CHOLESKY-LIKE FACTORIZATION OF SYMMETRIC INDEFINITE MATRICES AND ORTHOGONALIZATION WITH RESPECT TO BILINEAR FORMS

M. ROZLOŽNíK [†¶], F. OKULICKA-DłUŻEWSKA [§], AND A. SMOKTUNOWICZ [§]

Abstract. It is well-known that orthogonalization of column vectors in a rectangular matrix B with respect to the bilinear form induced by a nonsingular symmetric indefinite matrix A can be seen as its factorization $B = QR$ that is equivalent to the Cholesky-like factorization in the form $B^TAB = R^T\Omega R$, where Ω is some signature matrix. Under the assumption of nonzero principal minors of the matrix $M = B^TAB$ we give bounds for the conditioning of the triangular factor R in terms of extremal singular values of M and of only those principal submatrices of M, where there is a change of sign in Ω. Using these results we study the numerical behavior of two types of orthogonalization schemes and we give the worst-case bounds for quantities computed in finite precision arithmetic. In particular, we analyze the implementation based on the Cholesky-like factorization of M and the Gram-Schmidt process with respect to the bilinear form induced by the matrix A. To improve the accuracy of computed results we consider also the Gram-Schmidt process with reorthogonalization and show that its behavior is similar to the scheme based on the Cholesky-like factorization with one step of iterative refinement.

Key words. Symmetric indefinite matrices, Cholesky-like factorization, orthogonalization techniques, indefinite bilinear forms, Gram-Schmidt process, rounding error analysis.

AMS subject classifications. 65F25, 65F35, 65F05, 65G50.

1. Introduction. For a real symmetric (in general indefinite) nonsingular matrix $A \in \mathscr{R}^{m,m}$ and for a full column rank matrix $B \in \mathscr{R}^{m,n}$ ($m \geq n$) we look for a factorization $B = QR$, where $Q \in \mathscr{R}^{m,n}$ is so-called left (A,Ω)-orthogonal, i.e. its columns are mutually orthogonal with respect to the bilinear form induced by the matrix A, with $Q^TAQ = \Omega \in \mathscr{R}^{n,n}$ being a signature matrix $\Omega \in \mathrm{diag}(\pm 1)$, and where $R \in \mathscr{R}^{n,n}$ is upper triangular with positive diagonal elements. Note that the full-column rank condition of the matrix B is not enough for the existence of the factors Q and R such that Q is left (A,Ω)-orthogonal and R is upper triangular with positive diagonal entries. It is also easy to see that if the factorization $B = QR$ exists, it can be regarded as an implicit Cholesky-like factorization of the symmetric indefinite matrix $M = B^TAB = R^T\Omega R$ (without its explicit computation), delivering the same upper triangular factor R. Conversely, given the Cholesky-like factorization of M, the left (A,Ω)-orthogonal factor Q can be then recovered as $Q = BR^{-1}$. Such problems appear explicitly [15] or implicitly in many applications such as eigenvalue problems, matrix pencils and structure-preserving algorithms [22, 26], saddle point problems and optimization with interior-point methods [13, 37, 30] or indefinite least squares problems [4, 9, 24, 25].

It is clear that for $A = I$ we get the standard QR factorization of B that corresponds to the left (I,I)-orthogonal Q satisfying $Q^TQ = I$ (see, e.g., [19]). In the case of symmetric positive definite A, this matrix induces a non-standard inner product and the (A,I)-orthogonal factor we look for can be still recovered from the (I,I)-orthogonal factor in the QR factorization of the matrix $A^{-1/2}B$, where $A^{1/2}$ denotes the matrix square root of A. In addition, the upper triangular factor R is the Cholesky factor of the matrix $M = B^TAB = R^TR$. The indefinite case with a diagonal $A \in \mathrm{diag}(\pm 1)$ has been studied intensively by several authors [5, 8, 10, 12, 32, 31, 29]. These concepts can be extended also to the case of a general symmetric indefinite (but still nonsingular) matrix A. The matrix $M = B^TAB$ is then also symmetric indefinite and there

[†]Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, CZ-182 07 Prague 8, Czech Republic. E-mail: miro@cs.cas.cz

[§]Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, Warsaw, 00-662 Poland. E-mail: F.Okulicka@mini.pw.edu.pl, smok@mini.pw.edu.pl

[¶]This research is supported by the Grant Agency of the Czech Republic under the project 108/11/0853 and by the international collaboration support of the Academy of Sciences of the Czech Republic.

exists its LDL^T factorization $P^TMP = LDL^T$, where P is a permutation matrix representing some pivoting strategy, L is unit lower triangular, and D is block diagonal with diagonal blocks of dimension 1 and 2. For details we refer to the papers of Bunch [6] or Bunch and Parlett [7]. Considering the eigenvalue decomposition of D in the form $D = S\Lambda S^T = S|\Lambda|^{1/2}\Omega|\Lambda|^{1/2}S^T$, where S is also block diagonal with diagonal blocks of dimension 1 and 2, Λ is diagonal and Ω is its signature matrix, the LDL^T factorization of M can be rewritten as $P^TB^TABP = R^T\Omega R$ with $R = LS|\Lambda|^{1/2}$ being now block upper triangular with diagonal blocks of dimension 1 and 2. Indeed, there exists a left (A,Ω)-orthogonal factor Q such that $BP = QR$. Note that the permutation matrix P can be interpreted here as a given permutation of column vectors stored in the matrix B. This approach was actually used by Slapničar in [32] and Singer in [29] who considered the case $A \in \mathrm{diag}(\pm 1)$ and more general factorization $BP = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$, where $Q \in \mathscr{R}^{m,m}$ is (A,Ω)-orthogonal with $\Omega \in \mathscr{R}^{m,m}$ and $R \in \mathscr{R}^{n,n}$ is again block upper triangular with diagonal blocks of dimension 1 and 2. It is clear that if we restrict the factor R to the class of upper triangular matrices, such factorization does not always exist. This situation has been called in [31, 29] the triangular case of indefinite QR factorization and its version without any column pivoting in B will be discussed in this contribution. For a given $A \in \mathrm{diag}(\pm 1)$ and under the assumption of nonzero principal minors of the matrix M it was shown in [8, 12] that each nonsingular matrix B can be factorized into a product of the so-called pseudo-orthogonal matrix Q and the upper triangular matrix with positive diagonal entries R. Such a matrix B is in [10] called a non-exceptional matrix and in [12] it is called decomposable in the group of all isometries with respect to the bilinear form induced by the matrix A.

These results also indicate that at least from a theoretical point of view the problem with a general symmetric nonsingular A can be transformed into a problem with A equal to a certain signature matrix. However, we are interested in applications, where A is not available explicitly, but it can be accessed by evaluating matrix-vector products, or in situations when m is significantly larger than n and where the approach based on the complete factorization of A (or transformation into a diagonal form) can be expensive even with the use of efficient sparse solvers. Therefore, throughout the paper we consider the case of a general symmetric but nonsingular matrix A. The existence of the decomposition $B = QR$, where Q is left (A,Ω)-orthogonal and R upper triangular with positive diagonal entries (and so also the existence of the Cholesky-like factorization of M) for a general symmetric and nonsingular A is discussed in the following Theorem 1.1. Its statement of is not new and it has been discussed in various forms by several authors [5, 8, 10, 12, 31, 29].

THEOREM 1.1. *Let $B \in \mathscr{R}^{m,n}$ be full-column rank and $A \in \mathscr{R}^{m,m}$ be symmetric indefinite. There exists a unique decomposition $B = QR$, where $Q \in \mathscr{R}^{m,n}$ is left (A,Ω)-orthogonal with $\Omega \in \mathrm{diag}(\pm 1)$ and the matrix $R \in \mathscr{R}^{n,n}$ is upper triangular with positive diagonal elements if and only if no principal minor of $M = B^TAB$ vanishes.*

Proof. The matrix M has all nonzero principal minors if and only if it has the LU factorization $M = \tilde{L}U$, where $\tilde{L}$ is unit lower triangular and U upper triangular. It is easy to check that the product of the first j diagonal elements of U coincides with the jth principal minor of M for all $j = 1,\ldots,n$ [8]. The factor $\Omega = \mathrm{diag}(\omega_1,\ldots,\omega_n)$ will be then a diagonal matrix with $\omega_i \in \{-1,1\}$ such that the product of its first j elements is equal to the sign of the jth principal minor of M for all $j = 1,\ldots,n$. Obviously ω_j is also the sign of the jth diagonal element of U and we can find a real diagonal matrix D such that $U = D\Omega D\tilde{U}$, where $\tilde{U}$ is unit upper triangular. As M is symmetric, we have

$$M = \tilde{L}U = \tilde{L}D\Omega D\tilde{U} = \tilde{U}^TD\Omega D\tilde{L}^T,$$

and the uniqueness of the LU decomposition of M implies that $\tilde{L} = \tilde{U}^T$. Defining $R = D\tilde{U}$

and $Q = BR^{-1}$ we have now $M = R^T \Omega R$ and $B = QR$ with $Q^T A Q = \Omega$. □

The organization of the paper is as follows. In Section 2 we give our basic results on the Cholesky-like factorization of a general symmetric indefinite matrix M. In particular, we develop bounds for the extremal singular values of the triangular factor R and the (A,Ω)-orthogonal factor Q in terms of the spectral properties of principal submatrices of the matrix M. Then in Section 3 we give a description of four schemes used for orthogonalization with respect to the bilinear form induced by the matrix A. Section 4 is devoted to the scheme for computing the factors Q and R that directly uses the Cholesky-like factorization of the matrix M. Section 5 recalls the classical Gram-Schmidt algorithm with the bilinear form induced by the matrix A. In both cases we also consider the corresponding algorithm with reorthogonalization or iterative refinement and focus on their rounding error analysis. We give the worst-case bounds for quantities computed in finite precision arithmetic and formulate our results on the factorization error and on the loss of $(A,\bar{\Omega})$-orthogonality (measured by $\|B - \bar{Q}\bar{R}\|$ and $\|\bar{Q}^T A \bar{Q} - \bar{\Omega}\|$) in terms of quantities proportional to the roundoff unit u, in terms of $\|A\|$, $\|B\|$ or $\|M\|$, and in terms of the extremal singular values of computed factors $\bar{Q}$ and $\bar{R}$. Finally, in Section 6 we present some numerical experiments that illustrate our theoretical results.

The symbol $\sigma_k(A)$ denotes the kth largest singular value of A and provided that A has a full column rank $\kappa(A) = \sigma_1(A)/\sigma_n(A)$ denotes the condition number of the matrix $A \in \mathscr{R}^{m,n}$. We use the notation $|A|$ and $|a|$ for the matrix and vector whose elements are the absolute values of corresponding elements of the matrix $A \in \mathscr{R}^{m,n}$ and the vector $a \in \mathscr{R}^n$, respectively. By $\langle a,b\rangle = a^T b$ we denote the Euclidean inner product of two vectors a and b. The term $\|a\|$ is the corresponding Euclidean norm of the vector a and $\|A\| = \sigma_1(A)$ stands for the 2-norm of the matrix A. The quantities computed in finite precision arithmetic will be denoted by the quantity with an extra upper-bar as e.g. $\bar{Q} = [\bar{q}_1,\ldots,\bar{q}_n]$, $\bar{\Omega}$ or $\bar{R}$. We assume the arithmetic with the standard rules for floating-point computations (see, e.g., [19]). We use the notation $c_k u = c_k(m,n)u$, $k = 1,\ldots,10$ denoting low-degree polynomials in the dimensions m and n multiplied by the unit roundoff u; they are independent of $\kappa(A)$, $\kappa(B)$ or $\kappa(M)$ but they do depend on details of the computer arithmetic. For simplicity we do not give their exact specification and we also omit the terms proportional to higher powers of u.

2. Cholesky-like factorization of symmetric indefinite matrices. We consider the general case of symmetric nonsingular A and we introduce the notation $B_j = [b_1,\ldots,b_j] \in \mathscr{R}^{m,j}$ and $M_j = B_j^T A B_j$. Assuming that M_j is nonsingular for $j = 1,\ldots,n$, in the following, we give bounds for the conditioning of factors Q_j and R_j such that $B_j = Q_j R_j$, where $Q_j = [q_1,\ldots q_j]$ is left (A,Ω_j)-orthogonal with $\Omega_j = \mathrm{diag}(\omega_1,\ldots,\omega_j)$ and R_j is upper triangular with positive diagonal entries.

For A positive definite one would have the signature matrix equal to $\Omega_j = I_j$ with the factors R_j and Q_j satisfying the bounds $\|R_j\| = \|A^{1/2}B_j\|$, $\|R_j^{-1}\| = 1/\sigma_j(A^{1/2}B_j)$ and $\|Q_j\| = 1/\sigma_j(A^{1/2}Q_{B,j}) \le \|A^{-1}\|^{1/2}$, $\sigma_j(Q_j) = 1/\|A^{1/2}Q_{B,j}\| \ge 1/\|A\|^{1/2}$, where $Q_{B,j}$ is the matrix with column vectors that form an orthonormal basis of the range of B_j. Then $\kappa(R_j) = \kappa(A^{1/2}B_j) = \kappa^{1/2}(M_j)$ and $\kappa(Q_j) = \kappa(A^{1/2}Q_{B,j}) \le \kappa^{1/2}(A)$. For details we refer, e.g., to papers [27] or [23].

For A indefinite it follows from $M_j = R_j^T \Omega_j R_j$ that $\|M_j\| \le \|R_j\|^2$ and $\|M_j^{-1}\| \le \|R_j^{-1}\|^2$ and thus the square root of the condition number of M_j is just a lower bound for the condition number of the factor R_j, i.e., we have only $\kappa^{1/2}(M_j) \le \kappa(R_j)$. The upper bound for $\kappa(R_j)$ seems to be more difficult to obtain. We set $w_1 = m_{1,1}$ and $r_{1,1} = \sqrt{|w_1|}$. For each $j = 2,\ldots,n$

we consider the factorization $M_j = R_j^T \Omega_j R_j$ in the bordered form

(2.1)
$$M_j = \begin{pmatrix} M_{j-1} & m_{1:j-1,j} \\ m_{1:j-1,j}^T & m_{j,j} \end{pmatrix} = \begin{pmatrix} R_{j-1}^T & 0 \\ r_{1:j-1,j}^T & r_{j,j} \end{pmatrix} \begin{pmatrix} \Omega_{j-1} & 0 \\ 0 & \omega_j \end{pmatrix} \begin{pmatrix} R_{j-1} & r_{1:j-1,j} \\ 0 & r_{j,j} \end{pmatrix},$$

where the off-diagonal entries $r_{1:j-1,j}$ in the factor R_j are given as

$$r_{1:j-1,j} = \Omega_{j-1}^{-1} R_{j-1}^{-T} m_{1:j-1,j}, \tag{2.2}$$

where $m_{1:j-1,j} = B_{j-1}^T A b_j$ and $m_{j,j} = b_j^T A b_j$. It appears from (2.1) that the diagonal entries $r_{j,j}$ are related to the Schur complement $w_j = M_j \backslash M_{j-1} := m_{j,j} - m_{1:j-1,j}^T M_{j-1}^{-1} m_{1:j-1,j}$. Indeed, we have

$$r_{j,j}^2 \omega_j = m_{j,j} - r_{1:j-1,j}^T \Omega_{j-1} r_{1:j-1,j} = w_j \tag{2.3}$$

with $r_{j,j} = \sqrt{|w_j|}$. Since the Schur complement w_j comes from the block factorization

$$M_j = \begin{pmatrix} I & 0 \\ m_{1:j-1,j}^T M_{j-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} M_{j-1} & 0 \\ 0 & w_j \end{pmatrix} \begin{pmatrix} I & M_{j-1}^{-1} m_{1:j-1,j} \\ 0 & 1 \end{pmatrix}, \tag{2.4}$$

it follows that $w_j = \det(M_j)/\det(M_{j-1})$. The lower bound $|w_j| \geq \sigma_j(M_j)$ can be obtained by considering the interlacing property $|w_j|^{-1} \leq \|M_j^{-1}\|$ from the inverse of the matrix M_j

$$M_j^{-1} = \begin{pmatrix} M_{j-1}^{-1} + H_{j-1} & -M_{j-1}^{-1} m_{1:j-1,j} w_j^{-1} \\ -w_j^{-1} m_{1:j-1,j}^T M_{j-1}^{-1} & w_j^{-1} \end{pmatrix}, \tag{2.5}$$

where $H_{j-1} = M_{j-1}^{-1} m_{1:j-1,j} w_j^{-1} m_{1:j-1,j}^T M_{j-1}^{-1}$. Note that this identity will play an important role in the further analysis.

It is also clear that if A is positive definite then the size of the Schur complement w_j is always bounded by the diagonal element $m_{j,j}$ in (2.1). In the general indefinite case it can be quite large and we have only the upper bound

$$|w_j| \leq |m_{j,j}| + |m_{1:j-1,j}^T M_{j-1}^{-1} m_{1:j-1,j}| \leq \|M_j\|(1 + \|M_{j-1}^{-1}\|\|M_j\|).$$

This is then reflected in the increasing size of the entries in the factor R_j. In the following we give upper bounds for the extremal singular values and the condition number of $\bar{R}_j$.

THEOREM 2.1. *Let $A \in \mathscr{R}^{m,m}$ be symmetric and $B \in \mathscr{R}^{m,n}$ be of full-column rank such that no principal minor of the matrix $M = B^T A B$ vanishes (i.e., M_j is nonsingular for all $j = 1, \ldots, n$). The condition number of the triangular factor R in the Cholesky-like factorization $M = R^T \Omega R$ is bounded as follows:*

$$\kappa(R) \leq \|M\| \left(\|M^{-1}\| + 2 \sum_{j;\ \omega_{j+1} \neq \omega_j} \|M_j^{-1}\| \right). \tag{2.6}$$

Proof. Using the identities (2.2) and (2.3) the inverse of the factor R_j is given as

$$R_j^{-1} = \begin{pmatrix} R_{j-1}^{-1} & -r_{j,j}^{-1} R_{j-1}^{-1} r_{1:j-1,j} \\ 0 & r_{j,j}^{-1} \end{pmatrix} = \begin{pmatrix} R_{j-1}^{-1} & -M_{j-1}^{-1} m_{1:j-1,j} / \sqrt{|w_j|} \\ 0 & 1/\sqrt{|w_j|} \end{pmatrix}.$$

Consequently, the product $R_j^{-1}R_j^{-T}$ can be expressed recursively in the form

$$(R_j^TR_j)^{-1} = \begin{pmatrix} (R_{j-1}^TR_{j-1})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \omega_j \begin{pmatrix} H_{j-1} & -M_{j-1}^{-1}m_{1:j-1,j}s_j^{-1} \\ -w_j^{-1}m_{1:j-1,j}^TM_{j-1}^{-1} & w_j^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} (R_{j-1}^TR_{j-1})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \omega_j \left[M_j^{-1} - \begin{pmatrix} M_{j-1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right]. \tag{2.7}$$

The identity (2.7) provides the basic insight into the relation between the minimum singular values of the factors R_j and the minimum singular values of principal submatrices of M_j. Observe that the recursive use of (2.7) leads to the expansion of the matrix $(R_j^TR_j)^{-1}$ in terms of M_j^{-1} and of only those inverses of principal submatrices M_i where there is a change of sign in the factor Ω, i.e. only for such $i = 1,\ldots,j-1$ where $\omega_{i+1} \neq \omega_i$. Then $|\omega_{i+1} - \omega_i| = 2$ and we have the bound

$$\|R_j^{-1}\|^2 \leq \|M_j^{-1}\| + 2 \sum_{\substack{i=1,\ldots,j-1 \\ \omega_{i+1}\neq\omega_i}} \|M_i^{-1}\|.$$

It follows also from $R_j = \Omega_j^{-1}R_j^{-T}(B_j^TAB_j)$ that the norm of the matrix R_j can be bounded as $\|R_j\| \leq \|M_j\|\|R_j^{-1}\|$ which completes the proof. □

COROLLARY 2.2. *The norm of R_j thus can be bounded in terms of the norms of the Schur complements $M_j\backslash M_i$ corresponding to principal submatrices M_i, but only for those $i = 1,\ldots,j-1$, where $\omega_{i+1} \neq \omega_i$, i.e.,*

$$\|R_j\|^2 \leq \|M_j\| + 2 \sum_{\substack{i=1,\ldots,j-1 \\ \omega_{i+1}\neq\omega_i}} \|M_j\backslash M_i\|, \tag{2.8}$$

whereas $\|M_j\backslash M_i\| \leq \|M_j\|(1 + \|M_j\|\|M_i^{-1}\|)$.

Proof. Since the coefficients $r_{1:j-1,j}$ satisfy $r_{1:j-1,j} = R_{j-1}M_{j-1}^{-1}m_{1:j-1,j}$, the bound for the norm of R_j can be also derived from a bound of the product $R_j^TR_j$ given as

$$R_j^TR_j = \begin{pmatrix} I & 0 \\ m_{1:j-1,j}^TM_{j-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} R_{j-1}^TR_{j-1} & 0 \\ 0 & \omega_jw_j \end{pmatrix} \begin{pmatrix} I & M_{j-1}^{-1}m_{1:j-1,j} \\ 0 & 1 \end{pmatrix}.$$

This can be also rewritten as $R_j^TR_j = L_j^T\mathrm{diag}(\omega_1w_1,\ldots,\omega_jw_j)L_j$, where $w_1 = m_{1,1}$ and L_j is unit upper triangular matrix. Taking into account the block factorization (2.4) we can formulate a similar factorization $M_j = L_j^T\mathrm{diag}(w_1,\ldots,w_j)L_j$. Indeed

$$\begin{aligned} R_j^TR_j &= \omega_1M_j + \sum_{i=1}^{j-1}(\omega_{i+1}-\omega_i)L_j^T\mathrm{diag}(0,\ldots,0,w_{i+1},\ldots w_j)L_j \\ &= \omega_1M_j + 2\sum_{\substack{i=1,\ldots,j-1 \\ \omega_{i+1}\neq\omega_i}} \begin{pmatrix} 0 & 0 \\ 0 & M_j\backslash M_i \end{pmatrix}, \end{aligned}$$

where $M_j\backslash M_i$ denotes the Schur complement of the principal submatrix M_i subject to M_j. □

The bound (2.6) that holds for a general signature matrix $\Omega \in \mathrm{diag}(\pm 1)$ can be reformulated also for symmetric quasi-definite matrices, i.e. the matrices M with the square symmetric diagonal blocks M_{11} and M_{22} such that M_{11} is positive definite, M_{22} is negative definite and $M_{21} = M_{12}^T$ [37, 30]. For such matrices we have the Cholesky-like factorization

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{2.9}$$

where R_{11} and R_{22} are upper triangular of appropriate dimensions. The condition number of the factor R can be then bounded as follows.

THEOREM 2.3. *Let $A \in \mathscr{R}^{m,m}$ be symmetric and $B \in \mathscr{R}^{m,n}$ be such that the matrix M is symmetric quasi-definite with the Cholesky-like factorization (2.9). The condition number of the factor R from the factorization (2.9) is bounded as*

$$\kappa(R) \leq \|M\|(\|M^{-1}\| + 2\|M_{11}^{-1}\|). \tag{2.10}$$

Proof. It follows immediately from (2.9) that $M_{11} = R_{11}^T R_{11}$, $M_{12} = R_{11}^T R_{12}$ and $M_{22} = R_{12}^T R_{12} - R_{22}^T R_{22}$. The corresponding Schur complement matrix $M\backslash M_{11}$ is negative definite and it can be expressed as $M\backslash M_{11} = M_{22} - M_{21}M_{11}^{-1}M_{12} = M_{22} - R_{12}^T R_{12} = -R_{22}^T R_{22}$. The bound on $\|R^{-1}\|$ can be obtained considering the following two identities

$$R^{-1} = \begin{pmatrix} R_{11}^{-1} & -R_{11}^{-1}R_{12}R_{22}^{-1} \\ 0 & R_{22}^{-1} \end{pmatrix} = \begin{pmatrix} R_{11}^{-1} & -M_{11}^{-1}M_{12}R_{22}^{-1} \\ 0 & R_{22}^{-1} \end{pmatrix},$$

$$(R^T R)^{-1} = \begin{pmatrix} M_{11}^{-1} - M_{11}^{-1}M_{12}(M\backslash M_{11})^{-1}M_{21}M_{11}^{-1} & M_{11}^{-1}M_{12}(M\backslash M_{11})^{-1} \\ (M\backslash M_{11})^{-1}M_{21}M_{11}^{-1} & -(M\backslash M_{11})^{-1} \end{pmatrix}.$$

It is clear from (2.9) that

$$M^{-1} + (R^T R)^{-1} = 2\begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

and therefore $\|M^{-1}\| \leq \|R^{-1}\|^2 \leq \|M^{-1}\| + 2\|M_{11}^{-1}\|$. Using (2.9) we can bound the norm of R from below and from above as $\|M\|^{\frac{1}{2}} \leq \|R\| \leq \|M\|\|R^{-1}\|$. We also see that

$$R^T R = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} - 2(M\backslash M_{11}) \end{pmatrix} = M - 2\begin{pmatrix} 0 & 0 \\ 0 & M\backslash M_{11} \end{pmatrix}.$$

which leads to

$$\|M\|^{\frac{1}{2}} \leq \|R\| \leq (\|M\| + 2\|M\backslash M_{11}\|)^{\frac{1}{2}}, \tag{2.11}$$

where $\|M\backslash M_{11}\| \leq \|M\| + \|M\|^2\|M_{11}^{-1}\|$. □

Note that similar results could be formulated also in the case which uses some pivoting strategy when the Cholesky-like factorization is applied to the permuted columns of B. Such techniques, where the size of entries in the factor R is monitored and kept on a reasonable level, could lead to more stable factorizations. For simplicity of our approach, we do not consider a column pivoting in B here.

The properties of the so called J-orthogonal matrices have been studied in [20], see also [33]. In our left (A,Ω_j)-orthogonal case we have $Q_j^T A Q_j = \Omega_j$. If we take the eigendecomposition $A = V\Lambda V^T = (V|\Lambda|^{1/2})J(V|\Lambda|^{1/2})^T$ then there exists a permutation matrix $P_j \in \mathscr{R}^{j,j}$ so that $\tilde{P}_j J_j P_j^T = \Omega_j$, where J_j is a principal submatrix of the matrix $J \in \mathrm{diag}(\pm 1)$. Then the matrix $\tilde{Q}_j = |\Lambda|^{1/2}V^T Q_j P_j$ represents the first j columns of some (J,J)-orthogonal (i.e., square) matrix. In our terminology $\tilde{Q}_j$ is left (J,J_j)-orthogonal. Then $\kappa(Q_j) \leq \kappa^{1/2}(A)\kappa(\tilde{Q}_j)$. It was shown in [20] that the eigenvalues and singular values of any (J,J)-orthogonal matrix $\tilde{Q}$ satisfying $\tilde{Q}^T J \tilde{Q} = J \in \mathrm{diag}(\pm 1)$ come in reciprocal pairs and so its condition number is given by the square of its norm $\kappa(\tilde{Q}) = \|\tilde{Q}\|^2$. As it was pointed out the norm of $\tilde{Q}$ can be in general

quite large. Therefore it seems more useful to relate the conditioning of Q_j to the conditioning of the factor R_j as follows. The singular values of the factor Q_j can be bounded from the definition as $\|Q_j\| \leq \|B_j\|\|R_j^{-1}\|$ and $\sigma_j(Q_j) \geq \sigma_j(B_j)/\|R_j\|$ giving rise to the bound for its condition number $\kappa(Q_j) \leq \kappa(B_j)\kappa(R_j)$.

EXAMPLE 2.4. Let $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ be the identity matrix in $\mathscr{R}^{2,2}$ and let the standard unit vectors be orthogonalized with respect to the bilinear form determined by the matrix $A = \begin{pmatrix} 1 & \sqrt{\varepsilon} \\ \sqrt{\varepsilon} & -\varepsilon \end{pmatrix}$, where ε is a small positive number. The matrix $A \in \mathscr{R}^{2,2}$ is ill-conditioned with singular values given as $\|A\| \approx 1+\varepsilon$ and $\sigma_{min}(A) = 2\varepsilon$, while the factors Q, R and Ω are given as

$$Q = R^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \end{pmatrix}, \quad R = Q^{-1} = \begin{pmatrix} 1 & \sqrt{\varepsilon} \\ 0 & \sqrt{\varepsilon} \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The singular values of the triangular factor R are given as $\|R\| \approx \sqrt{1+\varepsilon}$ and $\sigma_{min}(R) \approx \sqrt{\varepsilon}$ resulting in $\kappa(R) = \kappa(Q) \approx \frac{1}{\sqrt{\varepsilon}}$. The Schur complement is equal to $M\backslash M_{11} = -2\varepsilon$. The dominant quantity in the bound (2.6) for $\|R^{-1}\|$ is therefore $\|M^{-1}\| \approx 1/(2\varepsilon)$, while the norm of R remains bounded due to (2.11). In such cases (especially when the principal matrix M_{11} is well-conditioned) the bound $\|R\| \leq \|M\|\|R^{-1}\|$ is a large overestimate with respect to the bound (2.11) that is based on the Schur complement $M\backslash M_{11}$. Roughly speaking, in such cases the conditioning of R is similar to the conditioning of the standard Cholesky factor with the positive definite matrix A, where $\kappa(R) = \kappa^{1/2}(M)$.

EXAMPLE 2.5. Let $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ be the identity matrix in $\mathscr{R}^{2,2}$ and let the standard unit vectors be orthogonalized with respect to the bilinear form determined by the matrix $A = \begin{pmatrix} \varepsilon & 1 \\ 1 & -\varepsilon \end{pmatrix}$, where ε is a small positive number. Indeed, $A \in \mathscr{R}^{2,2}$ is well-conditioned with extremal singular values given as $\|A\| = \sigma_{min}(A) = \sqrt{1+\varepsilon^2}$, while the factors Q, R and Ω are given as follows

$$Q = R^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\varepsilon}} & -\frac{1}{\sqrt{\varepsilon(1+\varepsilon^2)}} \\ 0 & \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon^2}} \end{pmatrix}, \quad R = Q^{-1} = \begin{pmatrix} \sqrt{\varepsilon} & \frac{1}{\sqrt{\varepsilon}} \\ 0 & \frac{\sqrt{1+\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The singular values of the triangular factor R satisfy $\|R\| \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}}$ and $\sigma_{min}(R) \approx \frac{\sqrt{\varepsilon}}{\sqrt{2}}$ resulting in the identity $\kappa(R) = \kappa(Q) \approx \frac{2}{\varepsilon}$. The Schur complement $M\backslash M_{11} = -(1+\varepsilon^2)/\varepsilon$ is large and both $\|R\|$ and $\|R^{-1}\|$ are large in this case. We see that the dominant quantity is given by the factor $\|M_{11}^{-1}\| = 1/\varepsilon$ and the bounds (2.6) or (2.10) are quite sharp.

3. Orthogonalization with respect to bilinear forms. Formally, we start with a linearly independent set of column vectors $b_1, \ldots, b_n$ stored in the matrix $B = [b_1, \ldots, b_n]$, and if it exists, generate a set of formally (A,Ω)-orthonormal vectors $q_1, \ldots, q_n$ that form the columns of the factor $Q = [q_1, \ldots, q_n]$ and that span the same subspace as the vectors $b_1, \ldots, b_n$. This is done so that at each step $j = 1, \ldots, n$ the column vectors of the submatrix $Q_j = [q_1, \ldots, q_j]$ form an (A,Ω)-orthonormal basis for the span of column vectors of the submatrix $B_j = [b_1, \ldots b_j]$. Therefore any vector b_j is a linear combination just of the vectors $q_1, \ldots q_j$ with the off-diagonal entries $r_{i,j}$, $i = 1, \ldots, j-1$ (in a compact form denoted also as $r_{1:j-1,j}$) and the diagonal entries $r_{j,j}$ that define then the j-th column of the triangular factor R.

We begin with $m_{1,1} = \omega_1 r_{1,1}^2$ and form $r_{1,1} = \sqrt{|m_{1,1}|}$, $\omega_1 = \mathrm{sign}[m_{1,1}]$ and $q_1 = b_1/r_{1,1}$. From (2.2) it follows for $j = 2,\ldots,n$ that $r_{i,j}$, $i = 1,\ldots,j-1$ can be computed successively column-by-column as a solution of the row-scaled lower triangular system with the matrix $R_{j-1}^T \Omega_{j-1} = (\Omega_{j-1} R_{j-1})^T$ and the right-hand side vector $m_{i,j} = b_i^T A b_j$, $i = 1,\ldots,j-1$ (in a compact form denoted as $m_{1:j-1,j}$) as follows

$$r_{i,j} = \frac{m_{i,j} - \sum_{k=1}^{i-1} r_{k,i}\omega_k r_{k,j}}{\omega_i r_{i,i}}. \tag{3.1}$$

The diagonal entry $r_{j,j}$ and the signature entry ω_j are then given from (2.3) as $\omega_j = \mathrm{sign}[w_j]$ and $r_{j,j} = \sqrt{|w_j|}$, where w_j stands for the Schur complement $w_j = m_{j,j} - r_{1:j-1,j}^T \Omega_{j-1} r_{1:j-1,j}$. Given the entries $r_{1:j-1,j}$ and $r_{j,j}$ in the triangular factor the vector q_j is then computed as

$$q_j = u_j/r_{j,j}, \quad u_j = b_j - Q_{j-1} r_{1:j-1,j} = b_j - \sum_{k=1}^{j-1} r_{k,j} q_k. \tag{3.2}$$

The resulting algorithm (in this paper denoted as the M-QR implementation) is summarized as Algorithm 1.

Algorithm 1 Implementation based on the Cholesky-like factorization of M (M-QR)

for $j = 1,\ldots,n$ **do**
$m_{1:j,j} = B_j^T A b_j$
$r_{1:j-1,j} = \Omega_{j-1}^{-1} R_{j-1}^{-T} m_{1:j-1,j}$
$w_j = m_{j,j} - r_{1:j-1,j}^T \Omega_{j-1} r_{1:j-1,j}$
$\omega_j = \mathrm{sign}[w_j]$
$r_{j,j} = \sqrt{|w_j|}$
$u_j = b_j - Q_{j-1} r_{1:j-1,j}$
$q_j = u_j/r_{j,j}$
end for

To improve the accuracy of computed factors one can introduce the implementation with iterative refinement, where the Cholesky-like factorization is applied first to the matrix $M = (Q^{(0)})^T A Q^{(0)} = (R^{(1)})^T \Omega^{(1)} R^{(1)}$ with $Q^{(0)} = B$ in order to get the factors $R^{(1)}$ and $\Omega^{(1)}$. The factor $Q^{(1)}$ is then obtained as $Q^{(1)} = B(R^{(1)})^{-1}$. In the second stage the Cholesky-like factorization is applied to the matrix $(Q^{(1)})^T A Q^{(1)} = (R^{(2)})^T \Omega^{(2)} R^{(2)}$ to get the factors $R^{(2)}$ and $\Omega^{(2)}$. The resulting factors are then $Q = Q^{(2)} = Q^{(1)}(R^{(2)})^{-1}$ and $R = R^{(2)} R^{(1)}$. It is clear that in exact arithmetic one has $\Omega^{(2)} = \Omega^{(1)} = (Q^{(1)})^T A Q^{(1)}$ and $R^{(2)} = I$ that lead then to $Q = Q^{(1)}$ and $R = R^{(2)}$. Introducing the column-oriented notation for the factors $Q_j^{(0)} = B_j$, $Q_j^{(k)} = [q_1^{(k)},\ldots,q_j^{(k)}]$ and $\Omega_j^{(k)} = \mathrm{diag}(\omega_1^{(k)},\ldots,\omega_j^{(k)})$, for the off-diagonal entries $r_{1:j-1,j}^{(k)}$ and the diagonal entries $r_{j,j}^{(k)}$ of the jth column of the factor $R^{(k)}$, where $j = 1,\ldots,n$ and $k = 1,2$, we can formulate the following Algorithm 2.

The (A,Ω)-orthonormal basis of the span of the matrix B can be computed successively column-by-column via Gram-Schmidt process, where the jth step delivers the columns of $Q_j = (q_1,\ldots q_j)$ that are orthonormal in the B-bilinear form. Various Gram-Schmidt schemes with indefinite A have been considered and effectively used in the context of solving structured eigenvalue problems [22, 26]. The first vector is given as $q_1 = a_1/r_{1,1}$ with $r_{1,1} = \sqrt{|m_{1,1}|}$ and $\omega_1 = \mathrm{sign}[m_{1,1}]$. Provided that the vectors $q_1,\ldots q_{j-1}$ are already (A,Ω)-orthonormal the jth step of the procedure has the form (3.2) and it follows from (2.2) and the

Algorithm 2 Implementation based on the Cholesky-like factorization of M with iterative refinement (M-QR2)

for $j = 1,\ldots,n$ **do**
  $q_j^{(0)} = u_j^{(0)} = b_j$
end for
for $k = 1,2$ **do**
  for $j = 1,\ldots,n$ **do**
    $m_{1:j,j}^{(k)} = (Q_j^{(k-1)})^T A q_j^{(k-1)}$
    $r_{1:j-1,j}^{(k)} = (\Omega_{j-1}^{(k)})^{-1}(R_{j-1}^{(k)})^{-T} m_{1:j-1,j}^{(k)}$
    $w_j = m_{j,j}^{(k)} - (r_{1:j-1,j}^{(k)})^T \Omega_{j-1}^{(k)} r_{1:j-1,j}^{(k)}$
    $\omega_j^{(k)} = \mathrm{sign}[w_j]$
    $r_{j,j}^{(k)} = \sqrt{|w_j|}$
    $u_j^{(k)} = u_j^{(k-1)} - Q_{j-1}^{(k-1)} r_{1:j-1,j}^{(k)}$
    $q_j^{(k)} = u_j^{(k)} / r_{j,j}^{(k)}$
  end for
end for

definition of $m_{1:j-1,j}$ that

$$r_{1:j-1,j} = \Omega_{j-1}^{-1} R_{j-1}^{-T}(B_{j-1}^T A b_j) = \Omega_{j-1}^{-1}(B_{j-1}R_{j-1}^{-1})^T A b_j = \Omega_{j-1}^{-1} Q_{j-1}^T A b_j. \tag{3.3}$$

Thus, alternatively to (3.1), the off-diagonal entries $r_{i,j}$, $i = 1,\ldots,j-1$ in the factor R can be computed via $r_{i,j} = \omega_i^{-1} q_i^T A b_j$ The new vector q_j is computed as $q_j = u_j / r_{j,j}$, where $r_{j,j} = \sqrt{|u_j^T A u_j|} = \sqrt{|w_j|}$ and $\omega_j = \mathrm{sign}[w_j]$, where the Schur complement is computed as $w_j = c_{j,j} - r_{1:j-1,j}^T \Omega_{j-1}^{-1} r_{1:j-1,j}$. As also indicated by Theorem 1.1 the diagonal elements $r_{j,j}$ do not vanish assuming that all principal submatrices M_j are nonsingular and they are bounded from below by $r_{j,j} \geq \sqrt{\sigma_j(M_j)}$ for each $j = 1,\ldots,n$. In addition, from $[B_{j-1}, b_j] + [0, -u_j] = Q_{j-1}[R_{j-1}, r_{1:j-1,j}]$ one can show that the vector u_j represents a correction of the full rank matrix B_j that leads to the rank deficient matrix $Q_{j-1}[R_{j-1}, r_{1:j-1,j}]$ and therefore its norm can be bounded from below as $\|u_j\| \geq \sigma_j(B_j)$. The upper bound $\|u_j\| \leq \|B_j\|(1 + \|M_{j-1}^{-1} m_{1:j-1,j}\|$ for u_j can be obtained from the identity

$$u_j = b_j - B_{j-1}R_{j-1}^{-1} r_{1:j-1,j} = B_j \begin{bmatrix} -R_{j-1}^{-1} r_{1:j-1,j} \\ 1 \end{bmatrix} = B_j \begin{bmatrix} -M_{j-1}^{-1} m_{1:j-1,j} \\ 1 \end{bmatrix}.$$

In the following we consider the classical Gram-Schmidt process frequently used for orthogonalization of vectors with respect to the bilinear form induced by the matrix A. This algorithm (denoted here as A-CGS) is summarized as Algorithm 3.

We also consider the classical Gram-Schmidt process with reorthogonalization (i.e. classical Gram-Schmidt process where the (A,Ω)-orthogonalization of the current vector b_j with respect to previously computed vectors is performed exactly twice). Provided that we have already computed the vectors $Q_{j-1} = [q_1,\ldots,q_{j-1}]$ at the jth step we generate the vectors

$$u_j^{(1)} = u_j^{(0)} - Q_{j-1} r_{1:j-1,j}^{(1)}, \quad r_{1:j-1,j}^{(1)} = \Omega_{j-1}^{-1} Q_{j-1}^T A u_j^{(0)}, \tag{3.4}$$

$$u_j^{(2)} = u_j^{(1)} - Q_{j-1} r_{1:j-1,j}^{(2)}, \quad r_{1:j-1,j}^{(2)} = \Omega_{j-1}^{-1} Q_{j-1}^T A u_j^{(1)}, \tag{3.5}$$

Algorithm 3 Classical Gram-Schmidt process with respect to the bilinear form (A-CGS)

for $j = 1, \ldots, n$ **do**
$r_{1:j-1,j} = \Omega_{j-1}^{-1} Q_{j-1}^T A b_j$
$u_j = b_j - Q_{j-1} r_{1:j-1,j}$
$m_{j,j} = b_j^T A b_j$
$w_j = m_{j,j} - r_{1:j-1,j}^T \Omega_{j-1} r_{1:j-1,j}$
$\omega_j = \mathrm{sign}[w_j]$
$r_{j,j} = \sqrt{|w_j|}$
$q_j = u_j / r_{j,j}$
end for

where $u_j^{(0)} = b_j$. The new vector q_j is then the result of the normalization of $u_j^{(2)}$ given as $q_j = u_j^{(2)}/r_{j,j}$ with $r_{j,j} = \sqrt{|w_j|}$, where $w_j = (u_j^{(2)})^T A u_j^{(2)}$. The jth column of the triangular factor R_j is given by elements $r_{1:j-1,j} = r_{1:j-1,j}^{(1)} + r_{1:j-1,j}^{(2)}$. It is evident that in exact arithmetic one would have $u_j^{(2)} = u_j^{(1)}$. The resulting algorithm (denoted as A-CGS2) is summarized as Algorithm 4. As we will see also in numerical experiments, the reorthogonalization often improves the accuracy of factors computed in finite precision arithmetic.

Algorithm 4 Classical Gram-Schmidt process with reorthogonalization (A-CGS2)

for $j = 1, \ldots, n$ **do**
$u_j^{(0)} = b_j$
for $k = 1, 2$ **do**
$r_{1:j-1,j}^{(k)} = \Omega_{j-1}^{-1} Q_{j-1}^T A u_j^{(k-1)}$
$u_j^{(k)} = u_j^{(k-1)} - Q_{j-1} r_{1:j-1,j}^{(k)}$
end for
$r_{1,j-1,j} = r_{1:j-1,j}^{(1)} + r_{1:j-1,j}^{(2)}$
$w_j = (u_j^{(2)})^T A u_j^{(2)}$
$\omega_j = \mathrm{sign}[w_j]$
$r_{j,j} = \sqrt{|w_j|}$
$q_j = u_j^{(2)} / r_{j,j}$
end for

The numerical behavior of orthogonalization techniques with the standard inner product ($A = I$) has been studied extensively over the last several decades. For main results related to the Householder or Givens QR we refer to Subsections 19.1–19.6 of [19]. Numerical properties of the modified Gram-Schmidt (MGS) process has been analyzed in [3]. The classical Gram-Schmidt (CGS) algorithm and the Gram-Schmidt process with reorthogonalization have been studied much later in [14, 34, 1]. For a positive and diagonal A, the numerical behavior of the weighted Gram-Schmidt process was thoroughly studied by Gulliksson in [18]. It appears that it is similar to the behavior of the standard process applied to the row-scaled matrix $\mathrm{diag}^{1/2}(A)B$ (see also [17, 27]). Thomas and Zahar in [35, 36] considered the Gram-Schmidt process with the inner product in the factorized form $A = LL^T$ and under certain assumptions on the accuracy of computed inner products proved results analogous to the standard Gram-Schmidt applied to the transformed matrix $L^{-T}B$. Several orthogonalization schemes with a non-standard inner product have been studied in [27] and [23] including the analysis of the effect of the conditioning of A on the factorization error and the loss of (A, I)-

orthogonality between the vectors computed in finite precision arithmetic (for details we refer to [27] and [23]).

In the following two sections we analyze the numerical behavior of all four algorithms described above. Section 4 deals with algorithms based on the Cholesky-like factorization of M (Algorithm 1 and Algorithm 2) and Section 5 deals with algorithms that use the Gram-Schmidt process with respect to the bilinear form induced by the matrix A (Algorithm 3 and Algorithm 4), respectively. If we implement such orthogonalization techniques, due to rounding, the computed quantities do not satisfy the identities $B = QR$ and $Q^T AQ = \Omega \in \mathrm{diag}(\pm 1)$ exactly, and the question is what is the best we can get in finite precision arithmetic. We denote the factors computed in finite precision arithmetic by $\bar{Q}$, $\bar{\Omega}$ and $\bar{R}$. The factorization error is measured by the quantity $\|B - \bar{Q}\bar{R}\|$ and the quality of the computed factor $\bar{Q}$ is usually measured by the quantity $\|\bar{Q}^T A\bar{Q} - \bar{\Omega}\|$ which is called the loss of $(A,\bar{\Omega})$-orthogonality here. We analyze these quantities, derive their corresponding bounds in terms of constants proportional to the roundoff unit u, of the norms $\|A\|$, $\|B\|$ or $\|M\|$, and in terms of the extremal singular values of factors $\bar{Q}$ and $\bar{R}$. Based on the results in previous section we also formulate the bounds for the norms of latter quantities in terms of the spectral properties of the slightly perturbed matrix M and its principal submatrices M_j with the change of the sign in the corresponding signature factor $\bar{\Omega}$.

4. Orthogonalization schemes based on the Cholesky-like factorization. In this section we analyze the factorization error and the loss of $(A,\bar{\Omega})$-orthogonality for quantities computed by Algorithm 1 and Algorithm 2. We show that while the bounds for the factorization error are very similar, the bounds for the loss of $(A,\bar{\Omega})$-orthogonality is significantly better for Algorithm 2 and it is probably the best what one can get in finite precision arithmetic. For the results on the Cholesky factorization in the symmetric positive definite case we refer to Chapter 10 of [19] (see also the stability analysis of the block LU factorization in [11]). The case when A is symmetric indefinite but M is still positive definite has been studied by Chandrasekaran, Gu and Sayed in the context of solving indefinite least squares problems and it was shown that the approach using the Cholesky factorization of a certain indefinite matrix produces a backward stable approximate solution [9].

First we recall the basic result on the Cholesky-like factorization that was already proved as Theorem 3.1 in [32] in a more general setting with column pivoting in B and block upper triangular R with diagonal blocks of dimension 1 or 2. Here we use its slight reformulation assuming only diagonal blocsk of dimension 1 and we consider also the explicit floating-point computation of the matrix M, where the error of computing its entries satisfies only $|\mathrm{fl}(M) - M| \leq c_1 u |B|^T |A||B|$ and it may exceed the size of $c_1 u|M|$ that appears in the bound (3.37) of [32].

THEOREM 4.1. *Assuming that* $c_2 u\|A\|\|B\|^2 \kappa(M) \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1} \neq \bar{\omega}_j}} \|M_j^{-1}\| < 1$ *the Cholesky-like factorization applied to the symmetric indefinite matrix M runs to completion and the computed factors $\bar{R}$ and $\bar{\Omega}$ satisfy*

$$M + \Delta M = \bar{R}^T \bar{\Omega} \bar{R}, \quad |\Delta M| \leq c_2 u [|\bar{R}|^T |\bar{R}| + |B|^T |A||B|]. \tag{4.1}$$

Proof. Assuming that factorization has successfully completed $j-1$ steps, producing a nonsingular matrix $\bar{R}_{j-1}$ it is easy to see that at step j we will still have $\bar{R}_j^T \bar{\Omega}_j \bar{R}_j = M_j + \Delta M_j$, where $|\Delta M_j| \leq c_2 u [|\bar{R}_j|^T |\bar{R}_j| + |B_j|^T |A||B_j|]$. The matrix $\bar{R}_j$ is nonsingular if the matrix $M_j + \Delta M_j$ is nonsingular. Considering thus for each step $j = 1,\ldots,n$ the assumption $\sigma_j(M_j) > c_2 u\|A\|\|B_j\|^2 \|M_j\| \max_{\substack{i=1,\ldots,j-1 \\ \bar{\omega}_{i+1} \neq \bar{\omega}_i}} \|M_i^{-1}\|$, the Cholesky-like factorization of M_j will produce a nonsingular matrix $\bar{R}_j$ and we get the desired statement. □

COROLLARY 4.2. *Under assumption of Theorem 4.1 the triangular factor $\bar{R}$ computed in the Cholesky-like factorization of M satisfies*

$$\|\bar{R}\|^2 \leq \|M+\Delta M\| + 2 \sum_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|(M+\Delta M)\backslash(M_j+\Delta M_j)\|$$

$$\leq 2n\|M+\Delta M\|^2 \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|(M_j+\Delta M_j)^{-1}\| \tag{4.2}$$

$$\leq \frac{2n(1+c_2u)}{1-c_2u} \frac{\|A\|\|B\|^2\|M\| \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|M_j^{-1}\|}{1-c_2u\|A\|\|B\|^2\kappa(M) \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|M_j^{-1}\|}. \tag{4.3}$$

We see that the accuracy of the Cholesky-like factorization $M+\Delta M = \bar{R}^T\Omega R$ depends on the size of its triangular factor. In the general symmetric indefinite case with $\Omega \in \mathrm{diag}(\pm 1)$ the growth factor $\|\bar{R}\|^2/\|M\|$ can be quite large and it depends also on the conditioning of the worst-conditioned principal submatrix $\max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|M_j^{-1}\|$, where we have a change of the sign in the factor $\bar{\Omega}$. In the following theorem we consider Algorithm 1 and give bounds for the factorization error and the loss of $(A,\bar{\Omega})$-orthogonality of the computed factors $\bar{Q}$ and $\bar{R}$.

THEOREM 4.3. *Let $\bar{R}$ and $\bar{\Omega}$ be the computed triangular and signature factors from the Cholesky-like factorization of the matrix M and let $\bar{Q}$ be the computed solution of triangular systems with the matrix $\bar{R}$ in Algorithm 1. Then under assumptions of Theorem 4.1 these factors satisfy*

$$\|B-\bar{Q}\bar{R}\| \leq c_3u(\|B\|+\|\bar{Q}\|\|\bar{R}\|) \leq c_3u\|B\|\kappa(\bar{R}), \tag{4.4}$$

$$\|\bar{Q}^TA\bar{Q}-\bar{\Omega}\| \leq c_4u\left[\kappa^2(\bar{R})+\|\bar{R}^{-1}\|^2\|A\|\|B\|^2+2\|A\bar{Q}\|\|\bar{Q}\|\kappa(\bar{R})\right]. \tag{4.5}$$

Proof. If the Cholesky-like factorization applied to the symmetric indefinite matrix M runs to completion then the columns of the factor $\bar{Q}$ are just the computed results of triangular back-solves satisfying (3.2). The vectors $\bar{u}_j$ satisfy at each step $j=1,\ldots,n$ the recurrence with computed quantities

$$\bar{u}_j = b_j - \bar{Q}_{j-1}\bar{r}_{1:j-1,j} + \Delta u_j, \quad |\Delta u_j| \leq (j-1)u\left[|b_j| + \sum_{i=1}^{j-1}|\bar{r}_{i,j}||\bar{q}_i|\right]. \tag{4.6}$$

The recurrence (4.6) together with the identity for the vector $\bar{q}_j = \mathrm{fl}[\bar{u}_j/\bar{r}_{j,j}]$ implying $\bar{u}_j = \bar{r}_{j,j}\bar{q}_j - \Delta q_j$ with $|\Delta q_j| \leq u|\bar{r}_{j,j}||\bar{q}_j|$ gives the desired statement

$$B+\Delta B = \bar{Q}\bar{R}, \quad \|\Delta B\| \leq c_3u(\|B\|+\|\bar{Q}\|\|\bar{R}\|)$$

with the columns of the matrix $\Delta B = [\Delta b_1,\ldots,\Delta b_n]$ defined as $\Delta b_j = \Delta u_j + \Delta q_j$. For the loss of $(A,\bar{\Omega})$-orthogonality we consider (4.4), express the factor $\bar{Q} = (B+\Delta B)\bar{R}^{-1}$ and take into account (4.1) so that $\bar{Q}^TA\bar{Q} = \bar{\Omega} + \bar{R}^{-T}\Delta M\bar{R}^{-1} + (AB\bar{R}^{-1})^T(\Delta B\bar{R}^{-1}) + (\Delta B\bar{R}^{-1})^T(AB\bar{R}^{-1}) + (\Delta B\bar{R}^{-1})^TA(\Delta B\bar{R}^{-1}) = \bar{\Omega}+\Delta A$ and

$$\|\Delta A\| \leq c_4u\left[\kappa^2(\bar{R})+\|\bar{R}^{-1}\|^2\|A\|\|B\|^2+2\|AB\bar{R}^{-1}\|\|B\bar{R}^{-1}\|\kappa(\bar{R})\right].$$

Taking into account that $B\bar{R}^{-1} = \bar{Q} - \Delta B\bar{R}^{-1}$ we get the bound (4.5). □

Ideally we could expect that the computed factors $\bar{Q}$ and $\bar{R}$ satisfy the recurrences $B+\Delta B=\bar{Q}\bar{R}$ and $M+\Delta M=\bar{R}^T\bar{\Omega}\bar{R}$ with the factorization errors $\|\Delta B\|\le c_3u\|B\|$ and $\|\Delta M\|\le c_2u\|M\|$. Then the loss of $(A,\bar{\Omega})$-orthogonality can be bounded as

$$\|\Delta A\|\le c_4u\left[\|M\|\|\bar{R}^{-1}\|^2+\|A\bar{Q}\|\|B\|\|\bar{R}^{-1}\|\right].$$

Such bounds will be difficult to achieve since the bound for $\|\Delta M\|$ in (4.1) depends also on $\|\bar{R}\|^2$ which can be significantly larger than $\|M\|$ and also most methods compute the columns of $\bar{Q}$ explicitly using the elements of $\bar{R}$. Thus the bounds (4.4) and (4.5) seem more probable in practical situations.

As it will be illustrated later in numerical experiments the accuracy of the computed factors can be often improved by one step of iterative refinement. We will show that while Algorithm 2 produces the factors $\bar{Q}$ and $\bar{R}$ with the factorization error that remains approximately on the same level, the loss of $(A,\bar{\Omega})$-orthogonality of the computed orthogonal factor can be significantly better than corresponding quantities in Algorithm 1. The results are summarized in the following theorem.

THEOREM 4.4. *For a symmetric nonsingular A and for a full-column rank matrix B satisfying the assumption $c_7u\|A\|\|B\|^2\kappa(M)(\|M^{-1}\|+\max_{\substack{j=1,\dots,n-1\\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}}\|M_j^{-1}\|)<1$ the factorization error $B-\bar{Q}^{(2)}\bar{R}^{(2)}\bar{R}^{(1)}$ and the loss of $(A,\bar{\Omega}^{(2)})$-orthogonality $(\bar{Q}^{(2)})^TA\bar{Q}^{(2)}-\bar{\Omega}^{(2)}$ between the columns of computed factor $\bar{Q}^{(2)}$ in Algorithm 2 are bounded as*

$$\|B-\bar{Q}^{(2)}\bar{R}^{(2)}\bar{R}^{(1)}\|\le c_5u\left[\|B\|+(\|\bar{Q}^{(1)}\|+\|\bar{Q}^{(2)}\|)\|\bar{R}^{(1)}\|\right], \tag{4.7}$$

$$\|(\bar{Q}^{(2)})^TA\bar{Q}^{(2)}-\bar{\Omega}^{(2)}\|\le c_6u\left[\|A\|\|\bar{Q}^{(1)}\|^2+\|A\bar{Q}^{(2)}\|\|\bar{Q}^{(2)}\|\right]. \tag{4.8}$$

Proof. In Algorithm 2 we compute first the Cholesky-like factorization of M to get the factors $R^{(1)}$ and $\Omega^{(1)}$. Then we recover the factor $Q^{(1)}$ using B and $R^{(1)}$. In the second stage we compute the Cholesky-like factorization of $(Q^{(1)})^TBQ^{(1)}$ to get $R^{(2)}$, $\Omega^{(2)}$ and finally we recover $Q^{(2)}$ from $Q^{(1)}$ and $R^{(2)}$. From the statement of Theorem 4.1 for the computed triangular factors $\bar{R}^{(1)}$ and $\bar{R}^{(2)}$ we have the identities

$$M+\Delta M^{(1)}=(\bar{R}^{(1)})^T\bar{\Omega}^{(1)}\bar{R}^{(1)},\|\Delta M^{(1)}\|\le c_2u[\|\bar{R}^{(1)}\|^2+\|A\|\|B\|^2], \tag{4.9}$$

$$(\bar{Q}^{(1)})^TA\bar{Q}^{(1)}+\Delta M^{(2)}=(\bar{R}^{(2)})^T\bar{\Omega}^{(2)}\bar{R}^{(2)},\|\Delta M^{(2)}\|\le c_2u[\|\bar{R}^{(2)}\|^2+\|A\|\|\bar{Q}^{(1)}\|^2]. \tag{4.10}$$

The orthogonal factors $\bar{Q}^{(1)}$ and $\bar{Q}^{(2)}$ are computed in triangular solves satisfying the recurrences

$$B+\Delta B^{(1)}=\bar{Q}^{(1)}\bar{R}^{(1)},\quad \bar{Q}^{(1)}+\Delta B^{(2)}=\bar{Q}^{(2)}\bar{R}^{(2)}, \tag{4.11}$$

where $\|\Delta B^{(1)}\|\le c_3u(\|B\|+\|\bar{Q}^{(1)}\|\|\bar{R}^{(1)}\|)$ and $\|\Delta B^{(2)}\|\le c_3u(\|\bar{Q}^{(1)}\|+\|\bar{Q}^{(2)}\|\|\bar{R}^{(2)}\|)$. Then we have $B+\Delta B^{(1)}+\Delta B^{(2)}\bar{R}^{(1)}=\bar{Q}^{(2)}\bar{R}^{(2)}\bar{R}^{(1)}$. Substituting for $\bar{Q}^{(1)}$ into (4.10) we get

$$(\bar{Q}^{(2)}\bar{R}^{(2)}-\Delta B^{(2)})^TA(\bar{Q}^{(2)}\bar{R}^{(2)}-\Delta B^{(2)})+\Delta M^{(2)}=(\bar{R}^{(2)})^T\bar{\Omega}^{(2)}\bar{R}^{(2)}.$$

Multiplying this identity from the left and right by $(\bar{R}^{(2)})^{-T}$ and $(\bar{R}^{(2)})^{-1}$, respectively, we obtain the expression for the loss of orthogonality $(\bar{Q}^{(2)})^TA\bar{Q}^{(2)}-\bar{\Omega}^{(2)}$. Taking norms we get

$$\|B-\bar{Q}^{(2)}\bar{R}^{(2)}\bar{R}^{(1)}\|\le\mathscr{O}(u)[\|\bar{Q}^{(1)}\|\|\bar{R}^{(1)}\|+\|\bar{Q}^{(2)}\|\|\bar{R}^{(2)}\|\|\bar{R}^{(1)}\|],$$

$$\|(\bar{Q}^{(2)})^TB\bar{Q}^{(2)}-\bar{\Omega}^{(2)}\|\le\mathscr{O}(u)[\kappa^2(\bar{R}^{(2)})+\|B\|\|\bar{Q}^{(1)}\|^2\|(\bar{R}^{(2)})^{-1}\|^2+\|B\bar{Q}^{(2)}\|\|\bar{Q}^{(2)}\|\kappa(\bar{R}^{(2)})].$$

The identity (4.10) can be reformulated into $\bar{\Omega}^{(1)} + \Delta A^{(1)} + \Delta M^{(2)} = (\bar{R}^{(2)})^T \bar{\Omega}^{(2)} \bar{R}^{(2)}$, where $\Delta A^{(1)} = (\bar{Q}^{(1)})^T B \bar{Q}^{(1)} - \bar{\Omega}^{(1)}$. Under our assumptions it follows from (4.5) that $\|\Delta A^{(1)} + \Delta M^{(2)}\| < 1$ and we obtain $\|\bar{R}^{(2)} - I\| \leq \|\Delta A^{(1)} + \Delta M^{(2)}\|/(1 - \|\Delta A^{(1)} + \Delta M^{(2)}\|)$. Consequently, $\kappa(\bar{R}^{(2)}) \approx \|\bar{R}^{(2)}\| \approx \|(\bar{R}^{(2)})^{-1}\| \approx 1 + \mathcal{O}(u)$ and we get the statements of our theorem. □

The bound (4.7) is very similar to the bound (4.4) as $\|\bar{Q}^{(2)}\| \approx \|\bar{Q}^{(1)}\| = \|\bar{Q}\|$ and $\|\bar{R}^{(1)}\| = \|\bar{R}\|$. On the hand, under a somewhat more strict assumption than in Theorem 4.1 we have obtained the bound (4.8) that is significantly better than the bound (4.5) and that is probably the best one can expect in a practical algorithm. Note that due to $\|\mathrm{fl}(\bar{Q}^T A \bar{Q}) - \bar{Q}^T A \bar{Q}\| \leq 3mu\|A\|\|\bar{Q}\|^2$ any bound for the loss of $(A, \bar{\Omega})$-orthogonality can hardly be expected less than the bound for the error in its computation.

5. Orthogonalization schemes that use Gram-Schmidt process with respect to a bilinear form. Probably the most frequently used orthogonalization scheme is the Gram-Schmidt process. This is true also when we consider the orthogonalization with respect to a bilinear form in practical applications [22, 26]. In this section we study the numerical behavior of the classical Gram-Schmidt process with respect to a bilinear form induced by A (see Algorithm 3) and derive bounds that are similar to bounds developed for Algorithm 1 that is based on the Cholesky-like factorization of M. Then we consider the classical Gram-Schmidt process with reorthogonalization (Algorithm 4) and show that reorthogonalization leads to similar effects as the iterative refinement in the approach based on the Cholesky-like factorization M. Indeed we show that while the factorization error in Algorithm 4 remains approximately on the same level, the bound for the loss of $(A, \bar{\Omega})$-orthogonality is significantly better than in Algorithm 3 and it is similar to the bound developed for Algorithm 2.

THEOREM 5.1. *The computed triangular factor $\bar{R}$ in Algorithm 3 is the exact Cholesky-like factor of the perturbed matrix*

$$M + \Delta M = \bar{R}^T \bar{\Omega} \bar{R}, \quad \|\Delta M\| \leq c_8 u[\|\bar{R}\|^2 + \|A\|\|B\|^2 + \|A\|\|B\|\|\bar{Q}\|\|\bar{R}\|]. \tag{5.1}$$

The factors $\bar{Q}$, $\bar{R}$ and $\bar{\Omega}$ computed by the Classical Gram-Schmidt process with respect to the bilinear form induced by the matrix A satisfy

$$\|B - \bar{Q}\bar{R}\| \leq c_3 u(\|B\| + \|\bar{Q}\|\|\bar{R}\|) \tag{5.2}$$

$$\|\bar{Q}^T A \bar{Q} - \bar{\Omega}\| \leq c_4 u \left[\kappa^2(\bar{R}) + \|\bar{R}^{-1}\|^2 \|A\|\|B\|^2 + 3\|A\|\|B\|\|\bar{R}^{-1}\|\|\bar{Q}\|\kappa(\bar{R})\right]. \tag{5.3}$$

Proof. As Algorithm 3 and Algorithm 1 use the same recurrence (3.2) for computing the orthogonal factor, the proof of the bound (5.2) is identical to the proof of the bound (4.4). The coefficients $\bar{r}_{1:j-1,j}$ computed in Algorithm 3 satisfy

$$\Delta r_{1:j-1,j} = \bar{r}_{1:j-1,j} - \bar{\Omega}_{j-1}^{-1} \bar{Q}_{j-1}^T A b_j, \quad |\Delta r_{1:j-1,j}| \leq (j-1)u\|A\|\|b_j\|\|\bar{Q}_{j-1}\|. \tag{5.4}$$

Premultiplying the jth column of the identity (5.2) written as $\bar{r}_{j,j}\bar{q}_j = b_j - \bar{Q}_{j-1}\bar{r}_{1:j-1,j} + \Delta b_j$ by the quantity $b_k^T A$ for $k > j$ we get after some manipulation $b_k^T A \bar{Q}_j \bar{r}_{1:j,j} = m_{k,j} + b_k^T A \Delta b_j$. Taking also into account the bound (5.4) we obtain the identity $\bar{r}_{1:j,k}^T \bar{\Omega}_j \bar{r}_{1:j,j} = m_{k,j} + (\Delta r_{1:j-1,k})^T \bar{\Omega}_j \bar{r}_{1:j,j} + b_k^T A \Delta b_j$. As discussed in [34] the diagonal elements $\bar{r}_{j,j}$ should be computed in the classical Gram-Schmidt process so that they satisfy

$$\bar{\omega}_j \bar{r}_{j,j}^2 + \Delta m_{j,j} = m_{j,j} - (\bar{r}_{1:j-1,j})^T \bar{\Omega}_{j-1} \bar{r}_{1:j-1,j}, |\Delta m_{j,j}| \leq c_8 u(\|A\|\|b_j\|^2 + \|\bar{r}_{1:j-1,j}\|^2).$$

This completes the proof of the first statement. The third statement follows from (5.2) rewritten as $\bar{Q}_j = (B_j + \Delta B_j)\bar{R}_j^{-1}$. Then we have $\bar{Q}_j^T A \bar{Q}_j = \bar{\Omega}_j + \bar{R}_j^{-T} \Delta M_j \bar{R}_j^{-1} + (AB_j\bar{R}_j^{-1})^T \Delta B_j \bar{R}_j^{-1} + (\Delta B_j \bar{R}_j^{-1})^T A \bar{Q}_j$. □

The bound (5.1) is similar to the bound (4.1) obtained for Algorithm 1. Note that the term $\|A\|\|B\|\|\bar{Q}\|\|\bar{R}\|$ in (5.1) can be further bounded using the bound (4.2), the bound $\|\bar{Q}\| \leq \|B+\Delta B\|\|\bar{R}^{-1}\|$ and the bounds

$$\begin{aligned}\|\bar{R}^{-1}\|^2 &\leq \|(M+\Delta M)^{-1}\| + 2 \sum_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|(M_j+\Delta M_j)^{-1}\| \\ &\leq 2n \max_{\substack{j=1,\ldots,n \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|(M_j+\Delta M_j)^{-1}\|, \end{aligned} \tag{5.5}$$

with $\bar{\omega}_{n+1}$ defined as $\bar{\omega}_{n+1} = -\bar{\omega}_n$. Indeed then we have the bound in the form

$$\|A\|\|B\|\|\bar{Q}\|\|\bar{R}\| \leq 2n\|A\|\|B\|(\|B\|+\|\Delta B\|)\|M+\Delta M\| \max_{\substack{j=1,\ldots,n \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|(M_j+\Delta M_j)^{-1}\|.$$

This shows that the third term in the right-hand side of (5.1) is in the worst case as large as the first term proportional to $\|\bar{R}\|^2$ that dominates this bound, see also (4.2). The bound for the loss of $(A,\bar{\Omega})$-orthogonality (5.3) is even more similar to its counterpart (4.5) obtained for Algorithm 1. The only difference consists in the overestimate of the term $\|A\bar{Q}\| \leq \|A(B+\Delta B)\|\|\bar{R}^{-1}\|$ which can be very rough in the case of large $\|\bar{R}^{-1}\|$ but small $\|\bar{R}\|$ (see Example 2.4 or Problem 1 in Section 6), whereas in the cases with $\|\bar{R}^{-1}\| \sim \|\bar{R}\|$ the dominant term in (4.5) and (5.3) is proportional to $\kappa^2(\bar{R})$ (see Example 2.5 or Problem 2 in Section 6) .

As we will show in the following the $(A,\bar{\Omega})$-orthogonality between the computed vectors can be improved by Algorithm 4 with the classical Gram-Schmidt process with reorthogonalization (i.e. classical Gram-Schmidt process where the (A,Ω)-orthogonalization of the current vector b_j with respect to previous basis vectors is performed exactly twice). The computed factors in Algorithm 4 satisfy the following statement.

THEOREM 5.2. *The factors $\bar{Q}$ and $\bar{R}$ computed by the classical Gram-Schmidt algorithm with reorthogonalization in Algorithm 4 satisfy the recurrence*

$$B+\Delta B = \bar{Q}\bar{R}, \quad \|\Delta B\| \leq 2c_3u[\|B\| + \|\bar{Q}\|(\|\bar{R}^{(1)}\| + \|\bar{R}^{(2)}\|)]. \tag{5.6}$$

Proof. The vectors $\bar{u}_j^{(1)}$ and $\bar{u}_j^{(2)}$ computed in finite precision arithmetic satisfy

$$\bar{u}_j^{(1)} = b_j - \bar{Q}_{j-1}\bar{r}_{1:j-1,j}^{(1)} + \Delta u_j^{(1)}, \|\Delta u_j^{(1)}\| \leq 2(j-1)u(\|a_j\| + \|\bar{Q}_{j-1}\|\|\bar{r}_{1:j-1,j}^{(1)}\|), \tag{5.7}$$

$$\bar{u}_j^{(2)} = \bar{u}_j^{(1)} - \bar{Q}_{j-1}\bar{r}_{1:j-1,j}^{(2)} + \Delta u_j^{(2)}, \|\Delta u_j^{(2)}\| \leq 2(j-1)u(\|\bar{u}_j^{(1)}\| + \|\bar{Q}_{j-1}\|\|\bar{r}_{1:j-1,j}^{(2)}\|). \tag{5.8}$$

This leads to $b_j + \Delta u_j^{(1)} + \Delta u_j^{(2)} = \bar{Q}_{j-1}\bar{r}_{1:j-1,j} + \bar{r}_{j,j}\bar{q}_j + \Delta u_j^{(0)}$, where $\bar{r}_{1:j-1,j} = \bar{r}_{1:j-1,j}^{(1)} + \bar{r}_{1:j-1,j}^{(2)}$, $\bar{R} = \bar{R}^{(1)} + \bar{R}^{(2)}$ and $\|\Delta u_j^{(0)}\| \leq u\|\bar{q}_j\||\bar{r}_{j,j}|$ which completes the proof. □

It follows from Theorem 5.2 that the factorization error of vectors computed in Algorithm 4 is not improved with respect to Algorithm 3. As we will see later in experiments due to two recurrences (5.7) and (5.8) it can be slightly larger, but this effect is reflected only in the additive increase of the constant in the corresponding bound. On the other hand, we will show that the loss of $(A,\bar{\Omega})$-orthogonality in Algorithm 4 can be significantly better than that in Algorithm 3.

THEOREM 5.3. *For a symmetric nonsingular A and a full column rank B satisfying the assumption on M in the form $c_{10}u\|A\|\|B\|^2\|M\|(\|M^{-1}\| + \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}} \|M_j^{-1}\|)^2 < 1$ the loss of $(A,\bar{\Omega})$-orthogonality in the computed factor $\bar{Q}$ in Algorithm 4 is bounded as*

$$\|\bar{Q}^TA\bar{Q} - \bar{\Omega}\| \leq c_9u\|A\|\|\bar{Q}\|^2. \tag{5.9}$$

Proof. It follows from (3.4) that in exact arithmetic the vectors q_j satisfy

$$Q_{j-1}^T A q_j = Q_{j-1}^T A \frac{u_j^{(2)}}{r_{j,j}} = \left[\Omega_{j-1} - Q_{j-1}^T A Q_{j-1}\right]^2 \frac{r_{1:j-1,j}}{r_{j,j}}$$

and due to (2.6) for each $j = 1,\ldots,n$ we have the bound for the norm of the last term

$$\frac{\|r_{1:j-1,j}\|}{r_{j,j}} \leq \kappa(R_j) \leq \|M_j\|(\|M_j^{-1}\| + 2 \sum_{\substack{i=1,\ldots,j-1 \\ \omega_{i+1} \neq \omega_i}} \|M_i^{-1}\|).$$

Asuuming that $\|\Omega_{j-1} - Q_{j-1}^T A Q_{j-1}\| \|r_{1:j-1,j}\|/r_{j,j} < 1$ we get that $\|Q_{j-1}^T A q_j\| \leq \|\Omega_{j-1} - Q_{j-1}^T A Q_{j-1}\|$ and thus the (A,Ω)-orthogonality of the new vector q_j is not be amplified in the second sweep of the Gram-Schmidt process. The proof for the vectors $\bar{q}_j$ computed in finite precision arithmetic is quite similar. From (5.7), (5.8) and (5.4) we have the recurrences for the computed vectors $\bar{u}_j^{(1)}$ and $\bar{u}_j^{(2)}$

$$\bar{Q}_{j-1}^T A \bar{u}_j^{(1)} = \Delta A_{j-1} \bar{\Omega}_{j-1}^{-1} \bar{Q}_{j-1}^T A b_j + \Delta v_j^{(1)}, \|\Delta v_j^{(1)}\| \leq (j-1)u \|A\bar{Q}_{j-1}\| \|\bar{Q}_{j-1}\| \|\bar{r}_{1:j-1,j}^{(1)}\|,$$
$$\bar{Q}_{j-1}^T A \bar{u}_j^{(2)} = \Delta A_{j-1} \bar{\Omega}_{j-1}^{-1} \bar{Q}_{j-1}^T A \bar{u}_j^{(1)} + \Delta v_j^{(2)}, \|\Delta v_j^{(2)}\| \leq (j-1)u \|A\bar{Q}_{j-1}\| \|\bar{Q}_{j-1}\| \|\bar{r}_{1:j-1,j}^{(2)}\|,$$

where $\Delta A_{j-1} = \bar{\Omega}_{j-1} - \bar{Q}_{j-1}^T A \bar{Q}_{j-1}$. It also follows from Theorem 5.2 and the same approach as in the proof of (5.1) that $(\bar{R}_j^{(1)})^T \bar{\Omega}_j \bar{R}_j = M_j + \Delta M_j^{(1)}$, where $\|\Delta M_j^{(1)}\| \leq c_8 u[\|\bar{R}_j\|^2 + \|A\|\|B\|^2 + \|A\|\|B\|\|\bar{Q}_j\|\|\bar{R}_j\|]$ and $\bar{R}_j = \bar{R}_j^{(1)} + \bar{R}_j^{(2)}$. Then we get $\bar{R}_j^T \bar{\Omega}_j \bar{R}_j = M_j + \Delta M_j$ with $\Delta M_j = \Delta M_j^{(1)} + (\bar{R}_j^{(2)})^T \bar{\Omega}_j \bar{R}_j$, where the size of the strictly upper triangular factor $\bar{R}_j^{(2)}$ represents the loss of $(A,\bar{\Omega})$-orthogonality after the first weep of the Gram-Schmidt process and after diagonal scaling with $\mathrm{diag}(\bar{R}_j)^{-1}$ it satisfies the bound (5.3). Assuming that $\sigma_j(M_j) > \|\Delta M_j\|$ for $j = 1,\ldots,n$, the diagonal element $\bar{r}_{j,j}$ can be bounded from below by $\bar{r}_{j,j} \geq \sqrt{\sigma_j(M_j) - \|\Delta M_j\|}$. This inequality follows directly from the fact that $\bar{\omega}_j \bar{r}_{j,j}^2$ is equal to the Schur complement of the principal submatrix $M_{j-1} + \Delta M_{j-1}$ subject to $M_j + \Delta M_j$. Then we have $\bar{r}_{j,j}^2 \geq \sigma_j(M_j + \Delta M_j)$. The $(A,\bar{\Omega})$-orthogonality of the vector $\bar{u}_j^{(1)}$ computed after the first sweep of the Gram-Schmidt process with respect to the previously computed vectors $\bar{Q}_{j-1}$ can be bounded as follows

$$\|\bar{Q}_{j-1}^T A \frac{\bar{u}_j^{(1)}}{\bar{r}_{j,j}}\| \leq \frac{\left[\|\Delta A_{j-1}\| + 2(j-1)u\|A\bar{Q}_{j-1}\|\|\bar{Q}_{j-1}\|\right] \|\bar{r}_{1:j-1,j}^{(1)}\|}{[\sigma_j(M_j) - \|\Delta M_j\|]^{1/2}}. \tag{5.10}$$

Due to Theorem 5.2 we can write the bounds $\|\bar{Q}_{j-1}\| \leq (\|B_{j-1}\| + \|\Delta B_{j-1}\|)\|\bar{R}_{j-1}^{-1}\|$ and $\|\bar{r}_{1:j-1,j}^{(1)}\|/\bar{r}_{j,j} \leq \kappa(\bar{R}_j)$ with $\kappa(\bar{R}_j) \leq \|M_j + \Delta M_j\|(\|(M_j + \Delta M_j)^{-1}\| + 2\sum_{i;\ \omega_{i+1} \neq \omega_i} \|(M_i + \Delta M_i)^{-1}\|)$. The $(A,\bar{\Omega})$-orthogonality of the vector $\bar{u}_j^{(2)}$ computed after the second sweep satisfies the bound

$$\|\bar{Q}_{j-1}^T A \frac{\bar{u}_j^{(2)}}{\bar{r}_{j,j}}\| \leq \left[\|\Delta A_{j-1}\| + 2(j-1)u\|A\bar{Q}_{j-1}\|\|\bar{Q}_{j-1}\|\right] \|\bar{Q}_{j-1}^T A \frac{\bar{u}_j^{(1)}}{\bar{r}_{j,j}}\|. \tag{5.11}$$

Now under our assumption the term on the right-hand side of (5.10) will be still less than 1 and we get the statement of our theorem. □

Indeed under somewhat more strict assumption we have derived the bound (5.9) that is significantly better than the bound (5.3) and that is similar to the bound (4.8). The behavior of Algorithm 4 is thus very similar to the behavior of Algorithm 2, but as we will see in the following section, due to more strict assumption it seems somewhat less robust when solving extremely ill-conditioned problems. Another frequently used alternative is the modified Gram-Schmidt process (A-MGS) that we do not discuss here. It is known that while its factorization error is similar to all other schemes, the loss of $(A,\bar{\Omega})$-orthogonality for the A-MGS process can be better than that for the A-CGS process. On the other hand, the bound for A-MGS can be hardly better than the bounds (4.8) or (5.9) that do not explicitly depend on the spectral properties of the matrix B.

6. Numerical experiments. In the following we illustrate our results from previous sections. All experiments are performed in double precision arithmetic using MATLAB where $u = 1.1 \times 10^{-16}$. We consider two cases of symmetric quasi-definite systems, where $\kappa(M_{11}) \ll \kappa(M)$ and $\kappa(M_{11}) \gg \kappa(M)$, respectively, and look first at the dependence of extremal singular values of the factors $\bar{R}$ and $\bar{Q}$ with respect to the conditioning of the matrix M, the principal submatrix M_{11} and the corresponding Schur complement matrix $M\backslash M_{11} = M_{22} - M_{12}^T M_{11}^{-1} M_{12}$. Then we report the factorization error $\|B - \bar{Q}\bar{R}\|$ and the loss of $(A,\bar{\Omega})$-orthogonality $\|\bar{Q}^T A\bar{Q} - \bar{\Omega}\|$ with respect to the conditioning of the matrices M and M_{11} for all four algorithms considered and analyzed in this manuscript: Algorithm 1 based on the Cholesky-like factorization of M (denoted as M-QR), Algorithm 2 with one step of iterative refinement (M-QR2), Algorithm 3 with the classical Gram-Schmidt (A-CGS) process with respect to the bilinear form induced by A and Algorithm 4 with one step of reorthogonalization (A-CGS2). For simplicity in all experiments we consider the trivial square case $B = I$ of appropriate dimension leading to $M = A$.

The first set of problems with dimensions $m = n = 20$ (denoted in all tables as Problem 1) is constructed so that the principal submatrix M_{11} of dimension 10 is positive definite and its condition number is fixed to $\kappa(M_{11}) = 100$, while the condition numbers of the off-diagonal blocks M_{12} are prescribed so that $\kappa(M_{12}) = 10^i$ for $i = 0,\ldots,8$. We fix the norms to $\|M_{11}\| = \|M_{12}\| = 1$. All eigenvalues and singular values are computed using the $logspace(0,-i,10)$ function in MATLAB as logarithmically equally spaced points between 1 and 10^{-i} for appropriate values $i = 0,\ldots,8$. Given the diagonal matrix with the prescribed eigenvalue distribution D we multiply it from the left and right by a randomly generated orthogonal matrix $V = orth(randn(10))$ and its transpose, respectively, and obtain the desired matrix in the form $M_{12} = VDV^T$. The dimension of M_{22} is the same as that of M_{11} and we put $M_{22} = 0$. This construction corresponds to the standard form of the indefinite saddle-point problem. The eigenvalue inclusion set of M has been analyzed by several authors [2, 13, 28, 37]. It follows that

$$\sigma(M) \in \left[\frac{1}{2}\left(\|M_{11}^{-1}\|^{-1} + \sqrt{\|M_{11}^{-1}\|^{-1} + 4\|M_{12}\|^2}\right), \frac{1}{2}\left(\|M_{11}\| - \sqrt{\|M_{11}\|^2 + 4\sigma_{min}^2(M_{12})}\right)\right]$$
$$\cup \left[\sigma_{min}(M_{11}), \frac{1}{2}\left(\|M_{11}\| + \sqrt{\|M_{11}\|^2 + 4\sigma_{min}^2(M_{12})}\right)\right].$$

Consequently, we can expect that the condition number of M will be approximately $\kappa(M) \approx 10^{2i}$ for $i = 0,\ldots,8$. This is well demonstrated in Table 1 as $\|M^{-1}\|$ increases quadratically with respect to the increase of the condition number of the block M_{12}. Since the positive definite principal submatrix M_{11} is well conditioned with $\|M_{11}^{-1}\| = 10^2$ the norm of the Schur complement $M\backslash M_{11}$ does not play any significant role in the bound (2.11) which leads to $\|R\| \approx \|M\|^{1/2}$. Due to the same reason it follows from (2.10) that $\|R^{-1}\| \approx \|M^{-1}\|^{1/2}$. Indeed while the norm of R remains approximately constant, the norm of the factor Q is increasing

$\|M_{12}^{-1}\|$	$\|M^{-1}\|$	$\|M\backslash M_{11}\|$	$\|\bar{R}\| = \|\bar{Q}^{-1}\|$	$\|\bar{R}^{-1}\| = \|\bar{Q}\|$
10^0	1.6180e+00	1.0000e+02	1.4142e+01	1.4142e+01
10^1	1.0099e+02	1.0000e+02	1.4142e+01	1.4142e+01
10^2	1.0001e+04	1.0000e+02	1.4142e+01	1.0001e+02
10^3	1.0000e+06	1.0000e+02	1.4142e+01	1.0000e+03
10^4	1.0000e+08	1.0000e+02	1.4142e+01	1.0000e+04
10^5	1.0000e+10	1.0000e+02	1.4142e+01	1.0000e+05
10^6	1.0000e+12	1.0000e+02	1.4142e+01	1.0000e+06
10^7	9.9808e+13	1.0000e+02	1.4142e+01	1.0000e+07
10^8	1.8925e+16	1.0000e+02	1.4142e+01	1.0000e+08

TABLE 1
The spectral properties of computed factors with respect to the conditioning of the submatrix M_{12} for Problem 1.

$\|M_{12}^{-1}\|$	Algorithm 1 M-QR	Algorithm 2 M-QR2	Algorithm 3 A-CGS	Algorithm 4 A-CGS2
10^0	9.044e-16 (4.463e-14)	4.001e-14 (8.904e-14)	4.352e-15 (4.463e-14)	1.141e-14 (8.904e-14)
10^1	3.782e-15 (4.463e-14)	1.709e-14 (8.904e-14)	3.554e-15 (4.463e-14)	9.483e-15 (8.904e-14)
10^2	2.050e-15 (3.142e-13)	1.418e-14 (6.283e-13)	1.776e-15 (3.142e-13)	1.055e-14 (6.283e-13)
10^3	1.538e-15 (3.140e-12)	1.322e-14 (6.280e-12)	4.577e-16 (3.140e-12)	5.941e-15 (6.280e-12)
10^4	7.916e-16 (3.140e-11)	1.490e-14 (6.280e-11)	1.025e-15 (3.140e-11)	1.090e-14 (6.280e-11)
10^5	1.215e-15 (3.140e-10)	1.511e-14 (6.280e-10)	5.458e-16 (3.140e-10)	1.036e-14 (6.280e-10)
10^6	1.165e-15 (3.140e-09)	8.877e-15 (6.280e-09)	1.017e-15 (3.139e-09)	8.829e-15 (6.280e-09)
10^7	1.790e-15 (3.081e-08)	2.216e-14 (6.280e-08)	4.440e-16 (3.192e-08)	1.094e-14 (6.280e-08)
10^8	1.861e-15 (1.852e-07)	2.576e-14 (6.280e-07)	9.930e-16 (1.519e-07)	1.184e-14 (6.280e-07)

TABLE 2
The factorization error $\|B - \bar{Q}\bar{R}\|$ with respect to the conditioning of the submatrix M_{12} for Problem 1.

with increasing conditioning of the matrix M as also indicated in Table 1. Table 2 reports the norm of factorization error $\|B - \bar{Q}\bar{R}\|$ for all our orthogonalization schemes together with corresponding bounds (4.4), (4.7), (5.2) and (5.6) evaluated in brackets after each quantity. The results show that the factorization error remains close to the roundoff unit for all schemes and actually is much better than the bound $c_3 u\|\bar{Q}\|\|\bar{R}\|$ that indicates that it should increase as $\|\bar{Q}\|$ increases. In Table 3 we give the norms for the loss of orthogonality $\|\bar{Q}^T A\bar{Q} - \bar{\Omega}\|$ and evaluations of their bounds. We see that the behavior of A-CGS is similar to M-QR as well as the behavior of A-CGS2 to M-QR2, while the former two schemes are significantly less accurate than the latter two schemes. Again the predicted bound $c_4 u\|A\|\|\bar{Q}\|^2$ (since $\|R\|$ is a moderate constant the bounds (4.8) and (5.9) are approximately the same, the bound (5.3) differs by a factor of $\|\bar{R}^{-1}\|$ from the bound (4.5) due to $\|\bar{R}\| \approx \|\bar{Q}^{-T}\| \approx \|A\bar{Q}\| \ll \|A\|\|\bar{Q}\| \approx \|A\|\|\bar{R}^{-1}\|$) seems to be an overestimate for all considered schemes. This situation is therefore analogous to the case with the non-standard inner-product induced by positive definite A (see [27, 23]), where the singular values of R are given by the singular values of the matrix $A^{1/2}B$ leading to $\kappa(R) = \kappa(A^{1/2}B)$ and the worst-case bound for the loss of orthogonality for M-QR2 or A-CGS2 is also given as $c_4 u\|A\|\|\bar{Q}\|^2$. Note also that it is important to perform the normalization in A-CGS in the same way as it is done in M-QR, i.e., to compute $r_{j,j}$ and ω_j from $\omega_j r_{j,j} = m_{j,j} - r_{1:j-1,j}^T \Omega_{j-1} r_{1:j-1,j}$ (see Table 3). The use of more standard formulas $r_{j,j} = \sqrt{|u_j^T A u_j|}$ and $\omega_j = \mathrm{sign}[u_j^T A u_j]$ can lead to significantly different results [34].

In the second set of problems (denoted as Problem 2) we take the positive definite block M_{11} of dimension 10 with prescribed singular values so that $\kappa(M_{11}) = 10^i$ for $i = 0, \ldots, 15$.

$\|M_{12}^{-1}\|$	Algorithm 1 M-QR	Algorithm 2 M-QR2	Algorithm 3 A-CGS	Algorithm 4 A-CGS2
10^0	6.976e-15 (2.671e-11)	3.137e-15 (1.162e-13)	4.173e-15 (5.206e-11)	3.195e-15 (7.185e-14)
10^1	8.594e-14 (2.669e-11)	6.651e-15 (8.926e-14)	4.739e-14 (3.583e-11)	7.155e-15 (4.485e-14)
10^2	1.898e-12 (1.334e-09)	5.640e-14 (2.545e-12)	3.172e-12 (9.915e-09)	3.300e-14 (2.231e-12)
10^3	4.826e-10 (1.334e-07)	3.242e-13 (2.263e-10)	1.713e-10 (9.512e-06)	4.418e-13 (2.231e-10)
10^4	2.959e-08 (1.334e-05)	4.963e-12 (2.234e-08)	1.987e-08 (9.472e-03)	3.583e-12 (2.231e-08)
10^5	1.562e-06 (1.334e-03)	3.782e-11 (2.231e-06)	1.443e-06 (9.468e+00)	3.520e-11 (2.231e-06)
10^6	2.408e-05 (1.334e-01)	2.033e-10 (2.231e-04)	3.104e-04 (9.463e+03)	2.471e-10 (2.231e-04)
10^7	3.703e-02 (1.285e+01)	2.520e-09 (2.231e-02)	3.410e-02 (9.952e+06)	2.688e-09 (2.231e-02)
10^8	6.524e-01 (4.644e+02)	2.060e-08 (2.231e+00)	7.661e-01 (1.073e+09)	2.490e-08 (2.231e+00)

TABLE 3
The loss of $(A,\bar{\Omega})$-orthogonality $\|\bar{\Omega}-\bar{Q}^TA\bar{Q}\|$ with respect to the conditioning of the submatrix M_{12} for Problem 1.

This is done in the same way as in Problem 1 with the exception that we set now the norm $\|M_{11}\| = 1/2$. We construct the blocks M_{12} and $M_{22} = -M_{11}$ so that the resulting indefinite matrix M of dimension $m = 20$ is perfectly well-conditioned with 10 positive and 10 negative unit eigenvalues with $\kappa(M) = 1$. Provided that the matrix M_{11} is generated from $M_{11} = VDV^T$ then following the arguments of Theorem 4.3.10 from the book of Horn and Johnson [21] one can show that the off-diagonal block M_{12} can be generated as $M_{12} = V(I-D^2)^{1/2}V^T$. Clearly, it follows from Table 4 that this set of problem corresponds to completely different situation where the unit singular values of M do not play any significant role. It is easily seen that the norm of Schur complement $M\backslash M_{11}$ increases as the norm of M_{11}^{-1} increases and due to (2.11) and (2.10) the norms of R and $Q = R^{-1}$ are given approximately as $\|R\| \approx \|Q\| \lessapprox \sqrt{2}\|M_{11}^{-1}\|^{1/2}$ as it is also visible in Table 4. This is then also reflected in Table 5 where the factorization error $\|B-\bar{Q}\bar{R}\|$ increases as the norms of $\bar{Q}$ and $\bar{R}$ increase, whereas all four orthogonalization schemes behave very similarly. This well corresponds to the bounds (4.4), (4.7), (5.2) and (5.6) evaluated in brackets. The norms for the loss of orthogonality $\|\bar{\Omega}-\bar{Q}^TA\bar{Q}\|$ are reported in Table 6. We see that all schemes generate vectors with similar level of orthogonality, whereas the predicted bounds (4.8) and (5.9) are quite sharp for the Cholesky A-QR2 and A-CGS2 algorithms, while the bounds (4.5) and (5.3) are large overestimates due since they contain the term proportional to $\kappa^2(\bar{R})$.

7. Conclusions. In this paper we have considered the case of symmetric indefinite A and assuming that all principal minors of $M = B^TAB$ are nonzero we have analyzed the conditioning of the triangular factor R from Cholesky-like factorization of $M = R^T\Omega R$. It appears that the inverse of the matrix R^TR can be expanded in terms of M^{-1} and in terms of only those inverses of principal submatrices of M where there is a change of the sign in the factor Ω. Similarly, the norm of R^TR can be bounded in terms of the norm of M and the norms of Schur complements of principal submatrices corresponding to the change of signature in M. Based on these results, we have analyzed two types of important schemes used for orthogonalization with respect to the bilinear form induced by A. For the M-QR implementation based on the Cholesky-like factorization of symmetric indefinite M we have shown that under reasonable assumption on the conditioning of M and its principal submatrices corresponding to all changes in the signature matrix such factorization runs to completion and the computed factors $\bar{R}$ and $\bar{\Omega}$ are the exact factors of the perturbed Cholesky-like factorization $M+\Delta M = \bar{R}^T\bar{\Omega}\bar{R}$ with $\|\Delta M\| \le c_3u[\|\bar{R}\|^2+\|A\|\|B\|^2]$. For the computed orthogonal factor $\bar{Q}$ it follows then that $\|\bar{Q}^TA\bar{Q}-\bar{\Omega}\| \le c_4u\left[\kappa^2(\bar{R})+\|A\|\|B\|^2\|\bar{R}^{-1}\|^2+2\|A\bar{Q}\|\|\bar{Q}\|\kappa(\bar{R})\right]$. The accuracy of this scheme can be improved by one step of iterative refinement when we apply the same factorization to the actual $\bar{Q}^TA\bar{Q}$ and get the bound $\|\bar{Q}^TA\bar{Q}-\bar{\Omega}\| \le c_5u\|A\|\|\bar{Q}\|^2$.

$\|M_{11}^{-1}\|$	$\|M^{-1}\|$	$\|M\backslash M_{11}\|$	$\|\bar{R}\| = \|\bar{Q}^{-1}\|$	$\|\bar{R}^{-1}\| = \|\bar{Q}\|$
10^0	1.0000e+00	2.0000e+00	1.9319e+00	1.9319e+00
10^1	1.0000e+00	2.0000e+01	6.3226e+00	6.3226e+00
10^2	1.0000e+00	2.0000e+02	2.0000e+01	2.0000e+01
10^3	1.0000e+00	2.0000e+03	6.3246e+01	6.3246e+01
10^4	1.0000e+00	2.0000e+04	2.0000e+02	2.0000e+02
10^5	1.0000e+00	2.0000e+05	6.3246e+02	6.3246e+02
10^6	1.0000e+00	2.0000e+06	2.0000e+03	2.0000e+03
10^7	1.0000e+00	2.0000e+07	6.3246e+03	6.3246e+03
10^8	1.0000e+00	2.0000e+08	2.0000e+04	2.0000e+04
10^9	1.0000e+00	2.0000e+09	6.3246e+04	6.3246e+04
10^{10}	1.0000e+00	2.0000e+10	2.0000e+05	2.0000e+05
10^{11}	1.0000e+00	2.0000e+11	6.3246e+05	6.3246e+05
10^{12}	1.0000e+00	2.0000e+12	2.0000e+06	2.0000e+06
10^{13}	1.0000e+00	1.9999e+13	6.3245e+06	6.3245e+06
10^{14}	1.0000e+00	2.0004e+14	2.0188e+07	2.0520e+07
10^{15}	1.0000e+00	2.0011e+15	6.6349e+07	5.2040e+07

TABLE 4
The spectral properties of factors with respect to the conditioning of the principal submatrix M_{11} for Problem 2.

$\|M_{11}^{-1}\|$	Algorithm 1 M-QR	Algorithm 2 M-QR2	Algorithm 3 A-CGS	Algorithm 4 A-CGS2
10^0	2.220e-16 (1.050e-15)	3.415e-31 (1.879e-15)	2.220e-16 (1.050e-15)	2.220e-16 (1.879e-15)
10^1	1.857e-15 (9.098e-15)	4.440e-15 (1.797e-14)	9.220e-16 (9.098e-15)	2.579e-15 (1.797e-14)
10^2	9.558e-15 (8.903e-14)	2.541e-14 (1.778e-13)	3.662e-15 (8.903e-14)	2.865e-14 (1.778e-13)
10^3	8.163e-14 (8.884e-13)	5.696e-13 (1.776e-12)	5.684e-14 (8.884e-13)	2.299e-13 (1.776e-12)
10^4	4.839e-13 (8.882e-12)	2.773e-12 (1.776e-11)	2.343e-13 (8.882e-12)	1.688e-12 (1.776e-11)
10^5	6.712e-12 (8.881e-11)	3.480e-11 (1.776e-10)	3.668e-12 (8.881e-11)	1.199e-11 (1.776e-10)
10^6	4.389e-11 (8.881e-10)	2.865e-10 (1.776e-09)	1.627e-11 (8.881e-10)	2.256e-10 (1.776e-09)
10^7	3.253e-11 (8.881e-09)	5.162e-09 (1.776e-08)	1.164e-10 (8.881e-09)	2.476e-09 (1.776e-08)
10^8	1.991e-09 (8.881e-08)	3.829e-08 (1.776e-07)	2.328e-10 (8.881e-08)	9.028e-09 (1.776e-07)
10^9	1.903e-08 (8.881e-07)	4.751e-07 (1.776e-06)	3.093e-08 (8.881e-07)	3.577e-07 (1.776e-06)
10^{10}	9.490e-08 (8.881e-06)	3.141e-06 (1.776e-05)	5.999e-11 (8.881e-06)	1.543e-06 (1.776e-05)
10^{11}	2.037e-06 (8.881e-05)	3.182e-05 (1.776e-04)	2.385e-07 (8.881e-05)	2.080e-05 (1.776e-04)
10^{12}	4.628e-06 (8.881e-04)	2.697e-04 (1.776e-03)	4.265e-06 (8.881e-04)	3.724e-04 (1.776e-03)
10^{13}	3.456e-04 (8.881e-03)	4.352e-03 (1.776e-02)	6.823e-05 (8.881e-03)	2.321e-03 (1.776e-02)
10^{14}	2.303e-03 (8.885e-02)	8.462e-02 (1.776e-01)	6.992e-05 (8.885e-02)	2.660e-02 (1.933e-01)
10^{15}	7.873e-03 (8.897e-01)	8.442e-01 (1.776e+00)	3.244e-02 (8.894e-01)	2.309e-01 (1.864e+00)

TABLE 5
The factorization error $\|B - \bar{Q}\bar{R}\|$ with respect to the conditioning of the principal submatrix M_{11} for Problem 2.

We have considered also the A-CGS algorithm and its version with reorthogonalization A-CGS2 and have shown that their numerical behavior is similar to M-QR decomposition and its variant with refinement M-QR2, respectively.

8. Acknowledgement. The authors would like to express their gratitude to the referees for their careful reading of the manuscript and numerous helpful comments and suggestions. The first author would like to thank Ivan Slapničar, Sanja Singer, Jesse Barlow, José Román, Angelika Bunse-Gerstner, Gérard Meurant, Miroslav Fiedler, Keiichi Morikuni and Nicola Mastronardi for useful comments and fruitful discussion.

$\|M_{11}^{-1}\|$	Algorithm 1 M-QR	Algorithm 2 M-QR2	Algorithm 3 A-CGS	Algorithm 4 A-CGS2
10^0	5.032e-16 (1.010e-14)	3.206e-16 (1.657e-15)	5.341e-16 (1.319e-14)	3.937e-16 (8.286e-16)
10^1	1.288e-15 (1.073e-12)	8.771e-16 (1.775e-14)	1.334e-15 (1.428e-12)	1.261e-15 (8.876e-15)
10^2	4.558e-15 (1.066e-10)	3.595e-15 (1.776e-13)	4.885e-15 (1.422e-10)	3.265e-15 (8.881e-14)
10^3	1.987e-14 (1.065e-08)	1.670e-14 (1.776e-12)	3.264e-14 (1.421e-08)	2.073e-14 (8.881e-13)
10^4	1.515e-13 (1.065e-06)	1.248e-13 (1.776e-11)	1.999e-13 (1.421e-06)	1.384e-13 (8.881e-12)
10^5	1.044e-12 (1.065e-04)	8.175e-13 (1.776e-10)	1.684e-12 (1.421e-04)	1.237e-12 (8.881e-11)
10^6	1.051e-11 (1.065e-02)	7.131e-12 (1.776e-09)	1.116e-11 (1.421e-02)	6.426e-12 (8.881e-10)
10^7	5.844e-11 (1.065e+00)	5.081e-11 (1.776e-08)	6.888e-11 (1.421e+00)	5.110e-11 (8.881e-09)
10^8	3.517e-10 (1.065e+02)	2.385e-10 (1.776e-07)	7.128e-10 (1.421e+02)	5.838e-10 (8.881e-08)
10^9	5.633e-09 (1.065e+04)	4.735e-09 (1.776e-06)	1.691e-08 (1.421e+04)	3.239e-09 (8.881e-07)
10^{10}	6.420e-08 (1.065e+06)	4.727e-08 (1.776e-05)	8.141e-08 (1.421e+06)	4.827e-08 (8.881e-06)
10^{11}	3.312e-07 (1.065e+08)	2.829e-07 (1.776e-04)	8.981e-07 (1.421e+08)	4.216e-07 (8.881e-05)
10^{12}	3.450e-06 (1.065e+10)	2.692e-06 (1.776e-03)	4.197e-06 (1.421e+10)	6.093e-06 (8.881e-04)
10^{13}	2.236e-05 (1.066e+12)	5.520e-05 (1.776e-02)	7.544e-05 (1.420e+12)	4.312e-03 (8.881e-03)
10^{14}	5.407e-04 (1.064e+14)	3.647e-04 (1.776e-01)	3.666e-04 (1.422e+14)	2.060e+00 (9.800e-02)
10^{15}	5.433e-03 (1.101e+16)	2.921e-03 (1.772e+00)	3.650e-03 (1.413e+16)	3.903e+00 (9.997e-01)

TABLE 6

The loss of $(A,\bar{\Omega})$-orthogonality $\|\bar{\Omega} - \bar{Q}^T A\bar{Q}\|$ with respect to the conditioning of the principal submatrix M_{11} for Problem 2.

REFERENCES

[1] J. BARLOW AND A. SMOKTUNOWICZ, *Reorthogonalized block classical Gram–Schmidt*, Numer. Math. 123 (3) (2013), pp. 395–423.

[2] M. BENZI, G. H. GOLUB AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[3] Å. BJÖRCK, *Solving linear least squares problems by Gram-Schmidt orthogonalization*, BIT 7(1) (1967), pp. 1–21.

[4] A. BOJANCZYK, N.J. HIGHAM AND H. PATEL, *Solving the indefinite least squares problem by hyperbolic QR factorization*, SIAM J. Matrix Anal. Appl. 24, 4 (2003), pp. 914–931.

[5] M.A. BREBNER AND J. GRAD, *Eigenvalues of $Ax = \lambda Bx$ for real symmetric matrices A and B computed by reduction to a pseudosymmetric form and the HR process*, Linear Algebra Appl. 43 (1982), pp. 99–118.

[6] J. R. BUNCH, *Analysis of the diagonal pivoting method*, SIAM J. Numer. Anal., 8 (1971), pp. 656–680.

[7] J. R. BUNCH AND B. N. PARLETT, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM J. Numer. Anal., 8 (1971), pp. 639–655.

[8] A. BUNSE-GERSTNER, *An analysis of the HR algorithm for computing the eigenvalues of a matrix*, Linear Algebra Appl. 35 (1981), pp. 155–173.

[9] S. CHANDRASEKARAN, M. GU AND A. H. SAYED, *A stable and efficient algorithm for the indefinite linear least-squares problem*, SIAM J. Matrix Anal. Appl. 20 (1998), pp. 354–362.

[10] J. DELLA-DORA, *Numerical linear algorithms and group theory*, Linear Algebra Appl., 10 (1975), pp. 267–283.

[11] J.W. DEMMEL, N.J. HIGHAM AND R.S. SCHREIBER, *Stability of block LU factorization*, Numer. Linear Algebra with Appl. 2(2) (1995), pp. 173–190.

[12] L. ELSNER, *On some algebraic problems in connection with general eigenvalue algorithms*, Linear Algebra Appl. 26 (1979), pp. 123–138.

[13] P.E. GILL, M.A. SAUNDERS AND J.R. SHINNERL, *On the stability of Cholesky factorization for symmetric quasidefinite systems*, SIAM J. Matrix Anal. Appl. 17, 1 (1996), pp. 35–46.

[14] L. GIRAUD, J. LANGOU, M. ROZLOŽNÍK AND J. VAN DEN ESHOF, *Rounding error analysis of the classical Gram-Schmidt orthogonalization process*, Numer. Math. 101(1) (2005), pp. 97–100.

[15] I. GOHBERG, P. LANCASTER AND L. RODMAN, *Matrices and Indefinite Scalar Products, Operator Theory: Advances and Applications*, vol. 8, Birkhäuser Verlag, Basel, Boston, Stuttgart, 1983.

[16] G.H. GOLUB AND C. VAN LOAN, *Unsymmetric positive definite linear systems*, Linear Algebra Appl. 28 (1979), pp. 85–97.

[17] M. GULLIKSSON, *Backward error analysis for the constrained and weighted linear least squares problem when using the weighted QR factorization*, SIAM J. Matrix Anal. Appl., 16(2) (1995), pp. 675–687.

[18] M. GULLIKSSON, *On the modified Gram-Schmidt algorithm for weighted and constrained linear least*

squares problems, BIT, 35(4) (1995), pp. 453–468.
[19] N.J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, Second edition. SIAM, Philadelphia, PA (2002).
[20] N.J. HIGHAM, *J-orthogonal matrices: properties and generation*, SIAM Review 45(3) (2003), pp. 504–519.
[21] R.A.HORN AND C.R. JOHNSON, *Matrix Analysis*, Cambridge University Press; Reprint edition (February 23, 1990).
[22] D. KRESSNER, M.M. PANDUR AND M. SHAO, *An indefinite variant of LOBPCG for definite matrix pencils*, 2013, Numer. Algorithms, Online First DOI 10.1007/s11075-013-9754-3.
[23] B.R. LOWERY AND J. LANGOU, *Stability analysis of QR factorization in an oblique inner product*, 2013, submitted.
[24] N. MASTRONARDI AND P. VAN DOOREN, *An algorithm for solving the indefinite least squares problem with equality constraints*, BIT 54(1) (2014), pp. 201–218.
[25] N. MASTRONARDI AND P. VAN DOOREN, *A structurally backward stable algorithm for solving the indefinite least squares problem with equality constraints*, IMA Journal on Numerical Analysis, accepted. DOI: 10.1093/imanum/dru004.
[26] E. ROMERO AND J. E. ROMAN. *A parallel implementation of Davidson methods for large-scale eigenvalue problems in SLEPc*, ACM Transactions on Mathematical Software, 40(2), Article 13, 2014.
[27] M. ROZLOŽNÍK, M. TŮMA, A. SMOKTUNOWICZ AND J. KOPAL, *Numerical stability of orthogonalization methods with a non-standard inner product*, BIT Numerical Mathematics 52 (2012), pp. 1035–1058.
[28] T. RUSTEN AND R. WINTHER, *A preconditioned iterative method for saddle-point problems*, SIAM Journal on Matrix Analysis and Applications 13(3) (1992), pp. 887–904.
[29] S. SINGER, *Indefinite QR Factorization*, BIT 46 (2006), pp. 141–161.
[30] S. SINGER AND S. SINGER, *Symmetric indefinite factorization of quasidefinite matrices*, Math. Commun. 4 (1999), 19–25.
[31] S. SINGER AND S. SINGER, *Rounding-error and perturbation bounds for the indefinite QR factorization*, Linear Alg. Appl. 309 (2000), pp. 103–119.
[32] I. SLAPNIČAR, *Componentwise analysis of direct factorization of real symmetric and Hermitian matrices*, Linear Alg. Appl., 272 (1998), pp. 227–275.
[33] I. SLAPNIČAR AND K. VESELIĆ, *A bound for the condition of a hyperbolic eigenvector matrix*, Linear Algebra Appl., 290 (1999), pp. 247–255.
[34] A. SMOKTUNOWICZ, J.L. BARLOW AND J. LANGOU, *A note on the error analysis of classical Gram-Schmidt*, Numer. Math., 105(2) (2006), pp. 299–313.
[35] S.J. THOMAS AND R.V.M. ZAHAR, *Efficient orthogonalization in the M-norm*, Congressus Numerantium 80 (1991), pp. 23–32.
[36] S.J. THOMAS AND R.V.M. ZAHAR, *An analysis of orthogonalization in elliptic norms*, Congressus Numerantium, 86 (1992), pp. 193–222.
[37] R.J. VANDERBEI, *Symmetric quasi-definite matrices*, SIAM J. on Optim., 5(1) 1995, pp. 100–113.