Available online at www.sciencedirect.com
SCIENCE DIRECT®
Linear Algebra and its Applications 397 (2005) 185–192
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

A direct proof and a generalization for a Kantorovich type inequality ☆

Jianguo Huang [a,*], Jieyong Zhou [b]

[a] *Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China*
[b] *Department of Applied Mathematics, Shanghai University of Finance and Economics, Shanghai 200433, China*

Received 28 May 2004; accepted 13 October 2004

Submitted by R.A. Brualdi

Abstract

A direct proof for a Kantorovich type inequality due to Bauer and Householder is presented. A generalization of the inequality is also established by the theory of compound matrices.
© 2004 Elsevier Inc. All rights reserved.

AMS classification: 15A42; 90C90

Keywords: Kantorovich inequality; Optimization; Compound matrices

1. Introduction

Let $A \in \mathbf{M}_n(\mathbf{R})$ be a symmetric positive definite matrix with eigenvalues $0 < \lambda_n \leqslant \cdots \leqslant \lambda_1$. Then for all $x \in \mathbf{R}^n \backslash \{0\}$,

$$\frac{(x^{\mathrm{T}}x)^2}{(x^{\mathrm{T}}Ax)(x^{\mathrm{T}}A^{-1}x)} \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}. \tag{1.1}$$

☆ The work was partly supported by NNSFC under the grant no. 10371076, and E-Institutes of Shanghai Municipal Education Commission, N. E03004.

* Corresponding author. Presently engaged with Division of Computational Science, E-Institute of Shanghai Universities, Shanghai Normal University, China.

E-mail addresses: jghuang@online.sh.cn (J. Huang), zhoujieyong@etang.com (J. Zhou).

doi:10.1016/j.laa.2004.10.009

This is the famous Kantorovich inequality (see [5,6]) and was used in estimating convergence rate of the steepest descent method for minimizing quadratic problems [7]. During the past decades, many researchers have presented various extensions of the Kantorovich inequality which have important applications in statistics. Basically, these inequalities generalize (1.1) in two ways: either the vector x is replaced by a matrix or the positive symmetric matrix A is replaced by a more general matrix (in this case A^{-1} is also replaced by some generalized inverse of A), we refer to [8,10] and references therein for details. However, the following Kantorovich type inequality due to Bauer and Householder [1] was established along a different line.

Theorem 1.1. *Let $x, y \in \mathbf{R}^n$ such that $\frac{|x^{\mathrm{T}}y|}{\|x\|_2\|y\|_2} \geqslant \cos\theta$ with $0 \leqslant \theta \leqslant \frac{\pi}{2}$. Then*

$$\frac{(x^{\mathrm{T}}y)^2}{(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax)} \geqslant \frac{4}{\kappa + 2 + \kappa^{-1}}, \tag{1.2}$$

where A is the same matrix as used in (1.1), and $\kappa = \frac{\lambda_1}{\lambda_n}\frac{1+\sin\theta}{1-\sin\theta}$.

In contrast to most generalized Kantorovich inequalities, this inequality involves two different vectors in different positions (A^{-1} is related to y while A related to x). It is easy to show by the Cauchy–Schwarz inequality that (1.2) reduces to the usual Kantorovich inequality (1.1) when $\theta = 0$. After a long time of its appearance, inequality (1.2) found its important applications in convergence analysis for inexact preconditioned steepest descent method and inexact preconditioned conjugate gradient method for solving linear systems, we refer to [4] for details.

The proof of (1.2) by Bauer and Householder [1] is creative. A generalized Wielandt's inequality was first obtained, and (1.2) was then derived after a very technical deduction. Considering its elegance and importance, in this paper we intend to give a direct proof for inequality (1.2). The basic idea behind the new proof is very simple, the result for $n = 2$ is first proved, and the general one is then obtained by this result and a thorough study about an auxiliary optimization problem. Finally, a generalization of (1.2) is also established by the theory of compound matrices.

2. A direct proof for a Kantorovich type inequality

Before presenting a new proof for inequality (1.2), we first note that by a spectral decomposition of A and a transformation of variables, there is no harm in assuming that $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Lemma 2.1. *Inequality (1.2) holds when $n = 2$.*

Proof. Without loss of generality, we assume that the two vectors x and y are normalized, that means, $x = (\cos\alpha, \sin\alpha)^{\mathrm{T}}$, $y = (\cos\beta, \sin\beta)^{\mathrm{T}}$ for some α and β with $\beta - \alpha = \pm\theta_0$, $0 \leqslant \theta_0 \leqslant \theta$. In the following, we only consider the case $\beta - \alpha =$

θ_0 for simplicity, the other one may be treated in the same manners. By a direct computation, it follows that

$$
\begin{aligned}
f(\alpha) &= (y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax)\\
&= (\lambda_1^{-1}\cos^2\beta + \lambda_2^{-1}\sin^2\beta)(\lambda_1\cos^2\alpha + \lambda_2\sin^2\alpha)\\
&= \cos^2\beta\cos^2\alpha + \sin^2\beta\sin^2\alpha + \frac{\lambda_2}{\lambda_1}(\sin\alpha\cos\beta)^2\\
&\quad + \frac{\lambda_1}{\lambda_2}(\cos\alpha\sin\beta)^2\\
&= \cos^2\theta_0 - 2\cos\beta\sin\alpha\cos\alpha\sin\beta\\
&\quad + \frac{\lambda_2}{\lambda_1}(\sin\alpha\cos\beta)^2 + \frac{\lambda_1}{\lambda_2}(\cos\alpha\sin\beta)^2.
\end{aligned}
\tag{2.1}
$$

Let $z = \sin(\beta + \alpha)$. It is clear that

$$
\cos\alpha\sin\beta = \frac{1}{2}(z + \sin\theta_0), \quad \cos\beta\sin\alpha = \frac{1}{2}(z - \sin\theta_0).
$$

Substituting them into (2.1) implies

$$
\begin{aligned}
f(\alpha) = F(z) &= \cos^2\theta_0 - \frac{1}{2}(z^2 - \sin^2\theta_0) + \frac{\lambda_1}{4\lambda_2}(z + \sin\theta_0)^2\\
&\quad + \frac{\lambda_2}{4\lambda_1}(z - \sin\theta_0)^2.
\end{aligned}
$$

Since $|z| \leqslant 1$, and the coefficient of z^2 is $\frac{\lambda_1}{4\lambda_2} + \frac{\lambda_2}{4\lambda_1} - \frac{1}{2} \geqslant 0$, $F(z)$ is a convex function and must take its maximum at $z = \pm 1$. By a simple computation we find the maximum is attained at $z = 1$, and hence

$$
\begin{aligned}
\max_{|z|\leqslant 1} F(z) &= \cos^2\theta_0 - \frac{1}{2}(1 - \sin^2\theta_0) + \frac{\lambda_1}{4\lambda_2}(1 + \sin\theta_0)^2 + \frac{\lambda_2}{4\lambda_1}(1 - \sin\theta_0)^2\\
&= \frac{1}{4}\cos^2\theta_0\left\{2 + \frac{\lambda_1}{\lambda_2}\frac{1 + \sin\theta_0}{1 - \sin\theta_0} + \frac{\lambda_2}{\lambda_1}\frac{1 - \sin\theta_0}{1 + \sin\theta_0}\right\}.
\end{aligned}
\tag{2.2}
$$

Therefore,

$$
\begin{aligned}
\frac{(x^{\mathrm{T}}y)^2}{(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax)} &= \frac{\cos^2\theta_0}{(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax)}\\
&\geqslant \frac{4}{2 + \kappa(\theta_0) + \kappa^{-1}(\theta_0)},
\end{aligned}
\tag{2.3}
$$

where $\kappa(\theta_0) = \frac{\lambda_1}{\lambda_2}\frac{1+\sin\theta_0}{1-\sin\theta_0}$. Observing that $\kappa(\theta_0)$ is increasing on $[0, \theta]$, and $z + z^{-1}$ is also increasing as $z \geqslant 1$, by virtue of (2.3) we know

$$\frac{(x^{\mathrm{T}}y)^2}{(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax)} \geqslant \frac{4}{2+\kappa+\kappa^{-1}}$$

with $\kappa = \frac{\lambda_1}{\lambda_2}\frac{1+\sin\theta}{1-\sin\theta}$. This completes the proof of Lemma 2.1. □

We next consider an auxiliary optimization problem:

$$\max \qquad F(x, y) = \left(\sum_{i=1}^{n} \lambda_i^{-1} y_i^2\right)\left(\sum_{i=1}^{n} \lambda_i x_i^2\right) \tag{2.4}$$

$$\text{subject to} \quad b_1(x, y) = \sum_{i=1}^{n} x_i^2 - 1 = 0, \quad b_2(x, y) = \sum_{i=1}^{n} y_i^2 - 1 = 0,$$

$$b_3(x, y) = \sum_{i=1}^{n} x_i y_i - \cos\theta_0 = 0, \tag{2.5}$$

where $\theta_0 \in (0, \theta]$ is some given constant.

Lemma 2.2. *Assume that $0 < \lambda_n < \lambda_{n-1} < \cdots < \lambda_1$. Let $x^* = (x_1^*, x_2^*, \ldots, x_n^*)^{\mathrm{T}}$ and $y^* = (y_1^*, y_2^*, \ldots, y_n^*)^{\mathrm{T}}$ be a solution to the problem (2.4) and (2.5). Then there must exist two indices i_1 and i_2, $1 \leqslant i_1 < i_2 \leqslant n$, such that for all $j \in \{1, 2, \ldots, n\} \setminus \{i_1, i_2\}$, we have $x_j^* = y_j^* = 0$.*

Proof. The Jacobian matrix of the vector-valued function $\mathbf{b} = (b_1, b_2, b_2)^{\mathrm{T}}$ is

$$B = \partial_{(x,y)}\mathbf{b} = \begin{bmatrix} 2x_1 & 2x_2 & \cdots & 2x_n & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 2y_1 & 2y_2 & \cdots & 2y_n \\ y_1 & y_2 & \cdots & y_n & x_1 & x_2 & \cdots & x_n \end{bmatrix}.$$

It is easy to check that $\operatorname{rank}(B) = 3$ for all points (x, y) satisfying conditions (2.5), so the constraint conditions of the above problem pass the linear independence constraint qualification. Therefore, by the theory of optimization [2,3], (x^*, y^*) must satisfy the KKT conditions (standing for the Karush–Kuhn–Tucker conditions which are the first-order necessary conditions for constrained optimization problems). In other words,

$$\nabla_x \mathscr{L}(x^*, y^*, \mu^*) = 0, \quad \nabla_y \mathscr{L}(x^*, y^*, \mu^*) = 0, \tag{2.6}$$

where for a Lagrange multiplier vector $\mu = (\mu_1, \mu_2, 2\mu_3)^{\mathrm{T}}$,

$$\mathscr{L}(x, y, \mu) = \left(\sum_{i=1}^{n} \lambda_i^{-1} y_i^2\right)\left(\sum_{i=1}^{n} \lambda_i x_i^2\right) - \mu_1\left(\sum_{i=1}^{n} x_i^2 - 1\right) - \mu_2\left(\sum_{i=1}^{n} y_i^2 - 1\right) - 2\mu_3\left(\sum_{i=1}^{n} x_i y_i - \cos\theta_0\right)$$

means the lagrangian function related to problem (2.4) and (2.5).

Let

$$u = \sum_{i=1}^{n} \lambda_i^{-1} (y_i^*)^2, \quad v = \sum_{i=1}^{n} \lambda_i (x_i^*)^2. \tag{2.7}$$

By a direct computation, (2.6) can be rewritten in the form

$$\lambda_i x_i^* u - \mu_3^* y_i^* - \mu_1^* x_i^* = 0, \ \ \lambda_i^{-1} y_i^* v - \mu_3^* x_i^* - \mu_2^* y_i^* = 0, \quad 1 \leqslant i \leqslant n. \tag{2.8}$$

Multiplying the first equation of (2.8) by x_i^* and taking the summation from 1 to n, and noting the conditions (2.5), we get

$$uv - \mu_3^* \cos\theta_0 - \mu_1^* = 0. \tag{2.9}$$

Similarly, it follows from the second equation of (2.8) that

$$uv - \mu_3^* \cos\theta_0 - \mu_2^* = 0.$$

This with (2.9) implies $\mu_1^* = \mu_2^*$, and so (2.8) can be recast as

$$(u\lambda_i - \mu_1^*) x_i^* = \mu_3^* y_i^*, \quad (v\lambda_i^{-1} - \mu_1^*) y_i^* = \mu_3^* x_i^*, \quad 1 \leqslant i \leqslant n. \tag{2.10}$$

If $\mu_1^* = 0$, then the first equation of (2.10) yields

$$\sqrt{\lambda_i} x_i^* = \frac{\mu_3^*}{u} (\sqrt{\lambda_i})^{-1} y_i.$$

In this case,

$$\left(\sum_{i=1}^{n} \lambda_i^{-1} (y_i^*)^2\right) \left(\sum_{i=1}^{n} \lambda_i (x_i^*)^2\right) = \left(\sum_{i=1}^{n} x_i^* y_i^*\right)^2 = \cos^2\theta_0$$

and hence by the Cauchy–Schwarz inequality, (x^*, y^*) should be a minimizer of $F(x, y)$ subject to (2.5). This is a contradiction, so $\mu_1^* \neq 0$. Moreover, we can also show that $\mu_3^* \neq 0$. Otherwise, since $\{\lambda_i\}_{i=1}^n$ are distinct, by (2.10) there is at most one nonzero member in the set $\{x_i^*\}_{i=1}^n$ (resp. $\{y_i^*\}_{i=1}^n$), which contradicts conditions (2.5). Therefore,

$$x_i^* = 0 \quad \text{implies } y_i^* = 0 \text{ and vice versa.} \tag{2.11}$$

Now assume that $x_i^* y_i^* \neq 0$ for some i, $1 \leqslant i \leqslant n$. Then (2.10) implies

$$(u\lambda_i - \mu_1^*)(v\lambda_i^{-1} - \mu_1^*) = (\mu_3^*)^2,$$

i.e.,

$$u\mu_1^* (\lambda_i)^2 + \{(\mu_3^*)^2 - (\mu_1^*)^2 - uv\}\lambda_i + v\mu_1^* = 0. \tag{2.12}$$

Since $u\mu_1^* \neq 0$ and $\{\lambda_i\}_{i=1}^n$ are distinct, there are at most two different λ_i such that (2.12) holds. In other words, there are at most two indices i such that $x_i^* y_i^* \neq 0$. This with (2.11) implies asserted result. □

Theorem 2.1. *The maximum of problem* (2.4) *and* (2.5) *is*

$$F^* = \frac{\cos^2\theta_0}{4}\{2+\kappa(\theta_0)+\kappa^{-1}(\theta_0)\}$$

with $\kappa(\theta_0) = \frac{\lambda_1}{\lambda_n}\frac{1+\sin\theta_0}{1-\sin\theta_0}$.

Proof. According to Lemma 2.2 and the deduction of Lemma 2.1 (the deduction of (2.2)), it is easy to show that the maximum of (2.4) and (2.5) should be

$$F^* = \max_{1\leqslant i_1<i_2\leqslant n} \kappa_{i_1 i_2}(\theta_0)$$

with $\kappa_{i_1 i_2}(\theta_0) = \frac{\lambda_{i_1}}{\lambda_{i_2}}\frac{1+\sin\theta_0}{1-\sin\theta_0}$. The desired result then follows directly by noting that $\frac{\lambda_{i_1}}{\lambda_{i_2}} \leqslant \frac{\lambda_1}{\lambda_n}$ and $z+z^{-1}$ is an increasing function as $z \geqslant 1$. □

Remark 2.1. Theorem 2.1 still holds for problem (2.4) and (2.5) with the function $b_3(x,y)$ replaced by $b_3(x,y) = \sum_{i=1}^n x_i y_i + \cos\theta_0$.

Now we are ready to give a new proof of Theorem 1.1 due to Bauer and Householder.

Proof of Theorem 1.1. When $\theta = 0$, then $|x^{\mathrm{T}}y| = \|x\|_2\|y\|_2$, and by the Cauchy–Schwarz inequality it follows that $x = ay$ for some real number a. Eq. (1.2) is therefore valid since it is the usual Kantorovich inequality. When $\theta \in (0, \frac{\pi}{2}]$, considering the form of inequality (1.2), we can assume that x and y are normalized, i.e.,

$$\sum_{i=1}^n x_i^2 - 1 = 0, \quad \sum_{i=1}^n y_i^2 - 1 = 0, \quad \sum_{i=1}^n x_i y_i \pm \cos\theta_0 = 0.$$

For simplicity we only consider the case that $\sum_{i=1}^n x_i y_i + \cos\theta_0 = 0$. The other one may be treated similarly.

If the eigenvalues of A are distinct, then it follows from Theorem 2.1 that

$$(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax) \leqslant \frac{\cos^2\theta_0}{4}\left\{2+\kappa(\theta_0)+\kappa^{-1}(\theta_0)\right\} \leqslant \frac{\cos^2\theta_0}{4}\{2+\kappa+\kappa^{-1}\},$$

which implies (1.2).

If A has multiple eigenvalues, we can get (1.2) by the usual perturbation method (see [9–p. 76]). In fact, we can find a sequence $\{\lambda_i(k)\}_{i=1}^n$ such that

$$0 < \lambda_n(k) < \lambda_{n-1}(k) < \cdots < \lambda_1(k) \quad \text{and} \quad \lim_{k\to\infty}\lambda_i(k) = \lambda_i, \quad 1 \leqslant i \leqslant n.$$

Let $A(k) = \mathrm{diag}(\lambda_1(k), \ldots, \lambda_n(k))$. Then we have by the previous argument that

$$(y^{\mathrm{T}}A(k)^{-1}y)(x^{\mathrm{T}}A(k)x) \leqslant \frac{\cos^2\theta_0}{4}\{2+\kappa(k)+\kappa(k)^{-1}\},$$

where $\kappa(k) = \frac{\lambda_1(k)}{\lambda_n(k)}\frac{1+\sin\theta}{1-\sin\theta}$.

However,

$$\lim_{k\to\infty}(y^{\mathrm{T}}A(k)^{-1}y)(x^{\mathrm{T}}A(k)x)=(y^{\mathrm{T}}A^{-1}y)(x^{\mathrm{T}}Ax),\qquad \lim_{k\to\infty}\kappa(k)=\kappa,$$

so the desired result follows immediately from the last inequality by letting $k\to\infty$. $\square$

3. A generalization for a Kantorovich type inequality

We now extend inequality (2) in a matrix form. To do so, let us first review some basic results about compound matrices [5]. For a given matrix $A\in\mathbf{M}_{m,n}(\mathbf{R})$, the $\binom{m}{k}$-by-$\binom{n}{k}$ matrix whose α,β entry is $\det A(\alpha,\beta)$ is called the kth compound matrix of A and is denoted by $C_k(A)$. Here, $\alpha\subseteq\{1,\ldots,m\}$ and $\beta\subseteq\{1,\ldots,n\}$ are index sets of cardinality $k\leqslant\min\{m,n\}$, ordered lexicographically. The next result comes from [5–pp. 19–20].

Lemma 3.1. *The following statements hold*:

(1) *If* $A\in\mathbf{M}_{m,k}(\mathbf{R})$ *and* $B\in\mathbf{M}_{k,n}(\mathbf{R})$, *then*

$$C_r(AB)=C_r(A)C_r(B),\quad r\leqslant\min\{m,k,n\},$$
$$C_k(I)=I\in\mathbf{M}_{\binom{n}{k}},\quad C_k(A^{\mathrm{T}})=C_k(A)^{\mathrm{T}},$$

where I *denotes an identity matrix*.

(2) *If* $A\in\mathbf{M}_n(\mathbf{R})$ *is nonsingular*, *then* $C_k(A^{-1})=C_k(A)^{-1}$.

(3) *Let* $\{\lambda_i\}_{i=1}^n$ *be the eigenvalues of* $A\in\mathbf{M}_n(\mathbf{R})$. *Then the eigenvalues of* $C_k(A)$ *consist of all products* $\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}$ *with* $i_j\in\{1,2,\ldots,n\}$, $1\leqslant j\leqslant k$.

Theorem 3.1. *Let* $Z,W\in\mathbf{M}_{n,m}(\mathbf{R})$ $(m\leqslant n)$ *be two full rank matrices which satisfy that*

$$\frac{|\det Z^{\mathrm{T}}W|}{(\det Z^{\mathrm{T}}Z)^{1/2}(\det W^{\mathrm{T}}W)^{1/2}}\geqslant\cos\theta,\quad 0\leqslant\theta\leqslant\frac{\pi}{2}. \tag{3.13}$$

Then there holds

$$\frac{(\det Z^{\mathrm{T}}W)^2}{(\det W^{\mathrm{T}}A^{-1}W)(\det Z^{\mathrm{T}}AZ)}\geqslant\frac{4}{2+\kappa'+\kappa'^{-1}}, \tag{3.14}$$

where A *is the same matrix as used in inequality* (1.2), *and*

$$\kappa'=\frac{\lambda_1\lambda_2\cdots\lambda_k}{\lambda_n\lambda_{n-1}\cdots\lambda_{n-k+1}}\frac{1+\sin\theta}{1-\sin\theta}.$$

Proof. Let $\widetilde{Z} = C_m(Z)$, $\widetilde{W} = C_m(W)$ and $\widetilde{A} = C_m(A)$. Then it follows from the definition of compound matrices and Lemma 3.1 that $\det Z^{\mathrm{T}}Z = \|\widetilde{Z}\|_2^2$, $\det W^{\mathrm{T}}W = \|\widetilde{W}\|_2^2$, $\det Z^{\mathrm{T}}W = \widetilde{Z}^{\mathrm{T}}\widetilde{W}$ and

$$\det W^{\mathrm{T}}A^{-1}W = \widetilde{W}^{\mathrm{T}}(\widetilde{A})^{-1}\widetilde{W}, \quad \det Z^{\mathrm{T}}AZ = \widetilde{Z}^{\mathrm{T}}\widetilde{A}\widetilde{Z}.$$

Therefore, inequality (3.14) is derived immediately by Theorem 1.1 and the fact that $\lambda_{\max}(\widetilde{A}) = \lambda_1\lambda_2\cdots\lambda_k$, $\lambda_{\min}(\widetilde{A}) = \lambda_n\lambda_{n-1}\cdots\lambda_{n-k+1}$, where $\lambda_{\max}(\widetilde{A})$ and $\lambda_{\min}(\widetilde{A})$ denote the maximal and minimal eigenvalues of $\widetilde{A}$, respectively. □

Remark 3.1. Let $\mathscr{R}(Z)$ denote the subspace of $\mathbf{R}^n$ spanned by the columns of Z. Then

$$\angle(\mathscr{R}(Z), \mathscr{R}(W)) = \arccos \frac{|\det Z^{\mathrm{T}}W|}{(\det Z^{\mathrm{T}}Z)^{1/2}(\det W^{\mathrm{T}}W)^{1/2}}$$

means the angle between a pair of subspaces $\mathscr{R}(Z)$ and $\mathscr{R}(W)$ geometrically [9–p. 96]. Therefore, the above theorem provides a relationship between the angle $\angle(\mathscr{R}(Z), \mathscr{R}(W))$ and the angle $\angle(\mathscr{R}(A^{1/2}Z), \mathscr{R}(A^{-1/2}W))$.

Acknowledgments

The authors are grateful to Prof. H. Munthe-Kaas who sent them a technical report entitled "The convergence rate of inexact preconditioned steepest descent algorithm for solving linear systems", which is helpful for the study in this paper. They also thank an anonymous referee whose suggestions improved the presentation of the paper.

References

[1] F.L. Bauer, A.S. Householder, Some inequalities involving the euclidean condition of a matrix, Numer. Math. 2 (1960) 308–311.
[2] R. Fletcher, Practical Methods of Optimization, second ed., John Wiley & Sons, England, 1987.
[3] A. Forsgren, P.E. Gill, M.H. Wright, Interior methods for nonlinear optimization, SIAM Rev. 44 (2002) 525–597.
[4] G.H. Golub, Q. Ye, Inexact preconditioned conjugate gradient method with inner-outer iteration, SIAM J. Sci. Comput. 21 (1999) 1305–1320.
[5] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, London, 1985.
[6] A.S. Househoulder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964.
[7] J. Nocedal, Theory of algorithms for unconstrained optimization, Acta Numer. 1 (1992) 199–242.
[8] J.E. Pecaric, S. Puntanen, G.P.H. Styan, Some further matrix extensions of the Cauchy-Schwarz and Kantorovich inequalities, with some statistical applications, Linear Algebra Appl. 237 (1996) 455–476.
[9] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory, Academic Press, Boston, 1990.
[10] S.G. Wang, W.C. Ip, A matrix version of the Wielandt inequality and its applications to statistics, Linear Algebra Appl. 296 (1999) 177–181.