Akademie věd
České republiky

Teze disertace

k získání vědeckého titulu „doktor věd"

ve skupině věd fyzikálně-matematických

Gram-Schmidt orthogonalization in presence of rounding errors

Komise pro obhajoby doktorských disertací v oboru

Matematická analýza a příbuzné obory

Jméno uchazeče: Doc. Dr. Ing. Miroslav Rozložník

Pracoviště uchazeče: Matematický ústav AV ČR

Místo a datum: Praha, 10. 10. 2019

Articles included in this thesis

The articles that are included in this thesis are listed in the order of their appearance in the corresponding sections and subsections.

1. L. Giraud, J. Langou, M. Rozložník, J. van den Eshof, Rounding error analysis of the classical Gram-Schmidt orthogonalization process, Numer. Math. 101(1), 2005, 87–100.

2. L. Giraud, J. Langou, M. Rozložník, The loss of orthogonality in the Gram-Schmidt orthogonalization process, Comput. Math. Appl. 50 (7), 2005, 1069–1075.

3. C.C. Paige, M. Rozložník, Z. Strakoš, Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES, SIAM J. Matrix Anal. Appl. 28 (1), 2006, 264–284.

4. J. Liesen, M. Rozložník, Z. Strakoš, Least squares residuals and minimal residual methods, SIAM J. Sci. Comput. 23 (5), 2002, 1503–1525.

5. P. Jiránek, M. Rozložník, M.H. Gutknecht, How to make Simpler GMRES and GCR more stable, SIAM J. Matrix Anal. Appl. 30 (4), 2008, 1483–1499.

6. P. Jiránek, M. Rozložník, Adaptive version of Simpler GMRES, Numer. Algorithms 53 (1), 2010, 93–112.

7. M. Rozložník, M. Tůma, A. Smoktunowicz, J. Kopal, Numerical stability of orthogonalization methods with a non-standard inner product, BIT 52, 2012, 1035–1058.

8. J. Kopal, M. Rozložník, M. Tůma, Factorized approximate inverses with adaptive dropping, SIAM J. Sci. Comput. 38 (3), 2016, A1807–A1820.

9. M. Rozložník, F. Okulicka-Dluzewska, A. Smoktunowicz: Cholesky-like factorization of symmetric indefinite matrices and orthogonalization with respect to bilinear forms, SIAM J. Matrix Anal. Appl. 36(2), 2015, 727–751.

10. F.J. Hall, M. Rozložník, G-matrices, J-orthogonal matrices, and their sign patterns Czechosl. Mathematical Journal 66 (3), 2016, 653–670.

11. H. Faßbender, M. Rozložník: On the conditioning of factors in the SR decomposition, Linear Algebra Appl. 505, 2016, 224–244.

Contents

The Story of this Thesis

The Gram-Schmidt orthogonalization is perhaps the most widely known representative of a broad class of orthogonalization techniques and strategies. Although the notion of orthogonalization has been around for almost 200 years, only the papers by Jorgen Pedersen Gram and Erhard Schmidt lead to their popularization [12] [37]. Brief biographies of Gram and Schmidt can be found in the survey paper [26].

In this thesis we focus on the numerical behavior of such schemes used for orthogonalization of column vectors. Throughout this text, the orthogonalized vectors are given in advance as the columns of the matrix denoted by A. In particular, we analyze the orthogonalization process of E. Schmidt who recognized that his variant, known nowadays as classical Gram-Schmidt (CGS) algorithm, is essentially the same as an earlier algorithm introduced by J.P. Gram. A slight change of this algorithm gives the modified Gram-Schmidt (MGS) algorithm, that already appeared in a certain form much earlier in the book of P.S. Laplace [25]. Although these two variants are mathematically equivalent, due to rounding errors the set of vectors computed by these two schemes can be far from orthogonal (see e.g. [4]). In various textbooks one can read such statements that CGS can be unstable, and it can quickly lose all semblance of orthogonality or the orthogonality is completely absent. Up to a quite recent time, with the exception of a conjecture without proof by Kielbasinski and Schwetlick [20] [21], there was no bound for the loss of orthogonality in the CGS algorithm. The mechanism of the loss of orthogonality in the CGS algorithm was not studied at all. On the other hand, it was known already some time that the MGS algorithm has much better numerical properties. Thanks to the famous result of seminal paper of A. Björck [3], later reinforced by A. Björck and C.C. Paige in [5], it can be shown that the loss of orthogonality caused by rounding errors in MGS is linked linearly to the condition number $\kappa(A)$ of the matrix A.

Having in mind the first intuitive attempt of Kielbasinski to give a bound for CGS [20], we started to look at this problem. We managed to give a proof of the bound where the loss of orthogonality between the computed vectors depends on the $(n-1)$-th power of $\kappa(A)$, whereas n is the dimension of the initial matrix. However, our extensive numerical experiments indicated that this bound can be a huge overestimate and the loss of orthogonality for all problems does not depend on higher powers of the condition number than 2. In collaboration with L. Giraud, J. Langou and J. van der Eshof we proved in [10] that the loss of orthogonality for CGS can be bounded in terms of the square of the condition number $\kappa(A)$. This is true for every ma-

trix A such that A^TA is numerically nonsingular. The key observation here is that the computed triangular factor is numerically similar to the triangular factor computed in the Cholesky factorization of the associated cross-product (or Gram) matrix A^TA. This result essentially leads to theoretical justification of CGS that was widely overlooked and considered as unreliable. We also illustrated through numerical experiments that this bound is sharp [10], [11]. These results were further reinforced by J. Barlow, J. Langou and A. Smoktunowicz in [2], who pointed out the importance of a stable computation of normalization coefficients in CGS.

Another important open problem was related to the reorthogonalization that is frequently used to improve the orthogonality of vectors computed by some Gram-Schmidt scheme. The orthogonalization step (either in the CGS or MGS algorithm) is iterated twice or several times and the ultimate goal is to produce a set of vectors whose orthogonality is close to the working precision. Extensive experiments with iterative versions of Gram-Schmidt process were performed by Rice [31] and various schemes have been studied by several authors, including Abdelmalek, Daniel, Gragg, Kaufmann, Stewart and Ruhe [1], [6] and [36]. The first simple analysis for the case of two given vectors due to Kahan was published by Parlett in [30], who showed that unless the vectors are linearly dependent to roundoff unit, one reorthogonalization suffices to achieve orthogonality also to the round off unit level. Unfortunately, such analysis cannot be extended to the case with more than two vectors. Later, Hoffmann taking into account experimental results on numerical nonsingular problems in [17], observed that a third iteration never occurred for both CGS and MGS. Hoffmann thus conjectured that two iterations are enough for obtaining orthogonality on the level of round off unit also for general problems with more than two vectors [17]. However, a theoretical foundation for this observation remained an open question. In collaboration with L. Giraud and J. Langou we succeeded to analyze the CGS algorithm with one reorthogonalization. Assuming numerical full rank of the matrix A, we proved that these two iterations are sufficient to guarantee the orthogonality between computed vectors to the roundoff unit level [10], [11]. The key ingredient for the proof is the fact that the norm of the computed projection after the first step cannot be infinitely small, and it is bounded from below by the minimal singular value of A. Our main contribution explains that the size of normalization coefficients in the Gram-Schmidt process is essentially controlled by the condition number of the matrix A. Thus the "two-steps-are-enough" conjecture is indeed true for any set of initial vectors with a numerical full-rank. The importance of having such result was even

more profound since many recent experimental results indicate that CGS with (one) reorthogonalization maybe faster than MGS despite the fact that it performs twice as many arithmetical operations.

The above mentioned results on the loss of orthogonality in the Gram-Schmidt algorithm can be applied in the context of the GMRES method that is one of the most important and widely used iterative methods for solving linear systems. The classical GMRES method for solving nonsymmetric linear systems is based on the Arnoldi process, where some Gram-Schmidt algorithm is applied for constructing an orthonormal basis of associated Krylov subspace. This process is significantly different from the standard orthogonalization where the vectors are given in advance, since in the Arnoldi process the new vectors in the basis are computed recursively from previously computed basis vectors. It was shown in [8] that it can be seen as a recursive column-oriented QR of a particular matrix with the conditioning closely related to the minimum residual least squares problem in the GMRES method.

Numerical behavior of several implementations of GMRES was analyzed some time ago by several authors [8], [13]. In collaboration with C.C. Paige and Z. Strakoš we have proved in [29] the backward stability of the MGS GMRES algorithm showing that it belongs to numerically stable and reliable iterative schemes. The concept of backward stability of an iterative method is rather different from the standard concept used for direct methods, see e.g. [8] or [29]. It assumes that at some iteration step that is less or equal than the dimension, the computed approximate solution satisfies the perturbed linear system, where the relative perturbations of the system matrix and the right-hand side are proportional to the roundoff unit.

We also studied numerical stability for several other implementations of GMRES. Variants of residual minimizing Krylov subspace methods were comprehensively described in the paper [34] that served as a starting point for later developments. In the joint paper with J. Liesen and Z. Strakoš [27] we explain that the choice of the Krylov subspace basis is substantial for the numerical stability of GMRES. This is the case also for Simpler GMRES proposed in [38]. By shifting the Arnoldi process to begin with a different vector, a non-orthogonal basis of the Krylov subspace is generated. This leads to the implementation of GMRES, where the associated least squares problem is replaced by an easier-to-solve triangular system, see [38]. Simpler GMRES is numerically unstable due to the fact that the chosen Krylov basis is getting ill-conditioned as soon as the approximate solution converges. Closer to the exact solution, higher condition number of the basis and upper triangular matrix leading to the poor accuracy of the

computed approximate solution. This result is quite counterintuitive, and it actually shows, that even the most stable orthogonalization technique used in Simpler GMRES does not ensure a high accuracy of computed approximate solution.

On the other hand, in the paper with M.H. Gutknecht and P. Jiránek [19] we proved that the Krylov subspace basis containing the normalized residuals from the GMRES method is well-conditioned as long as we have a reasonable residual norm decrease. These results lead to a new implementation of GMRES (called RB-SGMRES) which is conditionally backward stable. The notion of conditional backward stability is also quite different from the notion of conditional backward stability standardly used e.g. for Gaussian elimination. Another stable variant of Simpler GMRES was proposed in the joint work with P. Jiránek [18]. It is based on the adaptive choice of the Krylov subspace basis at a given iteration step with the use of a criterion monitoring the intermediate residual norm decrease.

Later we continued our research considering a more general setting assuming the orthogonalization of vectors not only with respect to the standard inner product, but also with respect to some non-standard inner product, symmetric indefinite or skew-symmetric bilinear form. Throughout this text, the matrix that induces the inner product or bilinear form is denoted by B. In collaboration with M. Tůma, A. Smoktunowicz and J. Kopal [35] we analyzed a numerical behavior of the most frequently used orthogonalization schemes with respect to some non-standard inner product. We looked at the effects of conditioning of B on the factorization and gave bounds for the loss of orthogonality in such Gram-Schmidt process showing a significant difference to the case with the standard inner product or with the diagonal weighting. It is given by the fact that the size of local rounding errors in the orthogonalization is not determined by a nonstandard norm of the basis vectors but by their Euclidean norm that can be much larger up to the factor related to the condition number $\kappa(B)$ of the matrix B that induces this inner product. Note that the problem of extension of results existing for the standard inner product is nontrivial especially in cases where the norm induced by B is not monotonic and where it may happen that minimization of the B-norm of the projections computed in the orthogonalization process may lead to the amplification of local rounding errors due to their large Euclidean norms.

The theory when the basis of standard unit vectors is orthogonalized with respect to some other inner product is even more developed as it is heavily used in such applications as approximate inverse preconditioning. An important class of preconditioners is based on computing an approximate factorization of the matrix inverse that is

sufficiently sparse and robust. A new approach to construct approximate inverses for a symmetric positive definite matrix was proposed in the joint paper with M. Tůma and J. Kopal [22]. This scheme is based on adaptive dropping that is orthogonalization step dependent and on monitoring the condition number of the triangular factor in the related Gram-Schmidt algorithm.

We analyzed also the numerical behavior of the Gram-Schmidt orthogonalization with respect to a symmetric indefinite bilinear form that is induced by a general symmetric but nonsingular matrix B. In contrast to the case of inner product where the accuracy of computed factors depends only on conditioning of the initial matrix and of the matrix B that induces this inner product, the accuracy of schemes with bilinear forms depends also on the conditioning of all principal submatrices of the Gram matrix $A^T BA$, where there is a change of the sign in the corresponding signature matrix. This result was shown in collaboration with F. Okulicka-Dluzewska and A. Smoktunowicz. Indeed, the orthogonalization with respect to a bilinear form is much more complicated. Nevertheless, several algorithms for computing such factorizations were analyzed in [33], including the classical Gram-Schmidt and Gram-Schmidt with reorthogonalization.

The matrices that are orthogonal with respect to the indefinite signature matrix are often called J-orthogonal matrices. In the paper with F.J. Hall several connections between the class of J-orthogonal and the class of the so-called G-matrices were established. Based on this relation, several results on the sign patterns of J-orthogonal matrices were developed in [9].

As it is discussed in the last part of this thesis, the situation is even more complicated in the case of orthogonalization with respect to a skew-symmetric bilinear form, since the normalization step is not uniquely defined and there is a freedom in the computation of corresponding semi-symplectic and triangular factors. Such structure preserving algorithms are very popular tool in numerical linear algebra and various implementations of orthogonalization schemes were introduced. In a joint paper with H. Fassbender we presented best choices for free parameters in such orthogonalization scheme, in particular the parameters that minimize the condition number of the diagonal blocks in the triangular factor, or the parameters that minimize the condition number of the corresponding block in the semi-symplectic factor. For details we refer to the paper [9].

The author of this thesis would like to thank the coauthors of all papers in this thesis for their invention, effort, patience and friendship throughout many years of our collaboration.

1 Orthogonalization with respect to the standard inner product

Let $A = (a_1, \ldots, a_n) \in \mathcal{R}^{m,n}$ be a real $m \times n$ matrix with a full column rank with $m \geq \text{rank}(A) = n$. Throughout this section we consider orthogonalization techniques with respect to the standard inner product that generate an orthogonal basis $Q = (q_1, \ldots, q_n) \in \mathcal{R}^{m,n}$ of span(A) such that $A = QR$, where $R = (r_{i,j}) \in \mathcal{R}^{n,n}$ is upper triangular, and it contains the off-diagonal orthogonalization coefficients $r_{i,j}$, $i = 1, \ldots j-1$ and positive orthonormalization coefficients $r_{j,j}$ on its diagonal for each $j = 1, \ldots, n$. It is well-known that if the diagonal entries are set positive, then Q and R are uniquely defined. It follows from $Q^TQ = I$ that the factor R is equal to the Cholesky factor of the Gram-matrix A^TA satisfying $A^TA = R^TR$. Moreover, the condition numbers of Q and R are equal to $\kappa(Q) = 1$ and $\kappa(R) = \kappa(A)$.

There is no doubt that Gram-Schmidt orthogonalization is the most widely known and used orthogonalization technique for computing the factors Q and R. In commemoration of the 100th anniversary of contributions of Gram and Schmidt, a comprehensive survey of results on the Gram-Schmidt orthogonalization was published in [26]. It starts with a brief biographies together with the discussion of relation of their works to the QR factorization and to least squares problems. Introductory sections of [26] are followed with such issues as the loss of orthogonality between the vectors computed in finite precision arithmetic, reorthogonalization, stable solution of least squares problem or applications or application of Gram-Schmidt to Krylov subspace methods. Several computational version of the Gram-Schmidt process has been derived and analyzed.

In this section we focus on numerical properties of the Gram-Schmidt orthogonalization and we study the effects of rounding errors on this orthogonalization technique. The Gram-Schmidt process has two basic computational variants: the classical Gram-Schmidt (CGS) algorithm and the modified Gram-Schmidt (MGS) algorithm (see e.g. [3, 4, 26]). Due to rounding errors the set of vectors produced by either of these two methods can be far from orthogonal and sometimes the orthogonality can even be completely absent [3, 31]. Generally it is agreed that the MGS algorithm has much better numerical properties than the CGS algorithm [31, 4]. It is also well-known that the orthogonality between the vectors computed either by CGS or MGS can be improved by reorthogonalization. Here we concentrate on the classical or modified Gram-Schmidt algorithms and the classical Gram-Schmidt algorithm with reorthogonalization. These three algorithms are summarized in Table 1.

classical Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$
for $i = 1, \ldots, j-1$
$r_{i,j} = \langle a_j, q_i \rangle$
$u_j = u_j - r_{i,j} q_i$
$r_{j,j} = \sqrt{\|a_j\|^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$
$q_j = u_j / r_{j,j}$

modified Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$
for $i = 1, \ldots, j-1$
$r_{i,j} = \langle u_j, q_i \rangle$
$u_j = u_j - r_{i,j} q_i$
$r_{j,j} = \sqrt{\|a_j\|^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$
$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:
for $j = 1, \ldots, n$
$u_j = a_j$
for $k = 1, 2$
$a_j^{(k)} = u_j$
for $i = 1, \ldots, j-1$
$r_{i,j}^{(k)} = \langle a_j^{(k)}, q_i \rangle$
$u_j = u_j - r_{i,j}^{(k)} q_i$
$r_{j,j} = \|u_j\|$
$q_j = u_j / r_{j,j}$

Table 1: Gram-Schmidt orthogonalization with respect to standard inner product: classical algorithm, modified algorithm, and classical algorithm with reorthogonalization.

As the main goals of this section, we give a bound for the loss of orthogonality between the computed vectors in CGS, and under certain assumption on numerical nonsingularity of A we prove that their orthogonality can be significantly improved by one step of reorthogonalization. These two results are perhaps our most important contributions to understanding the numerical behavior the classical Gram-Schmidt algorithm and its version with reorthogonalization.

1.1 Rounding error analysis of classical Gram-Schmidt algorithm

It was observed in numerical experiments that the classical Gram-Schmidt process can compute a set of vectors which is far from orthogonal and sometimes the orthogonality can be lost completely [3, 31, 21, 20]. The key observation is that the computed triangular factor is numerically similar to the triangular factor computed in the Cholesky factorization of the associated cross-product (or Gram) matrix A^TA. It was shown in [10] that the triangular factor $\bar{R}$ computed in CGS is an exact Cholesky factor of the perturbed matrix $A^TA + E$ satisfying

$$A^TA + E = \bar{R}^T\bar{R}, \quad \|E\| \leq \mathcal{O}(u)\|A\|^2, \tag{1}$$

where $\|\dot{\|}$ denotes the spectral matrix norm and $\mathcal{O}(u)$ denotes some low-degree polynomial in the dimensions m and n multiplied by the roundoff unit u.

This result essentially gives rise to the fact that the loss of orthogonality between the vectors computed by the classical Gram-Schmidt algorithm can be bounded by the term proportional to the square of condition number of the matrix A. As it was shown in [10, 2], provided that the Gram matrix A^TA is numerically nonsingular, i.e. assuming that $\mathcal{O}(u)\kappa^2(A) < 1$, the loss of orthogonality between the column vectors in the factor $\bar{Q}$ computed by the classical Gram-Schmidt process can be bounded by a term proportional to the square of the condition number of A. Indeed, we have

$$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa^2(A)}{1 - \mathcal{O}(u)\kappa(A)}. \tag{2}$$

For rigorous analysis and other details we refer to the papers [10, 2, 11].

1.2 Gram-Schmidt algorithm with reorthogonalization

In contrast to the modified Gram-Schmidt algorithm [3, 4, 5], where the loss of orthogonality depends linearly on the conditioning of initial vectors as

$$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa(A)}{1 - \mathcal{O}(u)\kappa(A)}, \tag{3}$$

in the case of the classical Gram-Schmidt algorithm we have the quadratic dependence. Depending on the condition number of A, it may be or may not be satisfactory in practical situations. It is well-known fact that the orthogonality between computed vectors can be improved by reorthogonalization [6, 17, **?**]. The key ingredient is the proof that two steps are enough

for ensuring the orthogonality on the optimal level, when we apply the classical or modified Gram-Schmidt algorithm on a set of numerically nonsingular vectors satisfying the assumption $\mathcal{O}(u)\kappa(A) < 1$. It is based on a result showing that the norm of the projection $\bar{u}_j$ computed in even finite precision arithmetic cannot be infinitely small and essentially it is is bounded from below by the minimal singular value of A so that

$$\|\bar{u}_j\| \geq \sigma_n(A) - \mathcal{O}(u)\|A\|. \tag{4}$$

Using previous bound, it was shown in [10] that one step of reorthogonalization is enough for preserving the orthogonality of computed vectors close to the unit roundoff level. Assuming the numerical nonsingularity of initial column vectors, the orthogonality of the vectors $\bar{Q}$ computed by the classical Gram-Schmidt process with one step of reorthogonalization can be bounded as

$$\|I - \bar{Q}^T\bar{Q}\| \leq \mathcal{O}(u). \tag{5}$$

This phenomenon is often called as "two-steps-are-enough". For details we refer to the papers [10], [1] and to the short survey paper [11].

2 Arnoldi algorithm and GMRES

Given a squared nonsingular matrix $A \in \mathcal{R}^{m,m}$ and a vector $b \in \mathcal{R}^m$, the j-th Krylov subspace generated by A and b is defined as

$$K_j(A, b) = \text{span}\{b, Ab, \ldots, A^{j-1}b\}, \quad j = 1, 2, \ldots.$$

The results on the Gram-Schmidt orthogonalization of vectors can also be used in the context of the Arnoldi algorithm for constructing an orthonormal basis $V_j = (v_1, \ldots, v_j)$ of the Krylov subspace $K_j(A, b)$. In the variants of the Arnoldi algorithm that are based on the Gram-Schmidt orthogonalization the first vector v_1 is taken as a normalized vector b and a new basis vector v_{j+1} is the normalized result of the orthogonalization of the vector Av_j with respect to the previously generated vectors $v_1, \ldots, v_j$. Thus Arnoldi algorithm can be seen as a column-oriented QR factorization

$$[b, AV_j] = V_{j+1}R_{j+1}, \tag{6}$$

where $R_{j+1} \in \mathcal{R}^{j+1,j+1}$ is upper triangular with orthogonalization coefficients in its strict upper triangular part and with normalization coefficients on its diagonal. On the other hand, it is well-known fact that the conditioning of

the matrix $[b, AV_j]$ is closely related to the associated minimum residual least squares problem

$$\|b - AV_j y_j\| = \min_{y \in \mathcal{R}^j} \|b - AV_j y\|, \tag{7}$$

where $x_j = V_j y_j$, $j = 1, 2, \ldots$, defines the sequence of approximate solutions to the linear system $Ax = b$ generated by the Generalized minimum residual (GMRES) method. The minimum residual principle is represented by the least squares problem (7), and thus the GMRES method is often described as a sequence of least squares problems of increasing dimension. Mathematically (in exact arithmetic), there are several algorithmic variants for generating this sequence. Computationally (in finite precision arithmetic), however, different algorithms for computing the same sequence may produce significantly different results [8].

2.1 Backward stability of MGS-GMRES

The most usual implementation is modified Gram-Schmidt GMRES (MGS-GMRES). Using the relationship (6) it was shown [13] that the loss of orthogonality between the vectors $\bar{V}_{j+1}$ computed by the modified Gram-Schmidt Arnoldi algorithm is bounded as

$$\|I - \bar{V}_{j+1}^T \bar{V}_{j+1}\| \leq \frac{\mathcal{O}(u)\kappa([\bar{v}_1, A\bar{V}_j])}{1 - \mathcal{O}(u)\kappa([\bar{v}_1, A\bar{V}_j])}. \tag{8}$$

In addition, assuming that $\min_y \|\bar{v}_1 - A\bar{V}_j y\| \geq \mathcal{O}(u)\kappa(A)$ the condition number of the matrix $[\bar{v}_1, A\bar{V}_j]$ can be bounded further as

$$\kappa([\bar{v}_1, A\bar{V}_j]) \leq \frac{\mathcal{O}(1)\kappa(A)}{\min_{y \in \mathcal{R}^j} \|\bar{v}_1 - A\bar{V}_j y\|}.$$

Thus the complete loss of orthogonality (resulting into loss of linear independence of computed vectors) in the modified Gram-Schmidt Arnoldi algorithm can occur only after the residual $r_j = b - A\bar{V}_j y_j$ reaches its final accuracy level $\mathcal{O}(u)\kappa(A)$. These results were reinforced in the paper [29], where it was shown that MGS-GMRES is a backward stable algorithm. Indeed, for some iteration step $j \leq m$ the computed approximate solution $\bar{x}_j$ in MGS-GMRES satisfies the perturbed system $(A + \Delta A_j)\bar{x}_j = b + \Delta b_j$, where the relative perturbations are proportional to the roundoff unit as $\|\Delta A_j\|/\|A\| \leq \mathcal{O}(u)$ and $\|\Delta b_j\|/\|b\| \leq \mathcal{O}(u)$. This result depends on a more general result on the backward stability of a variant of the MGS algorithm applied to solving the minimum residual least squares problem, and uses results on MGS and its loss of orthogonality, together with an important relation between least squares residual norms and singular values of matrices associated to the studied least squares problem. For details we refer to [29].

2.2 Simpler GMRES is inherently unstable

Finite precision analysis was performed for several important implementations of GMRES [29, 27, 19, 18]. Our results in [27, 19] explain why the choice of the Krylov subspace basis is fundamental for the numerical stability of some implementation. Instead of V_j we consider in general a nonorthogonal but normalized basis $Z_j = (z_1, \ldots, z_j)$ of the Krylov subspace $K_j(A, b)$, the approximate solutions x_j in GMRES can be written as $x_j = Z_j z_j$, whereas the coefficient vector z_j is given as the solution of the upper triangular system $U_j z_j = Q_j^T b$ and where $Q_j = (q_1, \ldots, q_j)$ denotes the orthonormal basis of the subspace $AK_j(A, b)$ obtained from the QR factorization $AZ_j = Q_j U_j$. As it was shown for the case of Simpler GMRES in [27], where $Z_j = [b/\|b\|, Q_{j-1}]$, this choice of the basis is not very suitable from the stability of point of view. The conditioning of $[b/\|b\|, Q_{j-1}]$ is related to the convergence of the GMRES method as

$$\frac{\|b\|}{\min_{y \in \mathcal{R}^j} \|b - AV_j y\|} \leq \kappa([b/\|b\|, Q_{j-1}]) \leq \frac{2\|b\|}{\min_{y \in \mathcal{R}^j} \|b - AV_j y\|}. \qquad (9)$$

Due to (9) small residuals in the GMRES method lead to the ill-conditioning of matrices $A[b/\|b\|, Q_{j-1}]$ and U_j and this affects negatively the accuracy of computed coefficient vectors z_j and approximate solutions x_j. Indeed, even the best possible orthogonalization technique used for computing the basis Q_j does not compensate for the loss of accuracy due to an inappropriate choice of the basis Z_j. Therefore, Simpler GMRES is inherently less numerically stable than the the usual implementation of GMRES that uses classical or modified Gram-Schmidt algorithm. The details can be found in [27] and [19].

2.3 How to make simpler GMRES and GCR method more stable

We have already indicated that a different choice of the basis can significantly influence the numerical behavior of the resulting implementation. While Simpler GMRES is less stable due to the ill-conditioning of the basis $[b/\|b\|, Q_{j-1}]$, the residual basis defined as $Z_j = \tilde{R}_j = (\tilde{r}_0, \ldots, \tilde{r}_{j-1})$, where $\tilde{r}_k = r_k/\|r_k\|$ for $k = 0, \ldots, j-1$ (with $r_0 = b$) is the normalized residual from the GMRES method, is well-conditioned as long as we have a reasonable residual norm decrease. It was shown in [19] that if $b \notin AK_{j-1}(A, b)$ and $\|r_k\| < \|r_{k-1}\|$ for $k = 1, \ldots j-1$, the condition number of $\tilde{R}_j$ satisfies

$$\kappa(\tilde{R}_j) \leq \sqrt{m \left(1 + \sum_{k=1}^{j-1} \frac{\|r_{k-1}\|^2 + \|r_k\|^2}{\|r_{k-1}\|^2 - \|r_k\|^2}\right)}. \qquad (10)$$

These results lead to a new implementation, which is conditionally backward stable [19]. They also explain the experimentally observed fact that another mathematically equivalent method called GCR delivers very accurate approximate solutions when it converges fast enough without stagnation.

Another stable variant of Simpler GMRES was proposed in [18] and it is based on the adaptive choice of the Krylov subspace basis at a given iteration step using the intermediate residual norm decrease criterion. The new direction vector in the basis Z_j is chosen as in the original implementation of Simpler GMRES or it is equal to the normalized residual vector $\tilde{r}_j$. We show that such an adaptive strategy leads to a well-conditioned basis of the Krylov subspace and chosen the appropriate criterion such implementation of GMRES computes very accurate approximate solutions. A detailed analysis can be found in [18].

3 Orthogonalization with respect to a nonstandard inner product

Considering the inner product $\langle \cdot, \cdot \rangle_B$ induced by some symmetric positive definite matrix $B \in \mathcal{R}^{m,m}$, we can look for the B-orthogonal basis $Q = (q_1, \ldots, q_n) \in \mathcal{R}^{m,n}$ of the range of $A = (a_1, \ldots, a_n) \in \mathcal{R}^{m,n}$ satisfying the condition $Q^T B Q = I$. The Gram-Schmidt orthogonalization with respect to such nonstandard inner product leads to the factors Q and R satisfying $A = QR$, where $R \in \mathcal{R}^{n,n}$ is upper triangular with positive diagonal entries. It is clear that if B is symmetric positive definite, then the Gram matrix $A^T B A$ is also symmetric positive definite and its Cholesky factor is exactly equal to the factor R. It follows also from $A^T B A = R^T R$ that extremal singular values and condition number of R satisfy

$$\|R\| = \|B^{1/2} A\|, \ \|R^{-1}\| = 1/\sigma_m(B^{1/2} A), \ \kappa(R) = \kappa(B^{1/2} A) = \kappa^{1/2}(A^T B A),$$

where $B^{1/2}$ stands for the square root of the matrix B. Although the column vectors in the factor Q are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_B$, they are no longer orthogonal with respect to the standard inner product. Depending on the conditioning of the matrix B, the matrix Q can be rather ill-conditioned. Its extremal singular values and condition number satisfy

$$\|Q\| \leq \|B^{-1}\|^{1/2}, \quad \sigma_m(Q) \geq 1/\|B\|^{1/2}, \quad \kappa(Q) \leq \kappa^{1/2}(B).$$

The corresponding classical Gram-Schmidt algorithm, modified Gram-Schmidt algorithm, and classical Gram-Schmidt algorithm with reorthogonalization are given in Table 2.

classical Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$
for $i = 1, \ldots, j-1$
$r_{i,j} = \langle a_j, q_i \rangle_B$
$u_j = u_j - r_{i,j} q_i$
$r_{j,j} = \sqrt{\|a_j\|_B^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$
$q_j = u_j / r_{j,j}$

modified Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$
for $i = 1, \ldots, j-1$
$r_{i,j} = \langle u_j, q_i \rangle_B$
$u_j = u_j - r_{i,j} q_i$
$r_{j,j} = \sqrt{\|a_j\|_B^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$
$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:
for $j = 1, \ldots, n$
$u_j = a_j$
for $k = 1, 2$
$a_j^{(k)} = u_j$
for $i = 1, \ldots, j-1$
$r_{i,j}^{(k)} = \langle a_j^{(k)}, q_i \rangle_B$
$u_j = u_j - r_{i,j}^{(k)} q_i$
$r_{j,j} = \|u_j\|_B$
$q_j = u_j / r_{j,j}$

Table 2: Gram-Schmidt orthogonalization with respect to nonstandard inner product: classical algorithm, modified algorithm, and classical algorithm with reorthogonalization.

3.1 Numerical properties of orthogonalization methods with a nonstandard inner product

In paper [35] we study the numerical properties of several orthogonalization schemes, where the inner product is induced by a nontrivial symmetric positive definite matrix B. We analyzed the effect of its conditioning on the factorization and the loss of orthogonality between vectors computed in finite precision arithmetic. We consider the reference implementation based on the backward stable eigen-decomposition, modified and classical Gram-Schmidt algorithms, classical Gram-Schmidt algorithm with reorthogonalization as well as the implementation motivated by the approximate inverse preconditioner called AINV.

It is shown that in the case of a diagonal (and positive definite) B is similar to the case with the standard inner product where $B = I$. The bounds for the loss of B-orthogonality between the computed vectors $\bar{Q}$ in the classical and modified Gram-Schmidt algorithm are analogous to the bounds (2) and (3), respectively, but the matrix A is replaced by the matrix $B^{1/2}A$. For a diagonal B the matrix $B^{1/2}A$ is just the matrix A scaled by rows. Thus application of the Gram-Schmidt algorithm with the B-inner product applied to A is thus numerically similar to the Gram-Schmidt algorithm with the standard inner product applied to the row-scaled matrix $B^{1/2}A$.

The situation is more complicated in the case of a general symmetric positive definite B, where the norm $\|\cdot\|_B$ induced by B is not monotonic. Rounding error analysis of the classical Gram-Schmidt algorithm or modified Gram-Schmidt algorithm can be extended also to such case of nonstandard inner product. However, the resulting bounds contain additional factors that depend explicitly or implicitly on the condition number $\kappa(B)$.

It is shown in [35] (see also [28]) that if $\mathcal{O}(u)\kappa(B)\kappa(B^{1/2}A)\kappa(A) < 1$, then the loss of orthogonality in $\bar{Q}$ computed by the classical Gram-Schmidt algorithm is bounded by

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\|B\|^{1/2}\|\bar{Q}\|\kappa(B^{1/2}A)\kappa^{1/2}(B)\kappa(A)}{1 - \mathcal{O}(u)\|B\|^{1/2}\|\bar{Q}\|\kappa(B^{1/2}A)\kappa^{1/2}(B)\kappa(A)}. \tag{11}$$

Indeed the loss of B-orthogonality in the classical Gram-Schmidt algorithm is bounded by a quantity proportional not only to $\kappa(B^{1/2}A)\kappa^{1/2}(B)\kappa(A)$ (that essentially means the square of the condition number of the matrix $B^{1/2}A$ or in other words the condition number of the Gram matrix A^TBA) but also to the additional factor $\|B\|^{1/2}\|\bar{Q}\|$ with the worst-case bound

$$\|B\|^{1/2}\|\bar{Q}\| \leq \kappa^{1/2}(B).$$

For the modified Gram-Schmidt algorithm it is proved in [35] that assuming $\mathcal{O}(u)\kappa(B^{1/2}A)\kappa(B) < 1$ the loss of orthogonality between the computed columns in $\bar{Q}$ is bounded by

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\|B\|^{1/2}\|\bar{Q}\|\kappa^{1/2}(B)\kappa(B^{1/2}A)}{1 - \mathcal{O}(u)\|B\|^{1/2}\|\kappa^{1/2}(B)\bar{Q}\|\kappa(B^{1/2}A)}. \tag{12}$$

The loss of B-orthogonality in the modified Gram-Schmidt algorithm is thus significantly better than in the classical Gram-Schmidt algorithm. Nevertheless, the bound (12) is proportional not only to the condition number of $B^{1/2}A$ but also to the quantity $\|B\|^{1/2}\|\bar{Q}\|\kappa^{1/2}(B)$ that can be in the worst-case equal to $\kappa(B)$ and that represents the effect of the nonstandard inner product induced by the matrix B.

The B-orthogonality between the vectors computed by the Gram-Schmidt algorithm can be thus significantly lost and it can be improved by reorthogonalization. Here we consider the classical Gram-Schmidt algorithm with (one step of) reorthogonalization, see Table 2. The key idea here is that the B-norm of the projection computed after the first orthogonalization step is not infinitely small, but it remains bounded from below by the minimal singular value of the matrix $B^{1/2}A$. Taking into account also the second orthogonalization step, this result leads to the bound for the B-orthogonality that does not depend on the matrix $B^{1/2}A$. Assuming $\mathcal{O}(u)\kappa^{1/2}(B)\kappa(B^{1/2}A) < 1$, the loss of B-orthogonality between the computed columns of $\bar{Q}$ in the classical Gram-Schmidt algorithm with reorthogonalization is bounded by

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \mathcal{O}(u)\|B\|\|\bar{Q}\|^2. \tag{13}$$

Note that although this bound does not depend on $B^{1/2}A$, it does depend on the condition number of B as $\|B\|\|\bar{Q}\|^2 \leq \kappa(B)$. For details we refer, e.g., to papers [35] or [28].

3.2 Approximate inverse preconditioning

For the particular case $A = I$ the situation is more developed and the idea of computing B-orthogonal vectors from standard unit basis vectors is heavily used in many applications. If $A = I$, then the matrix R is the Cholesky factor of B satisfying $B = R^T R$ and the factor $Q = R^{-1}$ is its upper triangular inverse satisfying $\kappa(Q) = \kappa(R) = \kappa^{1/2}(B)$. In addition, Q represents an inverse factor in the triangular factorization $A^{-1} = QQ^T$. One of the important preconditioning classes involves computing an approximate inverse factorization such that it is a sparse approximation of A^{-1}. Many efficient schemes with incomplete factorizations of this form have been proposed and they are frequently used for practical problems. Although the main motivation for their development comes from parallel processing, concerns on the robustness and accuracy of such schemes became very important. The initial techniques as the AINV algorithm use for (incomplete) orthogonalization various oblique projections or the classical Gram-Schmidt algorithm. The AINV algorithm is actually a modification of the modified Gram-Schmidt algorithm "backwards" to the classical Gram-Schmidt algorithm. However, the recent trend is to stabilize them in terms of orthogonalization technique used in the factorization. This has lead to the use of modified Gram-Schmidt algorithm in the preconditioner called SAINV together with the accurate computation of diagonal entries in R. More detailed description and appropriate references can be found in [35].

A new approach to construct approximate inverses for a symmetric positive definite matrix B is proposed in [22]. This scheme is based on adaptive dropping in the computation of factors $\hat{Q}$ and $\hat{R}$ in approximate factorizations $\hat{Q}\hat{R} \approx I$ and $\hat{R}\hat{Q} \approx I$ so that $\hat{R}^T\hat{R} \approx B$ and $\hat{Q}\hat{Q}^T \approx B^{-1}$. Indeed, using the approximate inverse $\hat{Q}\hat{Q}^T$ as a preconditioner for the system $Bx = b$, where $B \in \mathcal{R}^{m,m}$ and $b \in \mathcal{R}^m$, the preconditioned system is of the form

$$\hat{Q}^T B\hat{Q}y = \hat{Q}^T b, \quad x = \hat{Q}y. \tag{14}$$

The quality of the approximation is thus given by the loss of orthogonality $\hat{Q}^T B\hat{Q} - I$ between the column vectors in the factor $\hat{Q}$. The crucial idea of this approach is to drop entries in $\bar{Q}$ so that the size of the right residual $\bar{R}\bar{Q} - I$ is throughout the orthogonalization uniformly bounded by a drop tolerance τ. This strategy essentially means to consider orthogonalization step-dependent dropping and to introduce the parameter $\tau_k \leq \tau/\kappa(\hat{R}_k)$ that takes into account the conditioning of the k-th principal submatrix of the factor $\hat{R}$ (denoted as $\hat{R}_k$ for $k = 1, \ldots, m$ here). This dropping techniques is thus based on monitoring the condition number $\kappa(\hat{R}_k)$ that increases with the orthogonalization step and thus the sequence of drop tolerances τ_k decreases as $\kappa(\hat{R}_k)$ increases. A natural strategy is then to keep the increase of $\kappa(\hat{R}_k)$ as low as possible and this is achieved by the column pivoting in the Gram-Schmidt algorithm. Based on numerical experiments it is shown in [22] that this approximate inverse preconditioner can efficiently solve large difficult problems and it is rather robust in comparison to other non-adaptive approximate inverse preconditioners.

4 Orthogonalization with respect to an indefinite bilinear form

For a symmetric but indefinite and nonsingular matrix $B \in \mathcal{R}^{m,m}$ and for a full column rank matrix $A = (a_1, \ldots, a_n) \in \mathcal{R}^{m,n}$ we can also look for the decomposition $A = QR$, where the columns of $Q = (q_1, \ldots, q_n) \in \mathcal{R}^{m,n}$ are mutually orthogonal with respect to the bilinear form $\langle B\cdot, \cdot\rangle$ so that $Q^T BQ = \Omega = (\mathrm{diag}(\omega_j)) \in \mathrm{diag}(\pm 1)$, and where $R \in \mathcal{R}^{n,n}$ is upper triangular with positive diagonal entries. It follows that under assumption on nonzero principal minors of $A^T BA$ (or in other words, if $A^T BA$ is strongly nonsingular) such decomposition exists and the triangular factor R satisfies the Cholesky-like factorization $A^T BA = R^T \Omega R$. Conversely, given the Cholesky-like factorization $A^T BA = R^T \Omega R$, the factor Q can be recovered as $Q = AR^{-1}$.

Note that for positive definite B the signature matrix Ω is equal to $\Omega = I$, and the condition numbers are equal to or can be bounds as $\kappa(R) = \kappa^{1/2}(A^TBA)$ and $\kappa(Q) \leq \kappa^{1/2}(B)$. For indefinite B from $A^TBA = R^T\Omega R$ it follows only that $\|A^TBA\| \leq \|R\|^2$ and $\|(A^TBA)^{-1}\| \leq \|R^{-1}\|^2$. Thus we have just a lower bound for the condition number of R as $\kappa^{1/2}(A^TBA) \leq \kappa(R)$. It was shown in [33] that the condition numbers of R and Q can be bounded also from above in terms of the conditioning of A^TBA and in terms of only those its principal submatrices $(A^TBA)_j$ where there is a change of the sign in the factor Ω. The following upper bounds hold for the norm of R^{-1}, for the norm of R, and for the condition number $\kappa(R)$,

$$\|R^{-1}\| \leq \left(\|(A^TBA)^{-1}\| + 2 \sum_{j;\ \omega_{j+1} \neq \omega_j} \|(A^TBA)_j^{-1}\| \right)^{1/2}, \tag{15}$$

$$\|R\| \leq \|A^TBA\|\|R^{-1}\|, \tag{16}$$

$$\kappa(R) \leq \|A^TBA\| \left(\|(A^TBA)^{-1}\| + 2 \sum_{j;\ \omega_{j+1} \neq \omega_j} \|(A^TBA)_j^{-1}\| \right), \tag{17}$$

respectively. The norm and the condition number of the factor Q can be then bounded from above as $\|Q\| \leq \|A\|\|R^{-1}\|$ and $\kappa(Q) \leq \kappa(A)\kappa(R)$. The particular case of a saddle-point matrix is treated in Chapter 5 of [32].

Several algorithms for computing such factors Q and R exist, see e.g. Section 3 in [33]. The corresponding classical Gram-Schmidt algorithm, modified Gram-Schmidt algorithm, and classical Gram-Schmidt algorithm with reorthogonalization are also given in Table 3.

4.1 Cholesky-like factorization of symmetric indefinite matrices and Gram-Schmidt orthogonalization

A significant part of the paper [33] is devoted to the rounding error of four important schemes used for orthogonalization of vectors with respect to the bilinear form induced by a symmetric indefinite but nonsingular matrix B. Two of them use the Gram-Schmidt algorithm. The worst-case bounds on the factorization error and on the loss of B-orthogonality for quantities computed in finite precision arithmetic are given in terms of the factors proportional to the roundoff unit u, in terms of the norms of A, B and A^TBA, and in terms of the extremal singular values of computed factors $\bar{Q}$ and $\bar{R}$.

The loss of B-orthogonality of vectors computed by the Gram-Schmidt algorithm with respect to the bilinear form induced by B is measured by the quantity $\|\bar{Q}^TB\bar{Q} - \bar{\Omega}\|$, where $\bar{\Omega}$ denotes the computed signature matrix.

classical Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$

for $i = 1, \ldots, j-1$
$r_{i,j} = \omega_i \langle Ba_j, q_i \rangle$
$u_j = u_j - r_{i,j} q_i$

$\omega_j = \text{sign}\left[\langle Ba_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2 \right]$

$r_{j,j} = \sqrt{\left| \langle Ba_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2 \right|}$

$q_j = u_j / r_{j,j}$

modified Gram-Schmidt:
for $j = 1, \ldots, n$
$u_j = a_j$

for $i = 1, \ldots, j-1$
$r_{i,j} = \omega_i \langle Bu_j, q_i \rangle$
$u_j = u_j - r_{i,j} q_i$

$\omega_j = \text{sign}\left[\langle Ba_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2 \right]$

$r_{j,j} = \sqrt{\left| \langle Ba_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2 \right|}$

$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:
for $j = 1, \ldots, n$
$u_j = a_j$

for $k = 1, 2$
$a_j^{(k)} = u_j$

for $i = 1, \ldots, j-1$
$r_{i,j}^{(k)} = \omega_i \langle Ba_j^{(k)}, q_i \rangle$
$u_j = u_j - r_{i,j}^{(k)} q_i$

$\omega_j = \text{sign}[\langle Bu_j, u_j \rangle]$

$r_{j,j} = \sqrt{|\langle Bu_j, u_j \rangle|}$

$q_j = u_j / r_{j,j}$

Table 3: Gram-Schmidt orthogonalization with respect to symmetric indefinite bilinear form: classical algorithm, modified algorithm, and classical algorithm with reorthogonalization.

Assuming the numerical nonsingularity of the Gram matrix $A^T BA$ and the numerical nonsingularity of its principal submatrices, where there is a change of the sign in the signature matrix in the form

$$\mathcal{O}(u)\|B\|\|A\|^2 \kappa(A^T BA) \max_{\substack{j=1,\ldots,n-1 \\ \bar{\omega}_{j+1} \neq \bar{\omega}_j}} \|(A^T BA)_j^{-1}\| < 1,$$

the loss of B-orthogonality between the column vectors in the factor $\bar{Q}$ computed by the classical Gram-Schmidt algorithm satisfies the bound

$$\|\bar{Q}^T B\bar{Q}-\bar{\Omega}\| \leq \mathcal{O}(u)\left[\kappa^2(\bar{R}) + \|\bar{R}^{-1}\|^2\|B\|\|A\|^2 + 3\|B\|\|A\|\|\bar{R}^{-1}\|\|\bar{Q}\|\kappa(\bar{R})\right]. \tag{18}$$

It was also shown in [33] that the accuracy of computed factors in the classical Gram-Schmidt algorithm can be improved by reorthogonalization. Assuming the somewhat stronger assumption

$$\mathcal{O}(u)\|B\|\|A\|^2\|A^TBA\|\left[\|(A^TBA)^{-1}\| + \max_{\substack{j=1,\ldots,n-1\\ \bar{\omega}_{j+1}\neq\bar{\omega}_j}}\|(A^TBA)_j^{-1}\|\right]^2 < 1,$$

the loss of orthogonality of computed vectors $\bar{Q}$ in the classical Gram-Schmidt algorithm with reorthogonalization is bounded by

$$\|\bar{Q}^T B\bar{Q} - \bar{\Omega}\| \leq \mathcal{O}(u)\|B\|\|\bar{Q}\|^2. \tag{19}$$

Note that due to (16)-(17) and since $\|\bar{Q}\| \lessapprox \|A\|\bar{R}^{-1}\|$, the bound (19) is significantly better than the bound (18). The improvement is essentially by the factor proportional to the condition number $\kappa(\bar{R})$. However, even in this case the loss of B-orthogonality depends on the condition number of the Gram matrix A^TBA and the condition number of all its principal submatrices where there is a change of the sign in the signature matrix. For details we refer to [33].

4.2 Sign patterns of J-orthogonal matrices

As we have noted, the column vectors in the factor $Q \in \mathcal{R}^{m,n}$ are B-orthogonal satisfying the identity $Q^TBQ = \Omega \in \mathcal{R}^{n,n}$. Considering the square case $n = m$ and taking the eigen-decomposition $B = U^T\Lambda U = (|\Lambda|^{1/2}U)^T J(|\Lambda|^{1/2}U)$, there exists a permutation matrix $P \in \mathcal{R}^{m,m}$ so that $PJP^T = \Omega$, where $J \in \mathrm{diag}(\pm 1) \in \mathcal{R}^{m,m}$ is a signature matrix of the diagonal matrix $\Lambda = |\Lambda|^{1/2}J|\Lambda|^{1/2}$. Then the square matrix $\tilde{Q} \in \mathcal{R}^{m,m}$ defined as $\tilde{Q} = |\Lambda|^{1/2}UQP$ is J-orthogonal and satisfies $\tilde{Q}^TJ\tilde{Q} = J$. It is quite clear that, since B is nonsingular, the matrices J and $\tilde{Q}$ are also nonsingular. The matrix $\tilde{Q} \in \mathcal{R}^{m,m}$ is called a G-matrix if it is nonsingular and there exist nonsingular diagonal matrices D_1 and D_2 such that $\tilde{Q}^{-T} = D_1\tilde{Q}D_2$, where $\tilde{Q}^{-T}$ denotes the transpose of the inverse of $\tilde{Q}$. In the paper [16] several connections are established between these two classes of matrices. Based on the observation that a matrix is a G-matrix if and only if it is diagonally (with positive diagonals) equivalent to a column permutation of a J-orthogonal

matrix, several results on the sign patterns of J-orthogonal matrices were developed in [16]. The key ingredient of this work consists in the identification of sign patterns of $m \times m$ matrices that allow a J-orthogonal matrix for some fixed matrix J or for an arbitrary matrix J. This analysis attempts to extend a research of many authors performed for the particular case of orthogonal matrices, where $J = I$ and where the column vectors are orthogonal with respect to the standard inner product. The class of $m \times m$ matrices that allow an orthogonal matrices was partially characterized for small dimensions m using the concept of the so-called sign-potentially orthogonal conditions that are necessary conditions for sign pattern to belong to this class. For a comprehensive description of results on sign patterns we refer to [14].

5 Orthogonalization with respect to a skew-symmetric bilinear form

It is well-known that if the matrix B induces a skew-symmetric bilinear form, then every vector is isotropic. Since the eigenvalues of a real skew-symmetric matrix are purely imaginary, it is not possible to diagonalize it using a real diagonal basis. However, it is possible to bring every skew-symmetric and nonsingular matrix of even dimension $B \in \mathcal{R}^{2m,2m}$ to a block diagonal form using the Schur-like factorization

$$B = V^T \begin{pmatrix} 0 & \Sigma^2 \\ -\Sigma^2 & 0 \end{pmatrix} V = V^T \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} V, \quad (20)$$

where $V \in \mathcal{R}^{2m,2m}$ is an orthogonal matrix satisfying $V^T V = V V^T = I$, and where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m) \in \mathcal{R}^{m,m}$ is diagonal and nonsingular with positive entries. For a tall full column rank matrix $\tilde{A} \in \mathcal{R}^{2m,2n}$, where $m \geq n = \mathrm{rank}(\tilde{A})/2$ one can look for a decomposition in the form $\tilde{A} = \tilde{Q}R$, where $R \in \mathcal{R}^{2n,2n}$ is upper triangular with positive entries on its diagonal, and where $\tilde{Q} \in \mathcal{R}^{2m,2n}$ satisfies the block B-orthogonality relation

$$\tilde{Q}^T B \tilde{Q} = \tilde{J} \equiv \mathrm{diag}\Big(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \Big) \in \mathcal{R}^{2n,2n}. \quad (21)$$

It is clear from (20) that if we introduce a full-column rank matrix $A = \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} V \tilde{A} = (a_1, \ldots, a_{2n}) \in \mathcal{R}^{2m,2n}$, then we can look for the triangular factorization $A = QR$ with $Q = (q_1, \ldots, q_{2n}) \in \mathcal{R}^{2m,2n}$ satisfying $Q^T J Q = \tilde{J}$, whereas the orthogonal matrix $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathcal{R}^{2m,2m}$ induces a nondegenrate skew-symmetric bilinear form that is often called symplectic bilinear

form. Similarly, the factorization $A = QR$ is frequently referred as the SR factorization of A, and the factor Q is called semi-symplectic factor. It was shown by several authors in various contexts (see e.g. the references in [9]) that such factorization exists if all principal submatrices of the Gram matrix A^TJA with even dimension are nonsingular. Then the upper triangular factor R can be computed from the Cholesky-like factorization

$$A^TJA = R^T\tilde{J}R \tag{22}$$

and the semi-symplectic factor can be recovered as $Q = AR^{-1}$. Several algorithms for computing such factors Q and R are used in practical computations, see e.g. the introductory discussion in [9]. The corresponding classical Gram-Schmidt and modified Gram-Schmidt algorithms are summarized in Table 4.

5.1 Conditioning of factors in the SR factorization

As we have noted above, under certain assumptions on the Gram matrix A^TJA, a full-column rank matrix A can be factorized into the product of semi-symplectic matrix Q and an upper triangular matrix R. As it is also seen in Table 4 this factorization is not unique. Since the SR factorization can be seen as an orthogonalization process with respect to a bilinear form induced by the skew-symmetric matrix J, this freedom can be interpreted as a freedom in the normalization step, where we look for the 2×2 upper triangular matrix $\begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}$ that satisfies

$$\begin{aligned} \begin{pmatrix} r_{11} & 0 \\ r_{12} & r_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} &= (A^TJA)_{2j} \backslash (A^TJA)_{2(j-1)} \\ &\equiv \alpha_j \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \end{aligned}$$

where $(A^TJA)_{2j} \backslash (A^TJA)_{2(j-1)}$ denotes the Schur complement of the principal submatrix $(A^TJA)_{2(j-1)}$ of order $2(j-1)$ with respect to the principal submatrix $(A^TJA)_{2j}$ of order $2j$ for $j = 2, \ldots, n$. For $j = 1$ we set $(A^TJA)_{2j} \backslash (A^TJA)_{2(j-1)} = (A^TJA)_{2j}$. The main goal of the paper [9] is to analyze the freedom of choice in the semi-symplectic and the upper triangular factors in the SR decomposition in order to develop numerically stable algorithms. In particular, several widely used suggestions on how to choose the free parameters are interpreted in terms of the conditioning of certain blocks of the semi-symplectic factor Q or the triangular factor R. As a result, two important choices with local optimality properties are proposed.

classical Gram-Schmidt:

for $j = 1, \ldots, n$

$[u_{2j-1}, u_{2j}] = [a_{2j-1}, a_{2j}]$

for $i = 1, \ldots, j-1$

$$[u_{2j-1}, u_{2j}] = [u_{2j-1}, u_{2j}] - [q_{2i-1}, q_{2i}] \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} [q_{2i-1}, q_{2i}]^T J [a_{2j-1}, a_{2j}]$$

$$\begin{pmatrix} r_{11} & 0 \\ r_{12} & r_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} = [u_{2j-1}, u_{2j}]^T J [u_{2j-1}, u_{2j}]$$

$$[q_{2j-1}, q_{2j}] = [u_{2j-1}, u_{2j}] \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}^{-1}$$

modified Gram-Schmidt:

for $j = 1, \ldots, n$

$[u_{2j-1}, u_{2j}] = [a_{2j-1}, a_{2j}]$

for $i = 1, \ldots, j-1$

$$[u_{2j-1}, u_{2j}] = [u_{2j-1}, u_{2j}] - [q_{2i-1}, q_{2i}] \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} [q_{2i-1}, q_{2i}]^T J [u_{2j-1}, u_{2j}]$$

$$\begin{pmatrix} r_{11} & 0 \\ r_{12} & r_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} = [u_{2j-1}, u_{2j}]^T J [u_{2j-1}, u_{2j}]$$

$$[q_{2j-1}, q_{2j}] = [u_{2j-1}, u_{2j}] \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}^{-1}$$

Table 4: Gram-Schmidt orthogonalization with respect to skew-symmetric bilinear form: classical algorithm and modified algorithm.

The first choice leads to the minimization of the condition number of the 2×2 diagonal blocks in the upper triangular factor. It turns out that it is equivalent to the choice of Mehrmann who suggested to restrict the diagonal block of R only to diagonal matrix setting its entries equal in absolute value. The second choice leads to the minimization of the condition number of local blocks in Q with columns vectors that are orthogonal and equilibriated but in general not normalized due to restriction given by symplectic bilinear form. The worst-case bounds for the extremal values of the whole semi-symplectic or upper triangular factor in terms of the spectral properties of even-dimensioned principal submatrices of the Gram matrix $A^T J A$ for the

first choice are also developed in [9]. However, the strategy that would lead to global minimization of the condition number of either semi-symplectic or triangular factor is not known yet and it is a subject of current research.

6 References

[1] N. Abdelmalek, Round off error analysis for Gram-Schmidt method and solution of linear least squares problems, BIT 11, 1971, 345–368.

[2] J. Barlow, J. Langou, A. Smoktunowicz, A note on the error analysis of classical Gram-Schmidt, Numer. Math. 105(2), 2006, 299–313.

[3] Å. Björck, Solving linear least squares problems by Gram-Schmidt orthogonalization, BIT 7, 1967, 1–21.

[4] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, PA, 1996.

[5] Å. Björck and C. Paige, Loss and recapture of orthogonality in the Modified Gram-Schmidt algorithm, SIAM J. Matrix Anal. Appl. 13(1), 1992, 176–190.

[6] J.W. Daniel, W.B. Gragg, L. Kaufman, G.W. Stewart, Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR Factorization, Math. Comp. 30, 1976, 772–795.

[7] A. Dax, A modified Gram-Schmidt algorithm with iterative orthogonalization and pivoting, Linear Algebra and its Appl. 310 (2000), 25–42.

[8] J. Drkošová, A. Greenbaum, M. Rozložník, Z. Strakoš, Numerical Stability of GMRES, BIT 35 (3), 1995, 309–330.

[9] H. Faßbender, M. Rozložník: On the conditioning of factors in the SR decomposition, Linear Algebra Appl. 505, 2016, 224–244.

[10] L. Giraud, J. Langou, M. Rozložník, J. van den Eshof, Rounding error analysis of the classical Gram-Schmidt orthogonalization process, Numer. Math. 101(1), 2005, 87–100.

[11] L. Giraud, J. Langou, M. Rozložník, The loss of orthogonality in the Gram-Schmidt orthogonalization process, Comput. Math. Appl. 50 (7), 2005, 1069–1075.

[12] J.P. Gram, Über die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate, Journal fr die Reine und Angewandte Mathematik 94, 1883, 41–73.

[13] A. Greenbaum, M. Rozložník, Z. Strakoš, Numerical behaviour of the Modified Gram-Schmidt GMRES implementation, BIT 37 (3), 1997, 709–719.

[14] F.J. Hall, Z. Li, Sign patterns matrices. Handbook of Linear Algebra. Chapman and Hall/CRC Press, Boca Raton, 2013.

[15] F.J. Hall, Z. Li, C. Parnass, M. Rozložník, Sign patterns of J-orthogonal matrices, Special Matrices 5 (1), 2017, 225–241.

[16] F.J. Hall, M. Rozložník, G-matrices, J-orthogonal matrices, and their sign patterns Czechoslovak Math. J. 66 (3), 2016, 653–670.

[17] W. Hoffmann. Iterative algorithms for Gram-Schmidt orthogonalization. Computing 41 (1989), 335–348.

[18] P. Jiránek, M. Rozložník, Adaptive version of Simpler GMRES, Numer. Algorithms 53 (1), 2010, 93–112.

[19] P. Jiránek, M. Rozložník, M.H. Gutknecht, How to make Simpler GMRES and GCR more stable, SIAM J. Matrix Anal. Appl. 30 (4), 2008, 1483–1499.

[20] A. Kiełbasiński, Numerical analysis of the Gram-Schmidt orthogonalization algorithm (Analiza numeryczna algorytmu ortogonalizacji Grama-Schmidta) (in Polish). Roczniki Polskiego Towarzystwa Matematycznego, Seria III: Matematyka Stosowana II, 1974, 15–35.

[21] A. Kiełbasiński, H. Schwetlick, Numeryczna Algebra Liniowa (in Polish). Wydawnictwo Naukowo-Techniczne, Warszawa (1994), Second edition.

[22] J. Kopal, M. Rozložník, M. Tůma, Factorized approximate inverses with adaptive dropping, SIAM J. Sci. Comput. 38 (3), 2016, A1807–A1820.

[23] J. Kopal, M. Rozložník, M. Tůma, An adaptive multilevel factorized sparse approximate inverse preconditioning, Advances in Engineering Software 113, 2017, 19–24.

[24] J. Kopal, M. Rozložník, M. Tůma, Approximate Inverse Preconditioners with Adaptive Dropping Advances in Engineering Software 84, 2015, 13–20.

[25] P.S. Laplace, Thorie Analytique des Probabilits, Troisime dition, Premier Supplment, Sur l'Application du Calcul des Probabilits a la Philosophie Naturelle, Courcier: Paris, 1820.

[26] S.J. Leon, A. Björck, W. Gander, Gram-Schmidt orthogonalization: 100 years and more, Numer. Linear Algebra Appl. 20, 2013, 492–532.

[27] J. Liesen, M. Rozložník, Z. Strakoš, Least squares residuals and minimal residual methods, SIAM J. Sci. Comput. 23 (5), 2002, 1503–1525.

[28] B.R. Lowery and J. Langou, Stability analysis of QR factorization in an oblique inner product, 2014, arXiv:1401.5171.

[29] C.C. Paige, M. Rozložník, Z. Strakoš, Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES, SIAM J. Matrix Anal. Appl. 28 (1), 2006, 264–284.

[30] B.N. Parlett, The Symmetric Eigenvalue Problem, Prentice Hall: Englewood Cliffs, N. J., 1980.

[31] J. R. Rice. Experiments on Gram-Schmidt orthogonalization. Math. Comp. 20, 1966, 325–328.

[32] M. Rozložník, Saddle-Point Problems and Their Iterative Solution Birkhäuser, Cham, 2018.

[33] M. Rozložník, F. Okulicka-Dluzewska, A. Smoktunowicz: Cholesky-like factorization of symmetric indefinite matrices and orthogonalization with respect to bilinear forms, SIAM J. Matrix Anal. Appl. 36(2), 2015, 727-751.

[34] M. Rozložník, Z. Strakoš, Variants of the Residual Minimizing Krylov Space Methods, In: Proceedings of the XI.Summer School Software and Algorithms of Numerical Mathematics, I.Marek et al.(eds), Plzen, U. of West Bohemia, 1996, 208–225.

[35] M. Rozložník, M. Tůma, A. Smoktunowicz, J. Kopal, Numerical stability of orthogonalization methods with a non-standard inner product, BIT 52, 2012, 1035–1058.

[36] A. Ruhe, Numerical Aspects of Gram-Schmidt Orthogonalization of Vectors. Linear Algebra Appl. 52/53, 1983, 591–601.

[37] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen I. Teil: Entwicklung willkrlicher Funktionen nach Systemen vorgeschriebener, Mathematische Annalen 63, 1907, 433–476.

[38] H. Walker, L. Zhou, A Simpler GMRES, Numer. Linear Algebra with Appl. 1(6), 1994, 571–581.