Symmetric indefinite factorization of quasidefinite matrices*

SANJA SINGER† AND SAŠA SINGER‡

Abstract. *Matrices with special structures arise in numerous applications. In some cases, such as quasidefinite matrices or their generalizations, we can exploit this special structure. If the matrix H is quasidefinite, we propose a new variant of the symmetric indefinite factorization. We show that linear system $Hz = b$, H quasidefinite with a special structure, can be interpreted as an equilibrium system. So, even if some blocks in H are ill–conditioned, the important part of solution vector z can be accurately computed. In the case of a generalized quasidefinite matrix, we derive bounds on number of its positive and negative eigenvalues.*

Key words: *quasidefinite matrices, inertia, special linear systems, accurate solution*

AMS subject classifications: 15A23, 65F05, 65G06

1. Introduction

Quasidefinite matrices have the form

$$H = \begin{bmatrix} A & B \\ B^* & -D \end{bmatrix}, \tag{1}$$

where A and D are symmetric (Hermitian) and positive definite. This class of matrices arises in optimization when barrier or interior–point methods are applied (see [6], [1]).

For example, to solve the damped linear least squares problem

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

*This paper was presented at the Special Section: Appl. Math. Comput. at the 7^{th} *International Conference on OR, Rovinj, 1998.*

†Faculty of Mechanical Engineering and Naval Architecture, I. Lučića 5, HR-10 000 Zagreb, Croatia, e-mail: `ssinger@math.hr`

‡Department of Mathematics, University of Zagreb, Bijenička 30, HR-10 000 Zagreb, Croatia, e-mail: `singer@math.hr`

one needs to solve the linear system

$$\begin{bmatrix} A & B \\ B^{\tau} & -D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{2}$$

with $A = D = \delta I$, $\delta > 0$.

2. Factorization of quasidefinite matrices

It is a well known fact that a symmetric indefinite factorization of the matrix (1) can be computed using only 1×1 pivots. Here, we give a new, elementary and constructive proof of this fact.

Let $A = G_1^* G_1$ and $D = G_3^* G_3$ be Cholesky factorizations of A and D, respectively. The matrix H can be factorized as

$$H = \begin{bmatrix} G_1^* & 0 \\ T^* & G_2^* \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & -I_\ell \end{bmatrix} \begin{bmatrix} G_1 & T \\ 0 & G_2 \end{bmatrix} = \begin{bmatrix} G_1^* G_1 & G_1^* T \\ T^* G_1 & T^* T - G_2^* G_2 \end{bmatrix} \tag{3}$$

with

$$\begin{aligned} T &= G_1^{-*} B, \\ G_2^* G_2 &= D + B^* A^{-1} B = G_3^* G_3 + T^* T. \end{aligned} \tag{4}$$

From (4) we see that the matrix G_2 can be computed as the upper triangular factor from the QR factorization

$$\begin{bmatrix} T \\ G_3 \end{bmatrix} = Q \begin{bmatrix} G_2 \\ 0 \end{bmatrix}.$$

From (3) we can also conclude a somewhat surprising fact that the inertia of H does not depend on B – the matrix H has exactly k positive and ℓ negative eigenvalues.

To ensure a numerical stability we usually use Bunch–Kaufman or Bunch–Parlett pivoting. But, if the matrices A and D are both diagonal, this sparsity pattern cannot be preserved. In this case, we can use diagonal pivoting in blocks A and D to reduce the fill–in of nonzero elements.

Another reason why permutations should be avoided lies in parallel computation. An algorithm which includes pivoting or permutation processes is almost strictly sequential and cannot take advantage of parallel processing.

But, here lies a danger. If pivots in A are too small compared with elements of B, big elements can occur in T and G_2. Note, if A and D are diagonal, computed elements of G_1 and T have small relative errors. The diagonals of G_1 and G_3 are computed using only square roots and off diagonal elements of G_1 and G_3 are exactly 0. Elements of T are computed from B and G_1 using only one division for each of them. For errors in computed matrices we have

$$\begin{aligned} |(\delta G_1)_{ii}| &\leq \varepsilon |(G_1)_{ii}|, \\ |(\delta G_3)_{jj}| &\leq \varepsilon |(G_3)_{jj}|, \end{aligned} \quad |(\delta T)_{ij}| \leq \varepsilon |(T)_{ij}|, \quad 1 \leq i \leq k,\ 1 \leq j \leq \ell, \tag{5}$$

where ε is a rounding error of basic floating–point operations.

What happens with the computed G_2? Using Higham [2] Lemma 18.8., we may interpret the computed QR factorization of the matrix

$$M = \begin{bmatrix} \widetilde{T} \\ \widetilde{G}_3 \end{bmatrix} := \begin{bmatrix} T + \delta T \\ G_3 + \delta G_3 \end{bmatrix} \tag{6}$$

as an exact factorization of a slightly perturbed matrix $M + \delta M$. The backward error matrix δM satisfies

$$|\delta M| \leq f(m, \ell)\, \varepsilon\, m\, G\, |M|, \tag{7}$$

where $m = k + \ell$ and ℓ are dimensions of M, $f(m, \ell)$ is a moderate polynomial and $G = m^{-1} e e^{\tau}$, with $e = [1, 1, \ldots, 1]^{\tau}$.

The following theorem can be proved.

Theorem 1. *Let H be a quasidefinite matrix with A and D diagonal. Let G_2 in (3) be obtained by QR factorization. Suppose that in the floating point arithmetic the symmetric indefinite factorization produces matrix*

$$\widetilde{R} = \begin{bmatrix} \widetilde{G}_1 & \widetilde{T} \\ 0 & \widetilde{G}_2 \end{bmatrix}.$$

Then $\widetilde{R}$ is an exact factor of the perturbed matrix

$$\widetilde{H} := H + \delta H = \begin{bmatrix} A + \delta A & B + \delta B \\ B^* + \delta B^* & -D - \delta D \end{bmatrix}$$

with

$$\begin{aligned} |\delta A| &\leq \varepsilon_p A \\ |\delta B| &\leq \varepsilon_p |B| \\ |\delta D| &\leq \varepsilon_p D + \varepsilon_m |M^*|\, G\, |M| \\ &\leq \varepsilon_p D + (1+\varepsilon)^2 \varepsilon_m \begin{bmatrix} |B^*| G_1^{-1} & G_3 \end{bmatrix} G \begin{bmatrix} G_1^{-1} |B| \\ G_3 \end{bmatrix}, \end{aligned}$$

where $m = k + \ell$, $\varepsilon_p := 2\varepsilon + \varepsilon^2$, $\varepsilon_m := \varepsilon f(m, \ell)(2m + \varepsilon f(m, \ell) m^3)$.

Proof. Multiplication of computed factors gives

$$\widetilde{H} = \begin{bmatrix} \widetilde{G}_1^* \widetilde{G}_1 & \widetilde{G}_1^* \widetilde{T} \\ \widetilde{T}^* \widetilde{G}_1 & \widetilde{T}^* \widetilde{T} - \widetilde{G}_2^* \widetilde{G}_2 \end{bmatrix}.$$

From (5), it follows that $|\delta G_1| \leq \varepsilon |G_1|$, and

$$\widetilde{G}_1^* \widetilde{G}_1 = G_1^* G_1 + \delta G_1^* G_1 + G_1^* \delta G_1 + \delta G_1^* \delta G_1 := A + \delta A,$$

or

$$|\delta A| \leq (2\varepsilon + \varepsilon^2) A.$$

The bound for $|\delta B|$ can be obtained similarly, if we notice that $|B| = G_1|T|$.

Finally, from (6) it follows

$$\begin{aligned} -D - \delta D &= \widetilde{T}^*\widetilde{T} - \widetilde{G}_2^*\widetilde{G}_2 \\ &= \widetilde{T}^*\widetilde{T} - (M^* + \delta M^*)(M + \delta M) \\ &= -\widetilde{G}_3^*\widetilde{G}_3 - \delta M^* M - M^*\delta M - \delta M^*\delta M. \end{aligned}$$

By definition of $\widetilde{G}_3$, we have

$$\widetilde{G}_3^*\widetilde{G}_3 = D + \delta G_3^* G_3 + G_3^*\delta G_3 + \delta G_3^*\delta G_3,$$

and

$$\begin{aligned} |\delta D| &\leq |\delta G_3^* G_3 + G_3^*\delta G_3 + \delta G_3^*\delta G_3 + \delta M^* M + M^*\delta M + \delta M^*\delta M| \\ &\leq \varepsilon_p D + |\delta M^*|\,|M| + |M^*|\,|\delta M| + |\delta M^*|\,|\delta M|. \end{aligned}$$

The first inequality for $|\delta D|$ follows from (7). The second inequality is a consequence of

$$|M| = \begin{bmatrix} |T + \delta T| \\ |G_3 + \delta G_3| \end{bmatrix} \leq (1+\varepsilon) \begin{bmatrix} |T| \\ |G_3| \end{bmatrix} = (1+\varepsilon) \begin{bmatrix} G_1^{-1}|B| \\ G_3 \end{bmatrix}.$$

□

It is easy to see that the computed factorization of H is “good” if the elements of $|M^*|\,G\,|M|$ are comparable in magnitude with the elements of D.

3. Inertia of generalized quasidefinite matrices

Higham and Cheng [3] have studied the inertia of matrices arising in optimization. In view of this, it is an interesting question how the inertia of H depends on off–diagonal block B, if A and D from (1) are Hermitian, nonsingular, but possibly indefinite.

Definition 1. *A nonsingular matrix H is generalized quasidefinite if the blocks A and D from (1) are nonsingular and indefinite.*

In this case we can give some bounds on the inertia of H. The inertia of H is an ordered triple (i_+, i_-, i_0), where i_+, i_- and i_0 are numbers of positive, negative and zero eigenvalues of H, respectively.

Theorem 2. *Let H be a generalized quasidefinite matrix. Then*

$$\mathrm{inertia}(H) = (i_+(H), i_-(H), 0)$$

satisfies

$$\max\{i_+(A), \ell - i_+(D)\} \leq i_+(H) \leq \min\{\ell + i_+(A), k + i_-(D)\},$$

and

$$i_-(H) = k + \ell - i_+(H).$$

Proof. If A and D are nonsingular, they have symmetric indefinite factorizations

$$A = P_1^* G_1^* J_1 G_1 P_1, \quad D = P_3^* G_3^* J_3 G_3 P_3,$$

where, for $i = 1, 3$, P_i are permutation matrices, J_i are signature matrices and G_i are block upper triangular matrices with diagonal blocks of order 1 or 2.

Let $T = J_1 G_1^{-*} P_1 B$ and $S = -D - T^* J_1 T$. The matrix H can be written as

$$H = \begin{bmatrix} P_1^* G_1^* & 0 \\ T^* & I_\ell \end{bmatrix} \begin{bmatrix} J_1 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} G_1 P_1 & T \\ 0 & I_\ell \end{bmatrix} = \begin{bmatrix} P_1^* G_1 J_1 G_1 P_1 & P_1^* G_1^* J_1 T \\ T^* J_1 G_1 P_1 & T^* J_1 T + S \end{bmatrix}.$$

Since H is nonsingular, we conclude that S is nonsingular too, and it can be factorized as

$$S = P_2^* G_2^* J_2 G_2 P_2.$$

Then

$$-D = T^* J_1 T + S,$$

or

$$\begin{aligned} P_2^* G_2^* J_2 G_2 P_2 &= -P_3^* G_3^* J_3 G_3 P_3 - T^* J_1 T \\ &= \begin{bmatrix} P_3^* G_3^* & T^* \end{bmatrix} \begin{bmatrix} -J_3 & 0 \\ 0 & -J_1 \end{bmatrix} \begin{bmatrix} G_3 P_3 \\ T \end{bmatrix}. \end{aligned}$$

Let $\text{inertia}(A) = (i_+(A), i_-(A), 0)$ and $\text{inertia}(D) = (i_+(D), i_-(D), 0)$. Nonsingularity of S implies that G_2 can be obtained by the indefinite QR factorization of

$$\begin{bmatrix} G_3 P_3 \\ T \end{bmatrix}$$

where the indefinite inner product is generated by the matrix

$$J := \begin{bmatrix} -J_3 & 0 \\ 0 & -J_1 \end{bmatrix}$$

(see Singer [5]). The inertia of J is equal to

$$\text{inertia}(J) = (i_-(A) + i_-(D), i_+(A) + i_+(D), 0).$$

Construction of the indefinite factorization implies that

$$\text{inertia}(J_2) = (i_+(J_2), \ell - i_+(J_2), 0),$$

and

$$\begin{aligned} 0 \;\leq\; & i_+(J_2) \;\leq\; i_-(A) + i_-(D) \\ 0 \;\leq\; & \ell - i_+(J_2) \;\leq\; i_+(A) + i_+(D), \end{aligned}$$

or

$$\max\{0, \ell - i_+(A) - i_+(D)\} \leq i_+(J_2) \leq \min\{\ell, i_-(A) + i_-(D)\}.$$

Finally, we conclude that

$$\max\{i_+(A), \ell - i_+(D)\} \leq i_+(H) \leq \min\{\ell + i_+(A), k + i_-(D)\}.$$

□

It is easy to show that the bounds from the previous theorem are nontrivial.

Example 1. *Let H be a generalized quasidefinite matrix with indefinite blocks A and D, such that* $\operatorname{inertia}(A) = (1, 7, 0)$ *and* $\operatorname{inertia}(D) = (2, 4, 0)$. *From the previous theorem we have*

$$\max\{1, 6-2\} \le i_+(H) \le \min\{6+1, 8+4\},$$

or

$$4 \le i_+(H) \le 7, \quad i_-(H) = 14 - i_+(H).$$

4. Accurate solution of special linear systems

Finally, we consider a special case of quasidefinite linear system (2), when A and D are diagonal matrices, with emphasis on accurate computation of the y part of the solution z.

Such a system is a generalization of the so-called equilibrium system

$$\begin{bmatrix} \widehat{A} & \widehat{B} \\ \widehat{B}^\tau & 0 \end{bmatrix} \begin{bmatrix} \widehat{x} \\ \widehat{y} \end{bmatrix} = \begin{bmatrix} \widehat{b} \\ 0 \end{bmatrix}, \tag{8}$$

with diagonal positive definite $\widehat{A}$ and full column rank $\widehat{B}$.

This system can be written as

$$\begin{aligned} \widehat{A}\widehat{x} + \widehat{B}\widehat{y} &= \widehat{b} \\ \widehat{B}^\tau \widehat{x} &= 0. \end{aligned}$$

The first equation gives $\widehat{x} = \widehat{A}^{-1}(\widehat{b} - \widehat{B}\widehat{y})$. Substituting this into the second equation gives

$$\widehat{y} = (\widehat{B}^\tau \widehat{A}^{-1} \widehat{B})^{-1} \widehat{B}^\tau \widehat{A}^{-1} \widehat{b}. \tag{9}$$

So, if we can bound $\|(\widehat{B}^\tau \widehat{A}^{-1} \widehat{B})^{-1} \widehat{B}^\tau \widehat{A}^{-1}\|$ independent of $\widehat{A}$, $\widehat{y}$ cannot be much larger than $\widehat{b}$, no matter how ill–conditioned $\widehat{A}$ is.

Vavasis in [7] proves the following result (originally proved independently by Stewart and Todd).

Theorem 3. *Let $\mathcal{A}$ denote the set of all $k \times k$ positive definite real diagonal matrices. Let $\widehat{B}$ be a $k \times \ell$ real matrix of rank ℓ. Then there exist constants $\chi_{\widehat{B}}$ and $\bar{\chi}_{\widehat{B}}$ such that for any $\widehat{A} \in \mathcal{A}$*

(a) $\|(\widehat{B}^\tau \widehat{A}^{-1} \widehat{B})^{-1} \widehat{B}^\tau \widehat{A}^{-1}\| \le \chi_{\widehat{B}}$, *and*

(b) $\|\widehat{B}(\widehat{B}^\tau \widehat{A}^{-1} \widehat{B})^{-1} \widehat{B}^\tau \widehat{A}^{-1}\| \le \bar{\chi}_{\widehat{B}}$.

Here we assume that the matrix norm $\|\ \|$ is induced by some vector norm. □

Our quasidefinite problem

$$\begin{bmatrix} A & B \\ B^\tau & -D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \tag{10}$$

can be written in a componentwise form as

$$\begin{aligned} Ax + By &= b \\ B^\tau x - Dy &= 0. \end{aligned}$$

Similarly, from the first equation we have $x = A^{-1}(b - By)$, and

$$y = (B^\tau A^{-1} B + D)^{-1} B^\tau A^{-1} b. \tag{11}$$

If $\|(B^\tau A^{-1}B + D)^{-1}B^\tau A^{-1}\|$ is bounded and depends only on B and D, ill–condition of A should not destroy the components of y. To show this, we transform the system (10) into the form (8).

Let $D = \Delta^2$ be the Cholesky factorization of D. Define

$$\widehat{A} = \begin{bmatrix} A & 0 \\ 0 & I_\ell \end{bmatrix}, \quad \widehat{B} = \begin{bmatrix} B \\ \Delta \end{bmatrix}, \quad \widehat{b} = \begin{bmatrix} b \\ 0 \end{bmatrix}. \tag{12}$$

It is obvious that if A is positive definite, $\widehat{A}$ is positive definite too. Also, matrix $\widehat{B}$ has full column rank because Δ is a nonsingular diagonal matrix.

The connection between the solution $\widehat{y}$ of linear system (8) and the y–part of (10) is very simple.

Theorem 4. *If matrices $\widehat{A}$ and $\widehat{B}$ and vector $\widehat{b}$ are defined by (12) then $\widehat{y} = y$.*

Proof. From (9), using (11) and (12), it follows

$$\begin{aligned} \widehat{y} &= \left(\begin{bmatrix} B^\tau & \Delta \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & I_\ell \end{bmatrix} \begin{bmatrix} B \\ \Delta \end{bmatrix} \right)^{-1} \begin{bmatrix} B^\tau & \Delta \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I_\ell \end{bmatrix}^{-1} \begin{bmatrix} b \\ 0 \end{bmatrix} \\ &= (B^\tau A^{-1} B + D)^{-1} \begin{bmatrix} B^\tau A^{-1} & \Delta \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} \\ &= (B^\tau A^{-1} B + D)^{-1} B^\tau A^{-1} b = y. \end{aligned}$$

□

Vavasis [7] and Hough and Vavasis [4] have developed algorithms for stable computation of $\widehat{y}$ component for the linear system (8). *Theorem 4* shows that these algorithms are suitable for quasidefinite linear systems (10) as well.

References

[1] 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**(1996), 35–46.

[2] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.

[3] N. J. HIGHAM, S. H. CHENG, *Modifying the inertia of matrices arising in optimization*, Linear Algebra Appl. **275–276**(1998), 261–279.

[4] P. D. HOUGH, S. A. VAVASIS, *Complete orthogonal decomposition for weighted least squares*, SIAM J. Matrix Anal. Appl. **18**(1997), 369–392.

[5] S. SINGER, *Indefinite QR Factorization and Its Applications*, Ph.D. thesis, Department of Mathematics, University of Zagreb, 1997 (In Croatian).

[6] R. J. VANDERBEI, *Symmetric quasidefinite matrices*, SIAM J. Optim. **5**(1995), 100–113.

[7] S. A. VAVASIS, *Stable numerical algorithms for equilibrium systems*, SIAM J. Matrix Anal. Appl. **15**(1994), 1108–1131.