ELSEVIER Linear Algebra and its Applications 275–276 (1998) 261–279

LINEAR ALGEBRA AND ITS APPLICATIONS

Modifying the inertia of matrices arising in optimization

Nicholas J. Higham *, Sheung Hun Cheng [1]

Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

Received 9 October 1996; accepted 10 March 1997

Submitted by V. Mehrmann

Abstract

Applications in constrained optimization (and other areas) produce symmetric matrices with a natural block 2×2 structure. An optimality condition leads to the problem of perturbing the (1,1) block of the matrix to achieve a specific inertia. We derive a perturbation of minimal norm, for any unitarily invariant norm, that increases the number of nonnegative eigenvalues by a given amount, and we show how it can be computed efficiently given a factorization of the original matrix. We also consider an alternative way to satisfy the optimality condition based on a projection approach. Theoretical tools developed here include an extension of Ostrowski's theorem on congruences and some lemmas on inertias of block 2×2 symmetric matrices. © 1998 Elsevier Science Inc. All rights reserved.

AMS classification: 65F15; 15A42

Keywords: Inertia; Optimization; Nonlinear programming; unitarily invariant norm

1. Introduction

Optimization is a rich source of linear algebra problems. An example is the problem of modified Cholesky factorization arising in Newton methods for unconstrained optimization, in which a possibly indefinite symmetric matrix must

* Corresponding author. E-mail: higham@ma.man.ac.uk.

[1] E-mail: chengsh@ma.man.ac.uk.

PII: S 0 0 2 4 - 3 7 9 5 (9 7) 1 0 0 1 5 - 5

be perturbed to make it positive definite, while at the same time producing a Cholesky factorization of the perturbed matrix [5]; [11], Section 4.4.2.2; [23]. The work described here can be thought of as an attempt to extend the notion of modified Cholesky factorization to constrained optimization.

A block 2×2 partitioning

$$C = \begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix}$$

of a symmetric matrix C arises in a number of applications, including constrained optimization, least squares problems and Navier–Stokes problems, as explained in the next section. The matrix D is positive semidefinite, but H can be indefinite, depending on the application. In constrained optimization, a "second order sufficiency" condition leads to the problem of perturbing H so that C has a particular inertia. It is this problem that motivated our work.

In Section 3 we present some background material on congruence transformations, including an extension of Ostrowski's theorem to transformations with a rectangular matrix. In Section 4 we derive some useful inertia properties of the matrix C. How to make a minimal norm (full) perturbation to increase the number of nonnegative eigenvalues of a symmetric matrix by a given amount is shown in Section 5. The main result of the paper is in Section 6, in which we derive, for any unitarily invariant norm, a perturbation to H (only) of minimal norm that increases the number of nonnegative eigenvalues of C by a given amount. For the optimization application, another way of writing the second order sufficiency condition is based on projecting H into the null space of A. We use this approach in Section 7 to derive another expression for a minimal norm perturbation to H that achieves the sufficiency condition. Finally, in Section 8 we consider how to implement our results in the optimization application and show that directions of negative curvature are produced as a by-product of the computations.

2. A symmetric block 2×2 matrix and its applications

Any symmetric matrix C can be written in the form

$$C = \begin{array}{c} \\ n \\ m \end{array}\!\!\begin{array}{c} \begin{array}{cc} n & m \end{array} \\ \begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix} \end{array},$$

where $H \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{n\times n}$ are symmetric and $A \in \mathbb{R}^{n\times m}$. The reason for using a block 2×2 partitioning and for placing a minus sign in front of the (2,2) block is that C then conveniently represents some particular cases arising in applications, which we now describe in roughly decreasing order of generality.

1. When D is diagonal and positive definite, C is the "primal- dual" matrix arising in certain interior methods for the general nonlinear programming problem [8,9]. Here, H is the Hessian of the Lagrangian function and A^{T} is the Jacobian of the constraint functions. The matrix C also arises in penalty function methods for nonlinear programming, with D a positive multiple of the identity matrix [14]. In these applications both $m \leqslant n$ and $m \geqslant n$ are possible.
2. When $D = 0$, C is the Karush–Kuhn–Tucker (KKT) matrix, which arises when Newton's method or a quasi-Newton method is applied to the problem [2]

   $$\min_x F(x) \qquad \text{subject to} \quad A^{\mathrm{T}}x = b, \tag{2.1}$$

   where $F : \mathbb{R}^n \to \mathbb{R}$ and $m \leqslant n$ [6], p. 123; [10,15]. To be precise, Newton's method leads to the equations

   $$\begin{bmatrix} H_k & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} p_k \\ \lambda_k \end{bmatrix} = \begin{bmatrix} -g_k \\ 0 \end{bmatrix},$$

   where H is the Hessian of F or an approximation to it, g is the gradient of F, p is a search direction, and λ is a Lagrange multiplier, and where a subscript k denotes evaluation at the kth iterate.
3. If H and D are positive definite, then C matches precisely the definition of a symmetric quasi-definite matrix [27]. Such matrices arise in interior methods for linear and quadratic programming and much is known about the existence and stability of their LDL^{T} factorizations [12,27].
4. Matrices with H positive definite and $D = 0$ arise in discretized incompressible Navier-Stokes equations [24], and their spectral properties are important in the development of preconditioned iterative methods [7].
5. The matrix with $H = \delta I$ and $D = \delta I$ $(\delta > 0)$ appears in the augmented system corresponding to the damped least squares problem

   $$\min_x \|b - Ax\|_2^2 + \delta^2 \|x\|_2^2;$$

   see Saunders [22].
6. For H positive definite and $D = 0$, C is the augmented system matrix arising in the generalized least squares problem $\min(b - Ax)^{\mathrm{T}} H^{-1}(b - Ax)(m \geqslant n)$ [3], Section 4.3.2; $H = I$ gives the standard least squares problem.

In quasi-Newton methods for the linear equality constrained problem (2.1) it is desirable that the Hessian approximation H satisfy the "second order sufficiency" condition [15]

[2] In optimization references linear constraints are usually written $Ax = b$; we find the transposed form more natural notationally.

$$p^{\mathrm{T}}Hp > 0 \quad \text{for all nonzero } p \text{ such that} \quad A^{\mathrm{T}}p = 0. \tag{2.2}$$

One equivalent condition is that the projected Hessian $Z^{\mathrm{T}}HZ$ is positive definite, where the columns of Z form a basis for the nullspace null (A^{T}). Less obviously, the condition (2.2) is also equivalent to requiring the so-called KKT matrix

$$K = \begin{array}{c c} & \begin{array}{cc} n & m \end{array} \\ \begin{array}{c} n \\ m \end{array} & \begin{bmatrix} H & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \end{array} \tag{2.3}$$

to have a certain inertia, as shown by Gould [13]. Recall that the inertia of a symmetric matrix is an ordered triple (i_+, i_-, i_0), where i_+ is the number of positive eigenvalues, i_- the number of negative eigenvalues, and i_0 the number of zero eigenvlaues. We write

$$\mathrm{inertia}(A) = (i_+(A), i_-(A), i_0(A)).$$

Theorem 2.1 (Gould). *Let A be of full rank m. The condition (2.2) holds if and only if K has the inertia (n,m,0).*

Proof. Let A have the QR factorization

$$A = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = [\,Y \quad Z\,]\begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $Y \in \mathbb{R}^{n\times m}$, $Z \in \mathbb{R}^{n\times(n-m)}$, and $R \in \mathbb{R}^{m\times m}$. Then

$$\begin{aligned} K &= \begin{bmatrix} H & Q\begin{bmatrix} R \\ 0 \end{bmatrix} \\ [\,R^{\mathrm{T}} \quad 0\,]Q^{\mathrm{T}} & 0 \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} Q^{\mathrm{T}}HQ & \begin{bmatrix} R \\ 0 \end{bmatrix} \\ [\,R^{\mathrm{T}} \quad 0\,] & 0 \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}^{\mathrm{T}} \\ &\sim \begin{bmatrix} Y^{\mathrm{T}}HY & Y^{\mathrm{T}}HZ & R \\ Z^{\mathrm{T}}HY & Z^{\mathrm{T}}HZ & 0 \\ R^{\mathrm{T}} & 0 & 0 \end{bmatrix} =: \widetilde{K}, \end{aligned}$$

where $\sim$ denotes congruence (in fact, this first transformation is an orthogonal similarity). Now define the nonsingular matrix

$$W = \begin{bmatrix} I_m & 0 & -\frac{1}{2}Y^{\mathrm{T}}HYR^{-\mathrm{T}} \\ 0 & I_{n-m} & -Z^{\mathrm{T}}HYR^{-\mathrm{T}} \\ 0 & 0 & R^{-\mathrm{T}} \end{bmatrix}.$$

It is straightforward to verify that

$$W\widetilde{K}W^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & I_m \\ 0 & Z^{\mathrm{T}}HZ & 0 \\ I_m & 0 & 0 \end{bmatrix}.$$

The eigenvalues of $W^{\mathrm{T}}\widetilde{K}W$ are 1 and -1, each repeated m times, together with the $n-m$ eigenvalues of $Z^{\mathrm{T}}HZ$. Since Z spans the null space of A^{T}, $Z^{\mathrm{T}}HZ$ is positive definite if and only if (2.2) holds, which completes the proof. □

From the requirement (2.2) and Theorem 2.1 the problem arises of perturbing H so that K achieves the desired inertia $(n, m, 0)$ [15]. The matrix A must not be perturbed, because this would correspond to changing the constraints in (2.1). The same problem is relevant for the primal-dual matrix with D diagonal and positive semidefinite [8]. We find a minimal-norm solution to a more general version of this inertia perturbation problem in Section 6. In Section 7 we consider an alternative approach to perturbing H to satisfy (2.2), based on the projected Hessian. First, we develop some necessary background theory.

3. Rectangular congruence transformations

Sylvester's inertia theorem says that the inertia of a symmetric matrix is preserved under a congruence transformation. Ostrowski's theorem [18], Theorem 4.5.9; [20,28] goes further by explaining how much the magnitudes of the eigenvalues can change. In the following statement of Ostrowski's theorem [18], Corollary 4.5.11 the transforming matrix X is permitted to be singular, in which case the transformation $X^{\mathrm{T}}AX$ is not a congruence transformation and can change the inertia. Throughout this paper the eigenvalues of a symmetric $n \times n$ matrix are ordered $\lambda_1 \leqslant \cdots \leqslant \lambda_n$, and $\lambda_i(A)$ denotes the ith smallest eigenvalue of A.

Theorem 3.1 (Ostrowski). *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $X \in \mathbb{R}^{n\times n}$. Then*

$$\lambda_k(X^{\mathrm{T}}AX) = \theta_k \lambda_k(A), \quad k = 1:n,$$

where $\lambda_1(X^{\mathrm{T}}X) \leqslant \theta_k \leqslant \lambda_n(X^{\mathrm{T}}X)$.

We now generalize Ostrowski's theorem to "rectangular congruences", in which the transforming matrix X is nonsquare. Such transformations change the dimension and hence the inertia, but for full rank X the amount by which the inertia can change depends on the difference of the dimensions of X, as shown in the corollaries below. First, we consider matrices X with at least as many rows as columns.

Theorem 3.2. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $X \in \mathbb{R}^{n\times m}$ $(n \geqslant m)$. Then*

$$\lambda_k(X^{\mathrm{T}}AX) = \theta_k \mu_k, \quad k = 1:m,$$

where

$$\lambda_k(A) \leqslant \mu_k \leqslant \lambda_{k+n-m}(A), \quad k = 1:m,$$

and $\lambda_1(X^{\mathrm{T}}X) \leqslant \theta_k \leqslant \lambda_m(X^{\mathrm{T}}X)$.

Proof. Let

$$X = U\begin{bmatrix}\Sigma \\ 0\end{bmatrix}V^{\mathrm{T}}$$

be a singular value decomposition, where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ are orthogonal and $\Sigma \in \mathbb{R}^{m\times m}$ is diagonal. Then

$$X^{\mathrm{T}}AX = V[\Sigma^{\mathrm{T}}0]U^{\mathrm{T}}AU\begin{bmatrix}\Sigma \\ 0\end{bmatrix}V^{\mathrm{T}} = V(\Sigma^{\mathrm{T}}\tilde{A}_{11}\Sigma)V^{\mathrm{T}},$$

where $\tilde{A}_{11}$ is the leading principal submatrix of order m of $\tilde{A} = U^{\mathrm{T}}AU$. By Ostrowski's theorem,

$$\lambda_k(X^{\mathrm{T}}AX) = \lambda_k(\Sigma^{\mathrm{T}}\tilde{A}_{11}\Sigma) = \lambda_k(\tilde{A}_{11})\theta_k,$$

where

$$\lambda_1(X^{\mathrm{T}}X) = \lambda_1(\Sigma^{\mathrm{T}}\Sigma) \leqslant \theta_k \leqslant \lambda_m(\Sigma^{\mathrm{T}}\Sigma) = \lambda_m(X^{\mathrm{T}}X).$$

Cauchy's interlace theorem [21], p. 186 shows that

$$\lambda_k(A) = \lambda_k(\tilde{A}) \leqslant \lambda_k(\tilde{A}_{11}) \leqslant \lambda_{k+n-m}(\tilde{A}) = \lambda_{k+n-m}(A), \quad k = 1:m,$$

which yields the result. □

In the case where X has orthonormal columns (so that $\theta_k \equiv 1$), Theorem 3.2 reduces to the Poincaré separation theorem [18], Corollary 4.3.16; [25], Corollary 4.4, p. 198.

Corollary 3.3. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and let $X \in \mathbb{R}^{n\times m}(n \geqslant m)$ be of full rank. Then*

$$\begin{aligned}\text{inertia }(A) - (n-m, n-m, n-m) &\leqslant \text{ inertia }(X^{\mathrm{T}}AX)\\ &\leqslant \text{ inertia }(A) + (0, 0, n-m).\end{aligned}$$

The next result covers the case $n \leqslant m$.

Theorem 3.4. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $X \in \mathbb{R}^{n\times m}$ $(n \leqslant m)$. Then $X^{\mathrm{T}}AX$ has $m-n$ zero eigenvalues, which we number $\lambda_1, \ldots, \lambda_{m-n}$; the remaining eigenvalues satisfy*

$$\lambda_{m-n+k}(X^{\mathrm{T}}AX) = \theta_k \lambda_k(A), \quad k = 1:n,$$

where $\lambda_{m-n+1}(X^{\mathrm{T}}X) \leqslant \theta_k \leqslant \lambda_m(X^{\mathrm{T}}X)$.

Proof. The proof is similar to that of Theorem 3.2. □

Corollary 3.5. *Let* $A \in \mathbb{R}^{n\times n}$ *be symmetric, and let* $X \in \mathbb{R}^{n\times m}$ $(n \leqslant m)$ *be of full rank. Then*

$$\text{inertia}\,(X^{\mathrm{T}}AX) = \text{inertia}\,(A) + (0, 0, m-n).$$

4. Inertia properties of C

In this section we derive some inertia properties of the matrix

$$C = \begin{array}{c} \\ n \\ m \end{array} \!\!\begin{array}{c} \begin{array}{cc} n & m \end{array} \\ \begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix} \end{array}. \tag{4.1}$$

Assume that H is nonsingular. We have

$$\begin{bmatrix} I & 0 \\ -A^{\mathrm{T}}H^{-1} & I \end{bmatrix} \begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix} = \begin{bmatrix} H & A \\ 0 & -D - A^{\mathrm{T}}H^{-1}A \end{bmatrix}$$
$$= \begin{bmatrix} H & 0 \\ 0 & -D - A^{\mathrm{T}}H^{-1}A \end{bmatrix} \begin{bmatrix} I & H^{-1}A \\ 0 & I \end{bmatrix},$$

which shows that

$$C \sim \begin{bmatrix} H & 0 \\ 0 & -D - A^{\mathrm{T}}H^{-1}A \end{bmatrix}. \tag{4.2}$$

This congruence is the basis of the following lemmas, the first of which is contained in [16], Theorem 3.

Lemma 4.1. *If* H *is nonsingular,* $D = 0$*, and* A *has full rank, then* $\text{inertia}(C) \geqslant (m, m, 0)$ *if* $m \leqslant n$ *and* $\text{inertia}\,(C) = (n, n, m-n)$ *if* $m \geqslant n$.

Proof. Let $\text{inertia}(H) = (a, b, 0)$ and $\text{inertia}(-A^{\mathrm{T}}H^{-1}A) = (p, q, r)$. Then from (4.2) we have

$$\text{inertia}\,(C) = (a + p, b + q, r).$$

First, suppose $m \leqslant n$. By Corollary 3.3 we have $p \geqslant b - (n - m)$, so that $a + p \geqslant a + b - (n - m) = m$. Similarly, $b + q \geqslant m$. If $m \geqslant n$, then Corollary 3.5 shows that $p = b$, $q = a$, and $r = m - n$, and the result follows.

Lemma 4.2. *If H is positive definite and D is positive semidefinite, then* $\text{inertia}(C) = (n, m-p, p)$, *where* $0 \leqslant p \leqslant m$. *If A has full rank or D is positive definite then $p = 0$.*

Proof. The result is a direct consequence of (4.2). $\square$

The next lemma shows the somewhat surprising property that the inertia of C is independent of H when all the blocks are square, $D = 0$ and A is nonsingular. This result is given by Haynsworth and Ostrowski [16], who attribute it to Carlson and Schneider [4].

Lemma 4.3. *Let $m = n$ and $D = 0$. Then C is nonsingular if and only if A is nonsingular, and in this case* $\text{inertia}(C) = (n, n, 0)$.

Proof. The nonsingularity condition follows from

$$\det(C) = (-1)^n \det\left(\begin{bmatrix} A & H \\ 0 & A^{\mathrm{T}} \end{bmatrix}\right) = (-1)^n (\det A)^2.$$

The inertia is obtained as a special case of Theorem 2.1, since (2.2) is trivially satisfied. $\square$

There does not seem to be any useful characterization of the eigenvalues of C. The most general matrix for which the eigenvalues are known explicitly is the matrix

$$B(\alpha, \beta) = \begin{bmatrix} \alpha I_n & A \\ A^{\mathrm{T}} & -\beta I_m \end{bmatrix}, \quad A \in \mathbb{R}^{n \times m}. \tag{4.3}$$

Saunders [22] shows that if A has rank p with nonzero singular values σ_i, $i = 1{:}p$, then

$$\lambda(B(\alpha, \beta)) = \begin{cases} \frac{1}{2}(\alpha - \beta) \pm \left(\sigma_i^2 + \frac{1}{4}(\alpha + \beta)^2\right)^{1/2}, & i = 1:p, \\ \alpha & n - p \text{ times}, \\ -\beta & m - p \text{ times}. \end{cases} \tag{4.4}$$

The conclusions of Lemmas 4.1–4.3 are readily verified for this matrix.

Finally, we give inequalities that bound the eigenvalues of C away from zero, which is of interest for investigating conditioning. This lemma is a restatement of the "separation theorem" of Von Kempen [26].

Lemma 4.4. *If H is positive definite and D is positive semidefinite then the eigenvalues λ_i of C satisfy*

$$\lambda_1 \leqslant \cdots \leqslant \lambda_m \leqslant -\lambda_{\min}(D) < \lambda_{\min}(H) \leqslant \lambda_{m+1} \leqslant \cdots \leqslant \lambda_{m+n}. \tag{4.5}$$

Proof. Let λ be an eigenvalue of C and x a corresponding eigenvector and write $Cx = \lambda x$ as

$$\begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \lambda \begin{bmatrix} y \\ z \end{bmatrix}.$$

Premultiplying the first equation of the pair by y^{T} and the second by z^{T}, and subtracting, yields

$$y^{\mathrm{T}}Hy - \lambda y^{\mathrm{T}}y = -z^{\mathrm{T}}Dz - \lambda z^{\mathrm{T}}z,$$

or

$$y^{\mathrm{T}}(H - \lambda I)y + z^{\mathrm{T}}(D + \lambda I)z = 0. \tag{4.6}$$

If $-\lambda_{\min}(D) < \lambda < \lambda_{\min}(H)$ then $H - \lambda I$ and $D + \lambda I$ are positive definite and (4.6) yields a contradiction since y and z are not both zero. The inequalities (4.5) now follow from Lemma 4.2. □

That the bounds on λ_m and λ_{m+1} in Lemma 4.4 are attainable is shown by (4.3) and (4.4). (For the interior eigenvalues, inequalities (4.5) can, of course, be improved by applying Cauchy's interlace theorem.)

A bound for the 2-norm condition number $\kappa_2(C) = \|C\|_2 \|C^{-1}\|_2$ is immediate.

Corollary 4.5. *If H and D are positive definite, then*

$$\kappa_2(C) \leqslant \|C\|_2 \max\{\|H^{-1}\|_2, \|D^{-1}\|_2\}.$$

5. Modifying the inertia: A general perturbation

Let $A \in \mathbb{R}^{n\times n}$ be symmetric. We denote by $\mu^{(k)}(A)$ the distance from A to the symmetric matrices with at least k more nonnegative eigenvalues than A (assuming that A has at least k negative eigenvalues):

$$\mu^{(k)}(A) = \min\{\|\Delta A\| : \Delta A = \Delta A^{\mathrm{T}}, i_+(A + \Delta A) + i_0(A + \Delta A) \geqslant i_+(A) + i_0(A) + k\}. \tag{5.1}$$

The distance is characterized by the following theorem, which generalizes a result giving the distance to the nearest symmetric positive semidefinite matrix [17]. Recall that a norm $\|\cdot\|$ is a unitarily invariant norm on $\mathbb{R}^{n\times n}$ if $\|UAV\| = \|A\|$ for all orthogonal U and V. We will need the characterization that any unitarily invariant norm is a symmetric gauge function on the singular values, that is, $\|A\| = \phi(\sigma_1, \ldots, \sigma_n)$, where ϕ is an absolute vector norm that is invariant under permutations of the entries of its argument [18], Theorem 7.4.24; [25], Theorem 3.6, p. 78.

Theorem 5.1. *Let the symmetric matrix $A \in \mathbb{R}^{n\times n}$ have the spectral decomposition $A = Q\Lambda Q^{\mathrm{T}}$, where Q is orthogonal and $\Lambda = \operatorname{diag}(\lambda_i)$ with*

$$\lambda_1 \leqslant \cdots \leqslant \lambda_p < 0 \leqslant \lambda_{p+1} \leqslant \cdots \leqslant \lambda_n,$$

and assume that $p \geqslant k$. Then for any unitarily invariant norm, an optimal perturbation in (5.1) is

$$\Delta A = Q \operatorname{diag}(\tau_i)\, Q^{\mathrm{T}}, \quad \tau_i = \begin{cases} -\lambda_i, & i = p-k+1{:}p, \\ 0, & \textit{otherwise}, \end{cases}. \tag{5.2}$$

and

$$\mu^{(k)}(A) = \phi(\tau_1, \ldots, \tau_n).$$

Proof. A generalization of the Wielandt-Hoffman theorem [18], Theorem 7.4.51; [25], p. 205 says that if A and $A + \Delta A$ are symmetric then

$$\|\Delta A\| \geqslant \| \operatorname{diag}(\lambda_i(A + \Delta A) - \lambda_i(A))\|$$

for any unitarily invariant norm. If ΔA is a feasible perturbation in (5.1) then

$$\begin{aligned} \|\Delta A\| &\geqslant \|\operatorname{diag}(0, \ldots, 0, \lambda_{p-k+1}(A + \Delta A) - \lambda_{p-k+1}(A), \ldots, \\ &\qquad \lambda_p(A + \Delta A) - \lambda_p(A), 0, \ldots, 0)\| \\ &\geqslant \|\operatorname{diag}(0, \ldots, 0, -\lambda_{p-k+1}(A), \ldots, -\lambda_p(A), 0, \ldots, 0)\|, \end{aligned}$$

where we have used $\lambda_p(A + \Delta A) \geqslant \cdots \geqslant \lambda_{p-k+1}(A + \Delta A) \geqslant 0$ and the gauge function property of the norms. It is easy to see that equality is attained for the perturbation given in the statement of the theorem and that this perturbation is feasible. □

6. Modifying the inertia: A structured perturbation

Returning to the partitioned matrix (4.1), we are interested in finding a perturbation ΔH such that

$$C + \Delta C = \begin{bmatrix} H + \Delta H & A \\ A^{\mathrm{T}} & -D \end{bmatrix}$$

has a given inertia. For the analysis in this section, C can be regarded as a general block 2×2 symmetric matrix – we will not need A to have full rank or the diagonal blocks to possess any definiteness properties, and m and n are arbitrary.

For the KKT matrix, practical interest is in increasing the number of positive eigenvalues (in view of Theorem 2.1 and Lemma 4.1), so we define, analogously to (5.1),

$$\psi^{(k)}(C) = \min\{\|\Delta H\|: \Delta H = \Delta H^{\mathrm{T}}, i_+(C+\Delta C) + i_0(C+\Delta C) \geqslant i_+(C) + i_0(C) + k\}. \tag{6.1}$$

Clearly, an optimal ΔH in (6.1) can be taken to be positive semidefinite and of rank k, hence of the form $\Delta H = VV^{\mathrm{T}}$ with $V \in \mathbb{R}^{n\times k} (k \leqslant n)$. Our solution to this problem is based on the following lemma. The lemma is not new; essentially the same result can be found in [1], Lemma 2.1 and [2], Corollary 2.2, for example.

Lemma 6.1. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and nonsingular and let $W \in \mathbb{R}^{n\times k}$. Then $i_+(A + WW^{\mathrm{T}}) + i_0(A + WW^{\mathrm{T}}) = i_+(A) + i_0(A) + k$ if and only if $-I_k - W^{\mathrm{T}}A^{-1}W$ is positive semidefinite.*

Proof. We have the congruences

$$B = \begin{array}{c} \\ n \\ k \end{array}\!\!\begin{array}{c} \begin{array}{cc} n & k \end{array} \\ \begin{bmatrix} A & W \\ W^{\mathrm{T}} & -I_k \end{bmatrix} \end{array} \sim \begin{bmatrix} A & 0 \\ 0 & -I_k - W^{\mathrm{T}}A^{-1}W \end{bmatrix}$$

and, for a suitable permutation Π,

$$\Pi^{\mathrm{T}}B\Pi = \begin{bmatrix} -I_k & W^{\mathrm{T}} \\ W & A \end{bmatrix} \sim \begin{bmatrix} -I_k & 0 \\ 0 & A + WW^{\mathrm{T}} \end{bmatrix}.$$

It follows that

$$\text{inertia}\ (A) + \text{inertia}\ (-I_k - W^{\mathrm{T}}A^{-1}W) = \text{inertia}\ (-I_k) + \text{inertia}\ (A + WW^{\mathrm{T}}),$$

that is,

$$\text{inertia}\ (A + WW^{\mathrm{T}}) = \text{inertia}\,(A) + \text{inertia}\ (-I_k - W^{\mathrm{T}}A^{-1}W) - \text{inertia}\ (-I_k).$$

The result is immediate. □

We apply Lemma 6.1 with A the matrix C (assumed to be nonsingular) and

$$W = \begin{array}{c} n \\ m \end{array}\!\!\begin{array}{c} k \\ \begin{bmatrix} V \\ 0 \end{bmatrix} \end{array}.$$

The lemma tells us that we need to minimize $\|VV^{\mathrm{T}}\|$ subject to

$$\begin{bmatrix} V^{\mathrm{T}} & 0^{\mathrm{T}} \end{bmatrix} C^{-1} \begin{bmatrix} V \\ 0 \end{bmatrix} \in \mathbb{R}^{k\times k} \tag{6.2}$$

having all its eigenvalues less than or equal to -1. Writing $G = C^{-1}(1{:}n, 1{:}n)$, this constraint is

$$\lambda_i(V^{\mathrm{T}}GV) \leqslant -1, \quad i = 1\ k. \tag{6.3}$$

By Corollary 3.3, a matrix V satisfying (6.3) exists only if G has at least k negative eigenvalues, which we assume to be the case. How to minimize $\|VV^{\mathrm{T}}\|$ for any unitarily invariant norm subject to (6.3) is shown by Corollary A.2 in Appendix A.

We summarize our findings in a theorem.

Theorem 6.2. *Let $H \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{m\times m}$ be symmetric and $A \in \mathbb{R}^{n\times m}$, and let*

$$C = \begin{bmatrix} H & A \\ A^{\mathrm{T}} & -D \end{bmatrix}.$$

Assume C is nonsingular, and let $G = C^{-1}(1\colon n, 1\colon n)$. There exists a feasible perturbation in the definition of $\psi^{(k)}(C)$ if and only if G has at least k negative eigenvalues. Let $G = Q\,\mathrm{diag}\,(\gamma_i)Q^{\mathrm{T}}$ be a spectral decomposition, where Q is orthogonal and $\gamma_1 \leqslant \cdots \leqslant \gamma_n$. Then, for any unitarily invariant norm, an optimal perturbation in (6.1) is

$$\Delta H = -Q\,\mathrm{diag}\,(\gamma_1^{-1}, \ldots, \gamma_k^{-1}, 0, \ldots, 0)Q^{\mathrm{T}} \tag{6.4}$$

and, in terms of the underlying gauge function ϕ,

$$\psi^{(k)}(C) = \phi(\gamma_1^{-1}, \ldots, \gamma_k^{-1}, 0, \ldots, 0). \tag{6.5}$$

The perturbation (6.4) is full, in general, so may not be a suitable perturbation when H is large and sparse. It is natural, therefore, to consider diagonal perturbations. The next result shows that a perturbation consisting of a suitable multiple of the identity matrix is also optimal in the 2-norm. This result can be deduced from Theorem 6.2, but we give an independent proof for completeness.

Theorem 6.3. *Under the same conditions as in Theorem 6.2, an optimal perturbation in (6.1) in the 2-norm is*

$$\Delta H = -\gamma_k^{-1} I. \tag{6.6}$$

Proof. Consider perturbations to C of the form $\Delta C = WW^{\mathrm{T}}$ with

$$W = \begin{matrix} n \\ m \end{matrix}\overset{n}{\begin{bmatrix} \alpha I \\ 0 \end{bmatrix}}. \tag{6.7}$$

It is straightforward to prove an analogue of Lemma 6.1 which says that if $A \in \mathbb{R}^{n\times n}$ is symmetric and nonsingular, $W \in \mathbb{R}^{n\times k}$, and $p \leqslant k$ then $i_+(A + WW^{\mathrm{T}}) + i_0(A + WW^{\mathrm{T}}) = i_+(A) + i_0(A) + p$ if and only if $-I_k - W^{\mathrm{T}} A^{-1}W$ has exactly p nonnegative eigenvalues. Applying this result to Eq. (6.7), we find that $\Delta H = WW^{\mathrm{T}}$ is a feasible perturbation in Eq. (6.1) if and only if $-I_n - \alpha^2 G$ has k nonnegative eigenvalues, where $G = C^{-1}$

$(1\colon n,\ 1\colon n)$. We are assuming that G has a least k nonnegative eigenvalues, so the minimal value of α^2 is $-1/\gamma_k$. This gives $\|\Delta C\|_2 = -1/\gamma_k$, which, in view of (6.5), shows that (6.7) is an optimal perturbation in the 2-norm. □

Note that whereas the perturbation (6.4) increases $i_+ + i_0$ by exactly k, the perturbation (6.6) will increase it by more than k if $\gamma_k = \gamma_{k+1} = \cdots = \gamma_{k+r}$ with $r \geqslant 1$.

7. A projected Hessian approach

For the matrix C with $m \leqslant n$, there is an alternative way to find a perturbation to H of minimal norm such that the second order sufficiency condition (2.2) is satisfied. As noted earlier, the condition (2.2) is equivalent to the projected Hessain $Z^{\mathrm{T}}HZ$ being positive definite, where the columns of $Z \in \mathbb{R}^{n\times(n-m)}$ form a basis for $\mathrm{null}(A)^{\mathrm{T}}$, which we will take to be orthonormal. Therefore we are interested in solving the problem

$$\min\{\|\Delta H\|\colon Z^{\mathrm{T}}(H+\Delta H)Z \text{ is positive semidefinite}\}. \tag{7.1}$$

From Theorem 5.1 we know that an optimal *arbitrary* perturbation E that makes $Z^{\mathrm{T}}HZ + E$ positive semidefinite is, for any unitarily invariant norm,

$$E = U\,\mathrm{diag}\,(\max(-\mu_i, 0))U^{\mathrm{T}}, \tag{7.2}$$

where $Z^{\mathrm{T}}HZ = U\mathrm{diag}(\mu_i)U^{\mathrm{T}}$ with $\mu_1 \leqslant \cdots \leqslant \mu_{n-m}$ is a spectral decomposition. Hence any feasible ΔH in (7.1) satisfies

$$\|E\| \leqslant \|Z^{\mathrm{T}}\Delta H\, Z\| \leqslant \|Z^{\mathrm{T}}\|_2\|\Delta H\|\ \|Z\|_2 \leqslant \|\Delta H\|,$$

using an inequality for unitarily invariant norms form [19], p.211. But the perturbation (7.2) is achieved in (7.1) by setting $\Delta H = ZEZ^{\mathrm{T}}$, and $\|\Delta H\| \leqslant \|Z\|_2\|E\|\ \|Z^{\mathrm{T}}\|_2 \leqslant \|E\|$. We conclude that

$$\Delta H = ZU\,\mathrm{diag}(\max(-\mu_i, 0))U^{\mathrm{T}}Z^{\mathrm{T}} \tag{7.3}$$

is a solution to (7.1) for any unitarily invariant norm. For the 2-norm, another solution is

$$\Delta H = \max(-\mu_1, 0)ZZ^{\mathrm{T}}. \tag{7.4}$$

For the special case of the KKT matrix, for which (2.2) is equivalent to inertia $(K) = (n, m, 0)$ by Theorem 2.1, the perturbation (7.3), is necessarily, of the same norm as (6.4) for $k = n - i_+(K)$ in Theorem 6.2, although this equivalence is not obvious from the formulae.

When D is positive definite, or D is positive semidefinite and A has full rank, Lemma 4.2 shows that we could achieve the desired inertia $(n, m, 0)$ by choosing ΔH to make $H + \Delta H$ positive definite. Theorem 5.1 with $k = p$ shows that the smallest value of $\|\Delta H\|_2$ for which $H + \Delta H$ is positive semidefinite is $\max(-\lambda_{\min}(H), 0)$. By definition, this perturbation is at least as large as the optimal ones (6.4) and (7.3), and from (7.3) we have $\|\Delta H\|_2 \leqslant \max(-\lambda_{\min}(Z^T HZ), 0)$, which can be arbitrarily smaller than $\max(-\lambda_{\min}(H), 0)$. We note, in particular, that the perturbations (6.4), (6.6), (7.3) and (7.4) all have 2-norms uniformly bounded by $\|H\|_2$, which is an important property for optimization applications [15].

We give a numerical example for illustration. Consider the KKT matrix

$$K = \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -100 & 1 \\ \hline 0 & 1 & 0 \end{array}\right],$$

$$\lambda(K) = \{-1.0002 \times 10^2, -9.9000 \times 10^{-1}, 1.0099 \times 10^{-2}\},$$

where the eigenvalues are given to five significant figures. Hence inertia $(K) = (1, 2, 0)$, and we want to change the inertia to (2,1,0). Since

$$K^{-1} = \left[\begin{array}{cc|c} -1 & 0 & 1 \\ 0 & 0 & 1 \\ \hline 1 & 1 & 99 \end{array}\right],$$

we find immediately from Theorem 6.2 with $k = 1$ that

$$\Delta H = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{7.5}$$

is a matrix of smallest norm, for any unitarily invariant norm, that changes the inertia of K to (1,1,1); indeed,

$$K + \Delta K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -100 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\lambda(K + \Delta K) = \{-1.0002 \times 10^2, 0, 1.9996 \times 10^{-2}\}.$$

For the projected Hessian approach we have $Z = [1 \quad 0]^T$, $Z^T HZ = -1$, and (7.3) again yields the perturbation (7.5). To achieve the inertia (2,1,0) that is required for the condition (2.2) to hold, we can replace ΔH by $(1 + \epsilon)\Delta H$ for any $\epsilon > 0$.

In order to perturb H to make it positive definite, which also produces the desired inertia, we must make a perturbation of 2-norm at least $-\lambda_{\min}(H) = 1.0001 \times 10^2$, which is two orders of magnitude larger than the minimal-norm perturbation (7.5).

8. Practical algorithm

We now turn to the optimization applications. We consider the situation where a linear system $Cx = b$ must be solved, but C needs to be perturbed in its (1,1) block, if necessary, to ensure that it has the inertia $(n, m, 0)$.

We assume that an LBL^{T} factorization of C is computed,

$$PCP^{\mathrm{T}} = LBL^{\mathrm{T}},$$

where L is unit lower triangular and B is block diagonal blocks of dimension 1 or 2; P is a permutation matrix that can be chosen according to one of various pivoting strategies. Since C and B have the same inertia, it is trivial to evaluate the inertia of C. If $i_+(C)$ is less than n, then Theorem 6.2 shows that to determine the optimal perturbation (6.4) we need to compute the $k = n - i_+(C)$ most negative eigenvalues of $G = C^{-1}(1\colon n, 1\colon n)$ and their corresponding eigenvectors; for the optimal 2-norm perturbation (6.6) it suffices to determine the kth most negative eigenvalue of G. To confirm that there are k negative eigenvalues of G, we apply Cauchy's interlace theorem, which yields

$$\lambda_i(G) \leqslant \lambda_{i+m}(C^{-1}), \quad i = 1 : n.$$

Hence if C has only $i_+(C) < n$ positive eigenvalues then G has at least $n - i_+(C)$ negative eigenvalues.

Since C may be large and sparse it is undesirable to form G explicitly. Therefore we suggest that the k most negative eigenvalues of G be computed using the Lanczos algorithm, which requires only the ability to form matrix-vector products with G. To form $y = Gx$ we note that

$$\begin{bmatrix} y \\ z \end{bmatrix} = C^{-1} \begin{bmatrix} x \\ 0 \end{bmatrix},$$

where $z \in \mathbb{R}^m$ is not of interest. Hence y is the first n components of the solution to the linear system

$$C \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix},$$

which can be solved using the LBL^{T} factorization.

Note that the perturbation (6.4) makes $C + \Delta C$ singular, since it moves k negative eigenvalues to the origin. Similarly, the perturbation (6.6) produces at least one zero eigenvalue. In practice a nonsingular $C + \Delta C$ is required, and the natural approach is to modify the perturbations so that the eigenvalues are moved to a positive tolerance δ instead of 0.

Having computed an optimal perturbation ΔH we have to refactorize $C + \Delta C$ in order to solve $(C + \Delta C)x = b$. It does not seem practical to apply updating techniques to the original factorization, since the update may not be of low rank.

In the case where $D = 0$, our algorithm provides, as a by-product, a direction of negative curvature, which is defined as a vector p for which [cf. (2.2)] $A^{\mathrm{T}}p = 0$ and $p^{\mathrm{T}}Hp < 0$. Such directions are needed in nonlinear programming to achieve convergence to points that satisfy second order necessary conditions for optimality. Writing the perturbation (6.4) as $\Delta H = VV^{\mathrm{T}}$, we know that the matrix (6.2), which we denote by S, is negative definite. Now

$$S = \begin{bmatrix} V^{\mathrm{T}} & 0^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = V^{\mathrm{T}}X, \quad \text{where} \quad \begin{bmatrix} H & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} V \\ 0 \end{bmatrix}.$$

Thus $HX + AY = V$ and $A^{\mathrm{T}}X = 0$, which implies $X^{\mathrm{T}}HX = S^{\mathrm{T}} = S$. The jth column x_j of X satisfies $x_j^{\mathrm{T}}Hx_j = s_{jj} < 0$, since S is negative definite, and $A^{\mathrm{T}}x_j = 0$. Thus, every column of X is a direction of negative curvature.

An alternative approach is to work with the projected Hessian $Z^{\mathrm{T}}HZ$ and to compute an optimal perturbation ΔH from (7.3) or (7.4). Again, the Lanczos algorithm can be used, this time to compute the negative eigenvalues of $Z^{\mathrm{T}}HZ$. This technique is already in use by some researchers (Gould, private communication).

Numerical experiments with the algorithms described above will be reported elsewhere.

Acknowledgements

Nick Gould suggested the problem of modifying the inertia of the KKT matrix by perturbing its (1,1) block, and some of the techniques we have used were inspired by his paper [15]. Philip Gill suggested extending our results to the more general primal-dual matrix with a negative definite (2,2) block. The first author thanks Ludwig Elsner for an invitation to visit the Department of Mathematics, University of Bielefeld, where some of this work was carried out with the support of SFB 343 (Diskrete Strukturen in der Mathematik). We thank Martin Hanke and Michael Saunders for suggesting improvements to the previous version of this paper.

Appendix A

Lemma A.1. *Let* $A \in \mathbb{R}^{n\times n}$ *be symmetric with the spectral decomposition* $A = Q\,\mathrm{diag}(\lambda_i)Q^{\mathrm{T}}$, *where* Q *is orthogonal and*

$$\lambda_1 \leqslant \cdots \leqslant \lambda_{p-1} \leqslant 0 < \lambda_p \leqslant \cdots \leqslant \lambda_n.$$

Let $X \in \mathbb{R}^{n\times k}$ *with* $k \leqslant n$, *and assume that* $p \leqslant n - k + 1$. *All matrices* X *that minimize all the singular values of* X *subject to satisfying the inequalities*

$$\lambda_i(X^{\mathrm{T}}AX) \geqslant 1, \quad i = 1\colon k, \tag{A.1}$$

are given by

$$X = Q(1\colon n, n-k+1\colon n)\mathrm{diag}\,(\lambda_{n-k+1}, \ldots, \lambda_n)^{-1/2}\ V, \tag{A.2}$$

where $V \in \mathbb{R}^{k\times k}$ *is an arbitrary orthogonal matrix.*

Proof. Let

$$X = U\Sigma V^{\mathrm{T}}, \qquad \Sigma = \begin{bmatrix} D \\ 0 \end{bmatrix}, \quad D = \mathrm{diag}(\sigma_i) \in \mathbb{R}^{k\times k}$$

be an SVD. Then

$$X^{\mathrm{T}}AX - I = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}AU\Sigma V^{\mathrm{T}} - I = V(DBD - I)V^{\mathrm{T}},$$

where $B = (U^{\mathrm{T}}AU)(1\colon k, 1\colon k)$. The constraint (A.1) is therefore equivalent to $DBD - I$ being positive semidefinite, which implies that

$$b_{ii} \geqslant \frac{1}{\sigma_i^2}, \quad i = 1 : k. \tag{A.3}$$

We wish to maximize the reciprocals σ_i^{-2}. Now the diagonal of the symmetric matrix B is largest when it contains the eigenvalues of B, that is, when B is diagonal, and the maximum over all U occurs when $B = \mathrm{diag}(\lambda_{n-k+1}, \ldots, \lambda_n)$. When B is diagonal, (A.3) is *equivalent* to (A.1). Hence for optimality we need to choose $U = Q(1\colon n, n-k+1\colon n)$ and then, to attain the bounds in (A.3), $\sigma_i = \lambda_{n-k+i}^{-1/2}$ (note that the σ_i are arranged in decreasing order). The matrix V is arbitrary. □

Corollary A.2. *Under the conditions of Lemma A.1 the matrix (A.2) minimizes* $\|XX^{\mathrm{T}}\|$ *subject to (A.1) for any unitarily invariant norm.*

Proof. The singular values of XX^{T} are the squares of the singular values of X, which are minimized by the matrix (A.2). The result follows from the gauge function property of unitarily invariant norms.

References

[1] P. Arbenz, W. Gander, G.H. Golub, Restricted rank modification of the symmetric eigenvalue problem: Theoretical considerations, Linear Algebra Appl. 104 (1988) 75–95.

[2] P. Arbenz, G.H. Golub, On the spectral decomposition of Hermitian matrices modified by low rank perturbations with applications, SIAM J. Matrix Anal. Appl. 9 (1) (1988) 40–58.

[3] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996, xvii + 408 pp., ISBN 0-89871-360-9.

[4] D. Carlson, H. Schneider, Inertia theorems for matrices: The semidefinite case, J. Math. Anal. Appl. 6 (3) (1963) 430–446.
[5] S.H. Cheng, N.J. Higham, A modified Cholesky algorithm based on a symmetric indefinite factorization, Numerical Analysis Report 289, Manchester Centre for Computational Mathematics, Manchester, England, 1996, pp. 18; SIAM J. Matrix Anal. Appl. (to appear).
[6] A.R. Conn, N.I.M. Gould, P.L. Toint, LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A), Springer, Berlin, 1992, pp. xviii+330, ISBN 0-387-55470-X.
[7] B. Fischer, A. Ramage, D.J. Silvester, A.J. Wathen, Minimum Residual Methods for Augmented Systems, Mathematics Research Report 15, Department of Mathematics, University of Strathclyde, Glasgow, Scotland, 1995, pp. 17.
[8] A. Forsgren, P.E. Gill, Primal-Dual Interior Methods for Nonconvex Nonlinear Programming, Technical Report NA 96-3, Department of Mathematics, University of California, San Diego, 1996, pp. 24, to appear in SIAM J. Optimization
[9] A. Forsgren, P.E. Gill, J.R. Shinnerl, Stability of symmetric ill-conditioned systems arising in interior methods for constrained optimization, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 187–211.
[10] A. Forsgren, W. Murray, Newton methods for large-scale linear equality-constrained minimization, SIAM J. Matrix Anal. Appl. 14 (2) (1993) 560–587.
[11] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization Academic Press, London, 1981, pp. xvi+401, ISBN 0-12-283952-8.
[12] P.E. Gill, M.A. Saunders, J.R. Shinnerl, On the stability of Cholesky factorization for symmetric quasidefinite systems, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 35–46.
[13] N.I.M. Gould, On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem, Math. Programming 32 (1985) 90–99.
[14] N.I.M. Gould, On the accurate determination of search directions for simple differentiable penalty functions, IMA J. Numer. Anal. 6 (1986) 357–372.
[15] N.I.M. Gould, Constructing appropriate models for large-scale, linearly constrained, nonconvex, nonlinear optimization algorithms, Report RAL-95-037, Atlas Centre, Rutherford Appleton Laboratory, Didcot, Oxon, UK, 1995, pp. 38.
[16] E.V. Haynsworth, A.M. Ostrowski, On the inertia of some classes of partitioned matrices, Linear Algebra Appl. 1 (1968) 299–316.
[17] N.J. Higham, Computing a nearest symmetric positive semidefinite matrix, Linear Algebra Appl. 103 (1988) 103–118.
[18] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985, pp. xiii+561, ISBN 0-521-30586-1.
[19] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991, pp. viii+607, ISBN 0-521-30587-X.
[20] A.M. Ostrowski, A quantitative formulation of Sylvester's law of inertia, Proc. Nat. Acad. Sci. 45 (1959) 740–744.
[21] B.N. Parlett, The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, 1980, pp. xix+348, ISBN 0-13-880047-2.
[22] M.A. Saunders, Solution of sparse rectangular systems using LSQR and CRAIG, BIT 35 (1995) 588–604.
[23] R.B. Schnabel, E. Eskow, A new modified Cholesky factorization, SIAM J. Sci. Statist. Comput. 11 (6) (1990) 1136–1158.
[24] D.J. Silvester, A.J. Wathen, Fast and robust solvers for time-discretised incompressible Navier-Stokes equations, in: D.F. Griffiths, G.A. Watson (Eds.), Numerical Analysis 1995, Proceedings of the 16th Dundee Conference Pitman Res. Notes Math. Ser. 344, Addison Wesley Longman, Essex, UK, 1996, pp. 154–168.
[25] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, London, 1990, pp. x+365, ISBN 0-12-670230-6.

[26] H.P.M. Von Kempen, Variation of the eigenvalues of a special class of Hermitian matrices upon variation of some of its elements, Linear Algebra Appl. 3 (1970) 263–273.
[27] R.J. Vanderbei, Symmetric quasi-definite matrices, SIAM J. Optim. 5 (1) (1995) 100–113.
[28] H.K. Wimmer, On Ostrowski's generalization of Sylvester's law of inertia, Linear Algebra Appl. 52/53 (1983) 739–741.