The Kantorovich inequality, with some extensions and with some statistical applications

by

Gülhan Alpargu

July 1996

Department of Mathematics and Statistics
McGill University, Montréal
Canada

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-19794-8

Canada

In memory of my aunt
Selver Dönmez

Acknowledgements

I wish to express my deepest gratitude to my advisor Professor George P. H. Styan. Without his significant contributions, invaluable discussions and assistance, this thesis could not be presented as it is. He also greatly helped in the compilation of the bibliography.

This thesis was motivated by papers presented at the Fourth International Wor shop on Matrix Methods for Statistics (Montréal, July 1995) by Shuangzhe Liu and Heinz Neudecker [92] and by Geoffrey S. Watson [169] (and the associated discussion). Professor Gene H. Golub kindly provided me with a copy of the English translation [73] of Kantorovich's original 1948 paper and drew my attention to the footnote by George E. Forsythe ([73], pp. 106–107).

Many thanks go to Professor S. W. Drury for helping me with the proof of the Bloomfield-Watson-Knott Inequality and to Dr. Shuangzhe Liu for providing me with copies of his Ph.D. thesis [89] and several other publications related to this thesis.

I am also very grateful to Professor Robert Vermes and to my student friends Jiabin Zhao, Ming-Yan Venus Chiu, Xiao-Jie Zhou and Levente Thomas Tolnai, who translated the articles which appear in the appendices of this thesis, and to Lassina Dembele and Marie-Ange Paget, who translated the abstract into French, and to Thomas Mattman, who helped with the grammar in the appendices. I also appreciate the help I received from Marika Asimakopulos and Julian di Giovanni.

I would like to express my deepest gratitude to Professor Heinz Neudecker, who read my thesis line by line and made some contributions and to Prof. Fikri Akdeniz, who introduced me to matrix methods for statistics and who helped me to find my direction in this area.

Just before the final deposit of this thesis Josip E. Pečarić drew my attention to the paper [50] by Roberto W. Frucht (1943), who established the "Kantorovich Inequality" quite explictly five years before Kantorovich (1948) [72]. Dr. Pečarić also provided me with several related references.

Part of this thesis was presented (as joint papers with George P. H. Styan) at the Sixth Lukacs Symposium (Bowling Green, Ohio, March 29–30, 1996, and published [1] in its *Proceedings*; other parts were presented [2] at the Fifth International Workshop on Matrix Methods for Statistics (Shrewsbury, England, July 18–19, 1996) and at the Conference in Honor of Shayle R. Searle on the Occasion of his Retirement (Cornell University, Ithaca, New York, August 9–10, 1996, and will appear [3] in its *Proceedings*; see also [4]. Furthermore, plans are underway to expand this thesis into a book [5], jointly with Josip E. Pečarić, Simo Puntanen and George P. H. Styan.

Very special thanks go to Evelyn Styan for giving me her warmest friendship during my studying and to my mother (Meryem), my brother (Selahattin) and my sisters (Gülfer, Gülsüm, Gülcan and Deniz) for giving their infinite encouragement to me for coming here.

My studies at McGill University are supported by a very generous scholarship from Harran University, ŞanlıUrfa, Turkey. In addition I acknowledge financial support from McGill University and from the Natural Sciences and Engineering Research Council of Canada.

Abstract

In this thesis we focus on the "Kantorovich Inequality":

$$\frac{t'At \cdot t'A^{-1}t}{(t't)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n},$$

where t is a real $n \times 1$ vector and A is a real $n \times n$ symmetric positive definite matrix, with λ_1 and λ_n, respectively, its (fixed) largest and smallest, necessarily positive, eigenvalues. We begin the thesis with five different proofs of the Kantorovich Inequality and continue by showing that it is equivalent to five closely related inequalities due, respectively, to Schweitzer (1914), Pólya-Szegö (1925), Krasnosel'skiĭ-Kreĭn (1952), Cassels (1955) and Greub-Rheinboldt (1959). We also examine several related inequalities which admit the Kantorovich Inequality as a special case, including the Bloomfield-Watson-Knott Inequality, for which we give a proof based on that presented by Bloomfield and Watson (1975). We also show that there appears to be a lacuna in the "brief proof" given by Yang (1990). Some statistical applications conclude the thesis with special emphasis on the efficiency of the Ordinary Least Squares Estimator in the Gauss-Markov linear statistical model.

Résumé

Dans ce mémoire nous nous intéressons à "l'inégalité de Kantorovitch":

$$\frac{t'At \cdot t'A^{-1}t}{(t't)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n},$$

où t est un vecteur réel $n \times 1$ et A une matrice $n \times n$ symétrique définie positive à coefficients réels, avec λ_1 et λ_n désignant respectivement la plus grande et la plus petite valeur propre de A, fixées. Nous commençons par donner cinq différentes preuves de l'inégalité de Kantorovitch pour ensuite établir qu'elle est équivalente à cinq autres inegalités respectivement dues à Schweitzer (1914), Pólya-Szegö (1925), Krasnosel'skiĭ-Kreĭn (1952), Cassels (1955) et Greub-Rheinboldt (1959), qui lui sont étroitement liées. Nous examinous également différentes inégalités relatives qui admettent celle de Kantorovitch comme cas particulier. Cela inclut l'inégalité Bloomfield-Watson-Knott dont nous donnons une preuve basée sur celle de Bloomfield et Watson (1975). Il nous apparaît aussi qu'il y a une lacune dans la "courte preuve" de Yang (1990). Nous concluons ce mémoire par quelques applications statistiques. Un accent particulier est mis sur l'estimateur des moindres carrés dans le modèle linéaire de Gauss-Markov.

Contents

Introduction

Our main focus in this thesis is on the Kantorovich Inequality

$$t'At \cdot t'A^{-1}t \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{0.1}$$

where t is a real $n \times 1$ vector and A is a real $n \times n$ symmetric positive definite matrix, with λ_1 and λ_n, respectively, its (fixed) largest and smallest, necessarily positive, eigenvalues. The inequality (0.1) is named after the Nobel Laureate and Academician Leonid Vital'evich Kantorovich (1912–1986) for the inequality he established in 1948 ([72], pp. 142–144; cf. also [73], pp. 106–107).

Another way of expressing (0.1) is in the "normalized reduced" form:

$$\sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{0.2}$$

with $\sum_{i=1}^{n} z_i^2 = 1$.

As far as we know, the first inequality of the type (0.2) to be published is the following inequality established in 1914 by Pál Schweitzer [145]:

$$\frac{1}{n^2}(x_1 + x_2 + \cdots + x_n)\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right) \leq \frac{(m + M)^2}{4mM}, \tag{0.3}$$

where the real positive numbers $x_1, \ldots, x_n$ satisfy $0 < m \leq x_i \leq M$ $(i = 1, \ldots, n)$. The "Schweitzer Inequality" (0.3) is a special case of (0.2) with $\lambda_i = x_i$, $z_i^2 = 1/n$, $\lambda_1 = M$ and $\lambda_n = m$.

In 1925 George Pólya and Gábor Szegö ([133],[134]) showed that

$$\frac{(a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2)}{(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2} \leq \frac{(ab + AB)^2}{4abAB}, \tag{0.4}$$

where

$$0 < a \leq a_i \leq A, \quad 0 < b \leq b_i \leq B \quad (i = 1, \ldots, n).$$

The only related work published before 1948 appears to be by József Kürschák who in 1914 posed the following problem in [82]: Prove that

$$\frac{\int_c^d f^2(x)\,dx \int_c^d g^2(x)\,dx}{\left[\int_c^d f(x)g(x)\,dx\right]^2} \leq \frac{(ab+AB)^2}{4abAB}, \tag{0.5}$$

where $f(x)$ and $g(x)$ are continuous functions in the interval (c, d) with $0 < a \leq f(x) \leq A$ and $0 < b \leq g(x) \leq B$. We observe that (0.5) is a continuous version of Pólya-Szegö Inequality (1925), but with $a = b\ (= m)$ and $A = B\ (= M)$.

In Chapter 1 we present five different proofs of the Kantorovich Inequality. The original proof by Kantorovich (1948) is followed by proofs by Anderson (1971) [6], Styan (1983) [152] and Bühler (1987) [23]. We end with the very recently published proof by Pták (1995) [135].

In Chapter 2 we consider in detail five "named" inequalities. In addition to the Schweitzer Inequality [145] and the Pólya-Szegö Inequality ([133], [134]), we also consider the Krasnosel'skiĭ-Kreĭn Inequality [80], the Cassels Inequality [28] and the Greub-Rheinboldt Inequality [60]. We show that these five inequalities are all equivalent to the Kantorovich Inequality, extending results by Watson in [169].

In Chapter 3 we present several inequalities which are related to the Kantorovich Inequality and several which admit the Kantorovich Inequality as a special case. We concentrate on the following papers: Wielandt (1953), Newman (1959), Strang (1960), Bauer (1961), Marcus and Khan (1961), Cargo and Shisha (1962), Diaz and Metcalf (1963), Marcus and Cayford (1963), Marshall and Olkin (1964), Fan (1966) and Shisha and Mond (1967).

In Chapter 4 we present a proof of the Bloomfield-Watson-Knott Inequality

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

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the, necessarily positive, eigenvalues of positive definite matrix A, X is the $n \times k$ matrix of rank k and $n \geq 2k$. We follow the proof given by Bloomfield and Watson [21] with a modification due to Drury [47]. We also discuss the "brief proof" of the Bloomfield-Watson-Knott Inequality given in 1990 by Hu Yang [177] and find that there appears to be a lacuna in his proof. We end this chapter by presenting various extensions of the Bloomfield-Watson-Knott Inequality due to Khatri and Rao (1981) [76] and Wang and Shao (1992) [161], as well as some related matrix inequalities published quite recently by Liu (1995) [88], Liu and Neudecker (1996) [93], and Pečarić, Puntanen and Styan (1996) [131].

In Chapter 5 we present a variety of statistical applications of the Kantorovich Inequality and the Bloomfield-Watson-Knott Inequality. We concentrate on these four papers: Magness and McGuire (1961), Venables (1976), Cressie (1980) and Wang and Shao (1992).

We continue with three Appendices featuring translations into English of the 1914 paper in Hungarian by Schweitzer [145] and two interesting papers in Chinese by Lin [85] and Chen (1987) [32]. We end the thesis with an extensive bibliography of about 200 publications related to the Kantorovich Inequality. References to reviews in *Jahrbuch über die Fortschritte der Mathematik*, *Mathematical Reviews*, and *Zentralblatt für Mathematik und ihre Grenzgebiete*, are given with codes JFM, MR and Zbl, respectively.

Chapter 1

The Kantorovich Inequality

1.1 The Kantorovich Inequality (1948)

Our main focus in this thesis is on the so-called "Kantorovich Inequality":

$$\frac{t'At \cdot t'A^{-1}t}{(t't)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.1}$$

where t is a real $n \times 1$ vector and A is a real $n \times n$ symmetric positive definite matrix, with λ_1 and λ_n, respectively, its (fixed) largest and smallest, necessarily positive, eigenvalues.

Throughout this thesis we will assume that all vectors and matrices are real, and that a positive definite matrix is symmetric; we will also assume that $\lambda_1 > \lambda_n > 0$ and usually that both λ_1 and λ_n are known. If $\lambda_1 = \lambda_n$ then the matrix A becomes a multiple of the identity matrix and all our inequalities become equalities; if λ_1 and λ_n are not known but we know that $0 < m \leq \lambda_n < \lambda_1 \leq M$ then we can replace λ_1 and λ_n in the upper bound in (1.1) by M and m, respectively, since then

$$\frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \leq \frac{(m + M)^2}{4mM}, \tag{1.2}$$

cf. (1.11) below.

The "complementary" or "reversal" inequality:

$$1 \leq \frac{t'At \cdot t'A^{-1}t}{(t't)^2} \tag{1.3}$$

is a version of the well-known "Cauchy-Schwarz Inequality":

$$(x'y)^2 \le x'x \cdot y'y, \tag{1.4}$$

where x and y are $n \times 1$ (nonnull) vectors. Some further details on the Cauchy-Schwarz Inequality are given at the end of this section. [The term "complementary" in this context appears to be due to Diaz and Metcalf (1964) [40] and "reversal" to Marshall and Olkin (1964) [105].]

Another way of expressing the Kantorovich Inequality (1.1) is in the "reduced" form:

$$\frac{\sum_{i=1}^n \lambda_i u_i^2 \cdot \sum_{i=1}^n \lambda_i^{-1} u_i^2}{(\sum_{i=1}^n u_i^2)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.5}$$

which may be "normalized" by setting $z_i^2 = u_i^2 / \sum_{i=1}^n u_i^2$ so that

$$\sum_{i=1}^n \lambda_i z_i^2 \cdot \sum_{i=1}^n \frac{1}{\lambda_i} z_i^2 \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.6}$$

with $\sum_{i=1}^n z_i^2 = 1$. We will refer to (1.6) as the "normalized reduced" Kantorovich Inequality.

To see that the left-hand side of (1.1) equals the left-hand side in (1.5) we note that since the matrix A is real and symmetric it may be orthogonally diagonalized: $A = P\Lambda P'$, say, where $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ and P is orthogonal; and we then write $P't = u = \{u_i\}$.

When the vector t in the Kantorovich Inequality (1.1) is normalized with $t't = 1$ then (1.1) becomes the "normalized" (unreduced) Kantorovich Inequality:

$$t'At \cdot t'A^{-1}t \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}. \tag{1.7}$$

The upper bound in the Kantorovich Inequality (1.1),

$$\frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} = \frac{\lambda_1 + \lambda_n}{2} \cdot \frac{\lambda_1^{-1} + \lambda_n^{-1}}{2} = \left(\frac{\frac{1}{2}(\lambda_1 + \lambda_n)}{\sqrt{\lambda_1\lambda_n}} \right)^2, \tag{1.8}$$

is the square of the ratio of the arithmetic mean to the geometric mean of the largest and smallest eigenvalues of the matrix A. We may also express (1.8) as

$$\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} = \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}}+\sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2 = \frac{1}{4}\left(\sqrt{\kappa}+\frac{1}{\sqrt{\kappa}}\right)^2 = \frac{(\kappa+1)^2}{4\kappa}, \tag{1.9}$$

where

$$\kappa = \text{cond}(A) = \frac{\lambda_1}{\lambda_n} \tag{1.10}$$

is the so-called "condition number" of A and so $1 \le \kappa < \infty$. We note that the upper bound (1.8) = (1.9) is nondecreasing in κ, cf. (1.2) above, i.e.,

$$\frac{(\kappa_1+1)^2}{4\kappa_1} \le \frac{(\kappa_2+1)^2}{4\kappa_2} \quad \text{whenever} \quad \kappa_1 \le \kappa_2. \tag{1.11}$$

The Kantorovich Inequality[1] is named after the Nobel Laureate and Academician Leonid Vital'evich Kantorovich (1912–1986) for the inequality he established in 1948 ([72], pp. 142–144; cf. also [73], pp. 106–107) in a long survey article in Russian on functional analysis and applied mathematics. For a biography of Kantorovich see the paper [200] by Makarov and Sobolev (1990). The Kantorovich Inequality is well known, cf. e.g., Horn and Johnson ([69], pp. 444–445), Marcus and Minc ([104], pp. 110 & 117), and is useful in estimating convergence rate in descent methods in numerical analysis, cf. e.g., Borowski and Borwein ([22], page 319). Our interest in the Kantorovich Inequality, however, comes from statistics, where it may be used to identify a lower bound for the efficiency of the (ordinary) least squares estimator of a single parameter in the general linear (statistical) model, cf. e.g., Watson (1955) [165], Magness and McGuire (1961) [97], Golub (1963) [59], Hannan (1970) [64], and Puntanen (1987) [136]. Indeed this thesis was motivated by papers presented at the Fourth International Workshop on Matrix Methods for Statistics (Montréal, July 1995) by Shuangzhe Liu and Heinz Neudecker [92] and by Geoffrey S. Watson [169] (and the associated discussion). See also the recent

[1] *Note added in proof.* I am very grateful to Josip E. Pečarić for drawing my attention to the paper [50] by Roberto W. Frucht (1943) in which the "Kantorovich Inequality" (1.5) is established explicitly, five years before Kantorovich (1948) [72]. In this same paper [50] the continuous version of (1.5) is also established (by Beppo Levi [83]). For biographies of Roberto W. Frucht see [198] by Frank Harary and [197] by Carlos Gonzalez de la Fuente and for biographies of Beppo Levi (1875–1961) see [201] by Norbert Schappacher and René Schoof and [196] by Salvatore Coen.

Ph.D. thesis by Shuangzhe Liu [89] and the papers by Alpargu and Styan ([1], [2], [3] and [4]).

According to Pečarić and Mond ([129], page 384) the "Kantorovich Inequality" is originally due to Charles Hermite (1822–1901), but no reference is given[2]. The earliest form of (1.1) *per se* that we have found in the literature is in [60] by Greub and Rheinboldt (1959); cf. also [151] by Strang (1960).

Equality holds in the Kantorovich Inequality (1.1) when

$$t = \frac{1}{\sqrt{2}}(p_1 \pm p_n), \tag{1.12}$$

where p_1 and p_n are orthonormal eigenvectors of A corresponding, respectively, to λ_1 and λ_n. When λ_1 and λ_n both have multiplicity 1 then this condition is also necessary. When λ_1 and λ_n, however, have multiplicities $f \geq 1$ and $h \geq 1$, respectively, so that say

$$\lambda_1 = \cdots = \lambda_f > \lambda_{f+1} \geq \cdots \geq \lambda_{n-h} > \lambda_{n-h+1} = \cdots = \lambda_n \tag{1.13}$$

then for equality in (1.1) we need

$$t = \frac{1}{\sqrt{2}}(P_1 a_1 \pm P_n a_n),$$

where P_1 and P_n are matrices, respectively $n \times f$ and $n \times h$, with columns being orthonormal eigenvectors of A corresponding, respectively, to λ_1 and λ_n. The vectors a_1 and a_n are arbitrary except that $a_1'a_1 = a_n'a_n = 1$.

Equality holds in the "normalized reduced" version (1.6) of the Kantorovich Inequality whenever

$$z_1^2 = z_n^2 = \tfrac{1}{2} \quad \text{and} \quad z_2 = \cdots = z_{n-1} = 0; \tag{1.14}$$

when λ_1 and λ_n both have multiplicity 1 then this condition is also necessary. When λ_1 and λ_n, however, have multiplicities $f \geq 1$ and $h \geq 1$, respectively, as in (1.13), then for equality in (1.6) we need

[2]*Note added in proof.* Frucht (1943) [50] established the reduced form (1.5), while Kantorovich (1948) [72] proved the normalized reduced form (1.6).

$$z_{f+1} = \cdots = z_{n-h} = 0 \quad \text{and} \quad z_1^2 + \cdots + z_f^2 = \tfrac{1}{2} = z_{n-h+1}^2 + \cdots + z_n^2. \tag{1.15}$$

A continuous normalized version of the Kantorovich Inequality is

$$\int_c^d f(x) g^2(x) dx \cdot \int_c^d \frac{1}{f(x)} g^2(x) dx \leq \frac{(m+M)^2}{4mM}, \tag{1.16}$$

where $f(x)$, $1/f(x)$ and $g(x)$ are integrable functions on $[c, d]$ with $0 < m \leq f(x) \leq M$ and $\int_c^d g^2(x)dx = 1$, cf. Bătineţu-Giurgiu (1994) [13] [Mitrinović (1970) ([110], p. 60) observes that (1.16) is "known" but does not present it explicitly].

The inequality, cf. (1.4) above:

$$(x'y)^2 \leq x'x \cdot y'y, \tag{1.17}$$

is a vector version of the well-known Cauchy-Schwarz [or Buniakovski] inequality, cf. e.g., Borowski and Borwein ([22], page 73), Mitrinović ([110], pp. 30–32). The inequality (1.17) is named after [Baron] Augustin-Louis Cauchy (1789–1857) and Hermann Amandus Schwarz (1843–1921) [and Viktor Jakovlevich Buniakovski (1804–1899)], cf. ([22], pp. 62, 71 & 524).

Equality holds in (1.17) if and only if x and y are linearly dependent, i.e.,

$$(x'y)^2 = x'x \cdot y'y \Longleftrightarrow x \propto y.$$

Let t be an $n \times 1$ vector, and let A be an $n \times n$ positive definite matrix so that there exists an $n \times n$ nonsingular matrix F so that

$$A = FF'. \tag{1.18}$$

Substituting $x = F't$ and $y = F^{-1}t$ in (1.17), gives the following matrix version of the Cauchy-Schwarz Inequality, cf. (1.3):

$$1 \leq \frac{t'At \cdot t'A^{-1}t}{(t't)^2}. \tag{1.19}$$

Equality holds in (1.19) if and only if $At \propto t$, i.e., t is an eigenvector of A.

1.2 Five Proofs of the Kantorovich Inequality

In this section we present five different proofs of the Kantorovich Inequality, beginning with the original proof by Kantorovich (1948). Our next three proofs are due to Anderson (1971) [6], Styan (1983) [152] and Bühler (1987) [23]. We end this section (and chapter) with the very recently published proof by Pták (1995) [135]. There are many other proofs (see the extensive bibliography at the end of this thesis), including the proof by Chen (1987) [32] (translated from the Chinese as Appendix C in this thesis) and the interesting proof using linear programming by Raghavachari (1986) [138]; see also Schopf (1960) [144].

1.2.1 Kantorovich (1948)

Kantorovich (1948) (cf. pp. 142–144 in [72] and pp. 106–107 in [73]) established the "normalized reduced form" of the Kantorovich Inequality, cf. (1.6),

$$f = \sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{1.20}$$

where $\sum_{i=1}^{n} z_i^2 = 1$, using the method of Lagrange multipliers.

We assume, without loss of generality, that the λ_i are distinct and so $\lambda_1 > \cdots > \lambda_n > 0$. We equate to zero the derivatives, with respect to z_i $(i = 1, \ldots, n)$, of the function

$$F = f - \alpha\left(\sum_{i=1}^{n} z_i^2 - 1\right),$$

where α is a Lagrange multipler. Writing $f = \sigma \cdot \hat{\sigma}$, where $\sigma = \sum_{i=1}^{n} \lambda_i z_i^2$ and $\hat{\sigma} = \sum_{i=1}^{n} \lambda_i^{-1} z_i^2$, we obtain

$$\frac{\partial F}{\partial z_i} = 2\left(\sigma \frac{1}{\lambda_i} z_i + \hat{\sigma}\lambda_i z_i - \alpha z_i\right) = 0 \quad (i = 1, \ldots, n),$$

and so

$$z_i(\sigma + \hat{\sigma}\lambda_i^2 - \alpha\lambda_i) = 0, \quad (i = 1, \ldots, n). \tag{1.21}$$

The second factor in (1.21) is a polynomial of the second degree in λ_i and so at most two distinct values of λ_i can make this factor zero. Hence at most two values z_k and z_l, say, of z_i are non-zero. In this event

$$\begin{aligned} f &= (\lambda_k z_k^2 + \lambda_l z_l^2)\left(\frac{1}{\lambda_k} z_k^2 + \frac{1}{\lambda_l} z_l^2\right) \\ &= \frac{1}{4}\left(\sqrt{\frac{\lambda_k}{\lambda_l}} + \sqrt{\frac{\lambda_l}{\lambda_k}}\right)^2 (z_k^2 + z_l^2)^2 - \frac{1}{4}\left(\sqrt{\frac{\lambda_k}{\lambda_l}} - \sqrt{\frac{\lambda_l}{\lambda_k}}\right)^2 (z_k^2 - z_l^2)^2 \qquad (1.22) \\ &\leq \frac{1}{4}\left(\sqrt{\frac{\lambda_k}{\lambda_l}} + \sqrt{\frac{\lambda_l}{\lambda_k}}\right)^2 \\ &\leq \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2 \qquad (1.23) \\ &= \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \qquad (1.24) \end{aligned}$$

cf. (1.2), (1.9) and (1.11), and thus (1.20) follows. [On page 143 of [72] and on page 107 of [73] the factor

$$\left(\sqrt{\frac{\lambda_k}{\lambda_l}} - \sqrt{\frac{\lambda_l}{\lambda_k}}\right)^2$$

in (1.22) is given with a + sign rather than the correct minus sign.] □

1.2.2 Anderson (1971)

Our favourite proof of the Kantorovich Inequality may be the following simple probabilistic proof given in 1971 by T. W. Anderson ([6], Lemma 10.2.5, p. 569), see also Bühler [23] and Watson [168].

We may write the product f in (1.20) as

$$f = \sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 = \mathbf{E}(T)\mathbf{E}\left(\frac{1}{T}\right), \qquad (1.25)$$

the product of the expected value of a random variable T and the expected value of its reciprocal $1/T$, where T assumes the values $\lambda_i \in [m, M]$ with probabilities $p_i = z_i^2$ $(i = 1, \ldots, n)$; $0 < m < M < \infty$.

For $0 < m \leq T \leq M$,

$$0 \leq (M - T)(T - m) = (M + m - T)T - mM, \tag{1.26}$$

which implies

$$\frac{1}{T} \leq \frac{m + M - T}{mM}; \tag{1.27}$$

and so

$$\begin{aligned} \mathbf{E}(T) \cdot \mathbf{E}\left(\frac{1}{T}\right) &\leq \mathbf{E}(T) \cdot \frac{m + M - \mathbf{E}(T)}{mM} \\ &= \frac{(m+M)^2}{4mM} - \frac{1}{mM}[\mathbf{E}(T) - \tfrac{1}{2}(m+M)]^2 \\ &\leq \frac{(m+M)^2}{4mM}, \end{aligned} \tag{1.28}$$

and our proof is complete. □

1.2.3 Styan (1983)

Styan (1983) [152], cf. also [1], proved the Kantorovich Inequality using the following inequality due to Marshall-Olkin (1964) [105]:

$$\lambda_1\lambda_n \cdot z'\Lambda^{-1}z \leq \lambda_1 + \lambda_n - z'\Lambda z, \tag{1.29}$$

where $z'z = 1$ and $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, with $\lambda_1 \geq \cdots \geq \lambda_n$.

The (normalized reduced) "Marshall-Olkin Inequality" (1.29), cf. also (3.46) in Chapter 3, follows directly from

$$\begin{aligned} \lambda_1 + \lambda_n - z'\Lambda z - \lambda_1\lambda_n \cdot z'\Lambda^{-1}z &= z'(\mathrm{diag}\{\lambda_1 + \lambda_n - \lambda_i - \lambda_i^{-1}\lambda_1\lambda_n\})z \\ &= z'(\mathrm{diag}\{(\lambda_1 - \lambda_i)(\lambda_i - \lambda_n)/\lambda_i\})z \\ &\geq 0 \end{aligned}$$

since $\lambda_1 \geq \lambda_i \geq \lambda_n > 0$.

We may then prove the "normalized reduced" Kantorovich Inequality (1.20) as follows:

$$\begin{aligned}\lambda_1\lambda_n(z'\Lambda z)(z'\Lambda^{-1}z) &= z'\Lambda z(\lambda_1\lambda_n \cdot z'\Lambda^{-1}z) \\ &\leq z'\Lambda z(\lambda_1 + \lambda_n - z'\Lambda z) \\ &= z'\Lambda z(\lambda_1 + \lambda_n) - (z'\Lambda z)^2 \\ &= \tfrac{1}{4}(\lambda_1 + \lambda_n)^2 - \left(z'\Lambda z - \tfrac{1}{2}(\lambda_1 + \lambda_n)\right)^2 \\ &\leq \tfrac{1}{4}(\lambda_1 + \lambda_n)^2, \qquad (1.30)\end{aligned}$$

cf. (1.28) and (3.43) in Chapter 3, and so (1.20) is established. □

1.2.4 Bühler (1987)

The simple probabilistic proof given in 1987 by Wolfgang J. Bühler [23], cf. also [1], starts, as with the proof by Anderson [6] that we presented in Section 1.2.2, with

$$f = \sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 = \mathbf{E}(T)\mathbf{E}(1/T), \qquad (1.31)$$

the product of the expected value of a random variable T and the expected value of its reciprocal $1/T$, where T assumes the values $\lambda_i \in [m, M]$ with probabilities $p_i = z_i^2$ $(i = 1, \ldots, n)$; $0 < m < M < \infty$.

From the Cauchy-Schwarz or correlation coefficient inequality:

$$-1 \leq \frac{\mathbf{E}(T \cdot 1/T) - \mathbf{E}(T)\mathbf{E}(1/T)}{\sqrt{\mathrm{Var}(T)\mathrm{Var}(1/T)}} = \mathrm{corr}(T, 1/T)$$

it follows that

$$\mathbf{E}(T)\mathbf{E}(1/T) \leq 1 + \sqrt{\mathrm{Var}(T)\mathrm{Var}(1/T)}. \qquad (1.32)$$

However

$$\mathrm{Var}(T) = \mathrm{Var}\left(T - \frac{M+m}{2}\right) = \mathrm{Var}(U),$$

say. Since

$$0 < m \leq T \leq M$$

we have

$$-\frac{M-m}{2} \leq U = T - \frac{M+m}{2} \leq \frac{M-m}{2}$$

and so

$$U^2 \leq \left(\frac{M-m}{2}\right)^2. \tag{1.33}$$

Hence

$$\text{Var}(T) = \text{Var}(U) = E(U^2) - [(E(U)]^2 \leq E(U^2) \leq \left(\frac{M-m}{2}\right)^2. \tag{1.34}$$

Similarly, we have

$$\text{Var}(1/T) \leq \frac{1}{4}\left(\frac{1}{m} - \frac{1}{M}\right)^2. \tag{1.35}$$

Combining (1.34) and (1.35) with (1.32) yields

$$f = \mathbf{E}(T)\mathbf{E}(1/T) \leq 1 + \sqrt{\left(\frac{M-m}{2}\right)^2 \frac{1}{4}\left(\frac{1}{m} - \frac{1}{M}\right)^2} = \frac{(m+M)^2}{4mM},$$

which proves the Kantorovich Inequality

$$f = \sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 \leq \frac{(m+M)^2}{4mM}. \tag{1.36}$$

Equality occurs in (1.36) if and only if equality occurs in both (1.34) and (1.35), and this happens if and only if

$$\mathbf{E}(U) = 0 \quad \text{and} \quad U^2 = \left(\frac{M-m}{2}\right)^2,$$

i.e., $T = m = \lambda_n$ and $T = M = \lambda_1$, each with probability 1/2. When the eigenvalues λ_1 and λ_n are both simple, i.e., with multiplicity 1, then this translates into our earlier conditions for equality in the Kantorovich Inequality:

$$z_1^2 = z_n^2 = \tfrac{1}{2} \quad \text{and} \quad z_2 = \cdots = z_{n-1} = 0. \tag{1.37}$$

When the eigenvalues λ_1 and λ_n have multiplicities $f \geq 1$ and $h \geq 1$, respectively, so that

$$\lambda_1 = \cdots = \lambda_f > \lambda_{f+1} \geq \cdots \geq \lambda_{n-h} > \lambda_{n-h+1} = \cdots = \lambda_n$$

then for equality in (1.36) we need, cf. (1.15),

$$z_{f+1} = \cdots = z_{n-h} = 0 \quad \text{and} \quad z_1^2 + \cdots + z_f^2 = \tfrac{1}{2} = z_{n-h+1}^2 + \cdots + z_n^2.$$

□

1.2.5 Pták (1995)

As we observed in §1.1, cf. (1.6) and (1.8), the normalized reduced Kantorovich Inequality may be stated as follows:

$$\sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 \leq \frac{a^2}{g^2}, \tag{1.38}$$

where $\sum_{i=1}^{n} z_i^2 = 1$ and

$$a = \tfrac{1}{2}(\lambda_1 + \lambda_n) \quad \text{and} \quad g = \sqrt{\lambda_1 \lambda_n}$$

are, respectively, the arithmetic and geometric means of λ_1 and λ_n. Since (1.38) is invariant with respect to replacing each λ_i with a positive multiple $c\lambda_i$, we may assume that $g = 1$ or, equivalently, that $\lambda_n = 1/\lambda_1$. It then follows that

$$\lambda_i + \frac{1}{\lambda_i} \leq \lambda_1 + \frac{1}{\lambda_1} \quad (i = 1, \ldots, n),$$

and so

$$\begin{aligned}
\sqrt{\sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2} &\leq \tfrac{1}{2}\left(\sum_{i=1}^{n} \lambda_i z_i^2 + \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2\right) \\
&= \tfrac{1}{2}\sum_{i=1}^{n}\left(\lambda_i + \frac{1}{\lambda_i}\right) z_i^2 \\
&\leq \tfrac{1}{2}\left(\lambda_1 + \frac{1}{\lambda_1}\right)\sum_{i=1}^{n} z_i^2 \\
&= \tfrac{1}{2}\left(\lambda_1 + \frac{1}{\lambda_1}\right) = \tfrac{1}{2}(\lambda_1 + \lambda_n) = a = \frac{a}{g},
\end{aligned}$$

and (1.38) follows at once. □

Chapter 2

Five Inequalities Related to the Kantorovich Inequality: 1914–1959

In this chapter we consider the following five "named" inequalities:

- §2.1 The Schweitzer Inequality (1914)
- §2.2 The Pólya-Szegö Inequality (1925)
- §2.3 The Cassels Inequality (1951/1955)
- §2.4 The Krasnosel'skiĭ-Kreĭn Inequality (1952)/ Householder (1964)
- §2.5 The Greub-Rheinboldt Inequality (1959),

which were published in 1914–1959. As we will show in §2.6 these five inequalities are all equivalent to the Kantorovich Inequality.

2.1 The Schweitzer Inequality (1914)

As far as we know, the first inequality of the type (1.5) to be published was in 1914 by Pál Schweitzer [145]. [Schweitzer's original paper [145] was published in Hungarian; we present an English translation as Appendix A of this thesis.] Schweitzer

(1914) considered real positive numbers

$$x_1, \ldots, x_n; \quad 0 < m \le x_i \le M \quad (i = 1, \ldots, n),$$

and showed that

$$\frac{1}{n}(x_1 + \cdots + x_n) \cdot \frac{1}{n}\left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right) \le \frac{1}{2}(m + M) \cdot \frac{1}{2}\left(\frac{1}{m} + \frac{1}{M}\right) = \frac{(m+M)^2}{4mM}. \tag{2.1}$$

The "Schweitzer Inequality" (2.1) is a special case of the reduced Kantorovich Inequality (1.5) with $\lambda_i = x_i$, $u_i^2 = 1$, $\lambda_1 = M$ and $\lambda_n = m$. And so we may consider the reduced Kantorovich Inequality as being a "weighted" version of the Schweitzer Inequality. An interesting proof of the Schweitzer Inequality using majorization is given by Marshall and Olkin ([105], p. 71).

It was shown in 1961 by Peter Henrici in [68] that surprisingly the Schweitzer Inequality (2.1) also implies the Kantorovich Inequality (1.5). For details see §2.6.1.

The complementary inequality

$$1 \le \frac{1}{n}(x_1 + \cdots + x_n) \cdot \frac{1}{n}\left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right),$$

which follows at once from the Cauchy-Schwarz Inequality (1.4), is the well-known arithmetic-harmonic mean inequality, cf. e.g., Mitrinović ([110], pp. 206–207).

From the conditions for equality in the Kantorovich Inequality (1.1) we see that equality can hold in the Schweitzer Inequality (2.1) only if n is even and then if and only if

$$x_1 = \cdots = x_{\frac{n}{2}} = m \quad \text{and} \quad x_{\frac{n}{2}+1} = \cdots = x_n = M. \tag{2.2}$$

In 1972 Alexandru Lupaş [96] gave the following version of the Schweitzer Inequality, which is stronger than the original Schweitzer Inequality (2.1) when n is odd (and identical to (2.1) when n is even):

$$\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} \frac{1}{x_i} \le \frac{1}{mM}\left(\left[\frac{n}{2}\right]M + \left[\frac{n+1}{2}\right]m\right)\left(\left[\frac{n+1}{2}\right]M + \left[\frac{n}{2}\right]m\right), \tag{2.3}$$

where [.] denotes the integral part. Equality is attained in (2.3) when the smallest $[n/2]$ of the numbers $x_1, \ldots, x_n$ are equal to m and the $[n/2]$ largest are equal to M, and when n is odd the "middle" x_i is equal to either m or M.

To prove (2.3) we follow Lupaş [96] by using a result in [20] by Biernacki, Pidek and Ryll-Nardzewski (1950) [cf. Mitrinović ([110], p. 205, §3.3.26)], who showed that if the real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ satisfy $m_1 \le a_i \le M_1$, $m_2 \le b_i \le M_2$ $(i = 1, \ldots, n)$, then

$$\left|\sum_{i=1}^{n} a_i b_i - \frac{1}{n}\sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i\right| \le (m_1 - M_1)(m_2 - M_2)w(n), \tag{2.4}$$

where

$$w(n) = \begin{cases} n/4 & \text{for } n \text{ even,} \\ \\ (n^2-1)/4n & \text{for } n \text{ odd.} \end{cases}$$

If we set $a_i = x_i > 0$, $b_i = 1/x_i$, $M_1 = M$, $m_1 = m$, $M_2 = 1/m$, $m_2 = 1/M$, we then obtain from (2.4) that

$$\frac{1}{n^2}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} \frac{1}{x_i} \le \frac{(m+M)^2}{4mM} - h(n),$$

where

$$h(n) = \begin{cases} 0 & \text{for } n \text{ even,} \\ \\ (M-m)^2/(4mMn^2) & \text{for } n \text{ odd.} \end{cases}$$

An easy computation shows that

$$\frac{(m+M)^2}{4mM} - h(n) = \frac{1}{mMn^2}\left(\left[\frac{n}{2}\right]M + \left[\frac{n+1}{2}\right]m\right)\left(\left[\frac{n+1}{2}\right]M + \left[\frac{n}{2}\right]m\right).$$

□

In the same paper [96] Lupaş stated (without proof) that the normalized reduced Kantorovich Inequality (1.6) could be improved to:

$$\sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 \le \frac{1}{mMn^2} \left(\left[\frac{n}{2}\right] M + \left[\frac{n+1}{2}\right] m \right) \left(\left[\frac{n+1}{2}\right] M + \left[\frac{n}{2}\right] m \right), \tag{2.5}$$

where $z_i^2 \ge 0$ $(i = 1, \dots, n)$ and $\sum_{i=1}^{n} z_i^2 = 1$. The inequality (2.5) reduces to the Kantorovich Inequality when n is even; it is, however, only valid for n odd when all the $z_i^2 > 0$ $(i = 1, \dots, n)$, and we allow the λ_i to vary (unless there are only two distinct values of the λ_i with multiplicities $[n/2]$ and $[n/2]+1$ respectively).

Also in [145] Schweitzer (1914) established a continuous analogue of (2.1). Let $f(x)$ and $1/f(x)$ be integrable functions on $[c, d]$ with $0 < m \le f(x) \le M$ on $[c, d]$. Then

$$\frac{1}{d-c} \int_c^d f(x)dx \cdot \frac{1}{d-c} \int_c^d \frac{1}{f(x)} dx \le \frac{(m+M)^2}{4mM}. \tag{2.6}$$

A quick proof of (2.6) was given in 1963 by Rennie [140] [for Schweitzer's original proof see Appendix A]: Integrating the inequality $(f-m)(f-M)/f \le 0$ from c to d gives

$$\int_c^d f(x)dx + mM \int_c^d \frac{1}{f(x)} dx \le (d-c)(m+M). \tag{2.7}$$

Multiplying (2.7) by $u = mM \int_c^d 1/f(x)\, dx$ gives

$$\begin{aligned} u \int_c^d f(x)dx &\le (d-c)(m+M)u - u^2 \\ &= -\left(\tfrac{1}{2}(d-c)(m+M) - u\right)^2 + \left(\tfrac{1}{2}(d-c)(m+M)\right)^2 \\ &\le \tfrac{1}{4}(d-c)^2(m+M)^2, \end{aligned}$$

which establishes (2.6). □

In 1961 E. Makai in [99] (cf. Mitrinović ([110], pp. 60–61) showed (details in §2.6.2 below) that the continuous Schweitzer inequality (2.6) implies the (discrete) Kantorovich Inequality (1.1).

2.2 The Pólya-Szegö Inequality (1925)

In 1925 George Pólya and Gábor Szegö, in the First Edition of Volume I (cf. [133], [134]) of their well-known book: *Aufgaben und Lehrsätze aus der Analysis–Problems* *and Theorems in Analysis*, showed that

$$\frac{\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2}{(\sum_{i=1}^n a_i b_i)^2} \le \frac{(ab+AB)^2}{4abAB}. \tag{2.8}$$

where

$$0 < a \le a_i \le A, \quad 0 < b \le b_i \le B \quad (i = 1, \ldots, n).$$

If we put $a_i^2 = x_i$ and $b_i^2 = 1/x_i$, with $a^2 = m, A^2 = M, b^2 = 1/M$ and $B^2 = 1/m$, in the "Pólya-Szegö Inequality" (2.8) then it becomes the Schweitzer Inequality (2.1). As we will observe at the end of §2.6.1, the Schweitzer Inequality (2.1) implies the Kantorovich Inequality, which in turns implies the Pólya-Szegö Inequality (2.8) since if we put $u_i^2 = a_i b_i$ and $\lambda_i = a_i/b_i$ in (1.5) then (1.5) becomes (2.8).

Equality holds in the Pólya-Szegö Inequality (2.8) if and only if

$$p = n\frac{A}{a} \Big/ \left(\frac{A}{a} + \frac{B}{b}\right) \quad \text{and} \quad q = n\frac{B}{b} \Big/ \left(\frac{A}{a} + \frac{B}{b}\right)$$

are positive integers and if p of the numbers $a_1, \ldots, a_n$ are equal to a and q of these numbers are equal to A, and if the corresponding numbers b_i are equal to B and b respectively.

A continuous version of (2.8), given in 1925 by Pólya-Szegö ([134], pp. 71–72, 254), is

$$\frac{\int_c^d f^2(x)\,dx \int_c^d g^2(x)\,dx}{\left[\int_c^d f(x)g(x)\,dx\right]^2} \le \frac{(ab+AB)^2}{4abAB}, \tag{2.9}$$

where $f(x)$ and $g(x)$ are continuous functions in the interval $[c,d]$ with $0 < a \le f(x) \le A$ and $0 < b \le g(x) \le B$.

The special case of (2.9) with $a = b$ and $A = B$:

$$\frac{\int_c^d f^2(x)\,dx \int_c^d g^2(x)\,dx}{\left[\int_c^d f(x)g(x)\,dx\right]^2} \le \frac{(a^2+A^2)^2}{4a^2A^2} \tag{2.10}$$

was posed as a "Problem" in 1914 by József Kürschák (1864–1933) [82] (in the same journal and volume as Schweitzer [145] but just over a hundred pages later!). [As far as we know there was no published solution *per se* to Kürschák's problem (2.10).] We have found no other similar inequalities published before 1948; an extensive bibliography is given at the end of this thesis.

2.3 The Cassels Inequality (1951/1955)

John William Scott Cassels (1922–...) [28] established[1] the following inequality as the Appendix in the Ph.D. thesis [163] by Watson (1951), see also [164] and [165]:

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \leq \max_{i,j} \frac{(a_i b_j + a_j b_i)^2}{4 a_i a_j b_i b_j}, \tag{2.11}$$

where $a_i > 0$, $b_i > 0$ and $w_i \geq 0$ $(i = 1, \ldots, n)$. By substituting

$$m = \min_i \frac{a_i}{b_i} \quad \text{and} \quad M = \max_i \frac{a_i}{b_i}, \tag{2.12}$$

the "Cassels Inequality" (2.11) becomes

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \leq \frac{(m+M)^2}{4mM}, \tag{2.13}$$

cf. (3.2) in [165].

The Cassels Inequality (2.13) is, however, the same as the Krasnosel'skiĭ-Kreĭn Inequality (2.22), as observed by Styan [153]. To see this, we substitute $a_i = b_i \lambda_i$, $b_i^2 w_i = u_i^2$, $m = \min \lambda_i = \min a_i/b_i = \lambda_n$ and $M = \max \lambda_i = \max a_i/b_i = \lambda_1$ in (2.13), which then becomes (2.22).

If we put the weights $w_i = 1$ in (2.11) then we obtain the "unweighted" Cassels Inequality:

$$\frac{\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2}{(\sum_{i=1}^n a_i b_i)^2} \leq \frac{(m+M)^2}{4mM}, \tag{2.14}$$

which is slightly stronger than the Pólya-Szegő Inequality (2.8):

[1] Watson [170] observed in 1950 that he asked "Henry Daniels who asked Cassels as they were putting on their gowns before lecturing for a reverse of the Cauchy-Schwarz Inequality; he just worked it out overnight". See Watson (1951) [163], (1952) [164], and (1955) [165].

$$\frac{\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2}{(\sum_{i=1}^{n} a_i b_i)^2} \le \frac{(ab + AB)^2}{4abAB}, \tag{2.15}$$

since, in general,

$$\frac{(m+M)^2}{4mM} \le \frac{(ab+AB)^2}{4abAB}, \tag{2.16}$$

where, cf. (1.2),

$$m = \min \frac{a_i}{b_i} \ge \frac{\min a_i}{\max b_i} = \frac{a}{B} \quad \text{and} \quad M = \max \frac{a_i}{b_i} \le \frac{\max a_i}{\min b_i} = \frac{A}{b}. \tag{2.17}$$

Equality holds in (2.16) if and only if equality holds throughout (2.17).

To establish (2.16) we rewrite it as, cf. (1.2),

$$\frac{(1+x)^2}{4x} \le \frac{(1+y)^2}{4y}, \tag{2.18}$$

where

$$x = \frac{M}{m} \quad \text{and} \quad y = \frac{AB}{ab}.$$

Since $x \le y$, cf. (2.17), the inequality (2.16) follows, cf. (1.2) and (1.11) in Chapter 1, as does the characterization for equality.

2.4 The Krasnosel'skiĭ-Kreĭn Inequality (1952)/ Householder (1964)

In 1952 Mark Aleksandrovich Krasnosel'skiĭ (1920-...) and Selim Griogor'evich Kreĭn in [80] showed that

$$\frac{t'A^2t \cdot t't}{(t'At)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{2.19}$$

where as above λ_1 and λ_n are the largest and smallest (fixed) eigenvalues of the positive definite matrix A and t is an $n \times 1$ vector.

The "Krasnosel'skiĭ-Kreĭn Inequality" (2.19), however, is just an alternate version of the Kantorovich Inequality (1.1). Since A is positive definite we may define a symmetric positive definite square root $A^{1/2}$ and substitute $t = A^{-1/2}u$ in (2.19) to obtain the Kantorovich Inequality

$$\frac{u'Au \cdot u'A^{-1}u}{(u'u)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}. \tag{2.20}$$

Indeed, as pointed out by Householder (1964) in his well-known book ([71], p. 83),

$$\frac{x'A^{v+1}x \cdot x'A^{v-1}x}{(x'A^v x)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}. \tag{2.21}$$

where v is an integer. Householder ([71], p. 83) shows that (2.21) remains valid with A complex Hermitian positive definite, x' the conjugate transpose of x and for any v.

To establish (2.21) we put $u = A^{v/2}x = (A^{1/2})^v x$ in (2.20). Clearly $v = 0$ in (2.21) yields the Kantorovich Inequality (2.20) while $v = 1$ yields the Krasnosel'skiĭ-Kreĭn Inequality (2.19).

Equality holds in (2.19) when

$$t = \frac{1}{\sqrt{\lambda_1}}p_1 \pm \frac{1}{\sqrt{\lambda_n}}p_n,$$

cf. (1.12), where p_1 and p_n are orthonormal eigenvectors of A corresponding, respectively, to λ_1 and λ_n. Another way of expressing the Krasnosel'skiĭ-Kreĭn Inequality (2.19) is in the "reduced" form[1]:

$$\frac{\sum_{i=1}^n \lambda_i^2 u_i^2 \cdot \sum_{i=1}^n u_i^2}{(\sum_{i=1}^n \lambda_i u_i^2)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}. \tag{2.22}$$

In the statistical theory of experimental design, Chakrabarti (1963) [195] proposed the following measure of imbalance:

$$\psi = \frac{\text{tr } C^2}{(\text{tr } C)^2} = \frac{\sum_{i=1}^s \gamma_i^2}{(\sum_{i=1}^s \gamma_i)^2}, \tag{2.23}$$

where the non-negative definite matrix C is the so-called C-matrix and the γ_i's are its s positive eigenvalues. Bartlett ([191], p. 99) showed that

$$\psi \leq 1 \tag{2.24}$$

[1] *Note added in proof.* A continuous version of (2.22) is given by Frucht (1943) [50].

improving upon the inequality $\psi \leq 2$ given by Chakrabarti [195]. We note, however, from (2.22) that

$$\psi \leq \frac{(\gamma_1 + \gamma_s)^2}{4s\gamma_1\gamma_s}, \tag{2.25}$$

where γ_1 and γ_s are respectively the largest and smallest positive eigenvalues of C. Then (2.25) is an improvement on (2.24) if and only if

$$s \geq \frac{(\gamma_1 + \gamma_s)^2}{4\gamma_1\gamma_s} = \frac{(\kappa + 1)^2}{4\kappa}, \tag{2.26}$$

where the condition number $\kappa = \gamma_1/\gamma_s$. The right-hand side of (2.26) tends to infinity as $\kappa \to \infty$. For "moderate" κ, however, say $\kappa = 4 \Longrightarrow s \geq 2$ or $\kappa = 16 \Longrightarrow s \geq 5$ gives better bound than (2.24).

Thibaudeau and Styan (1985) [203] gave other upper-bound improvements to (2.24).

2.5 The Greub-Rheinboldt Inequality (1959)

In 1959 Werner Greub and Werner Rheinboldt [60] showed that:

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \leq \frac{(ab + AB)^2}{4abAB}, \tag{2.27}$$

where

$$0 < a \leq a_i \leq A, \quad 0 < b \leq b_i \leq B \quad (i = 1, \ldots, n).$$

We note that the "Greub-Rheinboldt Inequality" (2.27) is a "weighted" version of the Pólya-Szegö Inequality (2.8); here we mean weighted in the same sense that the reduced Kantorovich Inequality (1.5) is a weighted version of the Schweitzer Inequality (2.1).

The Greub-Rheinboldt Inequality (2.27) is *per se* a slightly weaker version of the Cassels Inequality (2.11):

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{(\sum_{i=1}^n a_i b_i w_i)^2} \leq \max_{i,j} \frac{(a_i b_j + a_j b_i)^2}{4 a_i a_j b_i b_j} = \frac{(m + M)^2}{4mM}, \tag{2.28}$$

in that the Greub-Rheinboldt upper bound may be greater than the Cassels upper bound, i.e.,

$$\frac{(m+M)^2}{4mM} \leq \frac{(ab+AB)^2}{4abAB},$$

cf. (2.12).

We will, however, now show following an observation by Styan [153], that a reparameterization of the Greub-Rheinboldt Inequality (2.27) makes it, in fact, equivalent to the Cassels Inequality. To see this we substitute $a_i = \lambda_i$, $a = \lambda_n$, $A = \lambda_1$, $b_i = b = B = 1$, $w_i = u_i^2$ in the Greub-Rheinboldt Inequality

$$\frac{\sum_{i=1}^n a_i^2 w_i \cdot \sum_{i=1}^n b_i^2 w_i}{\left(\sum_{i=1}^n a_i b_i w_i\right)^2} \leq \frac{(ab+AB)^2}{4abAB}, \tag{2.29}$$

which then becomes the Krasnosel'skiĭ-Kreĭn Inequality:

$$\frac{\sum_{i=1}^n \lambda_i^2 u_i^2 \cdot \sum_{i=1}^n u_i^2}{(\sum_{i=1}^n \lambda_i u_i^2)^2} \leq \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n},$$

which we have already shown to be equivalent to the Cassels Inequality, cf. §2.3.

An integral analogue of the Greub-Rheinboldt Inequality is

$$\frac{\int_c^d f^2(x)h^2(x)\,dx \cdot \int_c^d g^2(x)h^2(x)\,dx}{\left[\int_c^d f(x)g(x)h^2(x)\,dx\right]^2} \leq \frac{(ab+AB)^2}{4abAB},$$

where $f(x)$, $g(x)$ and $h(x)$ are continuous functions on the interval $[c,d]$ with $0 < a \leq f(x) \leq A$, $0 < b \leq g(x) \leq B$ and $\int_c^d h^2(x) < \infty$. [Mitrinović ([110], p. 60) observed that such an integral analogue was "known" but did not give it.]

2.6 Six Named Inequalities are Equivalent

In this thesis we have so far considered the following six named discrete inequalities:

- §1.1 The Kantorovich Inequality (1.1)
- §2.1 The Schweitzer Inequality (2.1)
- §2.2 The Pólya-Szegö Inequality (2.8)

- §2.4 The Krasnosel'skiĭ-Kreĭn Inequality (2.22)
- §2.3 The Cassels Inequality (2.11)
- §2.5 The Greub-Rheinboldt Inequality (2.27),

and, as we have already observed, it is easy to see that:

- Kantorovich (1.5) $\Longrightarrow$ Schweitzer (2.1)
- Pólya-Szegö (2.8) $\Longrightarrow$ Schweitzer (2.1), and that:
- Greub-Rheinboldt (2.27) $\Longrightarrow$ Pólya-Szegö (2.8).

Moreover, we have shown that:

- Kantorovich (1.5) $\Longleftrightarrow$ Krasnosel'skiĭ-Kreĭn (2.22)

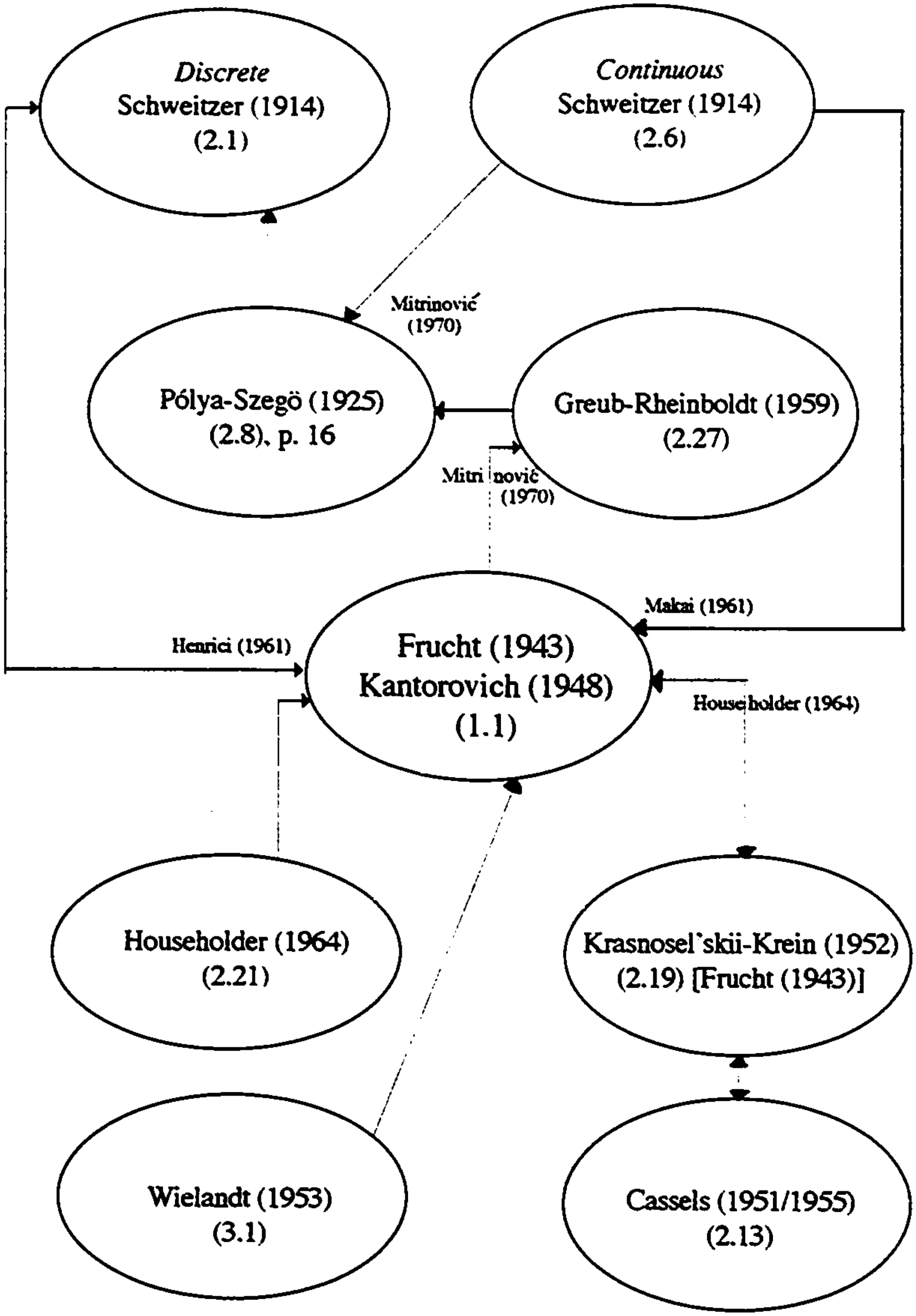
  $\Longleftrightarrow$ Cassels (2.11) $\Longleftrightarrow$ Greub-Rheinboldt (2.27).

And since Henrici (1961) [68] showed that:

- Schweitzer (2.1) $\Longrightarrow$ Kantorovich (1.5),

it follows that these six named inequalities are all equivalent, cf. Fig 1.

Fig. 1



Discrete
Schweitzer (1914)
(2.1)
Continuous
Schweitzer (1914)
(2.6)
Mitrinović
(1970)
Pólya-Szegö (1925)
(2.8), p. 16
Greub-Rheinboldt (1959)
(2.27)
Mitrinovič
(1970)
Makai (1961)
Henrici (1961)
Frucht (1943)
Kantorovich (1948)
(1.1)
Householder (1964)
Householder (1964)
(2.21)
Krasnosel'skii-Krein (1952)
(2.19) [Frucht (1943)]
Wielandt (1953)
(3.1)
Cassels (1951/1955)
(2.13)

We end this chapter by showing that both the discrete and continuous versions of the Schweitzer Inequality imply the discrete Kantorovich Inequality.

2.6.1 Henrici (1961): Discrete Schweitzer Implies Kantorovich

To see that the discrete Schweitzer Inequality

$$\frac{1}{n}(\lambda_1+\cdots+\lambda_n)\cdot\frac{1}{n}\left(\frac{1}{\lambda_1}+\cdots+\frac{1}{\lambda_n}\right)\leq\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right) \tag{2.30}$$

implies the normalized reduced Kantorovich Inequality

$$\sum_{i=1}^{n}\lambda_i z_i^2\cdot\sum_{i=1}^{n}\frac{1}{\lambda_i}z_i^2\leq\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \tag{2.31}$$

we follow the proof by Henrici (1961) [68]. It suffices to show that(2.31) holds for all rational z_i^2 with $\sum z_i^2=1$. Let us choose n to be "very large" so that each λ_i occurs "many times", and write

$$\lambda_{(1)}<\cdots<\lambda_{(d)}$$

for the d distinct λ's with multiplicities $m_1,\ldots,m_d$ and $\sum m_j=n$. Then the left-hand side of (2.30),

$$\left(\frac{\sum\lambda_{(j)}m_j}{\sum m_j}\right)\left(\frac{\sum\lambda_{(j)}^{-1}m_j}{\sum m_j}\right)=\sum\lambda_{(j)}z_j^2\cdot\sum\frac{1}{\lambda_{(j)}}z_j^2,$$

the left-hand side of (2.31), with $z_j^2=m_j/\sum m_j$, and the proof is complete. □

It was observed by Kantorovich (1948) in a footnote (on page 143 of [72] and page 106 of [73]) that his inequality (2.31) "is a special case of" the Pólya-Szegö Inequality (2.8)[2]. George E. Forsythe, however, who edited the 1952 English translation [73] of

[2] *Note added in proof.* Edwin F. Beckenbach makes the identical claim in his review [17] of the paper [50] by Roberto W. Frucht (1943). Neither this observation, however, nor any mention of the Kantorovich Inequality *per se* appears to be made by Beckenbach and Bellman in their famous book [19], first published in 1961 (cf. pp. 44–45); see also Beckenbach (1964) [18].

[72], noted (also on page 106 of [73]) that "it is not clear to me that Kantorovich's inequality really is a special case of that of Pólya and Szegö". Greub and Rheinboldt ([60], p. 407) found Forsythe's remark to be "well justified". The Pólya-Szegö Inequality, however, does imply the Kantorovich Inequality, albeit indirectly, since the discrete Schweitzer Inequality is a special case of the Pólya-Szegö Inequality and, as we have just seen, the discrete Schweitzer Inequality implies the Kantorovich Inequality.

2.6.2 Makai (1961): Continuous Schweitzer Implies Kantorovich

Makai (1961) [99] showed that the continuous Schweitzer Inequality

$$\frac{1}{b-a}\int_a^b f(x)dx \cdot \frac{1}{b-a}\int_a^b \frac{1}{f(x)}dx \le \frac{(m+M)^2}{4mM} \tag{2.32}$$

implies the normalized reduced Kantorovich Inequality (2.31). To see this, we put $a = 0,\ b = \sum_{i=1}^{n} z_i^2$ in (2.32) and

$$f(x) = \begin{cases} \lambda_1 & \text{for } 0 \le x < z_1^2 \\ \\ \lambda_i & \text{for } \sum_{j=1}^{i-1} z_j^2 \le x < \sum_{j=1}^{i} z_j^2 \quad (i = 2, \ldots, n), \end{cases}$$

where $0 < m \le \lambda_i \le M,\ (i = 1, \ldots, n)$ and (2.31) follows. □

Chapter 3

Inequalities Related to the Kantorovich Inequality: 1953–1967

In this chapter we present several inequalities which are related to the Kantorovich Inequality and several which admit the Kantorovich Inequality as a special case. We concentrate on the following eleven papers published from 1953 through 1967:

- §3.1 Wielandt (1953)
- §3.2 Newman (1959)
- §3.3 Strang (1960)
- §3.4 Bauer (1961)
- §3.5 Marcus and Khan (1961)
- §3.6 Cargo and Shisha (1962)
- §3.7 Diaz and Metcalf (1963)
- §3.8 Marcus and Cayford (1963)
- §3.9 Marshall and Olkin (1964)

- §3.10 Fan (1966)

- §3.11 Shisha and Mond (1967).

3.1 Wielandt (1953)/ Bauer and Householder (1960)

Bauer and Householder (1960) showed that for any two non-null vectors x and y and positive definite matrix A,

$$\frac{(x'y)^2}{x'x \cdot y'y} \leq \cos^2\phi \quad \text{and} \quad 0 \leq \phi \leq \frac{\pi}{2} \tag{3.1}$$

implies that

$$\frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \cos^2\theta,$$

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

When $\phi = \pi/2$ the vectors x and y must be orthogonal, cf. (3.1) and then

$$\frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq \frac{\kappa-1}{\kappa+1} = \frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}, \tag{3.2}$$

since now $\kappa = \cot^2(\theta/2)$.

Bauer and Householder [15] credit (3.2) to Wielandt (1953) [172] "and also private communication" [we have found it difficult to deduce (3.2) from the results in [172]]. Eaton (1976) [48] rediscovered (3.2), cf. Olkin [126].

Housholder ([71], pp. 83) observes that when

$$y = (x'x)A^{-1}x - (x'A^{-1}x)x$$

then the Wielandt Inequality becomes the Kantorovich Inequality.

Barnes and Hoffman [11], see also Wolkowicz [174] refer to the right-hand side of (3.2) as the "Kantorovich Ratio".

To prove the Wielandt Inequality we follow Householder ([71], pp. 81–85, §3.4) and start by introducing the 2×2 matrix

$$G = (u:v)'A(u:v) = \begin{pmatrix} u'Au & u'Av \\ v'Au & v'Av \end{pmatrix},$$

where $u = x/\sqrt{x'x}$ and $v = y/\sqrt{y'y}$, then $(u:v)'(u:v) = I_2$, the 2×2 identity matrix. Let $\gamma_1 \geq \gamma_2$ denote the eigenvalues of G and so the trace $\text{tr}G = \gamma_1 + \gamma_2$ and the determinant $\det G = \gamma_1\gamma_2$. Then by the Poincaré Separation Theorem, cf. e.g., Scott and Styan [202],

$$\lambda_1 \geq \gamma_1 \geq \gamma_2 \geq \lambda_n, \tag{3.3}$$

hence

$$\frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} = \frac{4\kappa}{(\kappa+1)^2}, \tag{3.4}$$

cf. (3.3) and (1.11) in chapter 1. Applying (3.4) yields

$$\begin{aligned} 1 - \frac{(x'Ay)^2}{x'Ax \cdot y'Ay} &= 1 - \frac{(u'Av)^2}{u'Au \cdot v'Av} \\ &= \frac{4\det G}{(\text{tr}G)^2 - (u'Au - v'Av)^2} \\ &= \frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2 - (u'Au - v'Av)^2} \\ &\geq \frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2}, \end{aligned}$$

with equality if and only if $x'Ax = y'Ay$. Hence

$$\frac{(x'Ay)^2}{x'Ax \cdot y'Ay} \leq 1 - \frac{4\kappa}{(\kappa+1)^2} = \left(\frac{\kappa-1}{\kappa+1}\right)^2 = \cos^2\theta, \tag{3.5}$$

and the Wielandt Inequality is established. Equality holds in the Wielandt Inequality whenever $x = p_1 + p_n$ and $y = p_1 - p_n$, where p_1 and p_n are, respectively, normalized eigenvectors corresponding to the eigenvalues λ_1 and λ_n of the matrix A. □

3.2 Newman (1959)

Let A be a symmetric matrix with eigenvalues λ in the closed interval $[m, M]$ with $0 < m < M$, and let $f(\lambda)$ and $g(\lambda)$ be real functions such that

$$0 < f(\lambda), g(\lambda) < \infty \quad (0 < m \leq \lambda \leq M), \tag{3.6}$$

and

$$f(\lambda), g(\lambda) \text{ are convex} \quad (0 < m \leq \lambda \leq M). \tag{3.7}$$

Then (3.6) implies that the matrices $F = f(A)$ and $G = g(A)$ are well defined and are positive definite. Let t be any vector normalized so that $t't = 1$ and set

$$h = t'Ft \cdot t'Gt.$$

Then Morris Newman (1959) [125] showed that

$$2h^{1/2} \leq \max\left(cf(m) + \frac{1}{c}g(m),\ cf(M) + \frac{1}{c}g(M)\right) \tag{3.8}$$

for every $c > 0$. Moreover, if in addition $f(M) - f(m)$ and $g(m) - g(M)$ have the same sign, then

$$2h^{1/2} \leq rf(m) + \frac{1}{r}g(m), \tag{3.9}$$

where

$$r = \left(\frac{g(m) - g(M)}{f(M) - f(m)}\right)^{\frac{1}{2}}.$$

If we choose $f(t) = t$ and $g(t) = t^{-1}$ then $r = 1/\sqrt{mM}$ and (3.9) reduces to the "normalized" Kantorovich Inequality, cf. (1.7) in Chapter 1,

$$t'At \cdot t'A^{-1}t \leq \frac{(m+M)^2}{4mM}, \tag{3.10}$$

with t normalized so that $t't = 1$.

A related inequality due to Ky Fan, included at the end of this paper [125] by Newman (1959) (and also later in the 1966 paper by Fan [49]), yields the following extension of the Kantorovich Inequality:

$$\sum_{i=1}^{k} t_i' A t_i \cdot \sum_{i=1}^{k} t_i' A^{-1} t_i \le \frac{(m+M)^2}{4mM}, \tag{3.11}$$

where the t_i are $n \times 1$ vectors normalized so that $\sum_{i=1}^{k} t_i' t_i = 1$.

When $k = 1$ the inequality (3.11) becomes (3.10). To see that (3.10) also implies (3.11), let the vector t be the $kn \times 1$ vector $(t_1', \dots, t_n')'$ and replace the $n \times n$ matrix A in (3.10) with the $kn \times kn$ Kronecker product $I_k \otimes A$, whose eigenvalues are just the n eigenvalues of the original $n \times n$ matrix A each repeated k times.

3.3 Strang (1960)

In 1960 W. Gilbert Strang [151] used the Cauchy-Schwarz Inequality to establish the following extended version of Kantorovich Inequality:

$$\frac{t' F' u \cdot u' F^{-1} t}{t' t \cdot u' u} \le \frac{(\sigma_1 + \sigma_n)^2}{4\sigma_1 \sigma_n}, \tag{3.12}$$

where t and u are $n \times 1$ vectors and F is an $n \times n$ nonsingular matrix with σ_1 and σ_n, respectively, its largest and smallest necessarily positive singular values.

When $t = u$ and $F = A$ is positive definite with singular values being its eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$ then (3.12) becomes

$$\frac{t' A t \cdot t' A^{-1} t}{(t' t)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}, \tag{3.13}$$

which is the Kantorovich Inequality.

To prove (3.12): Let $A = (F'F)^{1/2}$ be the positive definite square root of $F'F$ and so $P = F(F'F)^{-1/2} = FA^{-1}$ is an orthogonal matrix, i.e., $P'P = I$. Then by the generalized Cauchy-Schwarz Inequality:

$$t' A u \le \sqrt{t' A t \cdot u' A u}, \tag{3.14}$$

cf. e.g., Pečarić, Puntanen and Styan [131], where t and u are $n \times 1$ vectors and A is positive definite. Hence

$$t'F'u = t'AP'u = t'Az \leq \sqrt{t'At \cdot z'Az},$$

where $z = P'u$ and so $z'z = u'u$. Moreover,

$$t'F^{-1}u = t'A^{-1}P'u = t'A^{-1}z \leq \sqrt{t'A^{-1}t \cdot z'A^{-1}z},$$

and so

$$t'F'u \cdot t'F^{-1}u \leq \sqrt{t'At \cdot t'A^{-1}t \cdot z'Az \cdot z'A^{-1}z}.$$

Using the Kantorovich Inequality (1.1) we then obtain

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

since the eigenvalues of A coincide with the singular values of F, and our proof of (3.12) is complete. □

3.4 Bauer (1961)

In 1961 F. L. Bauer [14] showed that the Euclidean least-upper-bound norm

$$\text{lub} \parallel (T'AT)^{-1}T'A^2T(T'AT)^{-1} \parallel \leq \frac{(\kappa + 1)^2}{4\kappa}, \tag{3.15}$$

where the $n \times n$ matrix A is positive definite with condition number $\kappa = \text{cond}(A)$, and the $n \times k$ matrix T has orthonormal columns so that $T'T = I_k$. When $k = 1$ then T becomes a vector t, say, and (3.15) reduces to the normalized Krasnosel'skiĭ-Kreĭn Inequality, cf. (2.19) in Chapter 2, which we have already shown to be equivalent to the normalized Kantorovich Inequality (3.10).

Bauer [14] also established some related results involving partitioned matrices. Let the positive definite matrix A be partitioned as follows

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

with A_{11} of order $k \times k$. Then

$$\text{lub} \parallel A_{11}^{-1}A_{12} \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} - \frac{1}{\sqrt{\kappa}}\right), \tag{3.16}$$

where, as in (3.15), $\kappa = \text{cond}(A)$ is the condition number of the matrix A.

We also find, similarly that: If A_{22} is a nonsingular matrix then

$$\text{lub} \parallel A_{22}^{-1}A_{21} \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} - \frac{1}{\sqrt{\kappa}}\right). \tag{3.17}$$

If A_{12} is a nonsingular matrix then

$$\text{lub} \parallel A_{12}^{-1}A_{11} \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} - \frac{1}{\sqrt{\kappa}}\right). \tag{3.18}$$

If A_{21} is a nonsingular matrix then

$$\text{lub} \parallel A_{21}^{-1}A_{22} \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} - \frac{1}{\sqrt{\kappa}}\right). \tag{3.19}$$

Two results, somewhat more general than (3.15) and (3.16), were also established by Bauer [14]:

$$\text{lub} \parallel (T'AT)^{-1}T'AU \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} - \frac{1}{\sqrt{\kappa}}\right) \tag{3.20}$$

and

$$\text{lub} \parallel T'AT)^{-1}T'A \parallel \leq \frac{1}{2}\left(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\right), \tag{3.21}$$

where the $n \times (n-k)$ matrix U satisfies $TT' + UU' = I_n$. Choosing T to contain the leading k columns of I_n and U the other $n-k$ columns in (3.20) yields (3.16), while "squaring" both sides of (3.21) yields (3.15).

3.5 Marcus and Khan (1961)

In 1961 Marvin Marcus and N. A. Khan [103] observed that

$$A[i] \cdot A^{-1}[i] \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{3.22}$$

where $A[i]$ is the ith diagonal entry of the positive definite matrix A, while $\lambda_1 \geq \cdots \geq \lambda_n$ are the necessarily positive eigenvalues of A. If we put the vector t in the normalized Kantorovich Inequality (3.10) equal to e_i, the vector with all elements zero except for the ith which is 1, then (3.10) becomes (3.22).

Marcus and Khan [103] extended (3.22) by showing that the product of the determinants

$$|A[i_1, \ldots, i_k]| \cdot |A^{-1}[i_1, \ldots, i_k]| \leq \frac{1}{4}\left[\left(\frac{\prod_{j=1}^k \lambda_j}{\prod_{j=1}^k \lambda_{n-j+1}}\right)^{\frac{1}{2}} + \left(\frac{\prod_{j=1}^k \lambda_{n-j+1}}{\prod_{j=1}^k \lambda_j}\right)^{\frac{1}{2}}\right]^2, \tag{3.23}$$

where $A[i_1, \ldots, i_k]$ denotes the $k \times k$ principal submatrix of the positive definite matrix A comprised of the rows with indices $i_1, \ldots, i_k$, with $1 \leq i_1 < \cdots < i_k \leq n$ $(k = 1, \ldots, n)$. When $k = 1$ then (3.23) becomes (3.22).

The inequality (3.23), however, is weaker than the Bloomfield-Watson-Knott Inequality:

$$\frac{|X'AX| \cdot |X'A^{-1}X|}{|X'X|^2} \leq \prod_{i=1}^{\min(k,\, n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{3.24}$$

which we study in some detail in Chapter 4 of this thesis. If we put the $n \times k$ matrix X in (3.24) equal to the matrix whose k columns comprise the i_1th, ..., i_kth columns of I_n then the left-hand side of (3.24) becomes the left-hand side of (3.23). When $k = 1$ the right-hand sides of (3.23) and (3.24) are the same; when $k \geq 2$, however, we have, in general, that

$$\prod_{i=1}^{\min(k,\, n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} \leq \frac{1}{4}\left[\left(\frac{\prod_{j=1}^k \lambda_j}{\prod_{j=1}^k \lambda_{n-j+1}}\right)^{\frac{1}{2}} + \left(\frac{\prod_{j=1}^k \lambda_{n-j+1}}{\prod_{j=1}^k \lambda_j}\right)^{\frac{1}{2}}\right]^2. \tag{3.25}$$

We illustrate (3.25) for $k = 2$ (and $n \geq 4$), which may then be written as

$$\frac{(\kappa_1 + 1)^2}{4\kappa_1} \cdot \frac{(\kappa_2 + 1)^2}{4\kappa_2} \leq \frac{1}{4}\left[\sqrt{\kappa_1\kappa_2} + \frac{1}{\sqrt{\kappa_1\kappa_2}}\right]^2,$$

where

$$\kappa_1 = \lambda_1/\lambda_n \geq 1 \quad \text{and} \quad \kappa_2 = \lambda_2/\lambda_{n-1} \geq 1. \tag{3.26}$$

Then we see that (3.25) reduces to

$$(\kappa_1 + 1)^2(\kappa_2 + 1)^2 \leq 4(\kappa_1\kappa_2 + 1)^2; \tag{3.27}$$

taking square roots reduces (3.27) to

$$(\kappa_1 - 1)(\kappa_2 - 1) \geq 0, \tag{3.28}$$

which, in view of (3.26), is certainly true.

Furthermore, Marcus and Khan [103] generalized Fan's inequality (3.11), cf. Newman [125] and Fan [49], to:

$$\prod_{i=1}^{m}\left\{\sum_{j=1}^{k} t_j' A_i t_j\right\}^{\frac{1}{m}} \leq \frac{1}{m} \max_{1\leq r\leq n} \sum_{i=1}^{m} \lambda_r^{(i)}, \tag{3.29}$$

where the A_i $(i = 1, \ldots, m)$ are pairwise commutative $n\times n$ positive definite matrices with eigenvalues $\lambda_1^{(i)}, \ldots, \lambda_n^{(i)}$, respectively, for $i = 1, \ldots, m$, and $\sum_{j=1}^{k} t_j' t_j = 1$.

If we choose $m = 2$, $k = 1$, $A_1 = \sqrt{\lambda_1\lambda_n}A$ and $A_2 = (1/\sqrt{\lambda_1\lambda_n})A^{-1}$ in (3.29) then it becomes the normalized Kantorovich Inequality (3.10).

3.6 Cargo and Shisha (1962)

In 1962 G. T. Cargo and Oved Shisha [26], see also Rennie (1963) [25], Goldman (1964) [57], Marshall and Olkin (1964) [105] (see also §3.9 below), Mond (1966) [113], and Cargo (1972) [25], studied power means, which may be expressed in the form:

$$\mu_h = \left(t' A^h t\right)^{1/h}, \tag{3.30}$$

where t is an $n \times 1$ vector and A is an $n \times n$ positive definite matrix with eigenvalues λ_i such that $m \leq \lambda_i \leq M$, with $m < M$. Let $\kappa = M/m$. Then Cargo and Shisha [26] showed that for $r < s$ (with neither r nor s necessarily positive)

$$\frac{\mu_s}{\mu_r} \leq k, \tag{3.31}$$

where

$$k = \left(\frac{r(\kappa^s - \kappa^r)}{(s-r)(\kappa^r - 1)}\right)^{1/s} \Bigg/ \left(\frac{s(\kappa^r - \kappa^s)}{(r-s)(\kappa^s - 1)}\right)^{1/r} \quad \text{if} \quad rs \neq 0. \tag{3.32}$$

Moreover, when $rs = 0$ then

$$k = \begin{cases} \left(\frac{\kappa^{s/(\kappa^s - 1)}}{e \log[\kappa^{s/(\kappa^s - 1)}]}\right)^{1/s} & \text{if } r = 0 \\ \\ \left(\frac{\kappa^{r/(\kappa^r - 1)}}{e \log[\kappa^{r/(\kappa^r - 1)}]}\right)^{-1/r} & \text{if } s = 0. \end{cases}$$

Cargo and Shisha [26] also obtained a condition, albeit somewhat complicated, for equality in (3.31).

If we choose $r = -1$ and $s = 1$ in (3.31) then it reduces to the Kantorovich Inequality.

3.7 Diaz and Metcalf (1963)

In 1963 J. B. Diaz and F. T. Metcalf [39] showed that

$$\sum_{i=1}^{n} a_i^2 + mM \sum_{i=1}^{n} b_i^2 \leq (m+M) \sum_{i=1}^{n} a_i b_i, \tag{3.33}$$

where the numbers a_i and $b_i \neq 0$ satisfy

$$m \leq \frac{a_i}{b_i} \leq M \quad (i = 1, \ldots, n). \tag{3.34}$$

Equality holds in (3.33) if and only if in each of the n inequalities (3.34), at least one of the equality signs holds, i.e., either $a_i = mb_i$ or $a_i = Mb_i$ (where the equation may vary with i).

The inequality (3.33) follows easily by summing the inequality:

$$\left(\frac{a_i}{b_i} - m\right)\left(M - \frac{a_i}{b_i}\right) a_i^2 \geq 0 \tag{3.35}$$

over $i = 1, \ldots, n$.

Together with

$$\left[\left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}} - \left(mM\sum_{i=1}^{n} b_i^2\right)^{\frac{1}{2}}\right]^2 \geq 0, \tag{3.36}$$

the inequality (3.33) yields the unweighted Cassels Inequality (2.14) in Chapter 2:

$$\frac{\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2}{(\sum_{i=1}^{n} a_i b_i)^2} \leq \frac{(m+M)^2}{4mM}. \tag{3.37}$$

□

3.8 Marcus and Cayford (1963)

In 1963 Marvin Marcus and Afton Cayford [102] established the following generalization of the Kantorovich Inequality:

$$\frac{t'At \cdot t'A^{-p}t}{(t't)^2} \leq \frac{\lambda_n^{1-p}(\kappa^{p+1}-1)^2}{4\kappa^p(\kappa^p-1)(\kappa-1)}, \tag{3.38}$$

where $0 < p \leq 1$, and the positive definite matrix A has condition number $\kappa > 1$ and smallest eigenvalue $\lambda_n > 0$.

When $p = 1$ we see that (3.38) reduces to the Kantorovich Inequality:

$$\frac{t'At \cdot t'A^{-1}t}{(t't)^2} \leq \frac{(\kappa+1)^2}{4\kappa},$$

cf. (1.1) and (1.9) in Chapter 1.

Marcus and Cayford [102] also showed that (3.38) holds for $p > 1$ provided $\kappa \geq \kappa_p$, where κ_p is the unique root greater than 1 of $\kappa^{p+1} - 2\kappa^p + 1 = 0$. When $p > 1$ and $\kappa < \kappa_p$ then

$$\frac{t'At \cdot t'A^{-p}t}{(t't)^2} \leq \lambda_n^{1-p}.$$

3.9 Marshall and Olkin (1964)

In 1964 Albert W. Marshall and Ingram Olkin [105] gave several "Reversals of the Lyapunov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich Inequality". Let $t_1, \dots, t_n$ and $z_1^2, \dots, z_n^2$ be nonnegative real numbers with $\sum z_i^2 = 1$, and define $\mu_h = \sum_{i=1}^n z_i^2 t_i^h$; then

$$\mu_v^{w-u} \leq \mu_u^{w-v} \mu_w^{v-u} \quad (0 \leq u \leq v \leq w) \tag{3.39}$$

is the well-known "Lyapunov Inequality". Since the z_i^2 are nonnegative they may be considered as probabilities and we may write $\Pr(T = t_i) = z_i^2$ $(i = 1, \dots, n)$ and so

$$\mu_h = \sum_{i=1}^n z_i^2 t_i^h = \mathbf{E}(T^h)$$

is the hth moment of the random variable T.

In general there is no positive constant k, say, so that

$$\mu_v^{w-u} \geq k \mu_u^{w-v} \mu_w^{v-u} \quad (u \leq v \leq w)$$

but such a constant k does exist if we restrict the random variable T to be bounded and positive:

$$\Pr(m \leq T \leq M) = 1 \quad (0 < m < M). \tag{3.40}$$

Then for $r < s$ and $\Pr(Z > 0) = 1$ we have:

$$[\mathbf{E}(ZT^s)]^{1/s} / [\mathbf{E}(ZT^h)]^{1/r} \leq k[\mathbf{E}(Z)]^{\frac{1}{s}-\frac{1}{r}}, \tag{3.41}$$

where

$$k = \left(\frac{r(\kappa^s - \kappa^r)}{(s-h)(\kappa^h - 1)} \right)^{1/s} \Bigg/ \left(\frac{s(\kappa^s - \kappa^r)}{(s-r)(\kappa^s - 1)} \right)^{1/r}$$

and $\kappa = M/m$, cf. (3.32). Equality holds in (3.41) if and only if $\Pr(T = m$ or $T = M) = 1$ and

$$\mathbf{E}(ZT^s) = \frac{r(M^s m^r - M^r m^s)}{(M^r - m^r)(h - r)}\mathbf{E}(Z). \tag{3.42}$$

If we put $r = -1, s = 1$ and $Z = 1$ (with probability 1) in (3.42) then we obtain:

$$\sum_{i=1}^{n} \lambda_i z_i^2 \cdot \sum_{i=1}^{n} \frac{1}{\lambda_i} z_i^2 = \mathbf{E}(T) \cdot \mathbf{E}\left(\frac{1}{T}\right) \leq \frac{(\kappa+1)^2}{4\kappa} = \frac{(m+M)^2}{4mM}. \tag{3.43}$$

cf. (1.9) and (1.25) in Chapter 1.

Marshall and Olkin [105] also obtained several inequalities involving vectors and matrices. Let t be an $n \times 1$ vector and let A be an $n \times n$ positive definite matrix with, necessarily positive, eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then

$$(\lambda_1^{w-u}\lambda_n^{v-u} - \lambda_1^{v-u}\lambda_n^{w-u})t'A^u t \leq (\lambda_1^{w-u} - \lambda_n^{w-u})t'A^v t - (\lambda_1^{v-u} - \lambda_n^{v-u})t'A^w t, \tag{3.44}$$

and

$$k(t'A^u t)^{w-v}(t'A^w t)^{v-u} \leq (t'A^v t)^{w-u}, \tag{3.45}$$

where

$$k = \left(\frac{(\kappa^{w-u} - \kappa^{v-u})(w-u)}{(\kappa^{w-u} - 1)(w-v)}\right)^{w-u} \Bigg/ \left(\frac{(\kappa^{w-u} - \kappa^{v-u})(v-u)}{(\kappa^{v-u} - 1)(w-v)}\right)^{v-u}$$

with now $\kappa = \lambda_1/\lambda_n$, the condition number cond(A).

Let $u = -1$, $v = 0$ and $w = 1$. Then (3.45) becomes the Kantorovich Inequality:

$$\frac{t'At \cdot t'A^{-1}t}{(t't)^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n},$$

while (3.44) becomes the (unnormalized unreduced) Marshall-Olkin Inequality:

$$\lambda_1\lambda_n \cdot t'A^{-1}t \leq (\lambda_1 + \lambda_n)t't - t'At, \tag{3.46}$$

cf. (1.29) in Chapter 1.

3.10 Fan (1966)

In 1966 Ky Fan [49] showed that:

$$\frac{\sum_{i=1}^{k} t_i' A^p t_i}{\left[\sum_{i=1}^{k} t_i' A t_i\right]^p} \leq \frac{(M^p - m^p)^p (p-1)^{p-1}}{(M-m)(mM^p - Mm^p)^{p-1} p^p}, \tag{3.47}$$

where p is any integer not equal to 0 or 1 (and not necessarily positive), the t_i are normalized so that $\sum_{i=1}^{k} t_i' t_i = 1$, and A is an $n \times n$ positive definite matrix with its eigenvalues contained in the closed interval $[m, M]$, $0 < m < M$.

If we put $p = -1$ in (3.47) then we obtain

$$\sum_{i=1}^{k} t_i' A t_i \cdot \sum_{i=1}^{k} t_i' A^{-1} t_i \leq \frac{(m+M)^2}{4mM},$$

which was established by Fan in 1959 [125], cf. (3.11).

3.11 Shisha and Mond (1967)

So far in this thesis we have concentrated on inequalities concerning the product of two quadratic forms such as $t'At \cdot t'A^{-1}t$, where t is an $n \times 1$ vector and A is an $n \times n$ positive definite matrix with fixed, necessarily positive, eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. In 1967 Oved Shisha and Bertram Mond [148] (and independently and much later Styan (1983) [152] and Khatri (1984) [75]) obtained related inequalities which apply to differences of quadratic forms rather than to products. For example, Shisha and Mond [148] showed that, assuming t normalized with $t't = 1$,

$$t'At - \frac{1}{t'A^{-1}t} \leq \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2 \tag{3.48}$$

and

$$(t'A^2t)^{1/2} - t'At \leq \frac{(\lambda_1 - \lambda_n)^2}{4(\lambda_1 + \lambda_n)}. \tag{3.49}$$

Equality holds in (3.48) when

$$t = \left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1} + \sqrt{\lambda_n}}\right)^{\frac{1}{2}} p_1 \pm \left(\frac{\sqrt{\lambda_n}}{\sqrt{\lambda_1} + \sqrt{\lambda_n}}\right)^{\frac{1}{2}} p_n, \tag{3.50}$$

and in (3.49) when

$$t = \tfrac{1}{2}\left\{\left(\frac{\lambda_1 + 3\lambda_n}{\lambda_1 + \lambda_n}\right)^{\frac{1}{2}} p_1 \pm \left(\frac{3\lambda_1 + \lambda_n}{\lambda_1 + \lambda_n}\right)^{\frac{1}{2}} p_n\right\}. \tag{3.51}$$

cf. Styan [152], where p_1 and p_n are orthonormal eigenvectors of A corresponding, respectively, to λ_1 and λ_n. When the eigenvalues λ_1 and λ_n are both simple (i.e., each has multiplicity 1) then each of these conditions is also necessary.

To prove (3.48) we begin by diagonalizing $A = P\Lambda P'$ and writing $z = P't$ so that (3.48) may be written as

$$z'\Lambda z - \frac{1}{z'\Lambda^{-1}z} \le \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2. \tag{3.52}$$

We now prove (3.52), following [152], using the (normalized reduced) Marshall-Olkin Inequality:

$$\lambda_1\lambda_n \cdot z'\Lambda^{-1}z \le \lambda_1 + \lambda_n - z'\Lambda z,$$

cf. (3.46), so that

$$z'\Lambda z - \frac{1}{z'\Lambda^{-1}z} \le z'\Lambda z - \frac{\lambda_1\lambda_n}{\lambda_1 + \lambda_n - z'\Lambda z} = \lambda_1 + \lambda_n - \frac{\mu^2 + \lambda_1\lambda_n}{\mu},$$

where

$$\mu = \lambda_1 + \lambda_n - z'\Lambda z = z'(\mathrm{diag}\{\lambda_1 - \lambda_i + \lambda_n\})z > 0.$$

Hence

$$\begin{aligned} z'\Lambda z - \frac{1}{z'\Lambda^{-1}z} &\le \lambda_1 + \lambda_n - \frac{\mu^2 + \lambda_1\lambda_n}{\mu} \\ &= \lambda_1 + \lambda_n - 2\sqrt{\lambda_1\lambda_n} - \frac{(\mu - \sqrt{\lambda_1\lambda_n})^2}{\mu} \\ &\le \lambda_1 + \lambda_n - 2\sqrt{\lambda_1\lambda_n} \\ &= \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2, \end{aligned}$$

and the proof is complete. □

Another related inequality, announced by Styan and Zlobec (1982) [154], is:

$$t'A^2t - (t'At)^2 \le \left(\frac{\lambda_1 - \lambda_n}{2}\right)^2, \tag{3.53}$$

again with $t't = 1$; equality holds in (3.53) if and only if equality holds in the Kantorovich Inequality. We note that (3.53) is the "difference" analogue of the "product" Krasnosel'skiĭ-Kreĭn Inequality, cf. (2.19) in Chapter 2:

$$\frac{t'A^2t}{(t'At)^2} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{3.54}$$

normalized with $t't = 1$; equality holds in (3.54) if and only if equality holds in (3.53).

We prove (3.53), following Styan (1983) [152], by diagonalizing as before so that (3.53) becomes

$$z'\Lambda^2 z - (z'\Lambda z)^2 \le \left(\frac{\lambda_1 - \lambda_n}{2}\right)^2. \tag{3.55}$$

The special case of (3.55) with $z = e/\sqrt{n}$, where e is the $n \times 1$ vector with each element equal to 1, provides the following inequality for the variance of the λ_i

$$\frac{1}{n}\sum \lambda_i^2 - \left(\frac{1}{n}\sum \lambda_i\right)^2 \le \left(\frac{\lambda_1 - \lambda_n}{2}\right)^2,$$

established by Brauer and Mewborn (1959) [194].

We now prove (3.55), following Styan [152], as follows:

$$\begin{aligned} z'\Lambda^2 z - (z'\Lambda z)^2 &= z'\Lambda^2 z - \left(z'\Lambda z - \frac{\lambda_1 + \lambda_n}{2}\right)^2 + \frac{(\lambda_1 + \lambda_n)^2}{4} - (\lambda_1 + \lambda_n) z'\Lambda z \\ &\le \tfrac{1}{4}(\lambda_1 + \lambda_n)^2 + z'\Lambda^2 z - (\lambda_1 + \lambda_n) z'\Lambda z \qquad (3.56) \\ &= \tfrac{1}{4}(\lambda_1 - \lambda_n)^2 + \lambda_1\lambda_n + z'\Lambda^2 z - (\lambda_1 + \lambda_n) z'\Lambda z \\ &= \tfrac{1}{4}(\lambda_1 - \lambda_n)^2 - z'(\mathrm{diag}\{(\lambda_1 - \lambda_i)(\lambda_i - \lambda_n)\}) z \\ &\le \tfrac{1}{4}(\lambda_1 - \lambda_n)^2, \qquad (3.57) \end{aligned}$$

and the proof is complete. □

Equality holds in (3.53) if and only if equality holds in both (3.56) and (3.57), and when λ_1 and λ_n are both simple then this happens if and only if

$$z_1^2 = z_n^2 = \tfrac{1}{2} \quad \text{and} \quad z_2 = \cdots = z_{n-1} = 0,$$

cf. (1.37) in Chapter 1.

Chapter 4

The Bloomfield-Watson-Knott Inequality and Other Extensions of the Kantorovich Inequality

Let us consider the so-called Gauss-Markov linear statistical model

$$y = X\beta + \varepsilon, \quad \mathbf{E}(\varepsilon) = 0, \quad \mathrm{Cov}(\varepsilon) = \Sigma, \tag{4.1}$$

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

The Best Linear Unbiased Estimator (BLUE) of β is the generalized least squares estimator

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y, \tag{4.2}$$

which has covariance matrix

$$\mathrm{Cov}(\tilde{\beta}) = (X'\Sigma^{-1}X)^{-1}, \tag{4.3}$$

while the Ordinary Least Squares Estimator (OLSE)

$$\hat{\beta} = (X'X)^{-1}X'y \tag{4.4}$$

has covariance matrix

$$\operatorname{Cov}(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}, \tag{4.5}$$

and so the matrix

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1} - (X'\Sigma^{-1}X)^{-1} \tag{4.6}$$

is non-negative definite.

It is well known that $\hat{\beta} = \tilde{\beta}$ with probability one if and only if their covariance matrices are equal, i.e.,

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1} = (X'\Sigma^{-1}X)^{-1}. \tag{4.7}$$

For an extensive discussion of characterizations for the equality (4.7) see Puntanen and Styan [137] and Baksalary, Puntanen and Styan [10].

In general, however, equality does not hold in (4.7) and thus the matrix (4.6) is not zero. To measure how far away the matrix (4.6) is from the null matrix or equivalently to see how bad the OLSE is with respect to the BLUE, Geoffrey S. Watson in 1951 ([163], §3.3; see also [165], p. 330) introduced the "efficiency" of the OLSE $\hat{\beta}$ with respect to the BLUE $\tilde{\beta}$ as the ratio of their generalized variances defined as the determinants of the corresponding covariance matrices:

$$\operatorname{eff}(\hat{\beta}) = \frac{|\operatorname{Cov}(\tilde{\beta})|}{|\operatorname{Cov}(\hat{\beta})|} = \frac{|(X'\Sigma^{-1}X)^{-1}|}{|(X'X)^{-1}X'\Sigma X(X'X)^{-1}|} = \frac{|X'X|^2}{|X'\Sigma X| \cdot |X'\Sigma^{-1}X|}, \tag{4.8}$$

where $|.|$ denotes determinant.

The inequality

$$\frac{|X'X|^2}{|X'\Sigma X| \cdot |X'\Sigma^{-1}X|} \geq \prod_{i=1}^{k} \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}, \tag{4.9}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the, necessarily positive, eigenvalues of Σ, was originally conjectured in 1955 by James Durbin (cf. [165], p. 331) but first established for $k > 1$ only twenty years later in 1975 by Bloomfield and Watson [21] and Knott [79], and in 1981 by Khatri and Rao [76]; see also Yang [176]. When $k = 1$, however, the "Bloomfield-Watson-Knott Inequality" (4.9) reduces to the Kantorovich Inequality.

In 1981 Bartmann and Bloomfield [12] (see also Puntanen [136]) observed that

$$\mathrm{eff}(\hat{\beta}) = \prod_{i=1}^{p} \theta_i^2 = \prod_{i=1}^{p} (1 - \rho_i^2), \tag{4.10}$$

where p is the rank of X, while θ_i and ρ_i are, respectively, the ith largest canonical correlations between $\hat{\beta}$ and $\tilde{\beta}$ and between the ordinary least squares fitted values and residuals. The formula (4.9), therefore, remains valid for the efficiency of ordinary least squares when the design matrix X does not necessarily have full column rank.

In Section 4.1 below we present a proof of the Bloomfield-Watson-Knott Inequality (4.9) based closely on the proof given by Bloomfield and Watson [21] but with a modification due to Drury [47] and which avoids the Lagrange multipliers, while in Section 4.2 we indicate why we feel that the "brief proof" given in 1990 by Hu Yang [177] is incomplete. In Section 4.3 we present various extensions of the Bloomfield-Watson-Knott Inequality due to Khatri and Rao (1981) [76] and Wang and Shao (1992) [161], see also the paper by Lin (1986) [85] translated into English as Appendix B of this thesis.

4.1 Proof of the Bloomfield-Watson-Knott Inequality

Theorem 4.1 *Let Σ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n > 0$ and let X be an $n \times k$ matrix of rank k with $n > k$. Then*

$$f = \frac{|X'\Sigma X| \cdot |X'\Sigma^{-1}X|}{|X'X|^2} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}. \tag{4.11}$$

When $n \geq 2k$ then equality is attained in (4.11) *when*

$$X = \frac{1}{\sqrt{2}} \left(p_1 \pm p_n : p_2 \pm p_{n-1} : \cdots : p_k \pm p_{n-k+1} \right), \tag{4.12}$$

where $p_1, \ldots, p_n$ are normalized eigenvectors corresponding to $\lambda_1 \geq \cdots \geq \lambda_n$; when the eigenvalues are distinct then this condition is also necessary.

To prove the Bloomfield-Watson-Knott Inequality (4.11) we will use the following two lemmas:

Lemma 4.1 *Let P be an $n \times n$ skew-symmetric matrix. Then* $\exp(zP)$ *is an $n \times n$ orthogonal matrix.*

PROOF: When P is skew-symmetric, $P = -P'$ and so

$$\exp(zP) \cdot \exp(zP') = \exp(zP) \cdot \exp(-zP) = \exp(zP) \cdot [\exp(zP)]^{-1} = I,$$

cf. e.g., Horn and Johnson ([199], Theorem 6.2.38, p. 435), and thus $\exp(zP)$ is orthogonal. □

Lemma 4.2 *If* $\mathrm{tr}PA = 0$ *for every skew-symmetric matrix P then A is symmetric.*

PROOF: We may choose $P = A' - A$ and so

$$\mathrm{tr}PA = \mathrm{tr}(A' - A)A = \tfrac{1}{2}\mathrm{tr}(A - A')(A - A')' \geq 0 \tag{4.13}$$

and then $\mathrm{tr}PA = 0$ for all skew-symmetric matrix P implies $A = A'$. □

PROOF OF THEOREM 4.1: We assume, without loss of generality, that $n \geq 2k$ and that $\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 > \cdots > \lambda_n > 0$. We also assume that $X'X = I_k$ and so we may replace X in (4.11) by $X(z) = \exp(zP)X$, where P is skew-symmetric and thus $\exp(zP)$ is orthogonal (Lemma 4.1). Then

$$\begin{aligned} X(z)'\Sigma X(z) &= X' \exp(zP')\Sigma \exp(zP)X \\ &= X'(I - zP)\Sigma(I + zP)X + \mathrm{HOT} \\ &= X'(\Sigma + z[\Sigma, P])X + \mathrm{HOT} \\ &= (X'\Sigma X)(I + z(X'\Sigma X)^{-1}X'[\Sigma, P]X + \mathrm{HOT}), \end{aligned}$$

where HOT denotes terms of order z^2 and higher and the "commutator matrix" $[\Sigma, P] = \Sigma P - P\Sigma$. Since for small z the determinant

$$|I + zT + \mathrm{HOT}| = |\exp(zT + \mathrm{HOT})| = \exp[\mathrm{tr}(zT + \mathrm{HOT})] = 1 + z\mathrm{tr}T + \mathrm{HOT},$$

we see that

$$|X(z)'\Sigma X(z)| = |X'\Sigma X| \cdot \{1 + z\mathrm{tr}[(X'\Sigma X)^{-1}X'[\Sigma, P]X] + \mathrm{HOT}\} \tag{4.14}$$

and similarly

$$|X(z)'\Sigma^{-1}X(z)| = |X'\Sigma^{-1}X| \cdot \{1 + z\mathrm{tr}[(X'\Sigma^{-1}X)^{-1}X'[\Sigma^{-1}, P]X] + \mathrm{HOT}\}$$

so that

$$|X(z)'\Sigma X(z)| \cdot |X(z)'\Sigma^{-1}X(z)| = |X'\Sigma X| \cdot |X'\Sigma^{-1}X| \cdot \{1 + zh + \mathrm{HOT}\}.$$

For X to identify a critical point, it follows that

$$h = \mathrm{tr}\{(X'\Sigma X)^{-1}X'[\Sigma, P]X + (X'\Sigma^{-1}X)^{-1}X'[\Sigma^{-1}, P]X\} = 0 \tag{4.15}$$

for every skew-symmetric matrix P. Hence, from Lemma 4.2,

$$\begin{aligned} A &= X(X'\Sigma X)^{-1}X'\Sigma - \Sigma X(X'\Sigma X)^{-1}X' + X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} \\ &\quad -\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X' \end{aligned} \tag{4.16}$$

is symmetric; but A is also skew-symmetric, and so $A = 0$. Post-multiplying (4.16) by X yields

$$X - \Sigma X(X'\Sigma X)^{-1} + X - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} = 0$$

or

$$\Sigma X(X'\Sigma X)^{-1} + \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} = 2X. \tag{4.17}$$

Pre-multiplying (4.17) by $X'\Sigma$ we obtain

$$(X'\Sigma^2X)(X'\Sigma X)^{-1} + (X'\Sigma^{-1}X)^{-1} = 2X'\Sigma X. \tag{4.18}$$

Since $2X'\Sigma X$ and $(X'\Sigma^{-1}X)^{-1}$ are symmetric so is $(X'\Sigma^2X)(X'\Sigma X)^{-1}$, and hence $X'\Sigma^2X$ and $(X'\Sigma X)^{-1}$ commute and so are simultaneously diagonalizable; we may, therefore, suppose that both $X'\Sigma^2 X = D_1$ and $(X'\Sigma X)^{-1} = D_2$ are diagonal. Hence from (4.18)

$$(X'\Sigma^{-1}X)^{-1} = 2D_2^{-1} - D_1 D_2$$

is diagonal. Setting $X'\Sigma X = \text{diag}\{a_1, \ldots, a_k\}$ and $X'\Sigma^{-1}X = \text{diag}\{b_1, \ldots, b_k\}$ in (4.17), it becomes

$$\lambda_i x_{ij} a_j^{-1} + \lambda_i^{-1} x_{ij} b_j^{-1} = 2x_{ij} \quad (i = 1, \ldots, n, \;\; j = 1, \ldots, k)$$

or

$$x_{ij}(\lambda_i^2 a_j^{-1} + b_j^{-1} - 2\lambda_i) = 0 \quad (i = 1, \ldots, n, \;\; j = 1, \ldots, k). \tag{4.19}$$

Since the factor $\lambda_i^2 a_j^{-1} + b_j^{-1} - 2\lambda_i$ in (4.19) is a polynomial of degree two in λ_i it follows that for each $j = 1, \ldots, k$ at most two x_{ij} can be non-zero (cf. Kantorovich's proof in §1.2.1). The matrix X that maximizes

$$f = \frac{|X'\Sigma X| \cdot |X'\Sigma^{-1}X|}{|X'X|^2} \tag{4.20}$$

in (4.11) must, therefore, have the structure

$$X = \begin{pmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_m \end{pmatrix},$$

where X_h $(h = 1, \ldots, m)$ is $n_h \times k_h$ with $\sum n_h = n$ and $\sum k_h = k$ and $1 \le k_h \le n_h \le 2$. Hence each X_h must be of one of the following three types:

1. 1×1 orthogonal matrix (scalar) with value ± 1
2. 2×1 normalized vector, $X_h = \begin{pmatrix} \cos\theta_h \\ \sin\theta_h \end{pmatrix}$, say
3. 2×2 orthogonal matrix.

The contribution f_h, say, to f from each X_h of types 1 and 3 is, therefore, just 1. We concentrate, therefore, only on the X_h of type 2; let $\lambda_{h_1} > \lambda_{h_2}$ be the corresponding elements in Σ. Then the X_h of type 2 yields the contribution

$$\begin{aligned} f_h &= (\lambda_{h_1}\cos^2\theta_h + \lambda_{h_2}\sin^2\theta_h)(\lambda_{h_1}^{-1}\cos^2\theta_h + \lambda_{h_2}^{-1}\sin^2\theta_h) \\ &= \cos^4\theta_h + \sin^4\theta_h + \left(\frac{\lambda_{h_1}}{\lambda_{h_2}} + \frac{\lambda_{h_2}}{\lambda_{h_1}}\right)\cos^2\theta_h\sin^2\theta_h \\ &\leq \cos^4\theta_h + (1-\cos^2\theta_h)^2 + \left(\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}\right)\cos^2\theta_h(1-\cos^2\theta_h) \\ &\leq \frac{1}{4}\left(2 + \frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}\right) = \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \end{aligned} \tag{4.21}$$

since

$$\left(2 + \frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}\right) - 4\left[\cos^4\theta_h + (1-\cos^2\theta_h)^2 + \left(\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}\right)\cos^2\theta_h(1-\cos^2\theta_h)\right]$$

$$= (1-2\cos^2\theta_h)^2\left(\frac{(\lambda_1+\lambda_n)^2}{\lambda_1\lambda_n} - 4\right) \geq 0,$$

with equality if and only if $\cos^2\theta_h = \frac{1}{2}$ (recall that $\lambda_1 > \lambda_n$). Equality holds throughout (4.21), therefore, if and only if

$$\lambda_{h_1} = \lambda_1, \quad \lambda_{h_2} = \lambda_n \quad \text{and} \quad \cos^2\theta_h = \tfrac{1}{2}. \tag{4.22}$$

In view of this, let us define a permutation matrix Π such that (with n even)

$$\Pi\Sigma\Pi' = \mathrm{diag}\{\lambda_1, \lambda_n, \lambda_2, \lambda_{n-1}, \ldots, \lambda_k, \lambda_{n-k+1}\},$$

with $n \geq 2k$ and $\lambda_1 > \cdots > \lambda_n$; then the matrix X which maximizes f in (4.9) must satisfy

$$\Pi X = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & & & 0 \\ 1 & & & \\ & 1 & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & & & 1 \end{pmatrix},$$

and the proof is complete. □

4.2 Hu Yang's "Brief Proof"

In 1990 Hu Yang [176] presented a "brief proof" of the Bloomfield-Watson-Knott Inequality using the arithmetic-geometric mean inequality for matrices and the Poincaré Separation Theorem for eigenvalues. Unfortunately there appears to be a lacuna in his proof.

The arithmetic-geometric mean inequality for matrices, cf. e.g., Ando ([189],[190]), Bhatia and Davis [192] and Bhatia and Kittaneh [193], may be written as

$$A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} \leq_{\mathrm{L}} \tfrac{1}{2}(A+B), \tag{4.23}$$

where A and B are $n \times n$ positive definite matrices and the Löwner partial ordering $F \leq_{\mathrm{L}} G$ means that the matrix $G - F$ is symmetric non-negative definite. It follows at once from (4.23) that

$$|AB|^{1/2} \leq \left|\tfrac{1}{2}(A+B)\right|, \tag{4.24}$$

where $|.|$ denotes determinant. When A and B commute, their geometric mean becomes

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} = (AB)^{1/2}.$$

The Poincaré Separation Theorem, which we used to prove the Wielandt Inequality, cf. (3.3) in Chapter 3, is

$$\mathrm{ch}_{n-k+i}(A) \le \mathrm{ch}_i(P'AP) \le \mathrm{ch}_i(A) \quad (i = 1, \ldots, k), \tag{4.25}$$

where $\mathrm{ch}_i(A)$ denotes the ith largest eigenvalue, A is an $n \times n$ symmetric matrix and P is an $n \times k$ matrix with $P'P = I_k$.

Let Σ be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n > 0$ and let X be an $n \times k$ matrix of rank k with $n \ge k$. Then

$$f = \frac{|X'\Sigma X| \cdot |X'\Sigma^{-1}X|}{|X'X|^2} \le \prod_{i=1}^{k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} \tag{4.26}$$

is the Bloomfield-Watson-Knott Inequality, cf. (4.9). As in §4.1 we will assume, without loss of generality, that $n \ge 2k, X'X = I_k$ and that $\Sigma = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, with $\lambda_1 \ge \cdots \ge \lambda_n > 0$.

Yang [176] uses the $n \times n$ matrix

$$K(\lambda) = \begin{pmatrix} D(\lambda) & 0 & 0 \\ 0 & (\lambda_k + \lambda_{n-k+1})^{-1} I_{n-2(k-1)} & 0 \\ 0 & 0 & JD(\lambda)J \end{pmatrix}, \tag{4.27}$$

where $(k-1) \times (k-1)$ diagonal matrix

$$D(\lambda) = \begin{pmatrix} (\lambda_1 + \lambda_n)^{-1} & 0 & & 0 \\ 0 & (\lambda_2 + \lambda_{n-1})^{-1} & & 0 \\ & & \ddots & \\ 0 & 0 & & (\lambda_{k-1} + \lambda_{n-k+2})^{-1} \end{pmatrix} \tag{4.28}$$

and the "flip matrix" (or "negative diagonal matrix")

$$J = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ & & \ddots & & \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{4.29}$$

so that the diagonal matrix $JD(\lambda)J$ is $D(\lambda)$ with the diagonal elements arranged in precisely the opposite order. Yang [176] also uses the $n \times n$ diagonal matrix

$$E(\lambda) = \Sigma K(\lambda) = \begin{pmatrix} E_1(\lambda) & & \\ & E_2(\lambda) & \\ & & E_3(\lambda) \end{pmatrix}, \tag{4.30}$$

where the diagonal matrices

$$E_1(\lambda) = \begin{pmatrix} \frac{\lambda_1}{\lambda_1+\lambda_n} & & & \\ & \frac{\lambda_2}{\lambda_2+\lambda_{n-1}} & & \\ & & \ddots & \\ & & & \frac{\lambda_{k-1}}{\lambda_{k-1}+\lambda_{n-k+2}} \end{pmatrix} \tag{4.31}$$

and

$$E_3(\lambda) = \begin{pmatrix} \frac{\lambda_{n-k+2}}{\lambda_{k-1}+\lambda_{n-k+2}} & & & \\ & \frac{\lambda_{n-k+3}}{\lambda_{k-2}+\lambda_{n-k+3}} & & \\ & & \ddots & \\ & & & \frac{\lambda_n}{\lambda_1+\lambda_n} \end{pmatrix} \tag{4.32}$$

are both $(k-1) \times (k-1)$, while the $(n-2k+2) \times (n-2k+2)$ diagonal matrix

$$E_2(\lambda) = \begin{pmatrix} \frac{\lambda_k}{\lambda_k+\lambda_{n-k+1}} & & & \\ & \frac{\lambda_{k+1}}{\lambda_k+\lambda_{n-k+1}} & & \\ & & \ddots & \\ & & & \frac{\lambda_{n-k+1}}{\lambda_k+\lambda_{n-k+1}} \end{pmatrix}. \tag{4.33}$$

We observe that the numerators of the diagonal elements in $E_1(\lambda)$, $E_2(\lambda)$ and $E_3(\lambda)$ and, therefore, the numerators of the diagonal elements in $E(\lambda)$, are the eigenvalues of Σ in monotonically decreasing order. The denominators of the diagonal elements in $E_1(\lambda)$ and $E_3(\lambda)$ are in precisely opposite orders to each other, while in $E_2(\lambda)$ the denominators of the diagonal elements are all the same. Then

$$E_1(\lambda) + E_1(\lambda^{-1}) = E_3(\lambda) + E_3(\lambda^{-1}) = I_{k-1},$$

since for $i = 1, \ldots, k-1$, and $i = n-k+2, \ldots, n$,

$$\frac{\lambda_i}{\lambda_i + \lambda_{n-i+1}} + \frac{\lambda_i^{-1}}{\lambda_i^{-1} + \lambda_{n-i+1}^{-1}} = \frac{\lambda_i}{\lambda_i + \lambda_{n-i+1}} + \frac{\lambda_{n-i+1}}{\lambda_i + \lambda_{n-i+1}} = 1.$$

Moreover,

$$E_2(\lambda) + E_2(\lambda^{-1}) \leq_{\mathrm{L}} I_{n-2(k-1)},$$

since for $i = k, k+1, \ldots, n-k+1$,

$$\begin{aligned} \frac{\lambda_i}{\lambda_k + \lambda_{n-k+1}} + \frac{\lambda_i^{-1}}{\lambda_k^{-1} + \lambda_{n-k+1}^{-1}} &= \frac{\lambda_i}{\lambda_k + \lambda_{n-k+1}} + \frac{\lambda_i^{-1}\lambda_k\lambda_{n-k+1}}{\lambda_k + \lambda_{n-k+1}} \\ &\leq \frac{\lambda_k}{\lambda_k + \lambda_{n-k+1}} + \frac{\lambda_{n-k+1}}{\lambda_k + \lambda_{n-k+1}} = 1. \end{aligned}$$

Hence

$$E(\lambda) + E(\lambda^{-1}) \leq_{\mathrm{L}} I_n. \tag{4.34}$$

Using (4.24) and (4.34) we then obtain that

$$|X'E(\lambda)X| \cdot |X'E(\lambda^{-1})X| \leq \left|X'\left[\tfrac{1}{2}\left(E(\lambda) + E(\lambda^{-1})\right)\right]X\right|^2 \leq \left|\tfrac{1}{4}I_k\right| = \frac{1}{4^k}. \tag{4.35}$$

Yang ([176], equation (7) on p. 4589) presents the stronger inequality with determinants replaced by matrices with the Löwner partial ordering:

$$X'E(\lambda)X \cdot X'E(\lambda^{-1})X \leq_{\mathrm{L}} \left(X'\left[\tfrac{1}{2}\left(E(\lambda) + E(\lambda^{-1})\right)\right]X\right)^2 \leq_{\mathrm{L}} \tfrac{1}{4}I_k. \tag{4.36}$$

The first inequality in (4.36) is valid in general, however, only when $X'E(\lambda)X$ and $X'E(\lambda^{-1})X$ commute, cf. (4.24) above.

We introduce the $n \times k$ matrices

$$P = \Sigma^{1/2}X(X'\Sigma X)^{-1/2} \quad \text{and} \quad Q = \Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-1/2}$$

so that $P'P = Q'Q = I_k$. Then using the Poincaré Separation Theorem we find that

$$|P'K(\lambda)P| = \prod_{i=1}^{k} \mathrm{ch}_i\left(P'K(\lambda)P\right) \geq \prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda)\right)$$

and

$$|Q'K(\lambda^{-1})Q| = \prod_{i=1}^{k} \mathrm{ch}_i \left(Q'K(\lambda^{-1})Q\right) \geq \prod_{i=1}^{k} \mathrm{ch}_{n-k+i} \left(K(\lambda^{-1})\right).$$

However

$$|P'K(\lambda)P| = \frac{|X'\Sigma^{1/2}K(\lambda)\Sigma^{1/2}X|}{|X'\Sigma X|} = \frac{|X'E(\lambda)X|}{|X'\Sigma X|}$$

and

$$|Q'K(\lambda^{-1})Q| = \frac{|X'\Sigma^{-1/2}K(\lambda^{-1})\Sigma^{-1/2}X|}{|X'\Sigma^{-1}X|} = \frac{|X'E(\lambda^{-1})X|}{|X'\Sigma^{-1}X|}.$$

Hence

$$\begin{aligned} |X'\Sigma X| \cdot |X'\Sigma^{-1}X| &= \frac{|X'E(\lambda)X|}{|P'K(\lambda)P|} \cdot \frac{|X'E(\lambda^{-1})X|}{|Q'K(\lambda^{-1})Q|} \\ &\leq \left[4^k \prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda)\right) \cdot \prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda^{-1})\right)\right]^{-1}. \end{aligned} \tag{4.37}$$

If we put, cf. Yang ([176], p. 4589, equation (8)),

$$\prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda)\right) = \prod_{i=1}^{k} \frac{1}{\lambda_i + \lambda_{n-i+1}} \tag{4.38}$$

and

$$\prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda^{-1})\right) = \prod_{i=1}^{k} \frac{1}{\lambda_i^{-1} + \lambda_{n-i+1}^{-1}} \tag{4.39}$$

in the last expression in (4.37), then it becomes

$$\left[4^k \prod_{i=1}^{k} \frac{1}{\lambda_i + \lambda_{n-i+1}} \cdot \prod_{i=1}^{k} \frac{1}{\lambda_i^{-1} + \lambda_{n-i+1}^{-1}}\right]^{-1} = \prod_{i=1}^{k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}},$$

which then yields the Bloomfield-Watson-Knott Inequality

$$|X'\Sigma X| \cdot |X'\Sigma^{-1}X| \leq \prod_{i=1}^{k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}.$$

Unfortunately, however, it appears that this deduction is not valid since the equalities in (4.38) and (4.39) do not, in general, hold. To see this let us consider

(4.38). Its left-hand side is the product of the k smallest eigenvalues of the diagonal matrix $K(\lambda)$. Inspection of (4.27) shows that the matrix $K(\lambda)$ has at most k distinct diagonal elements, which are indeed the k smallest! But the k smallest diagonal elements of $K(\lambda)$ are, in general, not distinct. For example, if $k = 3$ and $n = 6$ then

$$K(\lambda) = \begin{pmatrix} \frac{1}{\lambda_1+\lambda_6} & & & & & \\ & \frac{1}{\lambda_2+\lambda_5} & & & & \\ & & \frac{1}{\lambda_3+\lambda_4} & & & \\ & & & \frac{1}{\lambda_3+\lambda_4} & & \\ & & & & \frac{1}{\lambda_2+\lambda_5} & \\ & & & & & \frac{1}{\lambda_1+\lambda_6} \end{pmatrix}. \tag{4.40}$$

Suppose that in (4.40) we have

$$\frac{1}{\lambda_1+\lambda_6} < \frac{1}{\lambda_2+\lambda_5} < \frac{1}{\lambda_3+\lambda_4}.$$

Then the left-hand side of (4.38) is

$$\left(\frac{1}{\lambda_1+\lambda_6}\right)^2 \cdot \frac{1}{\lambda_2+\lambda_5},$$

which is strictly less than its right-hand side:

$$\frac{1}{\lambda_1+\lambda_6} \cdot \frac{1}{\lambda_2+\lambda_5} \cdot \frac{1}{\lambda_3+\lambda_4}.$$

In general, therefore, we find that

$$\left[4^k \prod_{i=1}^{k} \mathrm{ch}_{n-k+i}(K(\lambda)) \cdot \prod_{i=1}^{k} \mathrm{ch}_{n-k+i}\left(K(\lambda^{-1})\right)\right]^{-1} \geq \prod_{i=1}^{k} \frac{(\lambda_i+\lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}$$

and so does not allow us to deduce the Bloomfield-Watson-Knott Inequality (4.9) from (4.37).

4.3 Inequalities Related to the Bloomfield-Watson-Knott Inequality

In 1981 C. G. Khatri and C. R. Rao [76] (see also [77]) generalized the Bloomfield-Watson-Knott Inequality in several directions.

Let T and Z be $n \times k$ matrices such that $T'T = Z'Z = I_k$ and let A be an $n \times n$ non-singular matrix with (fixed) singular values $\sigma_1 \geq \cdots \geq \sigma_n > 0$. Then

$$|T'AZ| \cdot |Z'A^{-1}T| \leq \prod_{i=1}^{min(k,n-k)} \frac{(\sigma_i + \sigma_{n-i+1})^2}{4\sigma_i\sigma_{n-i+1}}. \tag{4.41}$$

When $k = 1$ then (4.41) reduces to the inequality obtained by Strang (1960) [151], cf. (3.12) in §3.3.

$$\begin{aligned} \operatorname{tr}(T'AZ \cdot Z'A^{-1}T) &\leq \sum_{i=1}^{k} \frac{(\sigma_i + \sigma_{n-i+1})^2}{4\sigma_i\sigma_{n-i+1}} && \text{if } 2k \leq n \\ &\leq \sum_{i=1}^{n-k} \frac{(\sigma_i + \sigma_{n-i+1})^2}{4\sigma_i\sigma_{n-i+1}} + 2k - n && \text{if } 2k > n. \end{aligned} \tag{4.42}$$

Let B and C be symmetric $n \times n$ non-singular matrices such that $BC = CB$ is positive definite, let T be an $n \times k$ matrix of rank k and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of BC^{-1}. Then

$$\frac{|T'B^2T| \cdot |T'C^2T|}{|T'BCT|^2} \leq \prod_{i=1}^{min(k,n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} \tag{4.43}$$

and

$$\begin{aligned} &\operatorname{tr}T'B^2T(T'BCT)^{-1}T'C^2T(T'BCT)^{-1} \\ &\qquad \leq \sum_{i=1}^{k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} && \text{if } 2k \leq n \\ &\qquad \leq \sum_{i=1}^{n-k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} + (2k - n) && \text{if } 2k > n. \end{aligned} \tag{4.44}$$

In 1992 Song-Gui Wang and Jun Shao [161], using additional information about the matrix T, obtained a sharper bound, which we will call the "Constrained Kantorovich Inequality", cf. (4.45) and (4.47) below.

Let T be an $n \times p$ matrix of rank r, let A and B be $n \times n$ positive definite matrices with $\lambda_1 \geq \cdots \geq \lambda_n > 0$ the eigenvalues of $B^{1/2}A^{-1}B^{1/2}$, and let $p_1, \ldots, p_n$ be corresponding orthonormal eigenvectors. Suppose that the column space (or range) $\mathcal{C}(B^{1/2}T) \subset \mathcal{C}(p_{i_1}, \ldots, p_{i_k})$ for some $1 \leq i_1 \leq \cdots \leq i_k \leq n$, $k \leq n$. Then

$$T'BA^{-1}BT \leq_L \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} T'BT(T'AT)^- T'BT. \tag{4.45}$$

If $k = r$ then

$$T'BA^{-1}BT = T'BT(T'AT)^- T'BT.$$

Let T and T^{-1} partitioned as

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} T^{11} & T^{12} \\ T^{21} & T^{22} \end{pmatrix}, \tag{4.46}$$

where T_{11} is $h \times h$ and T_{22} is $(n-h) \times (n-h)$ matrices. Suppose $\mathcal{C}(Z_h) \subset \mathcal{C}(p_{i_1}, \ldots, p_{i_k})$ for some $1 \leq i_1 \leq \cdots \leq i_k \leq n$, where $Z_h = (I_h, 0)'$ is an $n \times h$ matrix. Then

$$T^{11} \leq_L \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} T_{11}^{-1}. \tag{4.47}$$

4.4 Other Matrix Extensions of the Kantorovich Inequality

In 1995 Shuangzhe Liu [88] established several extensions of the Kantorovich Inequality using the Löwner partial ordering.

Let A_i be $n \times n$ symmetric non-negative definite matrices and let V_i be $n \times p$ matrices $(i = 1, \ldots, k)$. Then

$$\left(\sum_{i=1}^{k} V_i' A_i V_i\right)^+ \leq_L \sum_{i=1}^{k} V_i' A_i^+ V_i \leq_L \frac{(m+M)^2}{4mM} \left(\sum_{i=1}^{k} V_i' A_i V_i\right)^+, \tag{4.48}$$

cf. (3.11) in chapter3, where the non-zero, necessarily positive eigenvalues of the A_i lie in the closed interval $[m, M]$ $(0 < m < M)$ and $\sum_{i=1}^{k} V_i' A_i A_i^+ V_i$ is idempotent. The inequality (4.48) with $k = 1$ was obtained earlier by Baksalary and Puntanen [9].

In 1996 Liu and Neudecker [93] obtained related inequalities involving Hadamard (elementwise) and Kronecker products. Let A_1 and A_2 be $n \times n$ positive definite matrices with λ_1 and λ_n, respectively, the largest and smallest eigenvalues of the Kronecker product $A_1 \otimes A_2$, of which the Hadamard product $A_1 \odot A_2$ is a principal submatrix. Then Liu and Neudecker [93] showed that:

$$\begin{aligned} A_1^2 \odot A_2^2 - (A_1 \odot A_2)^2 &\leq_{\mathrm{L}} \tfrac{1}{4}(\lambda_1 - \lambda_n)^2 I_p, \\ (A_1^2 \odot A_2^2)^{1/2} &\leq_{\mathrm{L}} \frac{\lambda_1 + \lambda_n}{2\sqrt{\lambda_1 \lambda_n}} A_1 \odot A_2, \end{aligned}$$

while Lui ([89], p. 65) established

$$A_1^2 \odot A_2^2 \leq_{\mathrm{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} (A_1 \odot A_2)^2.$$

We also find, similarly, that

$$\begin{aligned} A_1 \odot A_2 - (A_1^{-1} \odot A_2^{-1})^{-1} &\leq_{\mathrm{L}} \left(\sqrt{\lambda_1} - \sqrt{\lambda_n}\right)^2 I_p, \\ (A_1^2 \odot A_2^2)^{1/2} - A_1 \odot A_2 &\leq_{\mathrm{L}} \frac{(\lambda_1 - \lambda_n)^2}{4(\lambda_1 + \lambda_n)} I_p, \end{aligned}$$

cf. ([89], p. 65).

Let A be an $n \times n$ positive definite matrix and let T be an $n \times k$ matrix such that $T'T = I_k$. Then Marshall and Olkin [107] established the following matrix version of the Kantorovich Inequality:

$$T'A^{-1}T \leq_{\mathrm{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} (T'AT)^{-1}. \tag{4.49}$$

Pečarić, Puntanen and Styan [131] obtained several other similar matrix inequalities, e.g.,

$$\begin{aligned} T'A^2T &\leq_{\mathrm{L}} \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1 \lambda_r} (T'AT)^2, \\ (T'A^2T)^{1/2} - T'AT &\leq_{\mathrm{L}} \frac{(\lambda_1 - \lambda_r)^2}{4(\lambda_1 + \lambda_r)} T'H_AT, \\ T'A^2T - (T'AT)^2 &\leq_{\mathrm{L}} \tfrac{1}{4}(\lambda_1 - \lambda_r)^2 T'H_AT, \\ T'AT - (T'A^+T)^+ &\leq_{\mathrm{L}} \left(\sqrt{\lambda_1} - \sqrt{\lambda_r}\right)^2 T'H_AT. \end{aligned} \tag{4.50}$$

Here A is an $n \times n$ nonnegative definite matrix of rank r with r positive eigenvalues $\lambda_1 \geq \cdots \geq \lambda_r > 0$, H_A is the orthogonal projector AA^+ and T is an $n \times k$ matrix such that $H_A T$ is a partial isometry, i.e., $T'H_A T$ is idempotent. Mond and Pečarić [114], Liu and Neudecker [93] and Baksalary and Puntanen [9] presented related results.

Furthermore Baksalary and Puntanen [9], cf. Pečarić, Puntanen and Styan [131], generalized (4.49): Let A, H_A and T be defined as above except that $H_A T$ is a partial isometry. Then

$$\lambda_1 \lambda_r T'A^+T \leq_L (\lambda_1 + \lambda_r) T'H_A T - T'AT. \tag{4.51}$$

When $H_A T$ is a partial isometry, then (4.51) implies that

$$T'A^+T \leq_L \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1 \lambda_r} (T'AT)^+.$$

If T is a vector t, say, and $t'H_A t = 1$ then (4.51) may be written as

$$\lambda_1 \lambda_r t'A^+t \leq \lambda_1 + \lambda_r - t'At.$$

Then

$$\begin{aligned} \lambda_1 \lambda_r t'At \cdot t'A^+t &\leq (\lambda_1 + \lambda_r) t'At - (t'At)^2 \\ &= \tfrac{1}{4}(\lambda_1 + \lambda_r)^2 - \left[t'At - \tfrac{1}{2}(\lambda_1 + \lambda_r)\right]^2 \\ &\leq \tfrac{1}{4}(\lambda_1 + \lambda_r)^2, \end{aligned}$$

we have the following extension of the Kantorovich Inequality:

$$t'At \cdot t'A^+t \leq \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1 \lambda_r},$$

where the matrix A is positive semidefinite with rank r and $t'H_A t = 1$, and H_A is the orthogonal projector AA^+. This leads directly to

$$\frac{u'\Lambda u \cdot u'\Lambda^{-1} u}{(u'u)^2} \leq \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1 \lambda_r}, \tag{4.52}$$

where the $r \times r$ diagonal matrix Λ contains the r necessarily positive eigenvalues of A, while the $r \times 1$ vector $u = S't$, with $A = S\Lambda S'$, $A^+ = S\Lambda^{-1}S'$ and $S'S = I_r$. Comparing (4.52) with (1.5), we see that (4.52) follows directly from the Kantorovich Inequality as observed by Neudecker [123].

Recently Shuangzhe Liu [90] proposed the following problem: Let B and C be two positive definite matrices with eigenvalues contained in the interval $[m, M]$, where $0 < m \leq M$. Let $0 \leq \lambda \leq 1$. Prove that

$$\lambda B^2 + (1-\lambda)C^2 - [\lambda B + (1-\lambda)C]^2 \leq_{\mathrm{L}} \tfrac{1}{4}(M-m)^2 I. \tag{4.53}$$

We note that this is the special case of (4.50) with

$$T = (\sqrt{\lambda} I_n : \sqrt{1-\lambda} I_n)'$$

and

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}. \tag{4.54}$$

Other proofs of (4.53) are given by Louis Kates, Serge Kruk and Henry Wolkowicz [74] and by Ingram Olkin [127].

We also note that (4.53) may be extended to positive semi-definite matrices. Let H_B and H_C denote the orthogonal projector BB^+ and CC^+, respectively. When

$$H = \lambda H_B + (1-\lambda) H_C$$

is idempotent, cf. (4.50), it follows that

$$\lambda B^2 + (1-\lambda)C^2 - [\lambda B + (1-\lambda)C]^2 \leq_{\mathrm{L}} \tfrac{1}{4}(M-m)^2 H,$$

and then $H_B = H_C = H$ and the column spaces of A and B are identical.

Chapter 5

Some Statistical Applications

In this chapter we present a variety of statistical applications of the Kantorovich Inequality and the Bloomfield-Watson-Knott Inequality. We concentrate on these four papers:

- Magness and McGuire (1961)
- Venables (1976)
- Cressie (1980)
- Wang and Shao (1992).

5.1 Magness and McGuire (1961): Efficiency of Weighted Least Squares

Let us consider the model (4.1). In 1961 T. A. Magness and J. B. McGuire [97], apparently unaware of the work by Watson in 1951/1955 ([163], [163] and [165]) cf. (4.9), measured the efficiency, with respect to the BLUE, of the so-called Weighted Least Squares Estimator (WLSE)

$$\beta^* = (X'W^2X)^{-1}X'W^2y,$$

where the $n \times n$ diagonal matrix of "weights"

$$W = \{\text{diag}(\Sigma)\}^{-1/2}. \tag{5.1}$$

The covariance matrix of the WLSE β^* is

$$\text{Cov}(\beta^*) = (X'W^2X)^{-1}X'W^2\Sigma W^2X(X'W^2X)^{-1}.$$

Let us pre-multiply the model (4.1) by W and set $z = Wy$, $\eta = W\epsilon$; then we have $z = WX\beta + \eta$. The new error η has covariance matrix $W\Sigma W = R$, say, the correlation matrix of ϵ. There exists a $k \times k$ positive definite matrix B, say, such that $B'X'W^2XB = I$; we now set $WXB = F$ and $B^{-1}\beta = \alpha$ to obtain the "canonical" equation

$$z = F\alpha + \eta,$$

where $F'F = BX'W^2XB = I$. Let α^* and $\tilde{\alpha}$ be the WLSE and BLUE of α; the covariance matrices are

$$\text{Cov}(\alpha^*) = F'RF \quad \text{and} \quad \text{Cov}(\tilde{\alpha}) = (F'R^{-1}F)^{-1}.$$

Then from the Kantorovich Inequality

$$\text{Cov}(\alpha^*) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}\text{Cov}(\tilde{\alpha}), \tag{5.2}$$

where λ_1 and λ_n are, respectively, the largest and smallest eigenvalues of R.

Let us consider the following example of a regression problem.

Example 5.1

Let

$$R = \begin{pmatrix} 1 & r & 0 & 0 \\ r & 1 & 0 & 0 \\ 0 & 0 & 1 & r \\ 0 & 0 & r & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix},$$

where the largest and smallest eigenvalues of R are $1 + r$ and $1 - r$. It is easy to show $F'F = I$, $\text{Cov}(\alpha^*) = F'RF = I$ and

$$\mathrm{Cov}(\tilde{\alpha}) = (F'R^{-1}F)^{-1} = \begin{pmatrix} (1-r^2)^{-1} & 0 \\ 0 & (1-r^2)^{-1} \end{pmatrix} = (1-r^2)^{-1}I.$$

Equality in (5.2) is attained since $(\lambda_1 + \lambda_n)^2/4\lambda_1\lambda_n = (1-r^2)^{-1}$. [When $r > 0$ then $1 + r = \lambda_1$ is the largest and $1 - r = \lambda_n$ is the smallest eigenvalue; when $r < 0$ then $1 - r = \lambda_1$ is the largest and $1 + r = \lambda_n$ is the smallest eigenvalue.]

5.2 Venables (1976): Testing Sphericity

In 1976 W. Venables [156] used the Bloomfield-Watson-Knott Inequality to derive a class of union-intersection tests for sphericity using the likelihood ratio (LR) statistic. Let S be a $p \times p$ sample covariance matrix with distribution $nS \sim W_p(n, \Sigma)$, where W denotes the Wishart distribution. We want to test the hypothesis

$$H_0 : \Sigma = \sigma^2 I_p, \quad \sigma^2 \text{ unknown, within}$$
$$H_1 : \Sigma \text{ unspecified.}$$

Let us define

$$H_0(Q) : Q \text{ is an invariant subspace (i.e., } \Sigma t \in Q \text{ forall } t \in Q\text{)},$$
$$H_1(Q) : Q \text{ is not necessarily invariant,}$$

where the arbitrary q-dimensional subspace $Q \subset \mathcal{R}^p$, Euclidean p-space. Then

$$H_0 = \bigcap_Q H_0(Q) \quad \text{and} \quad H_1 = \bigcup_Q H_1(Q),$$

where the union and the intersection (UI) are over all q-dimensional subspaces Q of $\mathcal{R}^p$.

Suppose that the columns of the $p \times q$ matrix X_1 and the columns of the $p \times (p-q)$ matrix X_2 form orthonormal bases for Q and $Q^\perp$, respectively. Testing $H_0(Q)$ within $H_1(Q)$ is the same as testing whether the two sets of variates $X_1'Z$ and $X_2'Z$ are uncorrelated, where the $p \times 1$ vector Z is normally distributed $N(0, \Sigma)$.

The LR test uses the statistic

$$U_q(Q) = |X'SX|/|X_1'SX_1|\ |X_2'SX_2|,$$

where $X = (X_1 \ : \ X_2)$ is a $p \times p$ orthogonal matrix. Now

$$|X'SX| = |X_2'SX_2|/|X_1'S^{-1}X_1|$$

since $(X'SX)^{-1} = X'S^{-1}X$. Venables [156], therefore, considers the statistics

$$u_q(Q) = |X_1'SX_1||X_1'S^{-1}X_1|$$

and so the UI statistic is

$$u_q^* = \max_Q u_q(Q).$$

Using the Bloomfield-Watson-Knott Inequality it follows that

$$u_q^* = \prod_{i=1}^{q} \frac{(l_i + l_{p-i+1})^2}{4l_i l_{p-i+1}},$$

where $l_1 > l_2 > \cdots > l_p > 0$ are the eigenvalues of S.

5.3 Cressie (1980)

5.3.1 Efficiency of an Unbiased Weighted Estimator

In 1980 Noel Cressie [35] measured the efficiency of an Arbitrary Positively Weighted Unbiased Estimator (APWUE) with respect to the Optimally Weighted Unbiased Estimator (OWUE).

Let the random variables Y_i $(i = 1, \ldots, n)$, be uncorrelated with the same mean μ and with variances σ_i^2 $(i = 1, \ldots, n)$. Then $\hat{\mu}_w$ is an APWUE and $\hat{\mu}_0$ is the OWUE of μ as defined by

$$\hat{\mu}_w = \sum_{i=1}^{n} w_i Y_i \quad \text{and} \quad \hat{\mu}_0 = \left(\sum_{i=1}^{n} Y_i/\sigma_i^2\right) \Big/ \left(\sum_{i=1}^{n} 1/\sigma_i^2\right), \tag{5.3}$$

respectively, where $w_i > 0$ $(i = 1, \ldots, n)$, and $\sum_{i=1}^{n} w_i = 1$. Then

$$\text{eff}(\hat{\mu}_w) = \frac{\text{Var}(\hat{\mu}_0)}{\text{Var}(\hat{\mu}_w)} \geq \frac{4\kappa}{(\kappa + 1)^2}, \tag{5.4}$$

where $\kappa = M/m$ and

$$m = \min\{w_i\sigma_i^2;\ i = 1, \ldots, n\}, \quad M = \max\{w_i\sigma_i^2;\ i = 1, \ldots, n\}.$$

5.3.2 Asymptotic Efficiency of a General Weighted Median Estimator

Cressie [35] also studied the efficiency of a General Weighted Median Estimator (GWME) with respect to the Optimally Weighted Median Estimator (OWME).

Suppose that different scale parameters cause random variables $Y_1, \ldots, Y_n$ to be made up of k samples with a common mean μ: $Y_{i1}, \ldots, Y_{in_i}$, which are each independently and identically distributed with variance σ_i^2 $(i = 1, \ldots, k)$, and $\sum_{i=1}^k n_i = n$. Then a GWME $\tilde{\mu}_w$ is the (midpoint) value of (those) μ which minimizes

$$\sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} |Y_{ij} - \mu| \tag{5.5}$$

or equivalently as the solution to

$$\sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} \text{sgn}(Y_{ij} - \mu) = 0. \tag{5.6}$$

In (5.5) and (5.6) we may choose any weights w_i and we do not here require that $\sum_{i=1}^k w_i = 1$.

As observed by Cressie [35], Tukey (1974) [204] proved that when

$$n_i/n \to \theta_i \quad \text{as} \quad n \to \infty \tag{5.7}$$

and under (other) certain regularity conditions $\sqrt{n}(\tilde{\mu}_w - \mu)$ is asymptotically normal with mean 0 and variance σ_w^2, say. The OWME $\tilde{\mu}_0$ is defined as the GWME that minimizes σ_w^2 and the corresponding optimal weights are

$$w_i^0 = k/\sigma_i$$

(with σ_i known). Then we define the efficiency of the GWME with respect to the OWME as the ratio of their asymptotic variances:

$$\text{eff}(\tilde{\mu}_w) = \frac{(\sum w_i \theta_i/\sigma_i)^2}{\sum w_i^2 \theta_i \cdot \sum \theta_i/\sigma_i^2} \geq \frac{4\kappa}{(\kappa+1)^2}, \tag{5.8}$$

where $\kappa = M/m$ and

$$m = \min\{w_i\sigma_i;\ i = 1, \dots, k\}, \quad M = \max\{w_i\sigma_i;\ i = 1, \dots, k\}.$$

We note that the lower bounds in (5.4) and (5.8) differ only by the definition of m and M.

5.4 Wang and Shao (1992): Constrained Kantorovich Inequalities and the Efficiency of Ordinary Least Squares

In 1992 Song-Gui Wang and Jun Shao [161] sharpened the classical Kantorovich Inequality when the underlying vector satisfies certain constraints. These "Constrained Kantorovich Inequalities" yield sharper lower bounds for the efficiency of ordinary least squares in the general linear statistical model.

Let us consider the so-called Gauss-Markov linear statistical model, cf. §4.1,

$$y = X\beta + \varepsilon, \quad \mathrm{E}(\varepsilon) = 0, \quad \mathrm{Cov}(\varepsilon) = \Sigma,$$

where y is an $n \times 1$ vector of observations, β is a $k \times 1$ vector of unknown parameters, X is an $n \times k$ known design matrix of rank $r \leq k$, ε is an $n \times 1$ vector of random errors and the covariance matrix Σ is $n \times n$ and positive definite.

The Best Linear Unbiased Estimator (BLUE) of the estimable function $c'\beta$ is the generalized least squares estimator

$$c'\tilde{\beta} = c'(X'\Sigma^{-1}X)^{-}X'\Sigma^{-1}y, \tag{5.9}$$

which has variance

$$\mathrm{Var}(c'\tilde{\beta}) = c'(X'\Sigma^{-1}X)^{-}c, \tag{5.10}$$

while the Ordinary Least Squares Estimator (OLSE)

$$c'\hat{\beta} = c'(X'X)^{-}X'y \tag{5.11}$$

has variance

$$\text{Var}(c'\hat{\beta}) = c'(X'X)^{-}X'\Sigma X(X'X)^{-}c. \tag{5.12}$$

Any generalized inverses $(X'X)^{-}$ and $(X'\Sigma^{-1}X)^{-}$ may be chosen.

Let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the, necessarily positive, eigenvalues of the covariance matrix Σ and let $p_1, \ldots, p_n$ denote corresponding orthonormal vectors. We constrain the matrix X by supposing that its column space $\mathcal{C}(X) \subset \mathcal{C}(p_{i_1}, \ldots, p_{i_k})$ for some integers i_l satisfying $1 \leq i_1 \leq \cdots \leq i_k \leq n, \quad k \leq n$. It then follows that

$$\begin{aligned} \text{eff}(c'\hat{\beta}) &= 1 \qquad\qquad \text{if } r = k \\ &\geq \frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1} + \lambda_{i_k})^2} \quad \text{if } r < k. \end{aligned} \tag{5.13}$$

Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ be a random error vector such that $\mathbf{E}(\varepsilon_i)$=0 and

$$\text{Cov}(\varepsilon_i, \ \varepsilon_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ \sigma^2\rho & \text{if } i \neq j. \end{cases}$$

The covariance matrix $\text{Cov}(\varepsilon)$ then has the simple eigenvalue $\lambda = \sigma^2[1 + (n-1)\rho]$ with eigenvector 1_n and the eigenvalue $\tau = \sigma^2(1-\rho)$ with multiplicity $n-1$. More generally, let us suppose that the covariance matrix Σ has the following block-diagonal form:

$$\Sigma = \text{diag}\{\Sigma_{n_1}, \ldots, \Sigma_{n_k}\}, \tag{5.14}$$

where each Σ_{n_i} is a positive definite matrix of order $n_i \times n_i$, which satisfies

$$\Sigma_{n_i} \text{ has the simple eigenvalue } \lambda_i \text{ with eigenvector } 1_{n_i}, \tag{5.15}$$

$$\Sigma_{n_i} \text{ has the eigenvalue } \tau_i \text{ with multiplicity } n_i - 1. \tag{5.16}$$

Furthermore,

1. Let $J_1 = (1'_{n_1}, 0, \ldots, 0)'$, $J_2 = (0, 1'_{n_2}, 0, \ldots, 0)'$, $\ldots, J_k = (0, 0, \ldots, 1'_{n_k})'$. If $\mathcal{C}(X) \subset \mathcal{C}(J_1, \cdots, J_k)$, then for any estimable function $c'\beta$ the efficiency

$$\text{eff}(c'\hat{\beta}) \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}, \tag{5.17}$$

where $\lambda_1 = \max \lambda_i$ and $\lambda_n = \min \lambda_i$. If $\lambda_1 = \lambda_n$ then $c'\hat{\beta}$ is the BLUE for any estimable $c'\beta$.

2. If $X'J_i = 0$ for all i then for any estimable function $c'\beta$ the efficiency

$$\text{eff}(c'\hat{\beta}) \geq \frac{4\tau_1\tau_n}{(\tau_1 + \tau_n)^2},$$

where $\tau_1 = \max \tau_i$ and $\tau_n = \min \tau_i$. If $\tau_1 = \tau_n$ then $c'\hat{\beta}$ is the BLUE for any estimable function $c'\beta$.

3. If in (5.13), $k = 1$ and 1_n is a column of X then $c'\hat{\beta}$ is the BLUE for any estimable function $c'\beta$.

Suppose now that $\mathcal{C}(X(X'X)^-C) \subset \mathcal{C}(p_{i_1}, \ldots, p_{i_k})$ for some $1 \leq i_1 \leq \cdots \leq i_k \leq n$, where $C = (c_1, \ldots, c_l)$ and $c_i'\beta$ $(i = 1, \ldots, l)$, are linearly independent estimable functions. Then for any $c \in \mathcal{C}(C)$

$$\text{eff}(c'\hat{\beta}) \geq \frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1} + \lambda_{i_k})^2}. \tag{5.18}$$

Consider the model

$$y = 1_n\beta + \varepsilon, \quad \mathbf{E}(\varepsilon) = 0, \quad \text{Cov}(\varepsilon) = \Sigma,$$

where 1_n is an $n \times 1$ vector of ones and β is an unknown parameter. We know that the OLSE $\overline{y} = \sum_{i=1}^n y_i/n$ is also the BLUE if and only if $\Sigma 1_n = c1_n$ for some scalar c, cf. e.g., Puntanen and Styan [137]. It is easy to see that here 1_n is an eigenvector of Σ. Let us look at the following example:

$$\Sigma_0 = \begin{pmatrix} \frac{1}{2}\lambda + 2 & \frac{1}{2}\lambda - 2 & 0 & 0 \\ \frac{1}{2}\lambda - 2 & \frac{1}{2}\lambda + 2 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\lambda + 1 & \frac{1}{2}\lambda - 1 \\ 0 & 0 & \frac{1}{2}\lambda - 1 & \frac{1}{2}\lambda + 1 \end{pmatrix}.$$

Then $\overline{y}$ is the BLUE since $\Sigma_0 1_4 = \lambda 1_4$. Now let us make a small perturbation in Σ_0 so that

$$\Sigma_\varepsilon = \begin{pmatrix} \frac{1}{2}\lambda+2 & \frac{1}{2}\lambda-2 & 0 & 0 \\ \frac{1}{2}\lambda-2 & \frac{1}{2}\lambda+2 & 0 & 0 \\ 0 & 0 & \frac{1}{2}(\lambda+\varepsilon)+1 & \frac{1}{2}(\lambda+\varepsilon)-1 \\ 0 & 0 & \frac{1}{2}(\lambda+\varepsilon)-1 & \frac{1}{2}(\lambda+\varepsilon)+1 \end{pmatrix},$$

where $\varepsilon > 0$. Then $\overline{y}$ is no longer the BLUE since 1_4 is no longer an eigenvector of Σ_ε. The matrix Σ_ε has eigenvalues λ, $\lambda+\varepsilon$, 2, and 4, with $p_1 = ((1/\sqrt{2})1_2', 0, 0)'$ and $p_2 = (0, 0, (1/\sqrt{2})1_2')'$ are orthonormal eigenvectors corresponding to the eigenvalues λ, $\lambda+\varepsilon$, respectively. Since $C(p_1, p_2)$ is the smallest eigen subspace containing 1_4, then by (5.13)

$$\mathrm{eff}(\overline{y}) \geq \frac{4\lambda(\lambda+\varepsilon)}{(2\lambda+\varepsilon)^2}. \tag{5.19}$$

Note that the efficiency $\mathrm{eff}(\overline{y}) \to 1$ as $\varepsilon \to 0$. We conclude that the small perturbation in this example makes $\overline{y}$ robust. If we use the lower bound in (4.9) and assume that $\lambda + \varepsilon < 2$ then we obtain

$$\mathrm{eff}(\overline{y}) \geq \frac{16\lambda}{(\lambda+4)^2},$$

which is smaller than the "constrained" lower bound (5.19).

Example 5.2

Consider the random effects model

$$y = 1_n\mu + U\xi + \varepsilon, \quad \mathbf{E}(\varepsilon) = 0, \quad \mathrm{Cov}(\varepsilon) = \sigma_\varepsilon^2 I_n,$$

where y is an $n \times 1$ vector of observations, μ is a non-random unknown scalar parameter, U is an $n \times k$ known matrix of full rank and ξ is a $k \times 1$ random vector with $\mathbf{E}(\xi) = 0$, $\mathrm{Cov}(\xi) = \sigma_\xi^2 I_n$ and ε is an $n \times 1$ error vector which is independent of ξ. Let

$$U = P \begin{pmatrix} \Lambda^{1/2} \\ 0 \end{pmatrix} Q'$$

be a singular value decomposition of U, where $P = (p_1, \ldots, p_n)$ such that $P'P = I_n$, $\Lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 \geq \cdots \geq \lambda_k > 0$, Q is a $k \times k$ matrix with $Q'Q = I_k$. Then the covariance matrix $\mathrm{Cov}(y) = \sigma_\xi^2 UU' + \sigma_\varepsilon^2 I_n$ has k eigenvalues $\sigma_\xi^2\lambda_1 + \sigma_\varepsilon^2, \ldots, \sigma_\xi^2\lambda_k + \sigma_\varepsilon^2$ each with multiplicity one and σ_ε^2 with multiplicity $n-k$ with corresponding orthonormal eigenvectors $p_1, \ldots, p_n$.

If $UU'1_n \neq c1_n$, then 1_n is equal neither to one of the p_i nor to a linear combination of $p_{k+1}, \cdots, p_n$. Assume that $1_n = \sum_{i \in I} \alpha_i p_i$ for a subset $I \subset \{1, \ldots, n\}$ and some constants α_i. Let $i_0 = \min\{i : i \in I\}$. Then by (5.13)

$$\mathrm{eff}(\overline{y}) \geq \frac{4(1 + \lambda_{i_0}\sigma_\xi^2/\sigma_\varepsilon^2)}{(2 + \lambda_{i_0}\sigma_\xi^2/\sigma_\varepsilon^2)^2}.$$

Example 5.3

Consider the one-way analysis of variance model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (j = 1, \cdots, n_i, \quad i = 1, \cdots, k).$$

If $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ for all $i \neq i'$ and

$$\mathrm{Cov}(\varepsilon_{ij},\ \varepsilon_{ij'}) = \begin{cases} \sigma_i^2 & \text{if } j = j' \\ \sigma_i^2\rho_i & \text{if } j \neq j' \end{cases}$$

then (5.17) applies here since the conditions (5.14)-(5.16) are satisfied.

Example 5.4

Consider the analysis of variance model

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (i = 1, \ldots, a, \quad j = 1, \ldots, b), \tag{5.20}$$

where α_i are the treatment effects with $\sum_{i=1}^a \alpha_i = 0$ and the β_j are the block effects with $\sum_{i=1}^a \beta_j = 0$. Between blocks, the erors ε_{ij} are independent, within the jth block, $\mathrm{Cov}(\varepsilon_{jl}, \varepsilon_{jk}) = \sigma_j^2\rho_j$ if $l \neq k$ and $\mathrm{Cov}(\varepsilon_{jl}, \varepsilon_{jk}) = \sigma_j^2$ if $l = k$, so that the error covariance matrix satisfies the conditions (5.14)-(5.16).

We consider an arbitrary contrast $\alpha_1, \ldots, \alpha_a$ of the treatment effects. Let $C = (c_1, \ldots, c_{a-1})$, where $c_i = (0, 1, \ldots, -1, 0, \ldots, 0)'$ is an $ab \times 1$ vector with -1 in the

$(i+2)$th component of c_i $(i = 1, \ldots, a-1)$. Let $\gamma = (\mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_b)'$ denote the vector of parameters. We are interested in estimating $c'\gamma$, $c \in \mathcal{C}(C)$.

Let $z_i = (0, 1, \ldots, -1, 0, \ldots, 0)'$ $(i = 1, \ldots, a-1)$ be an $ab \times 1$ vector, where -1 is in the $(ib+1)$th component. Then $C = X'Z$, where $Z = (z_1, \ldots, z_{a-1})$ and X is the design matrix in the model (5.20). Let us write

$$X(X'X)^{-}C = (s_1, \cdots, s_{a-1}),$$

where

$$\begin{aligned} s_1 &= (1_b', -1_b', 0, \ldots, 0)' \\ s_2 &= (1_b', 0, -1_b', \ldots, 0)' \\ &\vdots \\ s_{a-1} &= (1_b', 0, 0, \ldots, -1_b')'. \end{aligned}$$

Then $\lambda_1 = \sigma_1^2[1 + (b-1)\rho_1], \ldots, \lambda_a = \sigma_a^2[1 + (b-1)\rho_a]$ are eigenvalues of the covariance matrix Σ in the model (5.20) and

$$\begin{aligned} J_1 &= (1_b', 0, \ldots, 0)' \\ J_2 &= (0, 1_b', \ldots, 0)' \\ &\vdots \\ J_a &= (0, 0, \ldots, 1_b')'. \end{aligned}$$

are, respectively, their orthogonal eigenvectors.

Since $\mathcal{C}(X(X'X)^{+}C) \subset \mathcal{C}(J_1, \ldots, J_a)$ it follows from (5.18) that for any $c \in \mathcal{C}(C)$

$$\text{Eff}(c'\hat{\beta}) \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}, \tag{5.21}$$

where $\lambda_1 = \max\{\lambda_i\}$ and $\lambda_n = \min\{\lambda_i\}$ $(i = 1, \ldots, a)$. The lower bound in (5.21) is much sharper than the lower bound obtained in (4.9) which is

$$\text{Eff}(c'\hat{\beta}) \geq \frac{4\tau_n\lambda_1}{(\tau_n + \lambda_1)^2},$$

where $\tau_n = \min\{\sigma_1^2(1-\rho_1), \ldots, \sigma_a^2(1-\rho_a)\}$. In particular if $\sigma_i^2 = \sigma^2$ and $\rho_i = \rho$, then the efficiency $\text{eff}(c'\hat{\beta}) = 1$, i.e., for any contrast of $\alpha_1, \ldots, \alpha_a$ the OLSE is the BLUE.

Example 5.5

Consider the model (4.1) with $\Sigma = \sigma^2 I_n$. Suppose that $\beta = (\beta_1' : \beta_2')'$, where β_1 and β_2 are $q \times 1$ and $(p-q) \times 1$ vectors and the full rank matrix $X = (X_1 : X_2)$, where X_1 and X_2 are $n \times q$ and $n \times (p-q)$ matrices, respectively. If we assume that $\beta_2 = 0$ then

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$

is the BLUE since the covariance matrix $\text{Cov}(\varepsilon) = \sigma^2 I_n$. The covariance matrix $\text{Cov}(\tilde{\beta}_1) = \sigma^2(X_1'X_1)^{-1}$. But $\tilde{\beta}_1$ is not robust against a violation of the assumption $\beta_2 = 0$. On the other hand the OLSE $\hat{\beta}_1$ of β_1, with $\text{Cov}(\hat{\beta}_1) = \sigma^2(X'X)^{11}$, where $(X'X)^{11}$ is the upper left $q \times q$ submatrix of $(X'X)^{-1}$, is robust but not efficient. Therefore, one may be interested in studying the relative efficiency of $\hat{\beta}_1$ when $\beta_2 = 0$. If $\mathcal{C}(Z_m) \subset \mathcal{C}\ (p_{i_1}, \ldots, p_{i_k})$ for some $1 \le i_1 \le \cdots \le i_k \le n$, then by (4.47)

$$\text{Cov}(\hat{\beta}_1) \le \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} \text{Cov}(\tilde{\beta}_1),$$

where $Z_m = (I_m, 0)'$ and $p_1, \ldots, p_n$ are orthonormal vectors of $X'X$ corresponding to eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$. In particular, if $\mathcal{C}(Z_m) \subset \mathcal{C}(p_1, \ldots, p_n)$ then

$$\text{Cov}(\hat{\beta}_1) \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \text{Cov}(\tilde{\beta}_1).$$

Example 5.6

Consider the Generalized Linear Model (GLM)

$$\text{E}(y_i) = \mu(\theta_i), \quad \text{Var}(y_i) = \phi\dot{\mu}(\theta_i) \quad (i = 1, \ldots, n),$$

where μ is a function on the reals and $\dot{\mu}$ is its derivative, $g(\mu(\theta_i)) = x_i'\beta$ is a link function, x_i is a $p \times 1$ vector of known values and β is a $p \times 1$ vector of unknown parameters. In this GLM, no assumption is made on the joint distribution of $y = (y_1, \ldots, y_n)'$ except that the covariance matrix of y is in block-diagonal form with small block sizes.

The weighted least squares estimator $\hat{\beta}$ of β in the GLM is defined to be a solution of

$$X'\Delta S = 0,$$

where the full rank $n \times p$ matrix $X = (x_1, \ldots, x_n)'$, $\Delta = \text{diag}\{h(x_1'\beta), \ldots, h(x_n'\beta)\}$ with $h(t) = d(g(\mu))^{-1}/dt$ and $S = (y_1 - g^{-1}(x_1'\beta), \ldots, y_n - g^{-1}(x_n'\beta))'$. Under some regularity conditions, $\hat{\beta}$ is asymtotically normal with mean β and asymptotic covariance matrix

$$\Sigma = (X'\Delta\Lambda\Delta X)^{-1}X'\Delta\text{Cov}(y)\Delta X(X'\Delta\Lambda\Delta X)^{-1},$$

where $\Lambda = \text{diag}\{\dot{\mu}(\theta_1), \ldots, \dot{\mu}(\theta_n)\}$. Let $B = \Delta\Lambda\Delta$ and $A = \Delta\Lambda[\text{Cov}(y)]^{-1}\Lambda\Delta$. If $\mathcal{C}(B^{1/2}X) \subset \mathcal{C}(p_{i_1}, \ldots, p_{i_k})$, $1 \leq i_1 \leq \cdots \leq i_k \leq n$, then by (4.45)

$$\Sigma \leq_{\text{L}} \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}(X'\Delta\Lambda[\text{Cov}(y)]^{-1}\Lambda\Delta X)^{-1},$$

where $p_1, \ldots, p_n$ are orthonormal eigenvectors of $\text{Cov}(y)$ corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. In particular

$$\Sigma \leq_{\text{L}} \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(X'\Delta\Lambda[\text{Cov}(y)]^{-1}\Lambda\Delta X)^{-1}.$$

Appendix A

Pál Schweitzer (1914): An inequality about the arithmetic mean[1]

We will prove the following theorem: If any natural numbers fall between two positive bounds, then the product of the arithmetic mean of these numbers and the arithmetic mean of reciprocals of these numbers cannot exceed the product of the arithmetic mean of their two bounds and the arithmetic mean of the reciprocals of their two bounds:

$$z = \frac{1}{n}(t_1 + \cdots + t_n) \cdot \frac{1}{n}\left(\frac{1}{t_1} + \cdots + \frac{1}{t_n}\right) \leq \frac{1}{2}(m+M) \cdot \frac{1}{2}\left(\frac{1}{m} + \frac{1}{M}\right), \qquad \text{(A.1)}$$

where $0 < m \leq t_i \leq M \quad (i = 1, \ldots, n)$.

To prove (A.1), let us consider for the moment that all the t's, with the exception of t_i, are fixed, and find at which point t_i in the interval (m, M) the function

$$z = f(t_i) = \frac{1}{n^2}(t_i + A)\left(\frac{1}{t_i} + B\right)$$

attains its maximum value. Differentiating this function, we get

$$\frac{dz}{dt_i} = \frac{1}{n^2}\left(B - \frac{A}{t_i^2}\right),$$

[1] In Hungarian: Egy egyenlötlenség az aritmetikai középértékröl, *Matematikai és Physikai Lapok* **23**, 257–261 (1914). English translation by Levente T. Tolnai and Robert Vermes.

which vanishes only at $t_i = \sqrt{A/B}$ and z has a minimum, therefore, or maximum in the interval (m, M) at either $t_i = m$ or $t_i = M$ according as

$$\frac{1}{n^2}(M + A)\left(\frac{1}{M} + B\right) \gtrless \frac{1}{n^2}(m + A)\left(\frac{1}{m} + B\right)$$

or

$$A \cdot \frac{m + M}{mM} \gtrless B.$$

For any A and B, i.e., for any $t_1, \cdots, t_{i-1}, t_{i+1}, \ldots, t_n$, if we want z to be its maximum, then we choose the value of t_i as M or m according to

$$A \cdot \frac{m + M}{mM} \gtrless B.$$

We can apply this argument to every t_i to obtain the maximum value of z as the following

$$z_{\max} = \frac{1}{n^2}\left[aM + (n - a)m\right]\left[\frac{a}{M} + \frac{n - a}{m}\right],$$

where a and $n - a$ count the number of t's equal to M and m, respectively. We may then write

$$z_{\max} \leq \frac{1}{2}(m + M) \cdot \frac{1}{2}\left(\frac{1}{m} + \frac{1}{M}\right) - \frac{1}{n^2}\left(\frac{n}{2} - a\right)^2 \frac{(m - M)^2}{mM}$$

or

$$z_{\max} \leq \frac{1}{2}(m + M) \cdot \frac{1}{2}\left(\frac{1}{m} + \frac{1}{M}\right),$$

which proves the required inequality (A.1). Equality holds in (A.1) if and only if n is even and $a = n/2$, i.e., an equal number of t's are equal to m and to M.

This inequality (A.1) can be used to establish an upper bound for the integral of reciprocal functions. Consider the numbers $t_1, \ldots, t_n$ as the values of the positive function $t = f(x)$ corresponding to the equally spaced values $x_1, \ldots, x_n$. The left-hand side of (A.1) then becomes

$$\frac{1}{n}\left[f(x_1)+f(x_2)+\cdots+f(x_n)\right]\cdot\frac{1}{n}\left[\frac{1}{f(x_1)}+\frac{1}{f(x_2)}+\cdots+\frac{1}{f(x_n)}\right],$$

which may be viewed at as an approximation to

$$\frac{1}{b-a}\int_a^b f(x)dx\cdot\frac{1}{b-a}\int_a^b\frac{dx}{f(x)}. \tag{A.2}$$

If we now take limits in (A.2), we obtain

$$\lim_{n\to\infty}\frac{1}{n^2}\left(f(x_1)+\cdots+f(x_n)\right)\cdot\left(\frac{1}{f(x_1)}+\cdots+\frac{1}{f(x_n)}\right)\le\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right), \tag{A.3}$$

where m and M represent a lower and upper bound, respectively, for the values of $f(x)$. Replacing the terms on the left-hand side of (A.3) by integrals we obtain

$$\frac{1}{(b-a)^2}\int_a^b f(x)dx\cdot\int_a^b\frac{dx}{f(x)}\le\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right)$$

and hence

$$\int_a^b\frac{dx}{f(x)}\le\frac{(b-a)^2}{\int_a^b f(x)dx}\cdot\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right).$$

If we also take into account that

$$\int_a^b\frac{dx}{f(x)}\ge\frac{(b-a)^2}{\int_a^b f(x)dx}, \tag{A.4}$$

which comes from Cauchy-Schwarz inequality:

$$\int_a^b\varphi^2dx\cdot\int_a^b\psi^2dx\ge\left(\int_a^b\varphi\psi\right)^2\quad(\varphi>0,\ \ \psi>0)$$

by setting $\varphi=\sqrt{f(x)}$ and $\psi=1/\sqrt{f(x)}$, we obtain

$$\frac{(b-a)^2}{\int_a^b f(x)dx}\le\int_a^b\frac{dx}{f(x)}\le\frac{(b-a)^2}{\int_a^b f(x)dx}\cdot\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right). \tag{A.5}$$

Since

$$\frac{1}{2}(m+M)\cdot\frac{1}{2}\left(\frac{1}{m}+\frac{1}{M}\right)=\left(\frac{\frac{m+M}{2}}{\sqrt{mM}}\right)^2,$$

i.e., the square of the ratio of the arithmetic mean and the geometric mean of m and M, is near 1 if M does not much differ from m, the integral $\int_a^b dx/f(x)$ is squeezed between two close bounds. The formula (A.3) is also useful in approximating integrals of reciprocal functions. If we take the arithmetic mean of the lower and the upper bounds, which has value

$$\frac{1}{2}\cdot\frac{(b-a)^2}{\int_a^b f(x)dx}\left[\frac{(m+M)^2}{4mM}+1\right],$$

the δ-error we make is smaller in absolute value than half of the difference between the bounds, i.e.,

$$|\delta| < \frac{(b-a)^2}{\int_a^b f(x)dx}\cdot\frac{1}{4}\cdot\left(\frac{m^2+M^2}{2mM}-1\right). \tag{A.6}$$

This formula can be used to approximate logarithms. By setting $f(x)=x$, after simplifying the corresponding value in (A.6), we find by taking logarithms that

$$\log x + \frac{2}{2x+1}\left(1+\frac{1}{8x(x+1)}\right)$$

instead of $\log x + 1$, while the absolute value of the δ- error

$$|\delta| < \frac{1}{4x(x+1)(2x+1)}.$$

If we calculate the logarithm in this way, the error we make starting at $x=10$ is smaller than

$$\frac{1}{4\cdot 10\cdot 11\cdot 21}=\frac{1}{9240}$$

and starting at $x=20$ is smaller than

$$\frac{1}{4\cdot 20\cdot 21\cdot 41}=\frac{1}{68880}.$$

Appendix B

Chun-Tu Lin (1986): An Extension of the Kantorovich Inequality and Its Application to Estimating Parameters in Linear Models[1]

B.1 Introduction

In this paper we obtain upper bounds for the expressions

$$\frac{|X'B^{-1}AB^{-1}X| \cdot |X'A^{-1}X|}{|X'B^{-1}X|^2} \quad \text{and} \quad \frac{|X'B^{-1}AB^{-1}Y| \cdot |Y'A^{-1}X|}{|X'B^{-1}X| \cdot |Y'B^{-1}Y|},$$

where A and B are $n \times n$ positive definite matrices, X and Y are arbitrary $n \times k$ matrices of rank $k < n$. In addition, some applications are discussed. The results in this paper are an extension of results by Khatri and Rao (1981) [76].

If A is an $n \times n$ symmetric positive definite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n > 0$, then for any non-zero $n \times 1$ vector x, we have

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

the Kantorovich Inequality. This inequality has already had a number of generalizations, cf. e.g., Strang (1960) [151], Greub and Rheinboldt (1959) [60], Bloomfield

[1] In Chinese: *Xitong Kexue yu Shuxue-Journal of Systems Science and Mathematical Sciences* 6 (3), 217–220 (1986). English translation by Ming-Yan Venus Chiu.

and Watson (1975) [21], and Knott (1975) [79].

Khatri and Rao (1981) generalized the Kantorovich Inequality as follows, cf. our §4.3:

$$\frac{|X'AY| \cdot |Y'A^{-1}X|}{|X'X| \cdot |Y'Y|} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}}, \tag{B.1}$$

where A is an $n \times n$ non-singular matrix, not necessarily symmetric, with $\lambda_1 \geq \cdots \geq \lambda_n > 0$, and λ_i^2 $(i = 1, \ldots, n)$, the eigenvalues of $A'A$.

Moreover

$$\frac{|X'B^2X| \cdot |X'C^2X|}{|X'BCX|} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\mu_i + \mu_{n-i+1})^2}{4\mu_i \mu_{n-i+1}}, \tag{B.2}$$

where B and C are $n \times n$ non-singular symmetric matrices so that BC is positive definite, $BC = CB$ and $\mu_1 \geq \cdots \geq \mu_n > 0$ are the roots of $|B - \mu C| = 0$.

In this paper we mainly give these two inequalities:

$$\frac{|X'B^{-1}AB^{-1}X| \cdot |X'A^{-1}X|}{|X'B^{-1}X|^2} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\mu_i + \mu_{n-i+1})^2}{4\mu_i \mu_{n-i+1}} \tag{B.3}$$

and

$$\frac{|X'B^{-1}AB^{-1}Y| \cdot |Y'A^{-1}X|}{|X'B^{-1}X| \cdot |Y'B^{-1}Y|} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\mu_i + \mu_{n-i+1})^2}{4\mu_i \mu_{n-i+1}}, \tag{B.4}$$

where A, B are positive definite matrices and $\mu_1 \geq \cdots \geq \mu_n > 0$ are the roots of $|A - \mu B| = 0$.

It is easy to see that the main result of Bloomfield and Watson [21] and Knott [79], i.e., the Bloomfield-Watson-Knott Inequality (4.9) in our Chapter 4, is the special case of (B.3) when $B = I$. Furthermore, if we replace A^{-1} by C^2, B^{-1} by BC, and use $BC = CB$, we obtain the inequality (B.2) from (B.3). Therefore (B.3) is a generalization of (B.2).

In our §B.3 below, we will point out that in estimating parameters in linear models, the inequality (B.3) is more useful than the inequality (B.2). In our §B.2 we will give a proof of the main inequalities (B.1) and (B.4) our method is much simpler than that used by Khatri and Rao (1981) [76].

B.2 Main results

In order to obtain the main results of this paper, we give two lemmas.

Lemma B.1 *If X and Y are $n \times k$ matrices of rank k, then*

$$|X'Y|^2 \leq |X'X||Y'Y|.$$

PROOF: Since $I - X(X'X)^{-1}X'$ is a symmetric idempotent matrix, where I is the identity matrix of rank n, it follows that $Y'[I - X(X'X)^{-1}X']Y$ is non-negative definite. That is $Y'Y - Y'X(X'X)^{-1}X'Y$ is non-negative definite. This yields the required result.

Lemma B.2 *If A is an $n \times n$ positive definite matrix, for any $n \times k$ matrix X of rank k, then*

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

where $\lambda_1 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of A.

PROOF: The proof can be found in Bloomfield and Watson [21] and Knott [79]

Theorem B.1 *Let A, B be $n \times n$ positive definite matrices. For any nxk matrix X of rank k, the inequality* (B.3) *holds.*

PROOF: Since B is a positive definite matrix, there exist a nonsingular matrix P with $B^{-1} = P'P$. Let $Z = PX$. Then the left hand side of (B.3) can be written as:

$$\frac{|X'P'PAP'PX| \cdot |X'P'PA^{-1}P'PX|}{|X'P'PX|^2} = \frac{|Z'(PAP')Z)| \cdot |Z'(PAP')^{-1}Z)|}{|Z'Z|^2}.$$

Notice that $\lambda_i(PAP') = \lambda_i(AP'P) = \lambda_i(AB^{-1})$ $(i = 1, \ldots, n)$, where $\lambda_i(.)$ represents the ith largest eigenvalue. Using Lemma $B.2$, the proof is accomplished.

Corollary B.1 *If C^2 replaces A^{-1} and BC replaces B^{-1}, then the inequality* (B.3) *becomes* (B.2), *where B, C are as defined just below* (B.2).

Theorem B.2 . *Let A be an $n \times n$ matrix. Then A can be written as*

$$A = P'DQ,$$

where P, Q are $n \times n$ orthogonal matrices, $D = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, $\lambda_1 \geq \cdots \lambda_n > 0$, with λ_i^2 $(i = 1, \ldots, n)$, the eigenvalues of $A'A$ and inequality (B.1) *holds.*

PROOF: From $A = P'DQ$, Lemma B.1 and Lemma B.2, we have

$$\begin{aligned} \frac{|X'AY| \cdot |Y'A^{-1}X|}{|X'X| \cdot |Y'Y|} &= \frac{|X'P'DQY| \cdot |Y'Q'D^{-1}PX|}{|X'X| \cdot |Y'Y|} \\ &\leq \frac{(|X'P'DPX||X'P'D^{-1}PX||Y'Q'DQY||Y'Q'D^{-1}QY|)^{1/2}}{|X'P'PX||Y'Q'QY|} \\ &\leq \prod_{i=1}^{\min(k,n-k)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}. \end{aligned}$$

In particular, if A is positive definite, and λ_i $(i = 1, \ldots, n)$, are the eigenvalues of A, then the inequality (B.1) holds .

Theorem B.3 *Assume A and B are $n \times n$ positive definite matrices. For any X, Y, $n \times k$ matrices of rank k, the inequality* (B.4) *holds.*

Proof: The proof is similar to the proof of Theorem B.1. Let $B^{-1} = P'P$, $Z = PX$, $W = PY$. Then the left hand side of (B.4) can be written as

$$\begin{aligned} \frac{|X'B^{-1}AB^{-1}Y| \cdot |Y'A^{-1}X|}{|X'B^{-1}X| \cdot |Y'B^{-1}Y|} &= \frac{|X'P'PAP'PY| \cdot |Y'P'PA^{-1}P'PX|}{|X'P'PX| \cdot |Y'P'PY|} \\ &= \frac{|Z'(PAP')W| \cdot |W'(PAP')^{-1}Z|}{|Z'Z| \cdot |W'W|}. \end{aligned}$$

Since $\lambda_i(PAP') = \lambda_i(AB^{-1})$ and using the special case of Theorem B.2, the inequality (B.4) is obtained.

Corollary B.2 *Let B, C be as mentioned in* (B.2), *then*

$$\frac{|X'B^2Y| \cdot |Y'C^2X|}{|X'BCX| \cdot |Y'BCY|} \leq \prod_{i=1}^{\min(k,n-k)} \frac{(\mu_i + \mu_{n-i+1})^2}{4\mu_i\mu_{n-i+1}}, \tag{B.5}$$

where $\mu_1 \geq \cdots \geq \mu_n > 0$ are the roots of $|B - \mu C| = 0$.

By simply replacing A^{-1} by C^2 and B^{-1} by BC in the inequality (B.4) we obtain (B.5) immediately. This can be viewed as a generalization of the inequality (B.2).

B.3 Applications

Let us consider the linear model

$$y = X\beta + \varepsilon, \quad \mathrm{E}(\varepsilon) = 0, \quad \mathrm{Cov}(\varepsilon) = A, \tag{B.6}$$

where y is an $n \times 1$ observed vector, X is a $n \times k$ known matrix of rank k, β is an unknown $k \times 1$ vector, and ε is an $n \times 1$ error vector. Then the Gauss-Markov estimator of β is

$$\hat{\beta} = (X'A^{-1}X)^{-1}X'A^{-1}y.$$

If the covariance matrix of model (B.6) is mistakenly taken as $\mathrm{Cov}(\varepsilon) = B$, then β might be mistakenly estimated as

$$b = (X'B^{-1}X)^{-1}X'B^{-1}y$$

Then the relative efficiency of b with respect to $\hat{\beta}$, using the ratio of generalized variances, is

$$\phi = \frac{|\mathrm{Cov}(b)|}{|\mathrm{Cov}(\hat{\beta})|} = \frac{|X'B^{-1}AB^{-1}X| \cdot |X'A^{-1}X|}{|X'B^{-1}X|^2}.$$

The inequality (B.3), therefore, gives an upper bound for this relative efficiency.

We may compare this result with that in Khatri and Rao (1981): When the covariance matrix in model (B.6) is $A^{-1} = C^2$ and the mistaken covariance matrix used is $D^{-1} = BC$, they give the upper bound for ϕ under the condition $BC = CB$. This actually requires the condition

$$AD = (BC^3)^{-1} = DA.$$

However, this condition is not satisfied in general. For example, if

$$A = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix},$$

where $0 < \rho < 1$, then $AD \neq DA$. The upper bound of ϕ which we give does not require the condition $AD = DA$.

References

[1] P. Bloomfield and G. S. Watson. The inefficiency of least squares. *Biometrika* **62**, 121–128 (1975).

[2] W. Greub and W. Rheinboldt. On a generalization of an inequality of L. V. Kantorovich. *Proceedings of the American Mathematical Society* **10**, 407–415 (1959).

[3] C. G. Khatri and C. R. Rao. Some extensions of the Kantorovich inequality and statistical applications. *Journal of Multivariate Analysis* **11**, 498–505 (1981).

[4] M. Knott. On the minimum efficiency of least squares. *Biometrika* **62**, 129–132 (1975).

[5] W. G. Strang. On the Kantorovich inequality. *Proceedings of the American Mathematical Society* **11**, 468 (1960).

Appendix C

Yong-Lin Chen (1987): A simpler proof of the Kantorovich Inequality and of a generalization of it[1]

In this paper we give a simpler elementary proof of the Kantorovich Inequality:

$$\sum a_i\lambda_i \cdot \sum a_i\lambda_i^{-1} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{C.1}$$

where $a_i > 0$ $(i = 1, \ldots, n)$, $\sum a_i = 1$, $0 < \lambda_1 \leq \cdots \leq \lambda_n$, and $\sum$ denotes $\sum_{i=1}^{n}$.

Since

$$\frac{\sqrt{\lambda_i}}{\sqrt{\lambda_1}} - \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_i}} \geq 0, \quad \frac{\sqrt{\lambda_n}}{\sqrt{\lambda_i}} - \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_n}} \geq 0 \quad (i = 1, \ldots, n),$$

we have

$$\sum \left(\frac{\sqrt{\lambda_i}}{\sqrt{\lambda_1}} - \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_i}}\right)\left(\frac{\sqrt{\lambda_n}}{\sqrt{\lambda_i}} - \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_n}}\right) a_i \geq 0. \tag{C.2}$$

Expanding (C.2) and using $\sum a_i = 1$ yield:

$$\frac{\sqrt{\lambda_n}}{\sqrt{\lambda_1}} + \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_n}} \geq \sqrt{\lambda_1\lambda_n}\sum a_i\lambda_i^{-1} + \frac{1}{\sqrt{\lambda_1\lambda_n}}\sum a_i\lambda_i, \tag{C.3}$$

and so

[1] In Chinese: *Shuxue de Shijian yu Renshi - Mathematics in Practice and Theory* **1987 (4), 78–79 (1987). English translation by Jiabin Zhao.**

$$\sqrt{\lambda_1\lambda_n}\sum a_i\lambda_i^{-1} + \frac{1}{\sqrt{\lambda_1\lambda_n}}\sum a_i\lambda_i \ \geq \ 2\left(\sqrt{\lambda_1\lambda_n}\sum a_i\lambda_i^{-1}\right)^{\frac{1}{2}}\left(\frac{1}{\sqrt{\lambda_1\lambda_n}}\sum a_i\lambda_i\right)^{\frac{1}{2}}$$

$$= \ 2\left(\sum a_i\lambda_i \cdot \sum a_i\lambda_i^{-1}\right)^{\frac{1}{2}}. \qquad \text{(C.4)}$$

Therefore (C.3) and (C.4) imply that

$$\frac{\sqrt{\lambda_n}}{\sqrt{\lambda_1}} + \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_n}} \geq 2\left(\sum a_i\lambda_i \cdot \sum a_i\lambda_i^{-1}\right)^{\frac{1}{2}}. \qquad \text{(C.5)}$$

Then it is easy to obtain (C.1) by squaring both sides of the inequality (C.5), and our proof is complete.

It is obvious that (C.1) becomes an equality if $\lambda_1 = \lambda_n$. When $0 < \lambda_1 < \lambda_n$ then (C.1) is an equality if and only if among the λ_i $(i = 1, \ldots, n)$, k of them take the value λ_1, the remainder take the value λ_n, and also $a_1 + \cdots + a_k = a_{k+1} \cdots + a_n = 1/2$, where $1 \leq k \leq n-1$. [In this paper all $a_i > 0$ $(i = 1, \ldots, n)$. -G. A.]

Furthermore, by the same method we can prove the following inequality which is a generalization of (C.1) as presented by En-Wei Shi (1985):

$$\frac{\sum p_i a_i^{\alpha} \cdot \sum p_i b_i^{\alpha}}{[\sum p_i (a_i b_i)^{\alpha/2}]^2} \leq \frac{1}{4}\left[\left(\frac{AB}{ab}\right)^{\frac{\alpha}{4}} + \left(\frac{ab}{AB}\right)^{\frac{\alpha}{4}}\right]^2, \qquad \text{(C.6)}$$

where $0 < a \leq a_i \leq A$, $\ 0 < b \leq b_i \leq B$, $p_i > 0$ $(i = 1, \ldots, n)$; $\alpha > 0$. In fact, we note that

$$\left(\frac{a_i^{\alpha/2}}{b_i^{\alpha/2}} - \frac{a^{\alpha/2}}{B^{\alpha/2}}\right)\left(\frac{b_i^{\alpha/2}}{a_i^{\alpha/2}} - \frac{b^{\alpha/2}}{A^{\alpha/2}}\right) p_i \geq 0 \quad (i = 1, \ldots, n).$$

We then have

$$\sum\left(B^{\alpha/2}a_i^{\alpha/2} - a^{\alpha/2}b_i^{\alpha/2}\right)\left(A^{\alpha/2}b_i^{\alpha/2} - b^{\alpha/2}a_i^{\alpha/2}\right) p_i \geq 0. \qquad \text{(C.7)}$$

Details of the proof are omitted.

It is also obvious that (C.6) is an equality if and only if either $a/B = A/b$ and $a_i/b_i = a/B$, $i = 1, \ldots, n$; or when $a/B \neq A/b$ there exists an integer k with $1 \leq k \leq n-1$ such that k of the a_i/b_i, $i = 1, \ldots, n$, take the value a/B, while the remaining $n-k$ take the value A/b, and

$$(Aa)^{\alpha/2}\sum p_i b_i^\alpha = (Bb)^{\alpha/2}\sum p_i a_i^\alpha.$$

Remark: From (C.6) we see that (C.7) is true under the weaker conditions:

$$p_i > 0,\ \ a_i > 0,\ \ b_i > 0\ \ (i = 1,\dots,n):\ \ \alpha > 0,\ \ A > 0,\ \ B > 0,\ \ a > 0,\ \ b > 0,$$

with

$$\frac{a}{B} \le \frac{a_i}{b_i} \le \frac{A}{b}\quad (i = 1,\dots,n).$$

the necessary and sufficient conditions for equality in (C.6) are the same as before.

Reference

[1] En-Wei Shi. An elementary proof of the Kantorovich inequality [in Chinese]. *Shuxue de Shijian yu Renshi–Mathematics in Practice and Theory* **1985** (4), 56–61 (1985).

Bibliography

[1] Gülhan Alpargu and George P. H. Styan. Some remarks and a bibliography on the Kantorovich inequality. In *Multidimensional Statistical Analysis and Theory of Random Matrices: Proceedings of the Sixth Lukacs Symposium, Bowling Green, OH, USA, 29–30 March 1996* (Arjun K. Gupta and Vyacheslav L. Girko, eds.), VSP International Science Publishers, Zeist, The Netherlands, pp. 1–13 (1996).

[2] Gülhan Alpargu and George P. H. Styan. The Kantorovich inequality and five related inequalities: 1914–1959. Paper presented at the Fifth International Workshop on Matrix Methods for Statistics, Shrewsbury, England, July 18–19, 1996.

[3] Gülhan Alpargu and George P. H. Styan. Some remarks on the Bloomfield-Watson-Knott Inequality and on some other inequalities related to the Kantorovich Inequality. Paper presented at the Conference in Honor of Shayle R. Searle on the Occasion of his Retirement, Cornell University, Ithaca, New York, August 9–10, 1996, and to appear in its *Proceedings* (1996).

[4] Gülhan Alpargu and George P. H. Styan. Research papers in Chinese on the Kantorovich and related inequalities. *Image* **17**, 12–13 (1996).

[5] Gülhan Alpargu, Josip E. Pečarić, Simo Puntanen and George P. H. Styan. *The Kantorovich Inequality: Extensions and Applications.* In preparation (1996).

[6] Theodore W. Anderson. *The Statistical Analysis of Time Series*. Reprint Edition. Wiley, New York (1994). [Cf. §10.2, pp. 560–571. Original version (1971) [MR 44:1169].]

[7] Tsuyoshi Ando. Strong quasiorder in the cone of positive definite matrices: a technical note (July 19, 1994). Unpublished paper, Research Institute for Electronic Science, Hokkaido University, Sapporo, Japan, 8 pp. (1994).

[8] D. Andrica and C. Badea. Grüss' inequality for positive linear functionals. *Periodica Mathematica Hungarica (Budapest)* **19**, 155–167 (1988) [MR 89d:26018].

[9] Jerzy K. Baksalary and Simo Puntanen. Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities. *Aequationes Mathematicae* **41**, 103–110 (1991) [MR 91k:15038].

[10] Jerzy K. Baksalary, Simo Puntanen and George P. H. Styan. On T. W. Anderson's contributions to solving the problem of when the ordinary least-squares estimator is best linear unbiased and to characterizing rank additivity of matrices. In *The Collected Papers of T. W. Anderson: 1943–1985* (George P. H. Styan, ed.), Wiley, New York, vol. 2, pp. 1579–1591 (1990).

[11] Earl R. Barnes and Alan J. Hoffman. Bounds for the spectrum of normal matrices. *Linear Algebra and Its Applications* **201**, 79–90 (1994) [MR 95c:15037].

[12] Flavio C. Bartmann and Peter Bloomfield. Inefficiency and correlation. *Biometrika* **68**, 67–71 (1981) [MR 83c:62093].

[13] D. M. Bătineţu-Giurgiu. În legătură cu inegalitatea L. V. Kantorovici [in Romanian: "Connections with the inequality of L. V. Kantorovich"]. *Gazeta Matematică (Bucharest)* **99** (2), 51–61 (1994) [Zbl 805:26014].

[14] Friedrich L. Bauer. A further generalization of the Kantorovič inequality. *Numerische Mathematik* **3**, 117–119 (1961).

[15] Friedrich L. Bauer and Alston S. Householder. Some inequalities involving the euclidean condition of a matrix. *Numerische Mathematik* **2**, 308–311 (1960) [Zbl 104:34502].

[16] Eugen Beck. Komplementäre Ungleichungen bei vergleichbaren Mittelwerten [in German]. *Monatshefte für Mathematik* **73**, 289–308 (1969) [MR 41:3691].

[17] Edwin F. Beckenbach. Review of [50]. *Mathematical Reviews* **4**, 286 (1943).

[18] Edwin F. Beckenbach. On the inequality of Kantorovich. *American Mathematical Monthly* **71**, 606–619 (1964) [MR 29:5971].

[19] Edwin F. Beckenbach and Richard Bellman. *Inequalities.* Fourth [Revised] Printing. Springer-Verlag, Berlin (1983) [MR 33:236]. [Cf. pp. 44–45. Original version (1961) [MR 28:1266].]

[20] Mieczysław Biernacki, H. Pidek and Czesław Ryll-Nardzewski. Sur une inégalité entre des intégrales définies [in French]. *Annales Universitatis Mariae Curie-Skłodowska, Sectio A: Mathematica (Lublin)* **4**, 1–4 (1950) [MR 13:118a].

[21] Peter Bloomfield and Geoffrey S. Watson. The inefficiency of least squares. *Biometrika* **62**, 121–128 (1975) [MR 51:9377].

[22] Ephraim J. Borowski and Jonathan M. Borwein. *Dictionary of Mathematics.* With the assistance of J. F. Bowers, A. Robertson and M. McQuillan. HarperCollins, London (1989) [MR 94b:00008].

[23] Wolfgang J. Bühler. Two proofs of the Kantorovich inequality and some generalizations. *Revista Colombiana de Matemáticas* **21**, 147–154 (1987) [MR 89k:15027].

[24] Peter S. Bullen, Dragoslav S. Mitrinović and Petar M. Vasić. *Means and Their Inequalities.* In English. Revised and Updated Edition. Mathematics and Its Applications: East European Series, vol. 31. D. Reidel, Dor-

drecht (1988) [MR 89d:26003]. [Cf. pp. 201–209. Original version: *Sredine i sa Njima Povezane Nejednakosti* [in Serbo-Croatian] (1977) [MR 80b:26001].]

[25] G. T. Cargo. An elementary, unified treatment of complementary inequalities. In *Inequalities-III: Proceedings of the Third Symposium on Inequalities held at the University of California, Los Angeles, September 1–9, 1969* (Oved Shisha, ed.), Academic Press, New York, pp. 39–63 (1972) [Zbl 306:26010].

[26] G. T. Cargo and Oved Shisha. Bounds on ratios of means. *Journal of Research of the National Bureau of Standards: Section B, Mathematics and Mathematical Physics* **66B**, 169–170 (1962) [Zbl 113:27102].

[27] John William Scott Cassels. Appendix 1: A proof of the inequality (3.3.2). In [163], pp. 138–139 (1951). [Published as [28] (1955).]

[28] John William Scott Cassels. Appendix: A proof of the inequality (3.2). In [165], pp. 340–341 (1955). [Original version: [27] (1951).]

[29] Sin-Chung Chang. On the product of convex combinations and the Kantorovich inequality. *Linear Algebra and Its Applications* **93**, 9–38 (1987) [MR 88d:15023].

[30] Yu-Chi Chang and Chun-Tu Lin. A generalization of Kantorovich's inequality with application to the measurement of inefficiency of least squares. Report B-84-6, Institute of Statistics, Academia Sinica, Taipei, Taiwan (1984).

[31] Jianbao Chen and T. Chen. The error ratio efficiency of the mean square in the general Gauss-Markov model. In *Proceedings of the Fourth China-Japan Symposium on Statistics* (Kunming, China), vol. 4, pp. 17–19.

[32] Yong-Lin Chen. A simpler proof of Kantorovich's inequality and of a generalization of it [in Chinese]. *Shuxue de Shijian yu Renshi–Mathematics in Practice and Theory* **1987** (4), 78–79 (1987) [MR 89c:26033]. [English translation by Jiabin Zhao is Appendix C of this thesis.]

[33] Chiou Wen Chiang. On a generalization of the Wielandt inequality [in English]. *Soochow Journal of Mathematics (Taipei)* **21**, 117–120 (1995) [MR 95m:47029].

[34] Achim Clausing. Kantorovich-type inequalities. *American Mathematical Monthly* **89**, 314, 327–330 (1982) [MR 83g:26026].

[35] Noel Cressie. M-estimation in the presence of unequal scale. *Statistica Neerlandica* **34**, 19–32 (1980) [MR 81i:62067].

[36] Daniel Culea. Inegalitățile lui P. Schweitzer și L. B. Kantorovici, aplicații [in Romanian: "Inequalities of P. Schweitzer and L. V. Kantorovich: applications"]. *Gazeta Matematică Seria Metodică (Bucharest)* no. 2, pp. 62–70 (1991).

[37] Malaya Das Gupta and Kali Charan Das. A closer look at the Kantorovich inequality. *Journal of Mathematical and Physical Sciences (Madras)* **27**, 335–346 (1993) [MR 95m:47030].

[38] Chandler Davis. Extending the Kantorovič inequality to normal matrices. *Linear Algebra and Its Applications* **31**, 173–177 (1980) [MR 81g:15020].

[39] Joaquín Basilio Diaz, and Frederic T. Metcalf. Stronger forms of a class of inequalities of G. Pólya-G. Szegö, and L. V. Kantorovich. *Bulletin of the American Mathematical Society* **69**, 415–418 (1963) [MR 26:3846].

[40] Joaquín Basilio Diaz and Frederic T. Metcalf. Complementary inequalities I: Inequalities complementary to Cauchy's inequality for sums of real numbers. *Journal of Mathematical Analysis and Applications* **9**, 59–74 (1964) [MR 30:4879]. [Published in part as [44].]

[41] Joaquín Basilio Diaz and Frederic T. Metcalf. Complementary inequalities II: Inequalities complementary to the Buniakowsky-Schwarz inequality for integrals. *Journal of Mathematical Analysis and Applications* **9**, 278–293 (1964) [MR 30:4880].

[42] Joaquín Basilio Diaz and Frederic T. Metcalf. Complementary inequalities IV: Inequalities complementary to Cauchy's inequ. lity for sums of complex numbers. *Rendiconti del Circolo Matematico di Palermo, Serie II* **13**, 291 328 (1964) [MR 33:1419].

[43] Joaquín Basilio Diaz and Frederic T. Metcalf. Complementary inequalities III: Inequalities complementary to Schwarz's inequality in Hilbert space. *Mathematische Annalen* **162**, 120–139 (1965) [MR 32:6235].

[44] Joaquín Basilio Diaz and Frederic T. Metcalf. Inequalities complementary to Cauchy's inequality for sums of real numbers. In *Inequalities: Proceedings of a Symposium held at Wright-Patterson Air Force Base, Ohio, August 19–27, 1965* (Oved Shisha, ed.). Academic Press, New York, pp. 73–77 (1967) [MR 36:5280]. [Based on the talk "Four inequalities in search of two authors" and on [40].]

[45] Joaquín Basilio Diaz, Alan J. Goldman and Frederic T. Metcalf. Equivalence of certain inequalities complementing those of Cauchy-Schwarz and Hölder. *Journal of Research of the National Bureau of Standards: Section B, Mathematics and Mathematical Physics* **68B**, 147–149 (1964) [MR 31:295].

[46] Zdeněk Dostál. Rennie's generalization of Kantorovich's inequality in a Hilbert space. *Sborník Prací Přírodovědečke Fakulty University Palackého v Olomouci, Matematika (Prague)* **19**, 101–102 (1980) [MR 82j:47025].

[47] S. W. Drury. Personal communication (1996).

[48] Morris L. Eaton. A maximization problem and its applications to canonical correlation. *Journal of Multivariate Analysis* **6**, 422–425 (1976) [MR 54:5272].

[49] Ky Fan. Some matrix inequalities. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* **29**, 185–196 (1966) [MR 35:2915].

[50] Roberto W. Frucht. Sobre algunas desigualdades: Observación relativa a la solución del Problema no. 21, indicada [143] por el Ing. Ernesto M. Saleme [in Spanish: On some inequalities: Observation concerning Problem no. 21, proposed [143] by Ernesto M. Saleme]. *Mathematicae Notae: Boletin del Instituto de Matemática "Beppo Levi" (Rosario)* **3**, 41–46 (1943) [MR 4:268, Zbl 060.14901]. [Includes an untitled appendix by Beppo Levi [83].]

[51] Masatoshi Fujii, Eizaburo Kamei and Akemi Matsumoto. Parameterized Kantorovich inequality for positive operators. *Nihonkai Mathematical Journal* **6**, 129–134 (1995).

[52] Dao De Gao. Estimation of the deviation between the least squares and the best linear unbiased estimators of the mean vector in a linear model [in Chinese]. *Yingyong Gailü Tongji–Chinese Journal of Applied Probability and Statistics* **8** (4), 391–397 (1992) [MR 93m:62150].

[53] Dao De Gao. A new class of Kantorovich-type inequalities and their applications in statistics [in Chinese]. *Xitong Kexue yu Shuxue–Journal of Systems Science and Mathematical Sciences* **13** (4), 331–337 (1993) [MR 94k:62102].

[54] Dao De Gao. Some extensions of the Kantorovich inequality and their applications in statistics [in Chinese]. *Gaoxiao Yingyong Shuxue Xuebao–A Journal of Chinese Universities, Series A (Hangzhou)* **10** (2), 181–188 (1995).

[55] Dao De Gao and Jing Long Wang. Efficiency of generalized least square estimates [in Chinese]. *Xitong Kexue yu Shuxue–Journal of Systems Science and Mathematical Sciences* **10** (2), 125–130 (1990) [MR 96c:62059].

[56] Şerban A. Gheorghiu. Note sur une inégalité de Cauchy [in French]. *Bulletin Mathématique de la Société Roumaine des Sciences (Bucharest)* **35**, 117–119 (1933) [Zbl 008:34504].

[57] Alan J. Goldman. A generalization of Rennie's inequality. *Journal of Research of the National Bureau of Standards: Section B. Mathematics and Mathematical Physics* **68B**, 59–63 (1964) [MR 29:5970].

[58] Allen A. Goldstein. A modified Kantorovich inequality for the convergence of Newton's method. In *Mathematical Developments arising from Linear Programming: Proceedings of the Joint Summer Research Conference, Bowdoin College, Brunswick, Maine, June 25–July 1, 1988* (Jeffrey C. Lagarias and Michael J. Todd, eds.), *Contemporary Mathematics*, vol. 114, pp. 285–294 (1990) [MR 92d:90078].

[59] Gene H. Golub. Comparison of the variance of minimum variance and weighted least squares regression coefficients. *The Annals of Mathematical Statistics* **34**, 984–991 (1963).

[60] Werner Greub and Werner Rheinboldt. On a generalization of an inequality of L. V. Kantorovich. *Proceedings of the American Mathematical Society* **10**, 407–415 (1959) [MR 21:37744].

[61] Gerhard Grüss. Über das Maximum des absoluten Betrages von $\frac{1}{b-a}\int_a^b f(x)g(x)dx - \frac{1}{(b-a)^2}\int_a^b f(x)dx \cdot \int_a^b g(x)dx$ [in German]. *Mathematische Zeitschrift* **39**, 215–226 (1934) [Zbl 010:01602].

[62] Shelby J. Haberman. How much do Gauss-Markov and least-square estimates differ? A coordinate-free approach. *The Annals of Statistics* **3**, 982–990 (1975) [MR 51:14437].

[63] Edward James Hannan. *Time Series Analysis*. Methuen, London (1960) [MR 22:5105]. [Cf. pp. 111–113.]

[64] Edward James Hannan. *Multiple Time Series*. Wiley, New York (1970) [MR 43:5673]. [Cf. pp. 420–423.]

[65] Zhi-Chuan Hao. Generalization and applications of Kantorovich inequality. Unpublished manuscript, Economic Information Department of Gui Zhou

Finance and Economic Institute of China. 13 pp. (c. 1992). [Results published, in part, in [111], pp. 122–123.]

[66] Godfrey Harold Hardy, John Edensor Littlewood and George Pólya. *Inequalities*. Second Edition. Cambridge University Press (1952) [MR 89d:26016]. [Cf. # 71. p. 62. Original version (1934); first paperback edition (1988), reprinted (1994).]

[67] John Z. Hearon. A generalized matrix version of Rennie's inequality. *Journal of Research of the National Bureau of Standards: Section B, Mathematics and Mathematical Physics* **71B**, 61–64 (1967) [MR 37:232].

[68] Peter Henrici. Two remarks on the Kantorovich inequality. *American Mathematical Monthly* **68**, 904–906 (1961) [MR 25:91].

[69] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Corrected Reprint Edition. Cambridge University Press (1990) [MR 91i:15001]. [Cf. pp. 442–445, 451–452. Original version (1985) [MR 87e:15001].]

[70] Alston S. Householder. The Kantorovich and some related inequalities. *SIAM Review* **7**, 463–473 (1965) [MR 32:6656].

[71] Alston S. Householder. *The Theory of Matrices in Numerical Analysis*. Reprint Edition. Dover, New York (1974) [MR 51:14539]. [Cf. §3.4, pp. 81–85. Original version: Blaisdell, New York (1964) [MR 30:5475].]

[72] Leonid Vital'evich Kantorovich. Funkcional'nyi analiz i prikladnaya matematika [in Russian: Functional analysis and applied mathematics]. *Uspekhi Matematičeskiĭ Nauk, Novaya Seriya* **3**, no. 6 (28), 89–185 (1948). [Cf. pp. 142–144. Translated into English as [73].]

[73] Leonid Vital'evich Kantorovich. Functional Analysis and Applied Mathematics [in English]. Translated from the Russian [72] by Curtis D. Benster and edited by George E. Forsythe. NBS Report no. 1509 (1101-10-5100), U.

S. Dept. of Commerce, National Bureau of Standards, Los Angeles (1952) [MR 14:766d]. [Cf. pp. 106–109.]

[74] Louis Kates, Serge Kruk and Henry Wolkowicz. Solution to Problem 16-4: Algebraic reverse of a convex matrix inequality. *Image* **17**, 31 (1996).

[75] Chinubhai Ghelabhai Khatri. An inequality concerning the difference of variances of two different estimates. *Mathematics Today (Ahmedabad)* **2**, 29–32 (1984) [MR 87g:62104].

[76] Chinubhai Ghelabhai Khatri and C. Radhakrishna Rao. Some extensions of the Kantorovich inequality and statistical applications. *Journal of Multivariate Analysis* **11**, 498–505 (1981) [MR 84e:15006].

[77] Chinubhai Ghelabhai Khatri and C. Radhakrishna Rao. Some generalizations of Kantorovich inequality. *Sankhyā Series A* **44**, 91–102 (1982) [MR 85m:15005].

[78] Konrad Knopp. Über die maximalen Abstände und Verhältnisse verschiedener Mittelwerte [in German]. *Mathematische Zeitschrift* **39**, 768–776 (1935) [Zbl 011:01203].

[79] Martin Knott. On the minimum efficiency of least squares. *Biometrika* **62**, 129–132 (1975) [MR 51:9378].

[80] Mark Aleksandrovich Krasnosel'skiĭ and Selim Griogor'evich Kreĭn. An iteration process with minimal residuals [in Russian]. *Matematičeskiĭ Sbornik, Novaya Seriya* **31** (73), 315–334 (1952) [MR 14:692d].

[81] Hiroshi Kurata and Takeaki Kariya. LUB for the covariance matrix of a GLSE in regression with applications to an SUR model and a heteroscedastic model. Paper presented at the Conference on Inference and Applications: 25th Anniversary Celebration of the School of Statistics, University of Minnesota, Minneapolis, June 21–22, 1996. Unpublished preprint, 16 pp. (1996).

[82] József Kürschák. Kitűzött feladat [in Hungarian: "Posed problem"]. *Matematikai és Physikai Lapok* **23**, 378 (1914).

[83] Beppo Levi. Untitled appendix [signed B.L.] in [50], pp. 45–46 (1943).

[84] Chun-Tu Lin. Extrema of quadratic forms and statistical applications. *Communications in Statistics-Theory and Methods* **13**, 1517–1520 (1984) [MR 86a:62083].

[85] Chun-Tu Lin. An extension of the Kantorovich inequality and its application to estimating parameters in linear models [in Chinese]. *Xitong Kexue yu Shuxue-Journal of Systems Science and Mathematical Sciences* **6** (3), 217–220 (1986) [MR 88a:62176]. [English translation by Ming-Yan Venus Chiu is Appendix B of this thesis.]

[86] Chun-Tu Lin. Extremes of determinants and optimality of canonical variables. *Communications in Statistics-Simulation and Computation* **19**, 1415–1430 (1990) [MR 91m:62091].

[87] Chun-Tu Lin. Extremes of ratios of determinants and canonical correlation variables [in English]. *Acta Mathematicae Applicatae Sinica (English Series)* **7** (3), 272–278 (1991) [MR 93e:62154].

[88] Shuangzhe Liu. Cauchy and Kantorovich inequalities involving sums of matrices. Unpublished paper, Institute of Actuarial Science and Econometrics, University of Amsterdam, 5 pp. (c. 1995).

[89] Shuangzhe Liu. *Contributions to Matrix Calculus and Applications in Econometrics*. Tinbergen Institute Research Series, no. 106. Thesis Publishers, Amsterdam (1995). [Cf. pp. 25–58.]

[90] Shuangzhe Liu. Problem 16-4: Reverse of a convex matrix inequality. *Image* **16**, 32 (1996).

[91] Shuangzhe Liu and Heinz Neudecker. Kantorovich inequalities and efficiency of estimators useful in panel data analysis. Technical Report no. AE 17/94, Institute of Actuarial Science and Econometrics, University of Amsterdam, 14 pp. (1994).

[92] Shuangzhe Liu and Heinz Neudecker. Kantorovich inequalities involving positive semidefinite matrices, and efficiency comparisons in linear models. Unpublished paper, Institute of Actuarial Science and Econometrics, University of Amsterdam, 21 pp. Submitted for publication (1995). [Presented at the Fourth International Workshop on Matrix Methods for Statistics (Montréal, July 1995), abstract no. MM-31.]

[93] Shuangzhe Liu and Heinz Neudecker. Several matrix Kantorovich-type inequalities, *Journal of Mathematical Analysis and Applications* **197**, 23–26 (1996).

[94] C. J. Lombard. Union-intersection tests for sphericity. *South African Statistical Journal* **17**, 165–175 (1983) [MR 86b:62089].

[95] Yu-Tong Lou. A generalization of Schweitzer's inequality and Grüss's inequality and their relationship [in Chinese]. *Qufu Shifan Daxue Xuebao Ziran Kexue Ban* **17** (4), 24–28 (1991).

[96] Alexandru Lupaş. A remark on the Schweitzer and Kantorovich inequalities. *Publikacije Elektrotehničkog Fakulteta Univerziteta u Beogradu, Serija Matematika i Fizika* **383**, 13–15 (1972) [MR 49:5278].

[97] T. A. Magness and J. B. McGuire. Comparison of least squares and minimum variance estimates of regression parameters. *The Annals of Mathematical Statistics* **33**, 462–470 (1961).

[98] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Revised Reprint Edition. Wiley, Chichester (1991) [MR 89g:15001]. [Cf. # 17, p. 238. Original version (1988).]

[99] Endre Makai. On generalizations of an inequality due to Pólya and Szegö. *Acta Mathematica Hungarica* **12**, 189–191 (1961).

[100] Marvin Marcus. A unified exposition of some classical matrix theorems. *Linear and Multilinear Algebra* **25**, 137–147 (1989) [MR 91a:15004].

[101] Marvin Marcus and Afton H. Cayford. Further results on the Kantorovich inequality [Abstract]. *Notices of the American Mathematical Society* **9**, 300 (1962).

[102] Marvin Marcus and Afton H. Cayford. Equality in certain inequalities. *Pacific Journal of Mathematics* **13**, 1319–1329 (1963) [MR 28:168].

[103] Marvin Marcus and Nisar A. Khan. Some generalizations of Kantorovich's inequality. *Portugaliae Mathematica* **20**, 33–38 (1961) [MR 24:A127].

[104] Marvin Marcus and Henryk Minc. *A Survey of Matrix Theory and Matrix Inequalities*. [Corrected] Reprint Edition. Dover, New York (1992). [Cf. §3.5.2, p. 110; §4.3, p.117. Original version: Allyn and Bacon, Boston (1964) [MR 29:112]; reprinted (with corrections): Prindle, Weber and Schmidt, Boston (1969).]

[105] Albert W. Marshall and Ingram Olkin. Reversal of the Lyapunov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich inequality. *Journal of Mathematical Analysis and Applications* **8**, 503–514 (1964) [MR 30:4881].

[106] Albert W. Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York (1979) [MR 81b:00002]. [Cf. # D.3.a, p. 71. Addenda by James V. Bondar: *Linear Algebra and Its Applications* **199**, 115–130.]

[107] Albert W. Marshall and Ingram Olkin. Matrix versions of the Cauchy and Kantorovich inequalities. *Aequationes Mathematicae* **40**, 89–93 (1990) [MR 91f:15045].

[108] Günter Meinardus. Über eine Verallgemeinerung einer Ungleichung von L. V. Kantorowitsch [in German]. *Numerische Mathematik* **5**, 14–23 (1963) [MR 28:3525].

[109] Lucia Migliaccio and Luciana Nania. The Kantorovich inequality under integral constraints. *Journal of Mathematical Analysis and Applications* **181**, 524–530 (1994) [MR 94m:26021].

[110] Dragoslav S. Mitrinović. *Analytic Inequalities*. In cooperation with Petar M. Vasić. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, vol. 165, Springer-Verlag, New York (1970) [MR 43:448]. [Cf. pp. 59–66. Addenda by Dragoslav S. Mitrinović and Petar M. Vasić: *Publikacije Elektrotehničkog Fakulteta Univerziteta u Beogradu, Serija Matematika i Fizika* **577–598**, 3–10 (1977) [MR 56:5818], and by Dragoslav S. Mitrinović, Ivan B. Lacković and Miomir S. Stanković: *ibid* **634–677**, 3–24 (1979) [MR 82c:26011].]

[111] Dragoslav S. Mitrinović, Josip E. Pečarić, and Arlington M. Fink (1993). *Classical and New Inequalities in Analysis*. Mathematics and Its Applications: East European Series, vol. 61, Kluwer, Dordrecht (1993) [MR 94c:00004]. [Cf. pp. 121–125, 222, 684–685.]

[112] Bertram Mond. A matrix version of Rennie's generalization of Kantorovich's inequality. *Proceedings of the American Mathematical Society* **16**, 1131 (1965) [MR 32:2423].

[113] Bertram Mond. A matrix inequality including that of Kantorovich. *Journal of Mathematical Analysis and Applications* **13**, 49–52 (1966) [MR 32:2423].

[114] Bertram Mond and Josip E. Pečarić. Matrix versions of some means inequalities. *Australian Mathematical Society Gazette* **20**, 117–120 (1993) [MR 94h:26025].

[115] Bertram Mond and Josip E. Pečarić. Inequalities with weights for powers of generalized inverses. *Bulletin of the Australian Mathematical Society* **48**, 7–12 (1993) [MR 94d:15002].

[116] Bertram Mond and Josip E Pečarić. A matrix version of the Ky Fan generalization of the Kantorovich inequality. *Linear and Multilinear Algebra* **36**, 217–221 (1994) [MR 95j:15017].

[117] Bertram Mond and Josip E. Pečarić. Inequalities involving powers of generalized inverses. *Linear Algebra and Its Applications* **199**, 293–303 (1994) [MR 95a:15004].

[118] Bertram Mond and Josip E. Pečarić. A matrix version of the Ky Fan generalization of the Kantorovich inequality II. *Linear and Multilinear Algebra* **38**, 309–313 (1995).

[119] Bertram Mond and Josip E. Pečarić. Generalization of a matrix inequality of Ky Fan. *Journal of Mathematical Analysis and Applications* **190**, 244–247 (1995) [MR 95m:15029].

[120] Bertram Mond and Oved Shisha. Ratios of means and applications. In *Inequalities: Proceedings of a Symposium held at Wright-Patterson Air Force Base, Ohio, August 19–27, 1965* (Oved Shisha, ed.), Academic Press, New York, pp. 191–197 (1967) [MR 39:2930a].

[121] Bertram Mond and Oved Shisha. Difference and ratio inequalities in Hilbert space. In *Inequalities-II: Proceedings of the Second Symposium on Inequalities held at the United States Air Force Academy, Colorado, August 14–22, 1967* (Oved Shisha, ed.), Academic Press, New York, pp. 241–249 (1970).

[122] Masahiro Nakamura. A remark on a paper of Greub and Rheinboldt. *Proceedings of the Japan Academy, Series A: Mathematical Sciences (Tokyo)* **36**, 198–199 (1960) [MR 22:12188].

[123] Heinz Neudecker. Personal communication (1996).

[124] Heinz Neudecker and Shuangzhe Liu. Alternative matrix versions of Kantorovich inequalities. Technical Report no. AE 5/94, Institute of Actuarial Science and Econometrics, University of Amsterdam, 7 pp., 15 January 1994.

[125] Morris Newman. Kantorovich's inequality. *Journal of Research of the National Bureau of Standards: Section B, Mathematics and Mathematical Physics* **64B**, 33–34 (1959).

[126] Ingram Olkin. Review of [48]. *Mathematical Reviews* **54**, no. 5272, p. 752 (1977).

[127] Ingram Olkin. Solution to Problem 16-4: Algebraic reverse of a convex matrix inequality. *Image* **17**, 31 (1996).

[128] Nobuo Ozeki. On the estimation of the inequalities by the maximum or minimum values [in Japanese]. *Journal of the College of Arts and Sciences, Chiba University* **5**, (1968) no. 2, 199–203 (1969) [MR 40:7408].

[129] Josip E. Pečarić and Bertram Mond. A matrix inequality including that of Kantorovich-Hermite II. *Journal of Mathematical Analysis and Applications* **168**, 381–384 (1992) [MR 93m:15031].

[130] Josip E. Pečarić and Bertram Mond. An inequality for operators in a Hilbert space, III. *Journal of Mathematical Analysis and Applications* **183**, 385–390 (1994) [MR 95g:47029].

[131] Josip E. Pečarić, Simo Puntanen and George P. H. Styan. Some further matrix extensions of the Cauchy-Schwarz and Kantorovich inequalities, with some statistical applications. *Linear Algebra and Its Applications* **237/238**, 455–476 (1996).

[132] W. V. Petryshyn. Direct and iterative methods for the solution of linear operator equations in Hilbert space. *Transactions of the American Mathematical Society* **105**, 136–175 (1962) [MR 26:3180].

[133] George Pólya and Gábor Szegö. *Aufgaben und Lehrsätze aus der Analysis, Band I: Reihen, Integralrechnung, Funktiontheorie* [in German]. Fourth Edition. Heidelberger Taschenbücher, vol. 73. Springer-Verlag, New York (1970). [Cf. # 92–93, pp. 57, 213–214. Original version: Julius Springer, Berlin (1925). Reprint: Dover, New York (1945) [MR7:418e]. Translated into English as [134] (1972).]

[134] George Pólya and Gábor Szegö. *Problems and Theorems in Analysis, Volume I: Series, Integral Calculus, Theory of Functions* [in English]. Translated from the German [133] by Dorothee Aeppli. Corrected printing of the revised translation of the fourth German edition. Grundlehren der Mathematischen Wissenschaften, vol. 193.Springer-Verlag, New York (1978) [MR 81e:00002]. [Cf. # 92–93, pp. 71–72, 253–255.]

[135] Vlastimil Pták. The Kantorovich inequality. *American Mathematical Monthly* **102**, 820–821 (1995) [MR 96f:26025].

[136] Simo Puntanen. On the relative goodness of ordinary least squares estimation in the general linear model. *Acta Universitatis Tamperensis, Series A* **216**, vi + 132 pp. (1987) [MR 88d:62115]. [Cf. Paper [1], §2.]

[137] Simo Puntanen and George P. H. Styan. The equality of the ordinary least squares estimator and the best linear unbiased estimator [with discussion]. *The American Statistician* **43**, 153–164 (1989) [MR 92e:62125].

[138] M. Raghavachari. A linear programming proof of Kantorovich's inequality. *The American Statistician* **40**, 136–137 (1986).

[139] C. Radhakrishna Rao. The inefficiency of least squares: extensions of the Kantorovich inequality. *Linear Algebra and Its Applications* **70**, 249–255 (1985) [MR 87d:62131].

[140] Basil C. Rennie. An inequality which includes that of Kantorovich. *American Mathematical Monthly* **70**, 982 (1963).

[141] Basil C. Rennie. On a class of inequalities. *Journal of the Australian Mathematical Society* **3**, 442–448 (1963) [MR 29:3590].

[142] Peter D. Robinson and Andrew J. Wathen. Variational bounds on the entries of the inverse of a matrix. *IMA Journal of Numerical Analysis* **12**, 463–486 (1992) [MR 93g:65044].

[143] Ernesto M. Saleme. Problema No. 21 [in Spanish]. *Mathematicae Notae: Boletin del Instituto de Matemática "Beppo Levi" (Rosario)* **2**, 197–199 (1942).

[144] Andreas H. Schopf. On the Kantorovich inequality. *Numerische Mathematik* **2**, 344–346 (1960) [MR 22:4723]. [Paper written by A. S. Householder after the death of Andreas H. Schopf on September 29, 1959.]

[145] Pál Schweitzer. Egy egyenlőtlenség az aritmetikai középértékről [in Hungarian: "An inequality about the arithmetic mean"]. *Matematikai és Physikai Lapok* **23**, 257–261 (1914) [JFM 45:1245]. [English translation by Levente T. Tolnai and Robert Vermes is Appendix A of this .]

[146] En-Wei Shi. An elementary proof of the Kantorovich inequality [in Chinese]. *Shuxue de Shijian yu Renshi–Mathematics in Practice and Theory* **1985** (4), 56–61 (1985) [MR 87g:15021].

[147] Oved Shisha and Bertram Mond. Differences of means. *Bulletin of the American Mathematical Society* **73**, 328–333 (1967) [MR 35:3024].

[148] Oved Shisha and Bertram Mond. Bounds on differences of means. In *Inequalities: Proceedings of a Symposium held at Wright-Patterson Air Force Base, Ohio, August 19–27, 1965* (Oved Shisha, ed.), Academic Press, New York, pp. 293–308 (1967) [MR 39:2930b].

[149] P. G. Spain. Operator versions of the Kantorovich inequality. *Proceedings of the American Mathematical Society* **124**, 2813–2819 (1996).

[150] James C. Spall. The Kantorovich inequality for error analysis of the Kalman filter with unknown noise distributions. *Automatica, A Journal of the International Federation of Automatic Control* **31**, 1513–1517 (1995) [MR 96e:93067].

[151] W. Gilbert Strang. On the Kantorovich inequality. *Proceedings of the American Mathematical Society* **11**, 468 (1960) [MR 22:2904].

[152] George P. H. Styan. On some inequalities associated with ordinary least squares and the Kantorovich inequality. In *Festschrift for Eino Haikala on his Seventieth Birthday, Acta Universitatis Tamperensis, Series A*, vol. 153, pp. 158–166 (1983) [MR 85g:62120].

[153] George P. H. Styan. Personal communication (1996).

[154] George P. H. Styan and Sanjo Zlobec. An inequality connected with the efficiency of the least-squares estimator [Abstract]. *The IMS Bulletin* **11**, 192 (1982). [Paper presented at the 181st IMS Meeting, San Diego, California, June 1982.]

[155] Ji-guang Sun. Extensions of the Kantorovich inequality and the Bauer-Fike inequality. *Journal of Computational Mathematics* **9**, 360–368 (1991) [MR 93e:65074].

[156] William N. Venables. Some implications of the union-intersection principle for tests of sphericity. *Journal of Multivariate Analysis* **6**, 185–190 (1976) [MR 54:3971].

[157] Chung-Lie Wang. Variants of the Hölder inequality and its inverses. *Canadian Mathematical Bulletin* **20**, 377–384 (1977) [MR 57:9929].

[158] Chung-Lie Wang. On development of inverses of the Cauchy and Hölder inequalities. *SIAM Review* **21**, 550–557 (1979) [MR 80g:26023].

[159] Jing-Long Wang and Dao De Gao. Measures of inefficiency for generalized least square estimates of the Euclidean norms of matrices [in Chinese]. *Yingyong Gailü Tongji-Chinese Journal of Applied Probability and Statistics* 7 (4), 361–366 (1991) [MR 93m:62160].

[160] Song-Gui Wang and Shein-Chung Chow. *Advanced Linear Models: Theory and Applications*. Marcel Dekker, New York (1994) [MR 96c:62123]. [Cf. pp. 44–46, 212–215.]

[161] Song-Gui Wang and Jun Shao. Constrained Kantorovich inequalities and relative efficiency of least squares. *Journal of Multivariate Analysis* **42**, 284–298 (1992) [MR 93j:62179].

[162] Song-Gui Wang and Hu Yang. Kantorovich-type inequalities and the measures of inefficiency of the GLSE. *Acta Mathematicae Applicatae Sinica* **5** (4), 372–381 (1989) [MR 91d:62104].

[163] Geoffrey Stuart Watson. Serial Correlation in Regression Analysis. Ph.D. thesis, Dept. of Experimental Statistics, North Carolina State College, Raleigh (1951). [Includes an appendix by J. W. S. Cassels [27], pp. 138–139. Results presented at the Econometric Society Conference, Louvain, Belgium, September 1951. Reprinted as [164]; published as [165] and [171].]

[164] Geoffrey Stuart Watson. Serial correlation in regression analysis. University of North Carolina, Mimeograph Series, no. 49 (1952). [Reprint of [163]; published as [165] and [171].]

[165] Geoffrey S. Watson. Serial correlation in regression analysis, I. *Biometrika* **42**, 327–341 (1955) [MR 17:382a]. [Includes an appendix by J. W. S. Cassels [28]. Part of [163] and [164].]

[166] Geoffrey S. Watson. Linear least squares regression. *The Annals of Mathematical Statistics* **38**, 1679–1699 (1967).

[167] Geoffrey S. Watson. *Selected Topics in Statistical Theory*. Mimeographed. Mathematical Association of America, Washington, DC (1971). [Cf. pp. 87–92. "Notes on lectures given at the 1971 MAA Summer Seminar, Williams College, Williamstown, Massachusetts."]

[168] Geoffrey S. Watson. A method for discovering Kantorovich-type inequalities and a probabilistic interpretation. *Linear Algebra and Its Applications* **97**, 211–217 (1987) [MR 89a:15018].

[169] Geoffrey S. Watson. Comments on some inequalities and the efficiency of least squares. Unpublished paper, Dept. of Mathematics, Princeton University, 9 pp. Submitted for publication (1995). [Presented at the Fourth International Workshop on Matrix Methods for Statistics (Montréal, July 1995), abstract no. MM-53.]

[170] Geoffrey S. Watson. Personal communication (1996).

[171] Geoffrey S. Watson and E. J. Hannan. Serial correlation in regression analysis, II. *Biometrika* **42**, 436–448 (1956) [MR 19:694h]. [Part of [163] and [164].]

[172] Helmut Wielandt. Inclusion theorems for eigenvalues. In *Simultaneous Linear Equations and the Determination of Eigenvalues* (Lowell J. Paige and Olga Taussky, eds.), National Bureau of Standards Applied Mathematics Series, U.S. Government Printing Office, Washington, D.C., vol. 29, pp. 75–78 (1953) [MR 15:496d].

[173] H. S. Witsenhausen. An inequality of Kantorovitch type for Stieltjes integrals. *Journal of Mathematical Analysis and Applications* **28**, 590–593 (1969) [MR 40:1839].

[174] Henry Wolkowicz. Bounds for the Kantorovich ratio. Research Report. The University of Alberta, Edmonton, 15 pp. (1981).

[175] Hu Yang. Extensions of the Kantorovich inequality and the error ratio efficiency of the mean square [in Chinese]. *Mathematica Applicata Yingyong Shuxue-Chinese Journal of Mathematics and its Applications* **1** (4), 85–90 (1988) [MR 90a:62186]. [First draft available of a translation into English by Xiao-Jie Zhou.]

[176] Hu Yang. The inefficiency of the least squares estimator and its bound [in Chinese]. *Applied Mathematics and Mechanics of China* **11** (11), 1087–1093 (1990).

[177] Hu Yang. A brief proof on the generalized [*sic*] variance bound of the relative efficiency in statistics. *Communications in Statistics-Theory and Methods* **19**, 4587–4590 (1990).

[178] Hu Yang. Two new classes of the generalized Kantorovich inequalities and their applications [in Chinese]. *Kexue Tongbao* **35**, 965–967 (1990) [MR 92m:15016]. [Translated into English as [181].]

[179] Hu Yang. Kantorovich Inequality and its Applications in Statistics [in Chinese]. M.Sc. thesis, Dept. of Statistics, Yunnan University, Kunming, 52 pp. (1990).

[180] Hu Yang. Some extensions on the upper bound of the Rayleigh quotient [in Chinese]. *Journal of Chongqing University* **9**, 16–23 (1990).

[181] Hu Yang. Two new classes of the generalized Kantorovich inequalities and their applications [in English, translated from the Chinese [178].]. *Chinese Science Bulletin* **36**, 1849–1851 (1991) [MR 92m:15016].

[182] Hu Yang. Matrix norm versions of the Kantorovich inequality and its applications [in English]. *Applied Mathematics: A Journal of Chinese Universities, Series B* **10B** (2), 133–140 (1995) [MR 96e:15035]. [Chinese summary: *Gaoxiao Yingyong Shuxue Xuebao–A Journal of Chinese Universities, Series A (Hangzhou)* **10** (2), 233 (1995).]

[183] Hu Yang. Efficiency matrix and the partial ordering of estimate. *Communications in Statistics-Theory and Methods* 25, 457–468 (1996).

[184] Hu Yang and Song-Gui Wang. Condition numbers, the spectral norm and precision of estimators [in Chinese]. *Yingyong Gailü Tongji-Chinese Journal of Applied Probability and Statistics* 7, 337–343 (1991).

[185] Lina Yeh. A note on Wielandt's inequality. *Applied Mathematics Letters* 8 (3), 29–31 (1995) [MR 96e:15029].

[186] Jinlong Zhan and Jianbao Chen. The inefficiency of least squares in Gauss-Markov and variance component models. Unpublished paper, Dept. of Basic Sciences, Kunming Institute of Technology, 12 pp. (1994).

[187] Shi-Qiong Zhou. On the weighted mean inequality [in Chinese]. *Changsha Jiaotong Xueyuan Xuebao-Journal of Changsha Communications Institute* 7 (3), 52–61 (1991) [MR 95g:26036].

[188] Shi-Qiong Zhou. The inverses of the Hölder inequality and their matrix forms [in Chinese]. *Changsha Jiaotong Xueyuan Xuebao-Journal of Changsha Communications Institute* 9, 10–16 (1993) [MR 95g:26036].

References Not Connected Directly to the Kantorovich Inequality

[189] Tsuyoshi Ando. Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra and Its Applications* **26**, 203–241 (1979) [MR 80f:15023].

[190] Tsuyoshi Ando. On the arithmetic-geometric-harmonic-mean inequalities for positive definite matrices. *Linear Algebra and Its Applications* **52/53**, 31–37 (1983) [MR 84j:15016].

[191] Sheryl Anne Bartlett. The Analysis of Two-way Cross-classified Unbalanced Data. M.Sc. thesis, Dept. of Mathematics, McGill University, Montréal (1980).

[192] Rajendra Bhatia and Chandler Davis. More matrix forms of the arithmetic-geometric mean inequality. *SIAM Journal on Matrix Analysis and Applications* **14**, 132–136 (1993) [MR 94b:15017].

[193] Rajendra Bhatia and Fuad Kittaneh. On the singular values of a product of operators. *SIAM Journal on Matrix Analysis and Applications* **11**, 272–277 (1990) [MR 90m:47033].

[194] Alfred Brauer and A. C. Mewborn. The greatest distance between two charcteristic roots of a matrix. *Duke Mathematical Journal* **26**, 653–661 (1959) [MR 22:10997].

[195] Manish C. Chakrabarti. On the C-matrix in design of experiments. *Journal of the Indian Statistical Association* **1**, 8–23 (1963) [MR 31:2804].

[196] Salvatore Coen. The life of Beppo Levi [in Italian]. In *Seminari di Geometria, 1991–1993* (Salvatore Coen, ed.), Universita degli Studi di Bologna, Dipartimento di Matematica, Bologna, pp. 193–232 (1994) [MR 95c:01034].

[197] Carlos Gonzalez de la Fuente. Roberto W. Frucht: The mathematician, the teacher, the man. *Scientia, Series A: Mathematical Sciences, Universidad Tecnica Federico Santa Maria, Valparaiso* **1**, iii–v (1988) [Zbl 679.01006].

[198] Frank Harary. Homage to Roberto Frucht. *Journal of Graph Theory* **6**, 97–99 (1982) [MR 83i:01065].

[199] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Corrected Reprint Edition. Cambridge University Press (1990) [MR 91ie:15001]. [Original version (1985) [MR 87e:15001].]

[200] Valeriĭ Leonidovich Makarov and Sergeĭ L'vovich Sobolev. Academician L. V. Kantorovich (19 January 1912 to 7 April 1986) [in English: translated from the Russian by Lev J. Leifman]. In *Functional Analysis, Optimization, and Mathematical Economics: A Collection of Papers Dedicated to the Memory of Leonid Vital'evich Kantorovich* (Lev J. Leifman, ed.), Oxford University Press, pp. 1–7 (1990) [MR 91i:01098].

[201] Norbert Schappacher and René Schoof. Beppo Levi and the arithmetic of elliptic curves. *The Mathematical Intelligencer* **18** (1), 57–69 (1996).

[202] Alastair J. Scott and George P. H. Styan. On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys. *Linear Algebra and Its Applications* **70**, 209–224 (1985) [MR 87i:62100].

[203] Yves Thibaudeau and George P. H. Styan. Bounds for Chakrabarti's measure of imbalance in experimental design. In *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications, Tampere, Finland, August 30–September 2, 1983* (T. Pukkila and S. Puntanen, eds.), Dept. of Mathematical Sciences, University of Tampere, Tampere, Finland, pp. 323–347 (1985).

[204] John W. Tukey. Statistics 411 Lecture Notes. Dept. of Mathematics, Princeton University, Princeton, New Jersey (1974).