers c_i, $i = 1, \ldots, n$, are arranged in ascending order of magnitude. Then $\sum_{i=1}^{n} c_i^{-2} \lambda_i$ is the greatest lower bound for the trace

$$\mathrm{tr}\,\{ (X'X)^{-1} X'\Lambda X (X'X)^{-1} \}$$

for any X in H, if and only if

$$c_i^{-2} \lambda_i^{1/2} \geq c_{i+1}^{-2} \lambda_{i+1}^{1/2}, \qquad i = 1, \ldots, n-1. \tag{4.4.6}$$

Proof : See Chan [24].

4.5 Method for discovering Kantorovich type inequalities and a probabilistic interpretation.

The Kantorovich inequality $(z^* A z)(z^* A^{-1} z) \leq \frac{(\alpha_1 + \alpha_n)^2}{4\alpha_1\alpha_n}$ can be proved using convexity and elementary geometry. The inequality

$$(z^* A z)(z^* A^{-1} z) \leq \frac{(\alpha_1 + \alpha_n)^2}{4\alpha_1\alpha_n} \tag{4.5.1}$$

where z is a unit vector in $\mathbb{C}^n$, and $A = A^*$ is a nonsingular $n \times n$ Hermitian matrix with eigenvalues $0 < \alpha_1 < \ldots < \alpha_n$, appeared in Kantorovich [62] and is now known as the Kantorovich inequality though an equivalent inequality is given in Hardy, Littlewood, and Polya [49].

Let A and B be commuting $n \times n$ hermitian matrices so we may write their spectral forms as $A = \sum \alpha_i P_i$, $B = \sum \beta_i P_i$ where $P_1, \ldots, P_r$ $(r < n)$ form an orthogonal resolution of the identity in $\mathbb{C}^n$ and where we will also be able to assume with no loss of generality that the α_i are the distinct eigenvalues of A. Let z be a unit vector, which may be written

$z = P_1 z + \ldots + P_r z = z_1 + \ldots + z_r$, so that $1 = z_1^* z_1 + \ldots + z_r^* z_r = \omega_1 + \ldots + \omega_r$,

where ω_i are non negative. Then $a = z^* A z = \sum \alpha_i \omega_i$ and $b = z^* B z = \sum \beta_i \omega_i$ are convex combinations of the eigenvalues of A and B. Thus the point (a, b) lies in the convex closure of the r points (α_i, β_i), a polygon. The points of this polygon may be written as $(\sum \alpha_i \omega_i , \sum \beta_i \omega_i)$.

To find lower and upper bonus for a function f of z^*Az and z^*Bz for all unit vector z. Draw the r points in the plane, their convex closure, and the curves $f(a,b) = k$ for various k. Since the point (a,b) must lie in or on the polygon, it is usually matter of inspection to find the smallest and largest possible values of k, which are the bound sought. Watson [108] remarked that the same idea applies to functions of more than two quadratic forms with commuting kernels, but the geometry becomes harder.

To get the Kantorovich inequality, take a nonsingular and $B = A^{-1}$. If $0 < \alpha_1 <\ldots< \alpha_r$, the polygon is the convex closure of r points on the positive branch of the hyperbola ab =1. Further take $f(a,b)= ab =k$, so f is constant on hyperbolas. The lower bound is clearly unity, which is also a consequence of the Cauchy inequalities. The positive branch of the hyperbola with the maximum k must have as a tangent the line joining $(\alpha_1, \frac{1}{\alpha_1})$ and $(\alpha_r, \frac{1}{\alpha_r})$. The points on this line have co-ordinates $a = \alpha_1 \omega + \alpha_r (1-\omega)$, $b = \alpha_1^{1} \omega + \alpha_r (1-\omega)$ where $\omega = \omega_1$ lies in $\{0,1\}$. Thus the maximum k may be found by differentiating ab with respect to ω. One finds $\omega = \frac{1}{2}$, and the maximum is the right hand side of (4.5.1). This bound is attained when $z = \frac{(a_1 + a_r)}{2^{1/2}}$, where a_1 and a_r are unit vectors in the invariant subspaces associated with P_1 and P_r. Thus (4.5.1) is proved. Notice that the right hand side of (4.5.1) is the maximum, over all nonnegative ω_i, ω_j adding to unity and overall pairs i and j, of $(\omega_i \alpha_i + \omega_j \alpha_j)(\omega_i \alpha_i^{-1} + \omega_j \alpha_j^{-1})$.

Now we have two extensions of (4.5.1).

If A is singular, set $B = A^-$ where $A^- = \sum \alpha_i^{-1} P_i$, the summation being over the nonzero eigenvalues of A, say $\alpha_s, \ldots, \alpha_r$. When $f(a,b) = ab$,

$$0 \leq (z^* A z)(z^* A^- z) \leq \frac{(\alpha_s + \alpha_r)^2}{4\alpha_s \alpha_r} \tag{4.5.2}$$

The lower bound is zero, because z can be an eigenvector corresponding to a zero eigenvalue.

Suppose now that the eigenvalues $\alpha_1, \ldots \alpha_{s-1}$ are negative and the remainder positive. the points $(\alpha_i, \alpha_i^{-1})$, $i = 1, \ldots, s-1$, lies on the negative branch of $ab = 1$, and the remainder on the positive branch. Their convex closure is always a quadrilateral, which may or may not cover the origin. In either case one sees that the lower bound is zero. Note that the Cauchy inequality argument does not apply here. To get the upper bound, we must consider the two hyperbolas touching, respectively, the sides of the quadrilateral defined by the join of $(\alpha_1, \alpha_1^{-1})$ and $(\alpha_{s-1}, \alpha_{s-1}^{-1})$ and the join of $(\alpha_s, \alpha_s^{-1})$ and $(\alpha_r, \alpha_r^{-1})$.

From the previous paragraphs the two k's will be $\frac{(\alpha_1 + \alpha_{s-1})^2}{4\alpha_1 \alpha_{s-1}}$ and $\frac{(\alpha_s + \alpha_r)^2}{4\alpha_s \alpha_r}$. Hence in this case

$$0 \leq z^* A z z^* A^{-1} z \leq \max\left\{\frac{(\alpha_1 + \alpha_{s-1})^2}{4\alpha_1 \alpha_{s-1}}, \frac{(\alpha_s + \alpha_r)^2}{4\alpha_s \alpha_r}\right\}.$$

4.6 Applications in information theory.

We prove an inequality concerning A and D_1AD_2, where A is a non-negative matrix and D_1 and D_2 are diagonal matrices with positive diagonal entries.

If x and y are nonnegative vectors of order n, and if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, then a well-known inequality asserts that $\prod_{i=1}^{n} x_i^{x_i} \geq \prod_{i=1}^{n} y_i^{x_i}$ with equality iff x = y.

Let

$$P^n = \left\{ x \in R^n : x \geq 0, \sum_{i=1}^{n} x_i = 1 \right\}$$

Here we have taken $0^0 = 1$ and $0\log 0 = 0$. The following inequality is well known and has applications in information theory.

Theorem 4.6.1 : If $x \in P^n$, $y \in P^n$, then

$$\prod_{i=1}^{n} x_i^{x_i} \geq \prod_{i=1}^{n} y_i^{x_i}$$

with equality iff x = y.

To describe various situations where the inequality of Theorem 4.6.1 can be applied to obtain inequalities concerning nonnegative matrices.

Let $A = (a_{ij})$ be a positive m×n matrix, and define a map $f : P^m \times P^n \rightarrow P^m \times P^n$ as follows :

$$f(x,y) = (\tilde{x}, \tilde{y}),$$

where

$$\tilde{x}_i = \frac{x_i \sum_{j=1}^{n} a_{ij} y_j}{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j}, \qquad i = 1, 2, \ldots, m,$$

$$\tilde{y}_j = \frac{y_j \sum_{i=1}^{n} a_{ij} x_i}{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j}, \qquad j = 1, 2, \ldots, n. \tag{4.6.1}$$

It is shown in [8] that f maps $P^m \times P^n$ onto itself. Also it is one-one.

The following inequality was proved by Atkinson, Watterson, and Moran [6]. For any $(x,y) \in P^m \times P^n$,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \tilde{x}_i \tilde{y}_j \geq \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j . \tag{4.6.2}$$

Lemma 4.6.1 : Let A be a nonnegative n×n matrix and let $(x,y) \in P^m \times P^n$. Then

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \left(x_i \sum_{s=1}^{n} a_{is} y_s \right) \left(y_j \sum_{r=1}^{m} a_{rj} x_r \right) \geq \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j \right)^3 .$$

Now we have a result which is stronger than Lemma 4.6.1.

Theorem 4.6.2 : Let A be a positive m×n matrix and let f be defined as in (4.6.1). Then for any $(x,y) \in P^m \times P^n$ and for any vectors $\lambda \geq 0$, $\mu \geq 0$,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \lambda_i \mu_j \geq \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j \right) \prod_{i=1}^{m} \left(\frac{\lambda_i}{x_i} \right)^{\tilde{x}_i} \prod_{j=1}^{n} \left(\frac{\mu_j}{y_j} \right)^{\tilde{y}_j} .$$

Proof : We will assume that x, y, λ, μ are positive, and the general case will be follow by a continuity argument. First note that the inequality of the theorem remains unchanged if each λ_i and each μ_j is multiplied by the same positive constant. So we assume without loss of generality that

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \lambda_i \mu_j = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j$$

But now the result follows, after a trivial simplification, from the following inequality, which is true in view of Theorem 4.6.1:

$$\prod_{i,j} (a_{ij} x_i y_j)^{a_{ij} x_i y_j} \geq \prod_{i,j} (a_{ij} \lambda_i \mu_j)^{a_{ij} x_i y_j} .$$

Bibliography

1. Aitken, A. C. (1959) *Determinants and matrices,* Oliver and Boyd.

2. Amir-Moez, A. R. (1956) *Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations,* Duke Math. J.23: 463-476.

3. Anderson, T. W. (1958) *An Introduction to Multivariate Analysis,* New York: Wiley.

4. Ando, T. (1979) *Concavity of certain maps on positive definite matrices and applications to Hadamard Products,* Linear Algebra Appl. 26: 203-241.

5. Ashley, J. (1987) *On the Perron -Frobenius eigenvector for nonnegative integrals matrices whose largest eigenvalue is integral,* Linear Algebra Appl. 94:103-108.

6. Atkinson, F. A. Watterson , G. A. and Moran, P.A.P. (1960) *A Matrix Inequality,* Quart. J. Math. Oxford Ser.(2),11:137-140.

7. Baksalary, J. K. and Pukelsheim, F. (1985) *A note on the Matrix Ordering of Special C-matrices,* Linear Algebra Appl. 70:263- 7.

8. Bapat, R. B. (1982) D_1AD_2 *Theorems for multidimensional matrices,* Linear Algebra Appl. 48: 437-442.

9. Bapat. R. B. (1986) *Applications of an inequality in information theory to matrices,* Linear Algebra Appl. 78: 107-117.

10. Bapat, R. B. and Kwong, M. K. (1987) *A generalization of* $A \circ A^{-1} \geq I$, Linear

Algebra Appl. 93: 107-112.

11. Barker, G. P. Berman, A. and Plemmons, R. J. (1978) *Positive diagonal solutions to the Lyapunov Equation,* Linear and Multilinear Algebra 5: 249-256.

12. Bartmann, F.C. and Bloomfield, P. (1981) *Inefficiency and correlation,* Biometrika, 68: 67-71.

13. Beckenbach, E. F. and Bellman, R. (1961) *Inequalities,* Springer-Verlag, Berlin.

14. Beckenbach, E. F. and Bellman, R. (1965) *Inequalities,* 2nd. rev. ed., Springer-Verlag, Berlin.

15. Bellman, H. (1970) *Introduction to Matrix Analysis,* 2nd. ed., McGraw-Hill, New York.

16. Bellman, R. (1982) *The largest and smallest characteristic roots of a positive definite matrix,* J. Math. Anal. Appl. 90: 207–208.

17. Bhatia, R. (1984) *Variation of symmetric tensor product and permanents,* Linear Algebra Appl. 62: 269-276.

18. Bialas, S. (1984) *Intervals of P-Matrices and Related Matrices,* Linear Algebra Appl. 58: 33-41.

19. Bloomfield, P. & Watson, G. S. (1975) *The inefficiency of least squares,* Biometrika 62: 121-128.

20. Botta, p. (1967) *Linear transformations that preserve the permanent,* Proc. Amer. Math. Soc. 18: 566-569.

21. Brockett, R. W. (1970) *Finite Dimensional Linear Systems,* New York: Wiley.

22. Chan, N. N. and Kawang, M. K. (1985) *Hermitian matrix inequalities and a*

conjecture, Amer. Math. Monthly 92: 533-541.

23. Chan, N. N. and Li, Kin-Hung (1983) *Diagonal elements and eigenvalues of a real symmetric matrix*, J. Math. Anal. Appl. 91: 562-566.

24. Chan, N. N. (1987) *Schur-Convexity for A-optimal Designs*, J. Math. Anal. Appl. 122: 1-6.

25. Chan, N. N. (1981) *Problems in a-optimal regression designs in "Conferen. Volume 43rd session of the International Statistical Institute"*, 39-42.

26. Chan, N. N. (1982) *A optimality for regression designs*, J. Math. Anal. Appl. 87: 45-50.

27. Chollet, J. (1982) *Is there a permanental analogue to Oppenheim's inequality?* Amer. Math. Monthly 89: 57-58.

28. Christif, K. and Pukelsheim, F. (1984) *Approximate design theory for a simpblock design with random block effects*, Tech Report No. 34.Inst. Fur Mathematik Univ. Augsburg.

29. Elsner, L. (1988) *A note on the variation of Permanents*, Linear Algebra Appl. 109: 37-39.

30. Fan, K. (1950) *On a theorem of Weyl concerning the eigenvalues of linear transformations*, II, Proc. Nat. Acad. Sci. 36: 31-35.

31. Fan, K. (1951) *Maximum properties and inequalities for the eigenvalue of completely continuous operators*, Proc. Nat. Acad. Sci. 37: 760-766.

32. Fan, K. and Hoffman, A. (1955) *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc. 6: 111-116.

33. Fiedler, M. (1957) *On a property of Hermitian matrices*, Mat. Fyz. Casopis

SAV 7:168-176.

34. Fiedler, M. (1961) *Uber eine Ungleichung fur positive definite Matrizen,* Math. Nachr. 23:197-199.

35. Fiedler, M. (1983) *A note on Hadamard Product of matrices,* Linear Algebra Appl. 49: 233-235.

36. Fischer, E. (1908) *Uber den Hadamardschen Determinantensatz,* Arch. Math. Phys. (3) 13: 32-40.

37. Gantmacher, F. (1959) *The Theory of Matrices,* Vol. II, Chelsea, New York.

38. Gantmacher, F. and Krein, M. (1960) *Oszillationskerne and kleine Schwingungen mechanischer Systeme,* Akademie Verlag, Berlin.

39. Garloff, J. (1982) *Majorization between the diagonal elements and the eigenvalues of an oscillating matrix,* Linear Algebra Appl. 47: 181-184.

40. Geldenhuys, G. and Sippel, C. (1987) *Bounds for the real eigenvalues of a cascade matrix,* Linear Algebra Appl. 93; 127-130.

41. Gibson, P. M. (1969) *An identity between permanents and determinants,* Amer. Math. Monthly 76: 270-271.

42. Gohberg, I. C. and Krein, M. G. (1969) *Introduction to the theory of Linear Nonselfadjoint Operators,* Translations of Math. Monographs 18, Amer.Math. Soc.

43. Goldberg, M. and Straus, E. G. (1979) *Norm properties of C-numerical radii,* Linear Algebra Appl. 24: 113-131.

44. Golub, G. and Kahan, W. (1965) *Calculating the singular values and*

pseudo-inverse of a matrix, SIAM J. Numer. Anal.2: 205-224.

45. Gregorac, R. J. and Hentzel, I. R. (1987) *A note on the Analogue of Oppenheim's inequality for permanents,* Linear Algebra Appl. 94: 109-112.

46. Grone, B. (1985) *An Inequality for the second immanant,* Linear and Multilinear Algebra 18: 147-152.

47. Grone, R. and Merris, R. (1987) *A Fischer inequality for the second immanant,* Linear Algebra Appl. 87: 77-83.

48. Haberman, S. (1975) *How mutch do Gauss-Markov and least squares estimate differ?* A coordinate free approach. Ann. Statist.3:982-90.

49. Hardy, G. H. , Littlewood, J. E. and Polya, G. (1934) *Inequalities,* Cambridge, Uiversity Press, New York.

50. Haynsworth, E. V. (1953) *Bounds for determinants with dominant main diagonal,* Duke Math. J. 20: 199–209.

51. Horn, A. (1954) *Doubly stochastic matrices and diagonal of a rotation matrix,* Amer. J. Math. 76: 620-630.

52. Horn, A. (1954) *On the eigenvalues of a matrix with prescribed singular values,* Proc. Amer. Math. Soc. 5: 4-7.

53. Horn, R. and Johnson, C. R. (1985) *Matrix Analysis,* Cambridge U. P., New York.

54. Householder, A. S. (1953) *Principles of Numerical Analysis,* McGraw-Hill, New York.

55. Ikebe, Y. , Inagaki, T. and Miyamoto, S. (1987) *The Monotonicity Theorem, Cauchy Interlace Theorem,* and the Courant-Fischer Theorem, Amer. Math.

Monthly 352-354.

56. Johnson, C. R. (1974) *Hadamard products of matrces,* Linear and Multilinear Algebra 1: 295-307.

57. Johnson, C. R. (1978) *Partioned and Hadamard product matrix inequalities,* J. Res. Nat. Bur. Standards 83: 585-591.

58. Johnson, C. R. and Shapiro, H. (1986) *The relative gain array, $A \,.\, A^{-1T}$,* SIAM J. Algebraic Discrete Methods 7: 627-644.

59. Johnson, C. R. and Elsner, L. (1987) *The relationship between Hadamard and conventional multiplication for positie definite matrices,* Linear Algebra Appl 92: 231-240.

60. Kankaanpaa, H. (1983) *Matriisiepayhtlo* $nS(A^{m+1}) \geq S(A)\ S(A^{m})$, Licentiate Thesis, Univ. of Tampere.

61. Kankaanpaa, H. , Merikoski, J. K. and Virtanen, A. (1987) *Improving an inequality for Sum of Elements of Matrix Powers,* Linear Aigebra Appl.92: 225-229.

62. Kantorovich, L. V. (1948) *Functional analysis and applied mathematics (in Russian) Uspekhi Math.* Nauk 3: 89-185; Transl. by Curtis D. Benster,Report 1509, National Bureau of Standards, Washington, D. C. 1952.

63. Karanam, V. R. (1982) *Eigenvalue Bounds for Algebraic Riccati and Lyapunov Equations,* IEEE Trans. Automat. Contr. 27: 461-463.

64. Kemp, M. C. and Kimura, Y. (1978) *Introduction to Mathematical Economics,* Springer, New York.

65. Khatri, C. G. (1978) *Some optimization problems with applications to canonical correlations and sphericity test,* J. Mult. Anal. 8:453-467.

66. Kimelfeld, B. (1985) *An inequality concerning minors of a semidefinite matrix,* Amer. Math. Monthly 92: 354-355.

67. Knott, M. (1975) *On the minimum efficiency of least squares,* Biometrika 62: 129-32.

68. Krauter, A. R. (1987) *On the greatest distance between two permanental roots of a matrix ,* Linear Algebra Appl. 93: 39-55

69. Kuharenko, N. W. (1984) *On a spectral property of a zero-trace matrix (in Russian),* Mat. Zametki 35(2): 149-151.

70. Lavoie, J. L. (1980) *A determinantal inequality involving the Moore-Penrose inverse,* Linear Algebra Appl. 31: 77-80.

71. Li, C. K. (1988) *Matrices with some extremal properties,* Linear Algebra Appl. 101: 255-267.

72. Lieb, E. (1966) *Proof of some conjectures on permanents,* J. Marh. Mech. 16: 127-134.

73. Littlewood, D. E. (1958) *The theory of group characters,* Oxford U. P., London.

74. London, D. (1966) *Inequalities in quardratic forms,* Duke Math. J. 33: 511-522.

75. Magnus, J. R. and Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.*

76. Marcus, M. (1963) *The permanent analogue of the Hadamard determinant theorem,* Bull. Amer. Math. Soc. 69: 494-496.

77. Marcus, M. (1979) *Variation on the Cauchy-Schwartz Inequality,* Linear

Algebra Appl. 27: 81-91.

78. Marcus, M. and May, F. C. (1962) *The permanent fuction,* Canad. J. Math.14: 177-189.

79. Markham, T. L. (1979) *Two properties of M-matrices,* Linear Algebra Appl. 28: 131-134.

80. Marshall, A. W. and Olkin, I. (1979) *Inequalities, Theory of Majorization and its applications,* Academic, New York.

81. Merris, R. and Watkins, W. (1985) *Inequalities and identities for generalized matrix functions,* Lnear Algebra Appl. 64: 223-242.

82. Milliken, G. A. and Akdeniz, F. (1977) *A theorem on the difference of the generalized inverse of two nonnegative matrices,* Comm. Statist. A-Theory Methods 6: 73-79.

83. Minc, H. (1978) *Permanents, in Encyclopedia of Mathematics and its Applications,* 6, Addison-Wesley.

84. Minkowski, H. (1905) *Discontinuitatssbereich fur arithmetische Aquivalenz,* J. f. Math. 129: 220-274.

85. Murphy, I. S. (1982) *A note on the product of complementary principal minors of a positive definite matrix,* Linear Algebra Appl.44: 169-172.

86. Neumaier, A. (1986) *On the Comparison of H-matrices with M-matrices,* Linear Algebra Appl. 83: 135-141.

87. Nishi, A. (1978/79) *Generalized Hadamard's Inequalities and their applications to statistics,* Bull. Math. Statistics V.18: 45-49.

88. Noble, B. and Daniel, J. W. (1977) *Applied Linear Algebra,* 2nd ed., Prentice-Hall, Englewood Cliffs.

89. Ortega, J. M. (1972) *Numerical Analysis: A second course,* Academic Press, New York.

90. Ostrowski, A. M. (1937) *Uber die Determinanten mit uberwiegender Hauptdiagonale,* Comment. Math. Helv. 10: 69-96.

91. Parodi, M. (1959) *La Localisation des valeurs characteristiques des matrices et ses applications,* Paris: Gauthier-Villars.

92. Pate, T. *Inequalities relating groups of diagonal products in a Gram matrix,* Linear and Multilinear Algebra.

93. Pease, M. C. (1965) *Methods of Matrix Algebra,* New York: Academic.

94. Penderson, G. K. (1972) *Some operator monotone functions,* Proc. Amer. Math. Soc. 36: 309-310.

95. Pukelseim, F. (1983) *On optimality properties of simple block designs in the approximate design theory,* J. Statist. Plann. Inference 8: 193-208.

96. Qi, Liqun (1984) *Some simple estimates for singular values of matrix,* Linear Algebra Appl. 56: 105-119.

97. Rao, C. R. (1973) *Linear Statistical Inference and its Applicatios,* 2nd ed. Wiley, New York.

98. Schur, J. (1909) *Uber die charakteristischen Wurzeln einer lineren Substitution mit einer Anwendung auf die Theorie der Integraleichungen,* Math. Ann. 66: 488-510.

99. Schur, I. (1918) *Uber endliche Gruppen und Hermitesche Formen,* Math. Z. 1:

184-207.

100. Smirnov, V. I. (1961) *Linear Algebra and Group Theory*, McGraw-Hill, New York.

101. Stewart, G. W. (1973) *Introduction to Matrix Computations*, Academic Press, New York.

102. Thompson, R. C. and Freede, L. J. (1971) *On the eigenvalues of sum of Hermitian matrices*, Linear Algebra Appl. 4: 369-376.

103. Varga, R. S. (1962) *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J.

104. Venables, W. (1976) *Some implications of the union-intersections principal for test of Sphericity*, J. Mult. Anal. 6: 185-190

105. Vitoria, J. (1979) *Generalization of inequalities by Carmichael and Mason and by Parodi*, Computing 22: 363-365.

106. Watson, G. S. (1967) *Linear least square regression*, Ann. Math. Statist.38: 1679-99.

107. Watson, G. S. (1972) *Prediction and the efficiency of least squares*, Biometrika 59: 91-8.

108. Watson, G. S. (1987) *A Method for Discovering Kantorovich-Type Inequalities and a Probabilistic Interpretation*, Linear Algebra Appl. 97: 211-7.

109. Wimmer, H. K. (1975) *Generalization of theorem of Lyapunov and Stein*, Linear Algebra Appl. 10: 139-146.

110. Yang, Y. (1988) *A matrix trace inequality*, J. Math. Anal. Appl. 133: 573-574.