NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2007; **14**:1–59
Published online 9 October 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.499

Review Article

Recent computational developments in Krylov subspace methods for linear systems

Valeria Simoncini[1,2,*,†] and Daniel B. Szyld[3,‡]

[1]*Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy*
[2]*IMATI-CNR, Pavia, and CIRSA, Ravenna*
[3]*Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, Pennsylvania 19122-6094, U.S.A.*

SUMMARY

Many advances in the development of Krylov subspace methods for the iterative solution of linear systems during the last decade and a half are reviewed. These new developments include different versions of restarted, augmented, deflated, flexible, nested, and inexact methods. Also reviewed are methods specifically tailored to systems with special properties such as special forms of symmetry and those depending on one or more parameters. Copyright © 2006 John Wiley & Sons, Ltd.

Received 25 September 2005; Revised 16 May 2006; Accepted 5 June 2006

KEY WORDS: Krylov subspaces; iterative methods; linear systems

Contents

*Correspondence to: Valeria Simoncini, Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy.
†E-mail: valeria@dm.unibo.it
‡E-mail: szyld@temple.edu

Contract/grant sponsor: U.S. National Science Foundation; contract/grant number: DMS-0207525
Contract/grant sponsor: U.S. Department of Energy; contract/grant number: DE-FG02-05ER25672
Contract/grant sponsor: Italian MIUR, PRIN 2004

1. INTRODUCTION

Krylov subspace methods are extensively used for the iterative solution of $n \times n$ linear systems of equations of the form

$$Ax = b, \tag{1}$$

and in particular those arising from discretizations of differential equations. This can be appreciated by a cursory look at any journal in science or engineering; see also the very recent book [100]. The availability of more powerful computers and of better methods has made it possible to solve larger and more complex problems, in application fields ranging from quantum chromodynamics [230] to air traffic control problems [3]. This creates the possibility of more detailed scientific models, which in turn serves as an inspiration for the study of even more effective approaches to their solution.

In some cases, direct methods based on sparse matrix factorizations can be used for fairly large problems [87–89], e.g. in the numerical treatment of some two-dimensional differential equations, or in problems coming from electrical power networks, where the fill-in is somewhat limited; see, e.g. [1]. Nonetheless, iterative methods are mandatory in the numerical solution of large three-dimensional differential equations, as well as in the solution of a large variety of application problems and in numerical discretization methods where the matrix itself is not explicitly available. For a recent paper on the use of direct methods in combination with iterative methods, see [86].

In this paper, we review many of the advances that have taken place within the Krylov subspace framework during the last 10–15 years. We collect many results not usually available in widely read books such as [13, 42, 53, 158, 228, 273, 333], surveys such as [92, 131, 171, 276], or the Templates on iterative methods [26]. We emphasize the fact that we mostly discuss recent computational developments. Nevertheless, in order to make this survey self-contained and also serve as an introduction to the subject, we include some material which is well known. We refer to the books and surveys just cited for the description of the historical evolution of these iterative methods before 1990, and for details on the basic methods beyond our brief presentation in Section 2. In a few instances, though, we dwell on some earlier results which we feel are not well known. Readers familiar with the basic methods may want to skip the first part of the paper.

Perhaps the two most widely used Krylov subspace methods for the solution of non-symmetric systems are GMRES (Generalized Minimal RESidual) and Bi-CGStab, described in Sections 2

DOI: 10.1002/nla

and 3, respectively. The advantage of the latter is that of limited storage needs, but there are many problems for which this method does not work well. For these problems, GMRES has become the method of choice, and this has led both to its ample study and to many extensions and variants. We discuss these developments in Sections 6–11, which in most cases apply to many methods.

In general, some methods described in this paper work best for some problems, and not for others. There is no one method which is recommended for all problems [238]. In the description of the methods, we often point out for which problems the method may work well. New ideas are commonly specific to certain situations, but not for all, and we also discuss this when appropriate. In Sections 12–14 we review many of the methods designed for specific classes of problems. Throughout the paper, we make references to the suitable literature, but we do not claim that our bibliography is complete. We tried to assemble many relevant references, so that the reader can go deeper into particular topics if he or she so desires.

While we report on much of the recent progress in the theory and practice of Krylov subspace methods, this survey cannot be exhaustive. Among the many topics not covered are the following (with some very limited pointers to the literature): the relation of Krylov subspace methods with orthogonal polynomials and Gaussian quadrature [115, 229], and with potential theory [85, 207]; special methods for model reduction [4, 21, 31, 128], although we briefly touch upon this in Section 14.2; for KKT systems or saddle point problems [33, 100]; or for integral operators [203]; considerations for parallel computers [82, 104, 178]; methods for singular matrices [134, 198, 262]; or for regularization of ill-posed problems [58, 179]; multiple right-hand side and block methods [172]; non-linear equations and optimization [35, 56, 205, 260]; or for matrix equations [93, 181, 183, 219], including the Sylvester equation [78, 185, 264, 286]. We also do not talk here about Krylov-based methods for eigenvalue calculations, such as Lanczos or Rational Krylov methods [20].

As is well known, an important ingredient that makes Krylov subspace methods work is the use of preconditioners, i.e. of a matrix or operator M used to convert problem (1) into another equivalent problem, e.g. into $M^{-1}Ax = M^{-1}b$ (left preconditioning) or

$$AM^{-1}y = b \quad \text{with } Mx = y \tag{2}$$

(right preconditioning). We refer the reader, e.g. to [26, 54, 68, 273], or the excellent survey [32] for more details on preconditioning. For completeness, we comment very briefly on some of the most simple preconditioning techniques in Section 7. In general, we assume that the matrix A is already preconditioned, except when we talk about variable or inexact preconditioning in Sections 10 and 11, and also in some cases in Sections 9 and 14. We do mention that the development of effective preconditioners is an active area of research, especially when these preconditioners are being designed for specific applications.

For simplicity of the exposition, and to concentrate on the computational developments, throughout most of the paper we assume exact arithmetic. In a few instances, though, we do refer to the influence of floating point arithmetic in the behaviour of the methods. We refer, e.g. to [90], [158, Chapter 4], [163, 215, 229, 266, 309, 320], for analysis of floating point arithmetic in the context of Krylov subspace methods, and for the important understanding that this analysis has provided. We remark that the use of floating point arithmetic may significantly deteriorate the performance of Krylov subspace methods. In particular, it is well known that (even simple) mathematically equivalent algorithms may have very different convergence behaviour when implemented in practice; see, e.g. [177, 211, 311], for recent analyses. We mention, in particular, that only very recently the modified Gram–Schmidt implementation of GMRES was shown to be backward stable, in the

DOI: 10.1002/nla

sense that the backward error is proportional to machine precision [249]. This was a longstanding problem since practitioners have used this implementation for a long time, although only the Householder implementation of GMRES was known to be robust [90].

We end this section by describing some of the notation we use. By x_* we denote the solution of (1). By $\langle x, y\rangle$ we denote an inner product between the vectors $x, y \in \mathbb{R}^n$. The transpose of a matrix A is denoted A^{T} and it depends on the underlying inner product, i.e. $\langle Ax, y\rangle = \langle x, A^{\mathrm{T}}y\rangle$. We use different inner products throughout the paper and in each case the induced vector norm is $\|x\| = \langle x, x\rangle^{1/2}$. In most cases though, the Euclidean inner product is used, i.e. $\langle x, y\rangle = x^{\mathrm{T}}y$, and the induced norm is the 2-norm. For any positive definite matrix M, the M-inner product is defined as $\langle x, y\rangle = \langle x, y\rangle_M = x^{\mathrm{T}}My$. For complex vectors, the conjugate transpose is denoted by x^*. The matrix or operator norm is induced from the just-defined vector norm in the usual manner $\|A\| = \max_{\{\|x\|=1\}} \|Ax\|$. The exception is the Frobenius norm $\|A\|_{\mathrm{Fr}} = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}$. The matrix I_m is the $m \times m$ identity matrix. When the dimension is clear from the context, we simply denote it by I. By $\sigma_{\max}$ and $\sigma_{\min}$ we denote the largest and smallest (nonzero) singular values of A. The condition number of A is $\kappa(A) = \sigma_{\max}/\sigma_{\min}$. The Euclidean vector e_j is the jth column of the identity of appropriate order. The range (or column space) of a matrix M is denoted by $\mathscr{R}(M)$.

A square matrix A is said to be normal if $A^{\mathrm{T}}A = AA^{\mathrm{T}}$. The algorithmic properties of many Krylov subspace methods and also certain bounds on their convergence vary depending on whether the matrix is normal, and we comment on this at various points in the paper. Clearly, symmetric matrices are normal. Other less trivial examples of normal matrices are skew-symmetric matrices (i.e. those satisfying $A = -A^{\mathrm{T}}$), orthogonal matrices, multiples of orthogonal matrices plus a complex shift, and also matrices of the form $A = M + \sigma I$, with M real symmetric and σ complex. This last case is of particular interest, as it arises in several applications; it is discussed in detail in Section 14.1.

Throughout the paper we assume real data, although most methods can be implemented in the complex case, and thus methods specifically designed for symmetric matrices apply to Hermitian ones as well, and we do not repeat the relevant issues. Specific comments for the general complex case are given in Section 12, and for more special cases in Section 13.3.

2. DESCRIPTION OF THE BASIC METHODS

Let x_0 be an initial approximation to the solution of (1), $r_0 = b - Ax_0$ be the initial residual and let

$$\mathscr{K}_m(A, r_0) = \mathrm{span}\{r_0, Ar_0, A^2r_0, \ldots, A^{m-1}r_0\} \tag{3}$$

be the Krylov subspace of dimension m defined by A and r_0. The short-hand notation $\mathscr{K}_m$ is used when the dependence on A and on the specific vector r_0 is clear from the context. Note that these subspaces are nested, i.e. $\mathscr{K}_m \subseteq \mathscr{K}_{m+1}$.

Krylov subspace methods are iterative methods in which at the mth step an approximation to the solution of (1), x_m, is found in $x_0 + \mathscr{K}_m$, i.e. this approximation is of the form $x_m = x_0 + q_{m-1}(A)r_0$, where q_{m-1} is a polynomial of degree at most $m-1$. If the system is real, then q_{m-1} can be chosen to have real coefficients. This natural expression implies that the residual $r_m = b - Ax_m$ is

DOI: 10.1002/nla

associated with the so-called *residual polynomial* p_m of degree at most m with $p_m(0) = 1$, since

$$r_m = b - Ax_m = r_0 - Aq_{m-1}(A)r_0 = p_m(A)r_0. \tag{4}$$

Analogously, the error satisfies $x_m - x_* = p_m(A)(x_0 - x_*)$, where x_* is the solution of (1). Let us denote by $\mathscr{P}_m$ the set of all polynomials p of degree at most m such that $p(0) = 1$. The approximation $x_m \in x_0 + \mathscr{K}_m$ (or equivalently, the corresponding polynomial) is often found by requiring x_m to be the minimizer of some functional. Different methods depend on the choice of this functional, on the characteristics of the matrix, and on some implementation details, and thus, each method defines implicitly a different polynomial $p_m \in \mathscr{P}_m$ (or q_{m-1}). For example, in the popular GMRES by Saad and Schultz [275], the approximation x_m is the one minimizing the 2-norm of the residual; see Section 2.2.

In the process of iteratively constructing a basis of $\mathscr{K}_m$, each method can be implemented so that at each iteration only one or two matrix–vector multiplications with A are required, in the form $z = Av$ (in some methods an additional operation of the form $y = A^{\mathrm{T}}w$ is needed). This fact is what makes these methods of practical application. In fact, the matrix itself is not needed, only its action as an operator on a vector is used, usually as a call to a subroutine.

In all cases treated here, the methods start with an initial vector x_0, with initial residual $r_0 = b - Ax_0$, and at the mth step obtain an element x_m of $x_0 + \mathscr{K}_m(A, r_0)$ satisfying a projection or minimizing condition of some kind. Let $r_m = b - Ax_m$ be the residual at the mth step. A general condition is the

- *Petrov–Galerkin condition*

$$r_m \perp \mathscr{L}_m, \tag{5}$$

where $\mathscr{L}_m$ is some m-dimensional subspace.

When $\mathscr{L}_m = \mathscr{K}_m$, (5) is called a

- *Galerkin condition*, i.e. we have

$$r_m \perp \mathscr{K}_m. \tag{6}$$

We also discuss the

- *Minimum residual condition*

$$\|r_m\| = \min_{x \in x_0 + \mathscr{K}_m} \|b - Ax\|. \tag{7}$$

It can be shown that (7) is a Petrov–Galerkin condition by choosing $\mathscr{L}_m = A\mathscr{K}_m$; see, e.g. [273].

We note that the nested property of the Krylov subspaces, imply that any method for which one of the conditions (5)–(7) holds will, in exact arithmetic, terminate in at most n steps. Of course, in practice one wants the methods to produce a good approximation to the solution of (1) in many fewer than n iterations.

In the rest of this section, we present the methods which are well known and commonly used. In several cases we review the development of more recent ideas and point to the relevant literature. We present here a basic description of the methods without giving full implementation details, which can be found, e.g. in the Templates [26]. For a historical perspective on the development of these methods, see the survey [276]. Further details on the basic methods can be found in the books by Greenbaum [158], Saad [273], or van der Vorst [333] already mentioned. More specialized, or more recent methods are presented in Sections 5 and 8–14.

DOI: 10.1002/nla

We assume from now on, without loss of generality, that unless otherwise specified $x_0 = 0$, and thus $r_0 = b$ throughout.

2.1. Arnoldi and Lanczos procedures

Before we describe the Krylov subspace methods, we present the Arnoldi procedure to construct an orthonormal basis of the Krylov subspace [9]. When the matrix is symmetric, this procedure simplifies and is due to Lanczos [208, 209].

Let $\beta = \|r_0\|$, and $v_1 = r_0/\beta$. An orthonormal basis $\{v_1, \ldots, v_m\}$ of $\mathcal{K}_m(A, r_0)$ is obtained one vector at a time by computing Av_k, orthogonalizing this vector with respect to the previous ones $v_1, \ldots, v_k$, and normalizing it. In other words, we have a relation of the form

$$v_{k+1} h_{k+1,k} = Av_k - \sum_{j=1}^{k} v_j h_{jk}, \tag{8}$$

where the coefficients $h_{jk} = \langle v_j, Av_k \rangle$, $j \leqslant k$, are such that orthogonality is achieved, and $h_{k+1,k}$ is positive and such that $\|v_{k+1}\| = 1$. If one collects the orthonormal vectors in the matrix $V_m = [v_1, \ldots, v_m]$, and the coefficients h_{jk} into the $(m+1) \times m$ upper Hessenberg matrix $H_{m+1,m}$, we can write the important Arnoldi relation

$$AV_m = V_{m+1} H_{m+1,m} \tag{9}$$

$$= V_m H_m + h_{m+1,m} v_{m+1} e_m^{\mathrm{T}}, \tag{10}$$

where H_m is the $m \times m$ matrix containing the first m rows of $H_{m+1,m}$, i.e.

$$H_{m+1,m} = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^{\mathrm{T}} \end{bmatrix}. \tag{11}$$

It follows from (9) that the rank of $H_{m+1,m}$ is the same as the rank of AV_m, i.e. $H_{m+1,m}$ has rank m whenever the new vector Av_m is linearly independent of the previous vectors $v_1, \ldots, v_m$. Observe also that if $h_{m+1,m} = 0$, these m vectors form an invariant subspace of A, and the solution of (1) belongs to this subspace. It also follows from (9)–(10) that

$$V_m^{\mathrm{T}} A V_m = H_m. \tag{12}$$

We note that in our description of the Arnoldi procedure, we used the standard Gram–Schmidt orthogonalization method. In actual implementations, one usually uses the modified Gram–Schmidt (MGS) variant which performs the same operations in a different order and it is more stable [36, 151].

The Lanczos procedure for symmetric A is similar, except that orthogonalization with respect to the last two previous vectors suffices. In other words, once Av_k is orthogonal to v_k and v_{k-1}, it is automatically orthogonal to v_j, for $j < k-1$. Thus, the right-hand side of (8) has only three terms, and the matrix (11) is tridiagonal. We denote it by $T_{m+1,m}$ and it has the form

$$T_{m+1,m} = \begin{bmatrix} T_m \\ t_{m+1,m} e_m^{\mathrm{T}} \end{bmatrix}, \tag{13}$$

where T_m is symmetric.

DOI: 10.1002/nla

There is also a (two-sided) Lanczos procedure for non-symmetric matrices by which a non-orthogonal basis $\{w_1, \ldots, w_m\}$ of $\mathscr{K}_m(A, r_0)$ is built [209]. Consider a (left) Krylov subspace defined by the transpose of A, and some auxiliary vector $\widehat{r}_0$ such that $\langle r_0, \widehat{r}_0\rangle \neq 0$, i.e. the subspace $\mathscr{K}_m(A^{\mathrm{T}}, \widehat{r}_0)$ (a common choice is $\widehat{r}_0 = r_0$). Let $\{\widehat{w}_1, \ldots, \widehat{w}_m\}$ be a basis for it. The (two-sided) Lanczos procedure progressively constructs these two bases so that they are bi-orthogonal, i.e. so that $\langle \widehat{w}_i, w_j\rangle = 0$ when $i \neq j$, and $\langle \widehat{w}_j, w_j\rangle \neq 0$. In other words, if $W_m = [w_1, \ldots, w_m]$ and $\widehat{W}_m = [\widehat{w}_1, \ldots, \widehat{w}_m]$, then $\widehat{W}_m^{\mathrm{T}} W_m$ is diagonal. Some freedom is left on how to scale the vectors $w_j, \widehat{w}_j$; it is customary to scale these two vectors so that $\langle \widehat{w}_j, w_j\rangle = 1,\ j = 1, \ldots, m$, i.e. $\widehat{W}_m^{\mathrm{T}} W_m = I$. The procedure just outlined is not always successful, i.e. it may break down as soon as a vector $\widehat{w}_j$ is found that is orthogonal to the corresponding w_j.

We call attention to a distinction between the breakdown just described due to the non-existence of the new basis vectors, sometimes called true breakdown, and the breakdown due to the implementation of the recurrences generating those vectors. The latter is sometimes called ghost breakdown, or pivot breakdown; see, e.g. Brezinski, Redivo Zaglia, and Sadok [48, 49], or Gutknecht [171]. In this last reference, a comparison of all the names of breakdowns is given.

Strategies that try to overcome breakdowns have been proposed in the literature, including the case of *near*-breakdown in which case $\langle \widehat{w}_j, w_j\rangle \approx 0$. A standard procedure, called *look-ahead Lanczos*, consists of relaxing the constraint of the matrix $\widehat{W}_m^{\mathrm{T}} W_m$ being diagonal. Indeed, intuitively, the look-ahead Lanczos *looks* ahead for the next basis vectors that will maintain $\widehat{W}_m^{\mathrm{T}} W_m$ non-singular, therefore only requiring that this be a *block* diagonal matrix, rather than simply diagonal. The look-ahead idea was first suggested by Parlett, Taylor, and Liu [256] for Lanczos breakdowns, only considering 2×2 diagonal blocks, while in Freund, Gutknecht, and Nachtigal [133] an implementation is presented with diagonal blocks of arbitrary size; see also [167, 170, 237], and [46, 47, 49] for further analyses of breakdown. Incurable breakdown occurs if the diagonal block cannot be 'closed'. In this case, the process needs to be restarted with the current approximate solution as initial guess, and a new auxiliary starting vector.

One advantage of the two-sided Lanczos procedure is that, assuming that no breakdown occurs, these bases can be constructed by a three-term recurrence, or two coupled two-term recurrences, and thus, only two to three previous vectors in each sequence need to be stored, and this was already noted by Lanczos [208]. The matrix collecting the orthogonality coefficients is thus tridiagonal of the form (13), as in the symmetric case, but here it is non-symmetric. Therefore, a relation such as (9)–(10) holds, namely,

$$A W_m = W_m T_m + t_{m+1,m} w_{m+1} e_m^{\mathrm{T}}; \tag{14}$$

we refer to the paper by Parlett [254], where context is given for this and other methods to reduce a matrix to tridiagonal form, and to Gutknecht [167, 170], where the relation of the Lanczos procedure to Padé approximations and formal orthogonal polynomials is fully developed. Further comments on breakdown and more references are given in Section 2.5.

We emphasize that the two-sided Lanczos procedure is based on a *short-term recurrence* and this is fundamental to keep storage requirements low. In contrast, the Arnoldi procedure requires that the whole basis be stored since a *full recurrence* is needed. Alternative strategies to remedy this are presented in Section 8.

We point out that on the negative side, in addition to the possible breakdown already mentioned, one needs to have access to both operators A and A^{T}, and A^{T} may be more expensive to apply than A. A typical situation occurs, for instance, when A is given only as an operator subroutine, as is the case in matrix-free Newton–Krylov methods [202].

DOI: 10.1002/nla

2.2. GMRES and GCR

We begin our description of the Krylov subspace methods with GMRES (Generalized Minimal RESidual), first proposed by Saad and Schultz [275]. The projection condition (7) is to minimize the residual over all possible vectors in the Krylov subspace $\mathcal{K}_m(A, r_0)$. That is, one obtains x_m such that

$$\|r_m\| = \|b - Ax_m\| = \min_{x \in \mathcal{K}_m(A,r_0)} \|b - Ax\|. \tag{15}$$

For GMRES, usually the 2-norm is used; see Section 13.1 for a discussion of minimization using other norms. We point out that the solution of the least squares problem (15) is unique as long as A has full rank [36].

The key to GMRES is the implementation of the solution of the least squares problem (15) using an orthonormal basis of the Krylov subspace produced by the Arnoldi procedure. We write the GMRES approximation at the mth step as

$$x_m = V_m y_m \tag{16}$$

for some $y_m \in \mathbb{R}^m$, so that using the Arnoldi relation (9) and the fact that $V_{m+1}e_1 = v_1 = b/\beta$ we have

$$r_m = b - Ax_m = b - AV_m y_m \tag{17}$$

$$= \beta v_1 - V_{m+1}H_{m+1,m}y_m = V_{m+1}(\beta e_1 - H_{m+1,m}y_m). \tag{18}$$

Since V_{m+1} has orthonormal columns, the least squares problem (15) can thus be rewritten as

$$\|r_m\| = \min_{y \in \mathbb{R}^m} \|\beta e_1 - H_{m+1,m}y\|. \tag{19}$$

Another key implementation feature of GMRES is the use of the QR factorization of

$$H_{m+1,m} = Q_{m+1}R_{m+1,m}, \tag{20}$$

where the $(m+1) \times (m+1)$ matrix Q_{m+1} is orthogonal, and

$$R_{m+1,m} = \begin{bmatrix} R_m \\ 0 \end{bmatrix}, \tag{21}$$

where the $m \times m$ matrix R_m is upper triangular. The QR factorization (20) is usually performed with Given rotations so that only two entries per step need to be computed and used to update the upper triangular matrix R_m; for details, see, e.g. [273, 275]. The least squares problem (19) can be replaced by

$$\|r_m\| = \min_{y \in \mathbb{R}^m} \|Q_{m+1}^{\mathrm{T}}\beta e_1 - R_{m+1,m}y\|. \tag{22}$$

Problem (22) has a unique solution when $H_{m+1,m}$ has full rank, i.e. rank m; and in this case, R_m is non-singular. Let

$$Q_{m+1}^{\mathrm{T}}\beta e_1 = \begin{bmatrix} t_m \\ \rho_{m+1} \end{bmatrix}. \tag{23}$$

DOI: 10.1002/nla

Then, the solution of (19) is $y_m = R_m^{-1} t_m$ and we can write

$$x_m = V_m(R_m^{-1} t_m). \tag{24}$$

Furthermore, it follows from (22) that $\|r_m\| = \|Q_{m+1}^{\mathrm{T}} \beta e_1 - R_{m+1,m} y_m\| = |\rho_{m+1}|$, and this is how the residual norm is checked in practical implementations. We note in passing that in finite precision arithmetic, the equality $\|b - Ax_m\| = |\rho_{m+1}|$ may not hold; see Greenbaum [157].

We remark that the sequence of residual norms $\|r_m\|$ generated by GMRES, like for all methods satisfying the minimum residual condition (7) on nested subspaces, is non-increasing. The main disadvantage of GMRES is that as the iterations proceed, i.e. as m grows, the storage requirements grow accordingly. One needs mn storage locations to store the matrix V_m. As we shall see, there are several alternatives to alleviate this, using, e.g. restarted or truncated methods. We describe them in Section 8.

An earlier method which also produces the approximation x_m defined by (15) is GCR by Eisenstat, Elman, and Schultz [96]. The difference with GMRES is in its implementation, since GCR uses a basis of $\mathscr{K}_m$ which is not orthogonal. Instead, the basis used, $\{p_1, p_2, \ldots, p_m\}$, is such that $p_i^{\mathrm{T}} A^{\mathrm{T}} A p_j = 0$, $i \neq j$, and these vectors can be obtained using the Arnoldi process on the subspace $A\mathscr{K}_m$. Since both the vectors p_i and Ap_i need to be stored, GCR is less attractive than GMRES and it is seldom used with the full recurrence. We mention it here since it is the basis for some methods described in Section 10. We also mention briefly two other methods not used nowadays in practical computations: Simpler GMRES by Walker and Zhou [342], which can be useful for certain estimates of convergence bounds [210], and ORTHODIR by Jea and Young [191], where the minimization is performed on $A\mathscr{K}_m$. See, e.g. [11, 158, 333], for descriptions of these and other earlier methods not discussed in this survey.

2.3. CG

The method of Conjugate Gradients (CG) was proposed by Hestenes and Stiefel [180], and it is the method of choice for symmetric positive definite linear systems. In our brief description of CG, we begin by considering the (symmetric) Lanczos procedure, and thus have a basis $\{v_1, \ldots, v_m\}$ of $\mathscr{K}_m$, as described in Section 2.1. If we consider the CG approximation as $x_m = V_m y_m$ for some $y_m \in \mathbb{R}^m$, the Galerkin condition (6) can be written as $0 = V_m^{\mathrm{T}}(b - Ax_m) = V_m^{\mathrm{T}} b - V_m^{\mathrm{T}} A V_m y_m = \beta e_1 - T_m y_m$, where T_m is symmetric tridiagonal, cf. (12). This implies that y_m is the solution of

$$T_m y_m = \beta e_1. \tag{25}$$

Since A is positive definite, so is $T_m = V_m^{\mathrm{T}} A V_m$, and therefore the linear system (25) is always solvable, and the Cholesky factorization always exists, as well as the factorization $T_m = L_m D_m L_m^{\mathrm{T}}$. It turns out that the diagonal matrix D_m and the unit bidiagonal matrix L_m are principal submatrices of the matrices D_{m+1} and L_{m+1} of the next step, respectively. We define the so-called search directions as the columns of $P_m = [p_1, \ldots, p_m]$, where $P_m = V_m L_m^{-\mathrm{T}}$. We then have that $p_1 = v_1$, and

$$p_m = v_m + \lambda_m p_{m-1}, \quad m > 1, \tag{26}$$

for some scalars λ_m; see, e.g. Reference [250]. With these search directions, one can obtain the CG approximation directly from the previous one, i.e. it holds that

$$x_m = x_{m-1} + \alpha_m p_m, \tag{27}$$

DOI: 10.1002/nla

for some scalar α_m, and this is how the CG approximation x_m is updated in actual implementations. In such implementations, one can obtain the Lanczos vectors and the search directions using two coupled two-term recurrences.

In this symmetric positive definite case, the Galerkin condition (6) can be shown to be equivalent to the minimizing condition $\min_{x \in \mathcal{K}_m} \psi(x)$, where $\psi(x) = \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b$, which is also equivalent to minimizing the A-norm of the error, i.e.

$$\min_{x \in \mathcal{K}_m} \|x - x_*\|_A, \tag{28}$$

where the A-norm is induced by the A-inner product $\langle x, y\rangle = x^{\mathrm{T}}Ay$. This norm is often referred to as the energy norm. It turns out that the search direction vectors are conjugate, i.e. they are orthogonal in the underlying A-inner product satisfying $p_i^{\mathrm{T}}Ap_j = 0$, $i \neq j$. The scalar α_m in (27) can be interpreted as the solution of the minimization problem $\min_{\alpha \in \mathbb{R}} \psi(x_{m-1} + \alpha p_m)$, and in fact

$$x_m = x_{m-1} + \alpha_m p_m = \arg\min_{x \in \mathcal{K}_m} \psi(x),$$

i.e. the one-dimensional minimization in the direction p_m is the same as the global minimization on the subspace $\mathcal{K}_m = \mathrm{span}\{p_1, \ldots, p_m\}$. For details, we refer, e.g. to the books by Axelsson [13], Fischer [115], Golub and Van Loan [151], or Luenberger [220], and also to the survey by Golub and O'Leary [148] which in addition includes an extensive bibliography on the subject.

Finally, we mention that CG can be used also for solving $Ax = b$ when A is non-symmetric or rectangular. Indeed, multiplying by A^{T} we obtain the system of normal equations $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$, whose coefficient matrix is symmetric and positive definite, if A has full column-rank. A good implementation of CG for the normal equations is LSQR by Paige and Saunders [251]. While for A symmetric and positive definite, convergence bounds of CG can be derived that only depend on $\kappa(A)$ (cf. Section 6), when using the system of normal equations $\kappa(A^{\mathrm{T}}A) = \kappa(A)^2$ is the key quantity, and much slower convergence may be observed; see Nachtigal, Reddy, and Trefethen [238] for an intriguing experimental comparison of CG on the system of normal equations with other methods designed specifically for non-symmetric systems.

2.4. MINRES and SYMMLQ

Paige and Saunders [250] proposed two methods for symmetric but indefinite linear systems. In the MINRES method, the minimal residual condition (7) is imposed, the Lanczos method is used to generate an orthonormal basis of $\mathcal{K}_m$, and, as outlined in the following, only two basis vectors are needed for the computation of the approximation x_m.

The implementation of the method relies on the QR factorization of the (rectangular) tridiagonal matrix $T_{m+1,m} = Q_{m+1}R_m$. Let $P_m = V_m R_m^{-1} = [p_1, \ldots, p_m]$. The columns of P_m can be computed one at a time, since $P_m R_m = V_m$ translates into a set of three-term recurrences for the columns of P_m owing to the fact that R_m has only three non-zero diagonals. It turns out that the minimum residual approximation (24) can be written as

$$x_m = P_m t_m = x_{m-1} + \tau_m p_m, \tag{29}$$

where $t_m^{\mathrm{T}} = (\tau_1, \ldots, \tau_m)$ i.e. only the last component of t_m changes from the previous step (see [250]); here, p_m is the mth direction vector, cf. (27).

A second approach consists of considering the Galerkin condition (6) as in CG, and the linear system (25). The tridiagonal matrix T_m may now be singular, or nearly singular. On the other hand,

DOI: 10.1002/nla

it is not hard to show that if T_m is singular, T_{m+1} is not, unless $t_{m+1,m} = 0$, in which case $r_m = 0$, cf. (13) and (8). Paige and Saunders [250] proposed to use the LQ factorization of T_m, producing the CG approximation if it exists, and another sequence of approximations. This is the SYMMLQ method. Paige and Saunders were not aware that this second sequence of approximations in fact minimizes the 2-norm of the error over the subspace $\mathscr{K}_m(A, Ar_0)$. Fridman [138] had suggested such an error-minimization algorithm. Fletcher [120] independently rediscovered Fridman's method and showed that SYMMLQ produces the same iterates. Stoer and Freund [319] showed that SYMMLQ can be considered a stable implementation of Fridman's method; see also [318, 323]. Other more recent error minimizing methods are described in Section 5.

2.5. FOM, Lanczos, and BiCG

Using the Arnoldi process for non-symmetric A, if one imposes the Galerkin condition (6), i.e. $V_m^{\mathrm{T}}(b - Ax_m) = 0$, one obtains the full orthogonalization method (FOM) [269]. As in CG, using $x_m = V_m y_m$, (17) and (12), we rewrite this condition as

$$0 = V_m^{\mathrm{T}} b - V_m^{\mathrm{T}} A V_m y_m = \beta e_1 - H_m y_m. \tag{30}$$

Thus, the FOM solution is obtained by solving at each step the $m \times m$ linear system

$$H_m y_m = \beta e_1.$$

A nice feature of this method is that the residual and its norm are easily available. Indeed, using the Arnoldi relation (10) we can write

$$r_m = b - AV_m y_m = \beta v_1 - V_m H_m y_m - h_{m+1,m} v_{m+1} e_m^{\mathrm{T}} y_m, \tag{31}$$

and since $V_m e_1 = v_1$, using (30) we have that

$$r_m = -h_{m+1,m} v_{m+1} e_m^{\mathrm{T}} y_m \tag{32}$$

and thus $\|r_m\| = h_{m+1,m} |e_m^{\mathrm{T}} y_m|$.

If instead of the Arnoldi process, one uses two-sided Lanczos and imposes the Petrov–Galerkin condition (5) with $\mathscr{L}_m = \mathscr{K}_m(A^{\mathrm{T}}, r_0)$, i.e. $\widehat{W}_m^{\mathrm{T}}(b - Ax_m) = 0$, one obtains the Lanczos method [209]. Considering $x_m = W_m y_m$, using (14) and scaling $\widehat{W}_m$ so that $\widehat{W}_m^{\mathrm{T}} W_m = I$, y_m is found by solving (25) with T_m non-symmetric and tridiagonal. The polynomial such that $r_m = p_m(A) r_0$ is then called the Lanczos (residual) polynomial.

Fletcher [120] suggested to use the LU factorization of the non-symmetric tridiagonal matrix T_m. This is called the Bi-Conjugate Gradient method (BiCG). Since BiCG is a different implementation of the Lanczos method, in exact arithmetic, the approximation x_m is the same as that of the Lanczos method, and thus $r_m = p_m(A) r_0$, with p_m the Lanczos polynomial. The problems of the non-symmetric Lanczos method are still present in BiCG, and in addition, the method may break down if the LU factorization without pivoting does not exist, whence this is called a pivot breakdown. QMR methods overcome this difficulty by solving the least squares problem instead of implicitly solving the $m \times m$ linear system with T_m. The Lanczos method (and thus BiCG) reduces to CG when $A = A^{\mathrm{T}}$ and it is positive definite. In this case, the Cholesky factorization of T_m always exists. The breakdown due to the possible singularity or near singularity of the tridiagonal matrix T_m can be fixed by skipping the singular step as proposed by Bank and Chan [24, 25], and goes with the name of 'composite step.' The idea is to use a 2×2 pivot when necessary and obtain an

DOI: 10.1002/nla

LDU factorization of T_m where D has either 1×1 or 2×2 diagonal blocks. We mention that the breakdowns of BiCG due to a Lanczos or a pivot breakdown in the Lanczos process may also be possibly solved by the use of look-ahead, as described in Section 2.1. Modified versions of the Lanczos process that mitigate the breakdown sensitivity have also been proposed by Joubert [195].

Hochbruck and Lubich [182] provide the following bound between the norm of the BiCG residual, which we denote by r_m^{B}, and that of GMRES at the previous step:

$$\|r_m^{\mathrm{B}}\| \leqslant \sqrt{m}\, \|g_m\| \|r_{m-1}^{\mathrm{G}}\|, \tag{33}$$

where g_m is the solution of $T_m^{\mathrm{T}} g_m = t_{m+1,m} e_m$, cf. (14). It follows from examples and a discussion in [182] that $\sqrt{m}\, \|g_m\|$ well represents the ratio of the residuals in a qualitative way.

There are several reasons why Lanczos or BiCG are not much used nowadays. In addition to the possible breakdown of two-sided Lanczos and the need to have access to both operators A and A^{T}, the Lanczos method is not very stable (see, e.g. Gutknecht and Strakoš [177]), and the residual norms may have large oscillations, sometimes referred to as irregular (or erratic) convergence. In fact, unlike the minimum residual methods, in the methods described here, the residual norms are not necessarily non-increasing. This should not be a problem as long as there is a downward trend, but many people prefer to see a smooth decreasing curve, and this fact has led to several suggestions on how to achieve some smoothing. This is described in Section 4. However, it should be kept in mind that smoothing the residual does not improve the numerical properties of the short-term Lanczos recurrence.

3. CGS, BI-CGSTAB, AND OTHER POLYNOMIAL PRODUCT METHODS

Sonneveld [313] developed Conjugate Gradient Squared (CGS) which is based on BiCG, but without the need of the transpose operator. The main idea is to exploit the fact that if p_m is any polynomial then $\langle p_m(A)v,\, p_m(A^{\mathrm{T}})w\rangle = \langle p_m^2(A)v,\, w\rangle$. While not requiring the transpose operator, the oscillatory behaviour of the convergence of BiCG is amplified.

van der Vorst [332] fixed some of the oscillatory behaviour by replacing the polynomial $p_m^2(A)$ with a product of polynomials $q_m(A)\, p_m(A)$, with p_m still being the Lanczos (or BiCG) polynomial, and choosing

$$q_m(\zeta) = (1 - \omega_1\zeta)(1 - \omega_2\zeta) \cdots (1 - \omega_m\zeta) = q_{m-1}(\zeta)(1 - \omega_m\zeta), \tag{34}$$

where the new root $1/\omega_m$ is chosen so that $r_m = q_m(A)\, p_m(A) r_0$ has minimum norm over all possible choices of this root. This method is called Bi-CGStab, and is one of the most popular methods in use today. Bi-CGStab is not very effective when the spectrum has large imaginary components, as is the case, e.g. for matrices stemming from advection dominated PDEs; see Gutknecht [169], and Sleijpen and Fokkema [303]. This problem motivated the introduction of methods such as Bi-CGStab2 and Bi-CGStab(ℓ) discussed next.

The idea of using a product of polynomials gave rise to several methods, appropriately called product methods. The first of such methods was Bi-CGStab2 [169], the odd step is as in (34), while at the next step, the last factor is replaced by a quadratic polynomial; see, also, Cao [61]. Computational experience using Bi-CGStab2 for VLSI design was reported by Pommerell and Fichtner [259]; the method is also successfully used in the numerical solution of

DOI: 10.1002/nla

advection–diffusion problems; see, for instance, the discussion in Elman, Sylvester, and Wathen [100, p. 176].

These ideas were further generalized to have q_m instead be a product of polynomials $\phi(\zeta)$ of degree ℓ such that $\phi(0) = 1$, i.e. $q_m = q_{k\ell} = \phi_k \phi_{k-1} \dots \phi_1$, and ϕ_k is chosen so that it minimizes the norm of $r_m = q_m(A)\, p_m(A) r_0$; see [303, 310] for more details.

Many other choices of the polynomial $q_m(\zeta)$ are possible in principle; see, e.g. Brezinski and Redivo Zaglia [44]; but the challenge is to find some that are better. As discussed later in Section 6, certain Krylov methods appear to have good convergence behaviour when the roots of the corresponding polynomial approximate well the eigenvalues of the matrix A. One of the advantages of these product methods is that, unlike Bi-CGStab, their polynomials can have complex roots, and thus have the possibility of working well with real matrices with complex spectrum. Thus, Zhang [350] chose $q_m(\zeta)$ using a polynomial three-term recurrence, while allowing for complex roots. The minimization step to find the next polynomial is over a two-dimensional search space; see also Röllin and Gutknecht [265] for alternative formulations. Numerical experiments indicate that this approach, called GPBi-CG, can outperform Bi-CGStab and CGS for matrices with complex spectrum [333, Section 9.3.1], [350].

Another product method is CGS2 by Fokkema, Sleijpen, and van der Vorst [121], where one of the polynomials is the Bi-CG polynomial p_m (corresponding to r_0), and the other is a Bi-CG polynomial corresponding to a different initial residual s_0. The two polynomials have roots which are approaching the same eigenvalues of the matrix as the degree increases (though in different ways), but their product produces a method with less erratic behaviour than CGS. The performance of CGS2 (and its convergence curve) is comparable to that of Bi-CGStab in some problems. In some cases, such as the linear systems at each step of a Newton method, CGS2 may perform better [121].

All product methods described in this section can be derived using a common theory of orthogonal polynomials; see Brezinski and Redivo Zaglia [45].

The concept of composite step already mentioned in Section 2.5 was also applied to Bi-CGStab by Chan and Szeto [65], thus eliminating a possible source of breakdown. Look-ahead strategies for product methods are also available in an attempt to avoid the other possible source of breakdown [44, 171, 173].

4. SMOOTHING PROCEDURES: QMR, TFQMR, QMR-CGStab, AND MORE

Freund and Nachtigal [135] proposed a procedure, called QMR (quasi-minimum residual), that replaces the Lanczos solution iterates with a new sequence of approximate solutions, such that the associated residuals satisfy a quasi-optimal minimization condition. In addition to overcoming the possible erratic behaviour of the BiCG residual, the proposed approach also avoids pivot breakdown. More precisely, let W_m contain the (non-orthogonal) basis of $\mathscr{K}_m(A, b)$, so that (14) holds. Then, the residual can be written as

$$r_m = b - A W_m y_m = \beta w_1 - W_{m+1} T_{m+1,m} y_m = W_{m+1}(\beta e_1 - T_{m+1,m} y_m). \tag{35}$$

Instead of minimizing $\|r_m\| = \|W_{m+1}(\beta e_1 - T_{m+1,m} y_m)\|$ as in GMRES, QMR minimizes the norm of the quasi-residual $\|\beta e_1 - T_{m+1,m} y_m\|$, i.e. y_m is found by solving the least squares problem

$$\min_{y \in \mathbb{R}^m} \|\beta e_1 - T_{m+1,m} y\|. \tag{36}$$

DOI: 10.1002/nla

The cost of solving (36) at each iteration is small, since $T_{m+1,m}$ is tridiagonal, though non-symmetric, so that the solution $W_m y_m$ can be easily updated without storing the whole basis matrix W_m. The additional cost, compared to the standard Lanczos method with three-term recurrence is quite irrelevant (a few scalar operations per step). Although the resulting method (Lanczos with the QMR procedure) does provide a smoother convergence history, a more sound numerical implementation is obtained after devising the *coupled* two-term recurrence version of the method, which was experimentally shown to be numerically more reliable in [136]. See also Cullum [72] for more numerical insight into the behaviour of QMR, compared with residual norm minimizing methods. We mention though the following well-known comparison between the norm of the QMR residual, and that of GMRES, denoted r_m^{Q} and r_m^{G}, respectively,

$$\|r_m^{\mathrm{Q}}\| \leqslant \kappa(W_{m+1}) \|r_m^{\mathrm{G}}\|, \tag{37}$$

which was first obtained by Nachtigal [237]. Since $\|r_m^{\mathrm{G}}\| \leqslant \|r_m^{\mathrm{Q}}\|$, if $W_{m+1}^{\mathrm{T}} W_{m+1} = I_{m+1}$ then it must hold $\|r_m^{\mathrm{Q}}\| = \|r_m^{\mathrm{G}}\|$. In general, the bound (37) suggests that the QMR residual norm may be very far from the optimal GMRES norm when the chosen basis is ill conditioned. However, it was recently shown in [300] that this type of bound may considerably underestimate the actual behaviour of quasi-optimal methods, in this case that of QMR. In particular, as long as the basis vectors remain linearly independent, convergence is not significantly different. We point out that in (37) the comparison of the two methods is done at the same iteration. An alternative measure of delay is obtained by monitoring the number of iterations required for the two methods to reach the same accuracy.

QMR is a successful implementation of more general residual smoothing schemes, which have received renewed interest after the work by Freund and Nachtigal. In fact, given a sequence of approximate solutions $\{x_k\}$ and associated residuals $r_k = b - Ax_k$, the following general residual smoothing technique can be considered (see, e.g. the presentation by Walker [341])

$$y_0 = x_0, \quad s_0 = r_0$$

$$y_k = y_{k-1} + \eta_k (x_k - y_{k-1}), \quad s_k = s_{k-1} + \eta_k (r_k - s_{k-1});$$

we refer to Brezinski and Redivo Zaglia [43] for more general procedures. Typically, the parameter η_k can be chosen so that the norms of the new sequence of residuals, $\{s_k\}$, have a smoother behaviour than their original counterparts. In particular, this is the case if η_k is chosen as $\eta_k = -s_{k-1}^{\mathrm{T}}(r_k - s_{k-1})/\|r_k - s_{k-1}\|^2$, which corresponds to the solution of the problem

$$\min_{\eta} \|s_{k-1} + \eta(r_k - s_{k-1})\|,$$

yielding the minimal $\|s_k\|$. As a consequence, $\|s_k\| \leqslant \|s_{k-1}\|$ and $\|s_k\| \leqslant \|r_k\|$. Several additional properties can be derived by exploiting further imposed conditions on the residual sequences, such as orthogonality among the original residual vectors. A full account can be found in Weiss [346]. With this choice of parameters, known relations between orthogonal residual and minimal residual methods may be restated, and they fully uncover the tight relation between, say, FOM and GMRES, as representatives of orthogonal and norm minimizing residual methods, respectively; cf. Brown [50], and Cullum and Greenbaum [73]. In particular, the following relation can be

DOI: 10.1002/nla

inferred from the general procedure above:

$$\|r_k^{\mathrm{F}}\| = \frac{\|r_k^{\mathrm{G}}\|}{\sqrt{1 - (\|r_k^{\mathrm{G}}\|/\|r_{k-1}^{\mathrm{F}}\|)^2}},$$

where r_k^{F}, r_k^{G} are the FOM and GMRES residuals after k iterations, respectively. This equality provides a fundamental argument in showing that peaks in the FOM residual norm correspond to plateaus in the GMRES residual norm, or, in other words, that the two methods converge hand in hand.

Another interesting choice in the smoothing sequence results from setting

$$\eta_k = \frac{\tau_k^2}{\|r_k\|^2},$$

where

$$\tau_0 = \|r_0\|, \quad \tau_k \quad \text{such that} \quad \frac{1}{\tau_k^2} = \frac{1}{\tau_{k-1}^2} + \frac{1}{\|r_k\|^2}.$$

In this case, and assuming exact arithmetic, it can be shown that if the original residual sequence is the one produced by BiCG, then the new sequence $\{s_k\}$ corresponds to the QMR residuals; see Zhou and Walker [352]. Considerations for the peaks/plateaus behaviour hold as for FOM and GMRES; see Cullum [71], and Cullum and Greenbaum [73]. A bound between QMR and GMRES residual norms similar to (33) can be found in Hochbruck and Lubich [182]. We also refer to Gutknecht and Rozložník [174] for further insight into relations between orthogonal residual methods and norm (quasi-)minimizing residual methods in a general framework.

With analogous devices [352], one can derive an additionally smoothed variant of Bi-CGStab, originally named QMRCGSTAB by Chan *et al.* [64]. A similar approach called TFQMR is used by Freund [127] to smooth the highly erratic convergence behaviour of CGS. A general theory of QMR smoothing for product methods is given by Ressel and Gutknecht [263] encompassing all these methods.

In [64, 127] the algorithms are implemented with a procedure that mimics that used for Lanczos-QMR described earlier in this section. The derivation relies on the fact that, by collecting the subsequent iterates, the approximate solution in the original methods Bi-CGStab and CGS, can be written as $x_k = x_0 + Z_k y_k$, where the full column-rank matrix Z_k satisfies a key relation of the form

$$AZ_k = U_{k+1} T_{k+1,k}. \tag{38}$$

Here the columns of U_{k+1} collect the first $k+1$ residuals, and $T_{k+1,k}$ is banded; in addition, note that (38) is even more general than (14) as it does not require that $Z_k = U_k$.

Then, following precisely the QMR derivation, the residual can be written as $r_k = r_0 - AZ_k y_k = U_{k+1}(\beta e_1 - T_{k+1,k} y_k)$, and a quasi-minimal residual norm procedure can be applied to determine a new vector y_k. It goes without saying that any Krylov subspace method where a relation of type (38) holds, can be equipped with a QMR smoothing. This fact suggests that smoothing may simply be viewed as a 'cosmetic' tool to let the convergence curve degrade more gently. A natural question is whether smoothing does provide a more accurate final solution or, in other words, whether the final attainable residual in finite precision arithmetic, is any better than the non-smoothed one.

DOI: 10.1002/nla

A thorough analysis performed by Gutknecht and Rozložník [175] answers negatively this question for most smoothing methods, that is, no gain is obtained by the smoothing algorithms in terms of final solution accuracy.

5. OTHER MINIMIZATION PROCEDURES

We begin by noting that some of the minimal residual methods described in the previous sections can be implemented in norms other than the 2-norm, and this is treated in some detail in Section 13.1; see also a special case of this in Section 10. When A is ill conditioned, small residuals do not necessarily imply accurate approximate solutions. Methods minimizing quantities other than the residual norm or the error A-norm have been proposed in the past few years. These include minimizing the 2-norm of the error, as we shall see next, and minimizing the joint backward perturbation norm, yielding an approximate solution in the Krylov subspace which is optimal in backward error sense.

A general error-minimization problem in a given subspace $\mathcal{S}_m$ of dimension m can be formulated as

$$\min_{x \in \mathcal{S}_m} \|x - x_*\|.$$

As mentioned in Section 2.4, algorithm SYMMLQ provides a stable implementation to solve this problem with $\mathcal{S}_m = \mathcal{K}_m(A, Ar_0)$, for symmetric, not necessarily positive definite matrices. For non-symmetric problems, Weiss [345, 346], more recently proposed a generalization of this approach by choosing $\mathcal{S}_m = A^{\mathrm{T}}\mathcal{K}_m(A^{\mathrm{T}}, r_0) \equiv \mathcal{K}_m(A^{\mathrm{T}}, A^{\mathrm{T}}r_0)$ and then applying a Galerkin condition to the error $x - x_*$. Although the approach minimizes the 2-norm of the error, this choice of approximation space does not ensure finite termination, unless the coefficient matrix is normal. Various alternative implementations of the method are investigated by Rozložník and Weiss [266], some of which are shown to be stable.

Methods minimizing the residual 2-norm determine x_m in $\mathcal{K}_m(A, r_0)$ satisfying $Ax_m = b - r_m$ with the minimum value of $\|r_m\|$. More generally, one can look for a vector x_m in $\mathcal{K}_m(A, r_0)$ satisfying $(A - \Delta_A)x_m = b + \Delta_b$, such that the *joint backward* perturbation matrix $[\Delta_A, \Delta_b]$ is minimized, in terms of the Frobenius norm. The approach is reminiscent of the total least squares problem [336], although in the present case the approximate solution is constrained to belong to the generated Krylov subspace. The problem can be formally stated as

$$\min_{x_m \in \mathcal{K}_m(A, r_0)} \|[\Delta_A, \Delta_b]\|_{\mathrm{Fr}}, \quad \text{subject to } (A - \Delta_A)x_m = b + \Delta_b. \tag{39}$$

A complete characterization of this problem and its solution was presented by Kasenally and Simoncini [201]. Moreover, the authors show that the solution to (39) is given by $x_m = V_m y_m$, where y_m satisfies

$$\begin{bmatrix} 1 \\ y_m \end{bmatrix} = \frac{1}{e_1^{\mathrm{T}} u_{m+1}} u_{m+1},$$

and where u_{m+1} is the left singular vector corresponding to the smallest singular value σ_{m+1} of the $(m+1) \times (m+1)$ matrix $[-\beta e_1, H_{m+1,m}]$. In addition, it holds that $\|[\Delta_A, \Delta_b]\|_{\mathrm{Fr}} = \sigma_{m+1}$. Clearly, the solution y_m is well defined only if the first component of u_{m+1} is non-zero, i.e. $e_1^{\mathrm{T}} u_{m+1} \neq 0$, and

DOI: 10.1002/nla

the solution is unique if σ_{m+1} is a simple singular value. These conditions are in agreement with the corresponding constraints in the total least squares setting [336]. The behaviour of the method for $e_1^{\mathrm{T}} u_{m+1} \approx 0$ is not analysed in [201], although a small factor $e_1^{\mathrm{T}} u_{m+1}$ may cause problems in finite precision arithmetic. It can be shown that the magnitude of $e_1^{\mathrm{T}} u_{m+1}$ is related to how close is $\delta = \sigma_{\min}([e_1, H_{m+1,m}])/\sigma_{\min}(H_{m+1,m})$ to one. A thorough analysis of the role of δ in the solution of least squares and total least squares problems can be found in [252, 253, 336].

An algorithm implementing the solution to (39) is in general computationally more expensive than GMRES or FOM, for it requires the solution of a singular value problem with an upper triangular matrix of size $m + 1$ at each iteration. On the other hand, its restarted version was shown to be more effective than restarted GMRES on certain sensitive matrices (restarted methods are described in Section 8). It is also worth noticing that an almost equivalent method could be devised by using the 2-norm instead of the Frobenius norm in the constrained minimization problem (39) [156]. Finally, we mention that the method above, referred to in [201] as Minpert, is a generalization of a method introduced in [200], where the perturbation in (39) is limited to the coefficient matrix, namely, it is assumed that $\Delta_b = 0$.

6. SPECTRAL TOOLS AND CONSIDERATIONS ON CONVERGENCE

In this survey, we do not analyse in detail the convergence properties of all the methods described in Sections 2–5. In this section, we present an introduction useful in understanding some of the convergence behaviour of the methods, and we introduce some spectral tools, such as Ritz and harmonic Ritz values, that have been used to enhance the convergence rate of some Krylov subspace methods.

The residual $r_m = r_0 - Ax_m$ satisfies (4). Therefore, the residual (and, similarly, the error) satisfies

$$\|r_m\| \leqslant \|p_m(A)\| \, \|r_0\|$$

for some $p_m \in \mathscr{P}_m$, where $\mathscr{P}_m$ is the set of all polynomials p of degree at most m such that $p(0) = 1$, and $\|p_m(A)\|$ is the induced matrix norm of $p_m(A)$. This simple bound shows that an estimate of the convergence of the residual norm may be obtained by analysing the behaviour of the associated polynomial on A. Note that this estimate does not take into account the action of r_0 on the matrix polynomial, therefore it is not sharp in most cases. For detailed studies on worst-case convergence of certain methods applied to some problems or for particular right-hand sides, see [107, 103, 216–218, 240, 241, 325, 349].

Assume next that A is diagonalizable, so that there exists a non-singular matrix X of eigenvectors of A and a diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ of corresponding eigenvalues such that $A = X\Lambda X^{-1}$. Therefore,

$$\|r_m\| \leqslant \max_{i=1,\ldots,n} |p_m(\lambda_i)| \|X\| \, \|X^{-1}\| \, \|r_0\|. \tag{40}$$

If the coefficient matrix A is normal (e.g. symmetric), then X is unitary, so that $\|X\| = \|X^{-1}\| = 1$, therefore an upper bound for the residual can be derived by only analysing the behaviour of the residual polynomial at the eigenvalues of A. When A is symmetric and positive definite and CG is used, a similar relation can be shown to hold for the A-norm of the error, which is the quantity

DOI: 10.1002/nla

that is minimized by the method. For CG we thus have

$$\|x_* - x_m\|_A \leqslant \min_{p_m \in \mathcal{P}_m} \max_{i=1,\ldots,n} |p_m(\lambda_i)| \|x_* - x_0\|_A$$

$$\leqslant \min_{p_m \in \mathcal{P}_m} \max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |p_m(\lambda)| \|x_* - x_0\|_A, \tag{41}$$

where in the last bound $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of the positive definite matrix A. The polynomial min–max problem has a classical solution [79], which yields the well-known bound

$$\|x_* - x_m\|_A \leqslant 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^m \|x_* - x_0\|_A, \tag{42}$$

where $\kappa = \lambda_{\max}/\lambda_{\min}$ is called the condition number of A. When the matrix is non-symmetric, a min–max polynomial problem can still be obtained by using a residual minimizing method such as GMRES. In this case, if r_m is the residual after m GMRES iterations, one obtains

$$\|r_m\| \leqslant \|X\| \, \|X^{-1}\| \min_{p_m \in \mathcal{P}_m} \max_{i=1,\ldots,n} |p_m(\lambda_i)| \, \|r_0\|.$$

Especially in the case of highly non-normal matrices, however, this bound may be a very poor estimate of the actual convergence, because $\|X\| \, \|X^{-1}\|$ may be very large irrespective of the value of $\|r_m\|$; see, also, Ipsen [187]. We also mention that a similar bound for non-diagonalizable matrices can be obtained, by using the Jordan canonical form of A; see Freund [125].

Other analyses have been proposed in the past few years, mostly aiming at sharpening the estimate of $\|p_m(A)\|$, see, e.g. [159–161, 165, 243]. Different approaches have used the field of values and pseudospectrum as possibly more representative tools for highly non-normal matrices [91, 107, 315, 327, 328]. Nonetheless, practical examples can be constructed where all these approaches fail to describe the actual residual behaviour [101], confirming that work is still needed to complete our understanding of the convergence when using a polynomial approximation technique.

Towards more realistic bounds, recent efforts have aimed at including the initial residual r_0 in the analysis. Some very insightful results have been obtained in [30, 213, 214], although sharp bounds for the residual in the most general cases are not available.

We also mention that when A is non-symmetric and positive real, i.e. when its symmetric part $(A + A^{\mathrm{T}})/2$ is positive definite, classical bounds can also be obtained using spectral information of the symmetric part of the matrix, and of $A^{\mathrm{T}}A$. In the case of GMRES, for instance, the following bound holds

$$\|r_m\| \leqslant \left(1 - \frac{\alpha^2}{\beta} \right)^{m/2} \|r_0\|, \tag{43}$$

where $\alpha = \lambda_{\min}((A + A^{\mathrm{T}})/2)$ and $\beta = \lambda_{\max}(A^{\mathrm{T}}A)$; see [96, 98, 275]. Although not sharp in general, bound (43) is commonly used in the context of discretization methods for partial differential equations to show that certain preconditioning schemes based on multilevel techniques provide convergence estimates that do not depend on the discretization parameters. This fact is usually established by showing that both α and β are bounded independently of the mesh size; see, e.g. [13, 57, 261, 281, 326, 348] and references therein. See, also, Starke [315] for a similar approach using the field of values and Eiermann and Ernst [92] for a complete derivation. We also mention

DOI: 10.1002/nla

the very recent paper by Beckermann, Goreinov, and Tyrtyshnikov [28], where bound (43) is improved.

Most convergence analyses of Krylov subspace methods, such as those described so far in this section—including those leading to the bounds (42) or (43), indicate a linear convergence rate, since they are based on asymptotic bounds. Nevertheless, in many occasions, these methods display faster convergence, and this has been called superlinear convergence. By this (non-traditional) superlinear convergence it is meant that the reduction of the residual norms is linear, but that the linear rate often increases, i.e. it accelerates as the iterations proceed; see [243, p. 9] for an alternative definition of superlinear convergence. This behaviour has been nicely characterized in the symmetric case, see, e.g. Axelsson and Lindskog [17], van der Sluis and van der Vorst [331]. In particular, it has been argued that once some of the roots of the polynomial defining the Krylov subspace method approximate certain eigenvalues of A, from then on the process behaves in the same manner as a new process with the same Krylov subspace method, where the initial residual has been stripped of all eigenvector components corresponding to these eigenvalues, and the rate changes. We also mention the work of Beckermann and Kuijlaars [29], where superlinear convergence for CG is studied using tools from potential theory; see also the recent paper by Kuijlaars [207].

In the non-symmetric case, different approaches have been proposed, including some generalizations of the concepts for the symmetric case. A bibliography on the subject can be found in the recent paper by Simoncini and Szyld [301]. In this last reference, a new general analytic model of superlinear convergence is presented. Briefly, this model says that as soon as the Krylov subspace $\mathscr{K}_m(A, r_0)$ approximates well an invariant subspace of A, from then on the process behaves in the same manner as a new process with the same Krylov subspace method, where the initial residual has been stripped of all components lying in that invariant subspace. One measure for the approximation of these subspaces is given by the so-called gap, which is related to the canonical angles between the subspaces.

We mention another recent contribution to the understanding of superlinear convergence of minimal residual methods by Kaporin [199], where a special conditioning measure is introduced using, among other parameters, the distance from the spectrum of A to the origin.

The relation between the residual polynomial and the convergence of Krylov subspace methods has motivated the study of these polynomials in greater detail. Information, either explicit or implicit, on the polynomial roots may be used to accelerate or to analyse convergence as the iterations proceed. It can be shown that the roots of the residual polynomial in FOM and CG coincide with the eigenvalues of the upper Hessenberg matrix H_m in the Arnoldi procedure; see, e.g. Saylor and Smolarski [284]. These are called Ritz values. For simplicity, let us consider the symmetric case. It can be shown that for m sufficiently large, some of the Ritz values tend to converge to the extreme eigenvalues of A. In other words, some of the roots of the residual polynomial of sufficiently large degree are very close to the extreme eigenvalues λ of A, so that $|p_m(\lambda)| \ll 1$. Analogously, it can be shown that the roots of the GMRES and MINRES residual polynomials coincide with the eigenvalues of the generalized eigenvalue problem

$$H_{m+1,m}^{\mathrm{T}} H_{m+1,m} z = \theta H_m^{\mathrm{T}} z$$

(see Freund [125], and Manteuffel and Otto [223]), or of the standard eigenvalue problem

$$(H_m + (H_m^{\mathrm{T}})^{-1} e_m h_{m+1,m}^2 e_m^{\mathrm{T}}) z = \theta z;$$

see Paige, Parlett, and van der Vorst [248], and Simoncini and Gallopoulos [295]. These latter eigenvalues have been first called harmonic Ritz values in [248], where it is shown that these

DOI: 10.1002/nla

eigenvalues derive from an orthogonal projection method for approximating eigenvalues of A^{-1}. It is interesting to observe that in terms of A, harmonic Ritz values are typically derived by means of an *oblique* projection method, whereas orthogonal (Galerkin) projection leads to Ritz eigenpairs [20, Section 3.2]; see, also, Morgan [231]. Relations between Ritz and harmonic Ritz values have been explicitly discussed by Goossens and Roose [155].

It is common practice to use Ritz or harmonic Ritz values as approximate eigenvalues in both the symmetric and non-symmetric case. In particular, these approximations are used in some of the acceleration procedures that we discuss in Section 9. On the other hand, in the non-normal case, eigenvalues of A, and thus their approximations, may not play a role in the convergence. In fact, it was shown by Greenbaum, Ptàk, and Strakoš [162, 164], that spectral information alone may provide misleading information in the non-normal case. These conflicting views indicate that the convergence analysis of Krylov subspace methods for general problems is a challenging area of research. Both the right-hand side and the invariant subspaces associated with A seem to provide more valuable information than the spectrum alone; the analytical model in [301] combines these two quantities in an insightful manner.

7. BRIEF COMMENTS ON PRECONDITIONING TECHNIQUES

We give here a brief introduction to preconditioning, mostly directed to the reader unfamiliar with this concept and its use. Again, we refer to [26, 32, 54, 68, 273], and references therein for details.

By preconditioning it is usually meant a transformation of the original problem (1) to the form

$$M_1^{-1} A M_2^{-1} \widehat{x} = M_1^{-1} b, \quad \widehat{x} = M_2 x,$$

where M_1, M_2 are two non-singular matrices with the following crucial properties: (i) their inverse should be cheaply applicable; (ii) their use should entail low memory requirements; (iii) the transformed problem should converge faster (less computational time) than the original problem. There is a clear conflict among these three requirements, especially for the construction of general-purpose preconditioners. The generic formulation above allows for left preconditioning ($M_2 = I$), right preconditioning ($M_1 = I$), or left–right preconditioning. Since only matrix–vector multiplications with the (preconditioned) coefficient matrix of the type $v = M_1^{-1} A M_2^{-1} z$ are required in Krylov subspace methods, the matrices M_1^{-1} and/or M_2^{-1} need not be explicitly known or computed. What is needed is a subroutine that computes, say, $M_1^{-1} w$ for any vector w. In some application-oriented preconditionining techniques, this feature allows to fully meet requirements (i) and (ii).

Bound (42) suggests that in the symmetric positive definite case M_1, M_2 may be chosen so that $\kappa(M_1^{-1} A M_2^{-1}) \ll \kappa(A)$. Generalizing this concept, a good preconditioner is such that in some sense $M_1^{-1} A M_2^{-1}$ is close to the identity, for example, with many eigenvalues clustered near 1. The simplest and cheapest preconditioner is, for instance, $M_1 = \mathrm{diag}(A)$, $M_2 = I$, i.e. a scaling so that $M_1^{-1} A$ has unit diagonal entries. This choice may be effective when A is strongly diagonally dominant.

We can identify two major approaches: those problem specific, often used in conjunction with a differential equation, and algebraic ones. A particularly successful approach in the latter class consists of approximately performing a possibly implicit factorization of the matrix A (or of its inverse), while maintaining low memory requirements. For example, an *incomplete* Cholesky

DOI: 10.1002/nla

factorization of a sparse positive definite symmetric matrix has the form LL^{T}, where L is derived in a way similar to a complete factorization, so as to either maintain a certain level of sparsity, and/or have values above certain threshold. The goodness of the approximation is measured in terms of some norm of $A-LL^{\mathrm{T}}$. The symmetric and indefinite case provides additional challenges.

In the non-symmetric case, such incomplete factorizations are also widely used. Even in the simplest cases, however, the amount of success often depends on the user ability to tune the fill-in and threshold parameters. Unfortunately, these algebraic approaches do not work well in all cases. As we discussed briefly in the previous section, part of the difficulty in the non-symmetric case is that the convergence of the Krylov subspace methods does not depend only on the eigenvalues, especially in the non-normal case. In particular, the convergence may not depend only on their relative position in the complex plane, e.g. clustering. Nevertheless, incomplete LU factorizations, especially using thresholds, and sparse approximate inverse factorizations can be very effective for many problems. Other widely used preconditioners are multigrid and domain decomposition techniques, which are commonly applied by means of a subroutine, since an explicit expression for them is not needed; see, e.g. [100, 326, 329]. For harder problems different types of preconditioners can be combined to optimize the trade-off between efficiency and computational constraints. We stress here that recent work on preconditioning has made iterative methods for certain problems closer to direct methods in terms of robustness [32, 100, 326].

8. REDUCING THE COST I: RESTARTED AND TRUNCATED METHODS

The methods based on the (full) Arnoldi recurrence for non-symmetric matrices are in general very expensive, in their original form. A large number of iterations may be required to achieve a sufficiently accurate solution, so that the Arnoldi matrix V_m becomes unacceptably large to be stored and to be kept orthonormal. The standard procedure consists in *restarting* the method when a maximum subspace dimension is reached. More precisely, after say, m iterations, the process is interrupted, and the current approximation x_m and associated residual r_m are computed. These become the starting quantities for the new recursion that is executed for at most m iterations. Clearly, any of the methods discussed in the previous sections could be used at each restart, such as GMRES, FOM, Minpert, and the restarted versions are denoted GMRES(m), FOM(m), Minpert(m), etc.; we also refer to Chronopoulos [69] for a different minimal residual-type implementation of the same idea. The overall procedure for a maximum 'maxit' number of restarts works as follows:

Given $A, x_0^{(i)}, b, m$, maxit
while $i <$ maxit
  Run m iterations of the chosen method and get $x_m^{(i)}$
  Test $\|r_m^{(i)}\| = \|b - Ax_m^{(i)}\|$. If satisfied then stop
  Set $x_0^{(i+1)} = x_m^{(i)}$, $i = i + 1$
end

The iteration above is a first example of nested iterations, where an (outer) iteration requires an (inner) iteration at each step; we discuss more advanced nested procedures later in this section and in Section 10. We also mention that in some cases, the efficiency of this type of nested iteration can be improved by judiciously playing with low- and high-accuracy computation; see Turner and Walker [330].

DOI: 10.1002/nla

The advantage of a restarted procedure is that at most m iterations of the Arnoldi method are carried out, so that both computational costs and memory allocations per cycle are under control. On the other hand, optimality properties of the process, as in GMRES or Minpert, are lost after the first restart. As a result, the overall process may not converge. In the case of GMRES, for instance, the outer recurrence may stagnate with $\|r_m^{(i+1)}\| \approx \|r_m^{(i)}\|$ for all i. Note that in GMRES the residual norm cannot increase in the outer iteration, since the optimality of the inner GMRES step ensures that $\|r_m^{(i)}\| \leqslant \|r_0^{(i)}\|$ for all i. If stagnation occurs, a simple cure is to enlarge the maximum allowed subspace dimension, m, to enrich the subspace information. We stress, however, that enlarging the subspace dimension does *not* always ensure faster convergence; see Eiermann, Ernst, and Schneider [94], and Embree [102], for some critical examples and for pointers to further numerical evidence. Moreover, choosing a larger m may not be possible in general, especially on large problems, since m is usually chosen as the maximum dimension affordable, although judiciously selecting a dynamic subspace dimension may provide some advantages; see Joubert [197]. When restarting is used, the GMRES residual may not be the most significant direction vector to carry over to the new restart. The residual $r_m^{(i)}$ is a linear combination of all $m+1$ basis vectors (cf. (18)); in particular, if $r_m^{(i)}$ is mostly in the direction of the first basis vector v_1, that is $r_m^{(i)}$ is almost a multiple of v_1, then at the next restart the starting vector $r_0^{(i+1)} \equiv r_m^{(i)}$ builds a Krylov subspace that is very close to the previous one, so that no significant approximation improvement is observed. This argument partially explains stagnation. In fact, if $v_{m+1}^{\mathrm{T}} r_m^{(i)} \neq 0$, then the overall generated space in two consecutive restarts of size m is the same as that obtained by performing $2m$ steps of the unrestarted method [94, 289]. Clearly, in this case the two procedures differ in the way they determine the solution within this subspace of dimension $2m$. A natural strategy to ensure that the overall subspace generated after restarting the method has maximum dimension (in exact arithmetic) is to enforce that the new direction vector has non-negligible component onto v_{m+1}. Such requirement is automatically satisfied by FOM, since $r_m^{(i)}$ is a multiple of v_{m+1}; see (32). In this sense, FOM may be more appealing than GMRES in a restarted context; see also the relevant discussion in [298, Section 5]. Other schemes can be coupled with GMRES at restart time, resulting in better restarting strategies than restarted GMRES; see, e.g. the strategy proposed in [288].

Recent efforts have focused on enriching the information carried over at restart time. The main motivation is that, except for the current approximation to the solution, at each restart all information generated during the inner Arnoldi iteration is completely lost. Numerical evidence showed that there is relevant information that, if kept, may help enhance later restarts. These issues are discussed in the following paragraphs. A recent experimental analysis by Zhong and Morgan [351] also pointed out that the eigencomponents of the GMRES residual vector may significantly vary at different restarts, showing a complementary role of various parts of the spectrum at restart time.

A simple way to maintain information from previously built vectors while limiting memory requirements, is to discard older vectors in the basis, that is, only the last j, say, vectors in the basis are kept orthogonal to each other. Then the summation in the basic recurrence (8) becomes

$$h_{k+1,k} v_{k+1} = A v_k - \sum_{i=\max\{1,k-j+1\}}^{k} h_{ik} v_i .$$

DOI: 10.1002/nla

The set of orthogonal Krylov subspace vectors is dynamically updated and after the first j steps, at each iteration the oldest vector leaves the set and the last computed vector enters the set. In this way, the original Arnoldi procedure is *truncated*; however, relation (10) remains valid, but only j consecutive columns of V_m are orthonormal, and the upper Hessenberg H_m is banded with upper bandwidth j. The fact that H_m is banded allows to progressively update the approximate solution x_m, while storing only the last j columns in the matrix V_m. In this manner restarting can be avoided, although the selection of j may require some tuning. Truncated versions of both GMRES and FOM can be derived, as shown by Saad and Wu [277], and Saad [269], respectively; in [270] a more efficient implementation of the original truncated FOM was also introduced. The advantages of these procedures over their restarted counterparts are not always apparent for general non-symmetric matrices; see, e.g. [273, Sections 6.4.2, 6.5.6]. However, a recent analysis in [300] shows that if the original full (non-truncated) method converges smoothly and reasonably fast, then the truncated scheme only experiences a small convergence delay; see, also, Jia [192] for a similar discussion specific for truncated FOM. A natural question when discussing truncation strategies is whether it is necessarily the wisest thing to keep the *latest* basis vectors; in general, some of the discarded vectors might have been more significant than the ones that are kept. Proposed strategies on how to choose the 'good' vectors are discussed in the next section.

9. REDUCING THE COST II: AUGMENTED AND DEFLATED METHODS

The general idea of deflation and augmentation methods is to determine an approximation space of dimension m as the direct sum of two spaces of smaller dimension, as $\text{span}\{v_1, \ldots, v_k, w_1, \ldots, w_{m-k}\}$: the first k vectors are determined using the standard Arnoldi procedure with the current residual, while the remaining vectors $w_1, \ldots, w_{m-k}$ contain relevant information saved from previous outer cycles.

The various techniques differ either on the strategy used to compute these latter vectors, or on the way these are included in the approximation procedure. A thorough analysis of these acceleration procedures in Hilbert spaces is presented in [94]. There, various approaches proposed in the literature are presented within the same framework, and some elegant relations among the discussed methods are uncovered. Below we review some of the principal techniques, while we refer to the original papers for a more detailed presentation and for performance evaluation.

In [232], Morgan proposes to compute $w_1, \ldots, w_{m-k}$ as approximate eigenvectors generated in the current approximation space, for example Ritz vectors in the case of FOM, or harmonic Ritz vectors in the case of GMRES. This strategy seems to work well when approximate eigenvectors associated with a group of small (in magnitude) eigenvalues are retained. The rationale behind the use of a nearly invariant subspace $\mathcal{U}$ is that the residual obtained after projection onto the so-called augmented space has little components onto $\mathcal{U}$; see, e.g. Eiermann, Ernst, and Schneider [94, Proposition 4.1].

It can be shown that if the approximate eigenvectors are selected as Ritz or harmonic Ritz vectors (see Section 6 for their definition), then the subspace $\text{span}\{v_1, \ldots, v_k, w_1, \ldots, w_{m-k}\}$ is still a Krylov subspace generated by A, but with a different starting vector. As an example, let w_1, w_2 be two harmonic Ritz vectors with θ_1, θ_2 associated harmonic Ritz values. We first notice that $Aw_j - \theta_j w_j = \gamma_j r_m,\ j = 1, 2$ for some γ_j, where r_m is the GMRES residual at the current cycle; in particular, this means that both eigenresiduals are multiples of the GMRES residual. Then

DOI: 10.1002/nla

setting $s = \gamma_2 w_1 - \gamma_1 w_2$ and letting $v_1 = r_m/\|r_m\|$ be the new cycle initial vector, we have (see, Morgan [233])

$$\text{span}\{s, As, \ldots, A^{m-1}s\} = \text{span}\{w_1, w_2, v_1, Av_1, \ldots, A^{m-3}v_1\}.$$

The idea of implicitly including the eigenvectors in restarted GMRES is first discussed by Morgan [233], while a more stable implementation is proposed in [234], where the inclusion of the enriching vectors is done within the Krylov subspace during each cycle. We also point to Baglama *et al.* [19] for an implementation of this idea using the Implicitly Restarted Arnoldi method of Sorensen [314]. This strategy seems to be particularly effective when *a priori* information on the problem confirms the presence of a group of small (or more generally outlying) eigenvalues. In particular, if spectral information is available from the application, then this should be directly injected into the process, possibly with an explicit deflation procedure; see below. We refer to Saad [274] for some theoretical results establishing bounds for the residual in the Krylov subspace augmented by nearly invariant subspaces, both in the symmetric and non-symmetric cases. The success of these augmented strategies also depends on the matrix being not too far from normal.

Available spectral information may be included in the restarted process in the form of acceleration procedures, such as polynomial preconditioning and explicit deflation strategies. Polynomial preconditioners are of the form $M^{-1} = p(A)$ for some polynomial p; see, e.g. [97, 123, 193, 196, 271]. The approach using polynomial preconditioning may be implemented by explicitly applying a conveniently chosen polynomial before each restart. Several strategies have been proposed, which mostly differ on the type of polynomial explicitly generated, either at each restart or once for all cycles. See [239, 285, 316], for some examples where the chosen polynomial $p = p(\lambda)$, $p(0) = 1$, is applied to the residual r_m as $\hat{r}_m = p(A)r_m$; the approximate solution is updated accordingly. The Krylov subspace method is then restarted with $\hat{r}_m$. Most commonly, spectral information is employed in the form of eigenvalue approximations, as polynomial roots, or by determining a least squares polynomial in a region possibly including the matrix eigenvalues. Such an approach is particularly convenient with symmetric problems, or with positive definite matrices having eigenvalues with small imaginary part. On the other hand, it was shown by Sorensen [314] that polynomial information computed within the Arnoldi procedure can be applied without explicitly performing matrix–vector multiplications with A; the resulting approach is the implicitly restarted method by Morgan [233] that we discussed in the previous paragraph.

If the spectral information carried around is limited, it may be convenient to explicitly *deflate* the eigenvector components and directly solve the *deflated* problem. We begin our description of this technique with the case of A symmetric and positive definite. Let the columns of matrix W be good eigenvector approximations of A. Let x_0 be a starting approximation such that $r_0 = b - Ax_0 \perp \mathscr{R}(W)$. Then an approximate solution can be determined as $x_0 + z$ where z solves

$$(A - AW(W^{\mathrm{T}}AW)^{-1}W^{\mathrm{T}}A)z = r_0, \quad r_0 \perp \mathscr{R}(W); \tag{44}$$

see, e.g. [105, 106, 122, 222, 244, 278]. The coefficient matrix needs not be computed explicitly. In practice, one can compute $\widehat{W} = AW$, factorize the small matrix $W^{\mathrm{T}}\widehat{W}$ once for all, and compute the matrix–vector multiply as

$$\hat{v} = Av, \quad y = (A - AW(W^{\mathrm{T}}\widehat{W})^{-1}W^{\mathrm{T}}A)v = \hat{v} - \widehat{W}((W^{\mathrm{T}}\widehat{W})^{-1}(W^{\mathrm{T}}\hat{v})).$$

A more general formulation is possible, that does not require the initial constraint $r_0 \perp \mathscr{R}(W)$. Indeed, the solution x can always be written as $x = x_0 + W(W^{\mathrm{T}}AW)^{-1}W^{\mathrm{T}}r_0 + P_D z$, where

DOI: 10.1002/nla

$P_D = I - AW(W^{\mathrm{T}}AW)^{-1}W^{\mathrm{T}}$ and z is any solution to the singular (but consistent) linear system $P_D Az = P_D b$; see Nabben and Vuik [235]. An iterative procedure following this approach reduces to (44) if $r_0 \perp \mathscr{R}(W)$.

It is shown by Saad *et al.* [278] that the approach in (44) can be nicely inserted into a CG-type iteration, and that the recurrence does not break down, in spite of the singular coefficient matrix. Moreover, the upper bound (42) for the convergence rate of CG holds with $\kappa = \lambda_{\max}/\lambda_k$, where λ_k is the smallest non-zero eigenvalue of A. In particular, if exact spectral information of A is available about some critical region of the spectrum, say the region closest to zero, convergence of the deflated method is only driven by the remaining (non-zero) eigenvalues of A. It should be noted that the procedure above amounts to finding an approximate solution to $Ax = b$ in the space spanned by W and the Krylov subspace vectors, making explicit the link between deflation and augmented procedures. In other words, the deflation procedure explicitly solves the linear system in the known eigenspace, while it uses CG to solve for the remaining invariant subspace. The matrix W may be either available from the application problem, or computed by means of a pre-processing with a (sparse) eigenvalue solver. We refer to Perotti and Simoncini [258] for an application with complex symmetric matrices in a structural dynamics problem, where the *known* approximate eigenvectors correspond to the rigid-body modes. There is an elegant and insightful connection between the explicit deflation method presented above and the so-called coarse grid correction preconditioning, a well-established approach in the domain decomposition literature; see, e.g. [261, 326]. This relation is uncovered by Nabben and Vuik [235]; see also [236, 247].

The procedure can be generalized to non-symmetric A. The problem to be solved becomes

$$(I - USW^*)Ax = (I - USW^*)b,$$

where $S = (W^*AU)^{-1}$ and the columns of U, W, span approximate right and left invariant subspaces, respectively, associated with a group of 'undesired' eigenvalues; see Chapman and Saad [66]. However, it was shown in [94, Theorem 4.6] that if an exact invariant subspace of A is available, the residual norm obtained with this approach is not smaller than that obtained by using the corresponding augmented technique. In case of nearly invariant subspaces the comparisons are less decisive.

Finally, an adaptive approach close to those described above is based on the idea of translating a group of small eigenvalues by means of a series of low-rank projections of the coefficient matrix of the form

$$\widetilde{A} = A(I + u_1 w_1^*) \cdots (I + u_k w_k^*),$$

where u_j, w_j are the right and left eigenvectors associated with the eigenvalues to be translated; see Kharchenko and Yeremin [204]. The Krylov subspace method is thus applied to $\widetilde{A}$. If a restarted method is employed, then $\widetilde{A}$ is updated at each restart, by possibly modifying the eigenvalues to be translated. We should mention that as described in previous sections, the whole procedure may be applied when A is already a preconditioned matrix, giving rise to a 'two-level' preconditioning strategy; see, e.g. Nicolaides [244], Padiy, Axelsson, and Polman [247], and Carpentieri, Duff, and Giraud [62]. In particular, in [62] the issues associated with an effective implementation of this approach on real application problems are explored; in Giraud, Gratton, and Martin [144] additional theoretical results are reported, together with the description of applications to the solution of sequences of linear systems. The accuracy with which spectral information is obtained is crucial for the performance of the method.

DOI: 10.1002/nla

A very promising generalization of augmentation procedures aims at determining augmentation spaces based on information other than spectral invariant subspaces, in the hope that performance will not crucially depend on the presence of well-identified eigenvalue clusters. At each restart, the strategy decides which subspace of the subspace available should be kept to augment the Krylov subspace at the next restart. The analysed space includes previously kept vectors, as well as the vectors generated in the current cycle. Usually, the new Krylov subspace vectors are orthogonalized with respect to the kept basis, so that completely fresh information is added.

In de Sturler [83] this general strategy, called GCROT, is employed within an inner–outer method; see Section 10 for a more detailed description of inner–outer methods including GCRO. The outer method used is restarted GCRO while the inner method is GMRES. At each iteration of GCRO, m steps of GMRES are carried out, an approximate solution and a Krylov subspace basis are generated. This inner basis is kept orthogonal also with respect to the outer GCRO basis built so far. At the end of the inner GMRES cycle, a subspace of dimension $p_1 < m$ is selected from the GMRES space, to complement the outer basis. The proposed strategy selects the subspace that provided the smallest inner residual norm. The actual algorithm is even more flexible, but for ease of presentation we restrict to this 'simpler' implementation. The outer process then proceeds one more step, and the described process is repeated until restart time for GCRO. If after the addition of the inner vectors the outer basis becomes larger than the maximum allowed dimension, truncation takes place and only the most significant vectors are retained; this step is carried out by comparing the residual components in the combined inner–outer space, and those in the inner space; we refer to [83] for a complete description of this approach. The results are in general quite satisfactory, and in some non-trivial instances the method may even compete with full (unrestarted) GMRES. Unfortunately, the current implementation requires the selection of as many as six parameters, which may be hard to tune. With some practice, however, some of these parameters may be fixed, as suggested by the author.

A simplified version of this approach, fully based on restarted GMRES has been recently proposed by Baker, Jessup, and Manteuffel [23]. In this case, at each restart an approximate solution is constructed by using the generated Krylov subspace and an additional basis, in which each of the k vectors contains 'error' information of each previously built k subspaces. In spite of the large number of numerical experiments on different benchmark problems, neither of these two methods seems to perform consistently better than the other. This is certainly a very promising area of research, i.e. to be able to substantially improve the performance of Krylov subspace method on large applications, when no *a priori* information on the problem is available.

10. FLEXIBLE METHODS WITH VARIABLE PRECONDITIONING

Flexible Krylov subspace methods refers to a class of methods in which the preconditioner is allowed to change from one step to the next. The ideas described in this section are most generally used with right preconditioning as in (2). Thus, at each step of the Krylov subspace method one needs to compute $AM^{-1}v_k$ (and then usually orthogonalize this vector as in (8)). This is performed in two steps: first solving the preconditioning equation

$$Mz = v_k, \tag{45}$$

and then computing Az.

DOI: 10.1002/nla

One of the motivations for these flexible methods with variable preconditioners is the need to solve each preconditioning equation (45) only approximately. In other words, when solving system (45) one may have a different matrix, say $M = M_k$, for each k. Often in these cases, an approximate solution $\widehat{z}$ of (45) is determined so that the associated residual falls below some prescribed (inner) tolerance, i.e. such that

$$\|v_k - M\widehat{z}\| < \varepsilon_k. \tag{46}$$

This is done, e.g. by Dolean and Lanteri [84] and Elman, Ernst, and O'Leary [99] using multigrid, or by Warsa and Benzi [343] using a two-stage preconditioner, one of which is variable; see, also, Carpentieri [2]. Furthermore, in certain cases, preconditioners can be updated with newly computed information, e.g. as in Eirola and Nevanlinna [95] (see also [51, 340, 346]), or when a preconditioner such as SOR uses a parameter which can be improved as the iterations proceed [257]. We discuss another motivation in more detail later in this section, where the preconditioner is itself a call to a Krylov subspace method.

We begin by briefly describing a flexible variant of GMRES, called FGMRES, proposed by Saad [272]. When we have a fixed right preconditioner, the approximation x_m is written as (cf. (16))

$$x_m = M^{-1} V_m y_m. \tag{47}$$

With right variable preconditioning, clearly $M^{-1}V_m \neq [M_1^{-1}v_1, \ldots, M_m^{-1}v_m]$, therefore the final approximate solution x_m cannot be recovered at convergence by means of (47). Instead, during the recurrence, one computes $z_k \approx M_k^{-1} v_k$, $k = 1, \ldots, m$, and collects these vectors in $Z_m = [z_1, \ldots, z_m]$, so as to write

$$x_m = Z_m y_m. \tag{48}$$

The corresponding Arnoldi relation is $AZ_m = V_{m+1} H_{m+1,m}$ (cf. (38)) and the vector y_m is obtained by the same minimization (19).

We make several observations. The subspace $\mathscr{R}(Z_m)$ is not necessarily a Krylov subspace. Unlike the situation in several methods described in Sections 8 and 9, there may not exist any Krylov subspace containing $\mathscr{R}(Z_m)$; for a special case though, see the end of the section. Nevertheless, as long as $\mathscr{R}(Z_m) \subset \mathscr{R}(Z_{m+1})$, and thus the subspace keeps growing, there is no breakdown, and the sequence $\|r_m\|$ is non-increasing; cf. the analysis by Eiermann and Ernst [92]. In this method, the storage is essentially doubled, since we have to store both Z_m (to get the iterates as in (48)) and V_m (to keep orthogonalizing against its columns).

Following analogous procedures, flexible variants of other preconditioned algorithms can be derived, such as variable preconditioned CG [153, 245], flexible CG [113], flexible QMR [322], or flexible BiCG and Bi-CGStab [337].

A case of special interest, both from a practical as well as from a theoretical point of view, is given by the possibility that the variable preconditioner M_k be a call to a Krylov subspace method, with either a larger tolerance, or a fixed number of iterations (or a combination of both). This preconditioner is referred to as the inner iterative method (or inner iterations) and the overall method as an inner–outer method, or sometimes as a nested method. Examples of these include GMRESR by van der Vorst and Vuik [335, 338, 339], where both the outer and the inner are the GMRES method. GMRESR is in fact a particular case of GMRES$\star$, a general inner–outer method where the outer method is GMRES [335]. Conceptually, GMRESR can be

DOI: 10.1002/nla

seen as FGMRES with an inner iterative method. In practice, the implementation is different since GMRESR is based on GCR (see Section 2.2), and furthermore in the original proposal for GMRESR [335], one or more additional steps of LSQR by Paige and Saunders [251] (see Section 2.3) are performed to guarantee that there is no breakdown, i.e. that the subspace keeps growing.

The first paper describing a Krylov subspace method for variable preconditioning was possibly by Axelsson and Vassilevski [18], where the Generalized Conjugate Gradient method (GCG) is used; see, also, [13]. GCG uses orthogonal directions z_i, $i = 1, \ldots, m$ (with respect to a particular inner product) to span a subspace in which the approximate solution x_m is taken so that the residual norm is minimized (in some norm induced by some—possibly different—inner product); see Axelsson [12].

It turns out then, that FGMRES and GMRESR are special cases of GCG with the appropriate choices of inner products, although van der Vorst [333, p. 91] says that the GMRES$\star$ implementation is a more efficient computational scheme. The Flexible CG of Notay [245] is also a particular case of GCG.

Another way of looking at these inner–outer methods is to view them as having polynomial preconditioners (as defined in Section 9), where the polynomial changes from one step to the next, as is implicitly defined by the inner Krylov subspace method. We mention in passing that a fixed polynomial preconditioning is often not competitive; see, e.g. Faber and Joubert [109].

de Sturler [81] made the observation that in the inner–outer methods just described, the inner iterative method does not take advantage of any of the information available from the subspace of the outer iteration. He proposed that the inner iteration take place in a subspace orthogonal to the (outer) Krylov subspace. In this manner, the inner iteration would minimize the residual over both the inner and the outer subspaces. Experiments in [81] indicate that in some cases, the additional cost of orthogonalizations may be justified, and the resulting method, the already discussed GCRO, may perform better than the inner–outer methods (without orthogonalizations).

In inner–outer methods, when the outer and the inner methods are the same, e.g. FGMRES-GMRES, FQMR-QMR, GMRESR, it is shown in [298] that the (global) iterate x_m lies in an m-dimensional subspace of a larger Krylov subspace, of dimension p, where p is the total number of inner iterations; cf. also [81] and [333, p. 93]. It should be clear that these methods cannot find a better approximation than the corresponding minimum residual method over $\mathscr{K}_p$, but of course the advantage is that many fewer vectors are kept in storage, and fewer calculations are performed as well. If one fixes the number of inner iterations, e.g. using FGMRES-GMRES(k), it is natural to compare the performance of these methods with the restarted methods of Section 8. An argument based on the minimization properties of the outer method can be made, indicating that one can expect that in general the inner–outer method would outperform the restarted method; see further details in [298].

11. INEXACT METHODS

In the first part of this section we return to considering the matrix A, and Krylov subspaces associated with it. This matrix may represent the already preconditioned operator AM^{-1} or $M^{-1}A$. As we have seen, this matrix is needed as an operator to produce a matrix–vector product $z = Av$. Inexact Krylov subspace methods is the name given to a class of methods where this matrix–vector

DOI: 10.1002/nla

product is not performed exactly. Instead, we have

$$z = (A + E)v = Av + f, \tag{49}$$

where E is some error matrix, which may change from one application of the operator A to the next. We remark that the magnitude of $\|E\|$ can be quite large, that is, we are not discussing here small perturbations, such as those arising in finite precision arithmetic.

There are many scientific applications where the inexact matrix–vector product (49) appears naturally. For example, when using approximately a Schur complement as in Mandel [221], Maryška, Rozložník, and Tůma [225], or Smith, Bjørstad, and Gropp [312], or other situations where the operator in question implies a solution of a linear system, such as in certain eigenvalue algorithms; see Golub, Zhang, and Zha [154], and Simoncini and Eldèn [293]. Other examples include cases when the matrix is very large (and/or dense), and a reasonable approximation can be used; see Carpentieri *et al.* [63], and Cundy *et al.* [76].

A series of experimental reports by Bouras, Frayssé, and Giraud [37–39], have brought to the fore interest in this class of methods, since the authors observed that the norm of the error matrix should be small during the first iterations of the Krylov subspace method, but that it can grow at later stages of the method. In other words, as the iterations proceed, the matrix–vector product can be more and more inexact. The authors posited that if one restricts the inexactness at the step m of the Krylov subspace method in (49) by

$$\|f\| \leqslant \frac{\varepsilon}{\|r_{m-1}\|}, \tag{50}$$

the overall method maintains its convergence properties. Their experiments show that this was true in some cases, but not in others.

Let E_k be the error matrix at the kth iteration. The Arnoldi relation (9) becomes

$$\begin{aligned} AV_m &= V_{m+1}H_{m+1,m} - [E_1v_1, E_2v_2, \ldots, E_mv_m] \\ &= V_{m+1}H_{m+1,m} - [f_1, \ldots, f_m]. \end{aligned} \tag{51}$$

This is another situation where the columns of V_m do not span a Krylov subspace generated by the matrix A. Nevertheless, the subspaces $\mathscr{R}(V_m)$ are nested, and as long as the new vector $(A+E_k)v_k$ is linearly independent of the vectors in $\mathscr{R}(V_m)$, a minimal residual method or a method with a Galerkin condition converges to the solution of (1); see, e.g. Eiermann and Ernst [92]. On the other hand, if for some k, $(A + E_k)v_k \in \mathscr{R}(V_k)$, then breakdown occurs. In Giraud, Gratton, and Langou [143, Theorem 1] sufficient conditions on $\|E_k\|$ are given so that only a benign breakdown occurs, with which the GMRES computed solution can still be obtained.

Analysis of inexact methods can be found in [299, 305], where bounds of the norm of the difference between the computed residual $\tilde{r}_m$ and the true residuals $r_m = b - Ax_m$, i.e. the residual gap, were given in terms of the magnitude of the error matrix. The computed residual is usually obtained from the method directly, for example, as in (32). Let $W_m = V_{m+1}H_{m+1,m}$, and let $y_m = [\eta_1^{(m)}, \ldots, \eta_m^{(m)}]^{\mathrm{T}}$. In [299] it is shown that for inexact GMRES

$$\|r_m - \tilde{r}_m\| \leqslant \sum_{k=1}^{m} |\eta_k^{(m)}| \, \|E_k\| \quad \text{and} \quad \|W_m^{\mathrm{T}} r_m\| \leqslant \|H_{m+1,m}\| \sum_{k=1}^{m} |\eta_k^{(m)}| \, \|E_k\|. \tag{52}$$

DOI: 10.1002/nla

From these bounds it can be seen that as long as the *products* $|\eta_k^{(m)}|\,\|E_k\|$ are small, then the residual gap and the projection of the residual onto the subspace $\mathscr{R}(W_m)$ are small, even if one of the factors, especially the norm of the error matrix is large. It turns out that the magnitude of the components of y_m is indeed decreasing, that is, $|\eta_1^{(m)}|\geqslant|\eta_2^{(m)}|\geqslant\cdots\geqslant|\eta_m^{(m)}|$, so that the magnitude of the error matrices can grow as the iterations progress. We refer to [302] for a detailed analysis of this decreasing pattern, while here we emphasize that it holds $|\eta_k^{(m)}|=O(\|e_1\beta-H_{k,k-1}y_{k-1}\|)$, where $\|e_1\beta-H_{k,k-1}y_{k-1}\|$ is the residual obtained at the $(k-1)$st iteration. We note that the second bound in (52) indicates how far one is from the exact situation, where one would have $W_m^{\mathrm{T}}r_m=0$. Bounds similar to those in (52) can also be obtained for inexact FOM.

In [299], it is also shown that if

$$\|f_k\|\leqslant\ell_m\frac{\varepsilon}{\|\tilde{r}_{k-1}\|}\quad\text{for } k\leqslant m, \tag{53}$$

then $\|r_m-\tilde{r}_m\|\leqslant\varepsilon$, where ℓ_m is a problem-dependent constant. Once this constant is obtained or estimated, one has a reliable method to reduce the cost of the matrix–vector product, dynamically changing how inexact it is, while controlling the attainable residual norm; see, also, Giraud, Gratton, and Langou [143], where some improvements of these criteria are given taking into account backward errors. This discussion also explains why criterion (50) failed in some cases in which the constant ℓ_m is less than one, so that specific information from the problem to be solved is needed to obtain a good estimate of ℓ_m. In most practical cases, if (53) is used, it was observed that the convergence curve for the inexact methods is identical to that of the exact method. No deterioration of convergence is noticed, although experiments with highly sensitive matrices or special right-hand sides, where convergence is in fact delayed have been reported; cf., e.g. [305]. For applications of these ideas to specific problems, see, e.g. [84, 119, 219, 282, 292, 306–308, 344].

When one uses a fixed matrix A, but a variable preconditioner, say such that $M_k^{-1}=M^{-1}+F_k$, then, as long as one can monitor F_k (or F_kv_k) we are in the situation described above since $AM_k^{-1}=AM^{-1}+AF_k$, and AF_k is the error matrix for this case. It follows from this observation that some flexible methods with variable preconditioning described in Section 10 can be recast in terms of inexact matrix–vector products. One of the distinctions is that here we need to monitor the inexactness so as to maintain the preconditioner effectiveness. We remark, however, that, as a key feature of flexible methods, one obtains the approximation directly from the preconditioned basis $z_1,\ldots,z_m$ as in (48). In particular, only the preconditioner is applied inexactly, and not the coefficient matrix, so that the true (unpreconditioned) residual can be evaluated exactly; see further the discussion in [299].

We mention that several authors have studied different aspects of the use of inexact matrix–vector multiplication in iterative methods, sometimes in the context of small perturbations, and in some other instances allowing for large tolerances (though not letting them grow); see, e.g. [149, 152, 153, 317]. Golub and Overton [149] were possibly the first to observe that the Conjugate Gradient method was robust in the sense that convergence can be achieved with inexact matrix–vector products, and in particular using inexact preconditioning. Golub and Ye [153] developed inner stopping criteria of the form (46) so as to maintain superlinear convergence of CG. This was improved by Notay [245]. These authors insisted on a small tolerance, and obtained a small 'delay' in the convergence, as compared to the exact matrix–vector product case; see also Reference [300], where an explanation is given for why this small delay is possible. Here

DOI: 10.1002/nla

we allow larger tolerances, but we lose the three-term recurrence of CG, and instead need the full orthogonalization of the new vector $(A + E_k)v_k$ with respect to all previous ones; cf. (51). In a sense it seems that for CG there is a choice: either one allows for small tolerances, keeping the three-term recurrence, and loosing the orthogonality of the basis, ore one allows for larger tolerances in the matrix–vector product, but as a consequence, one needs more storage for the orthogonal basis.

We conclude by mentioning that the general analytic model of convergence given in [301], and already mentioned in Section 6, applies to inexact Krylov methods as well, providing another explanation on why and how these methods work: at the beginning of the computations one needs the matrix–vector product to be more exact, and once the appropriate invariant subspace is well approximated, one can have more inexact matrix–vector products.

12. GENERAL COMPLEX MATRICES

Most methods described in previous sections can be employed with no change in case the linear system is complex, with the use of the complex inner product; this is done, for instance, by Chatfield *et al.* [67], Rusch [268], Joly and Meurant [194]. Note that often the methods and their implementation are naturally presented in complex arithmetic; see, e.g. Freund and Nachtigal [135], Gutknecht [167, 169, 170]. We refer to Saad [273, Section 6.5.9] for details on the implementation of GMRES.

In Jacobs [188] the idea of using BiCG with the conjugate initial residual as auxiliary vector is explored, in a way that anticipates its use for complex symmetric matrices (see Section 13.3), although this specific structure is not mentioned; see also [279] for a similar approach and [224] for an analysis of the functional associated with the CG method in complex arithmetic.

Complex matrices have in general complex eigenvalues, possibly with no specific pattern or axial symmetry. In particular, if the spectrum largely surrounds the origin, Krylov subspace methods may be very slow. In this case, if no preconditioning is available, one may want to solve the equivalent Hermitian system of normal equations $A^*Ax = A^*b$, by means of an implementation of CG in complex arithmetic. Convergence in this case depends on the spectral properties of A^*A, instead of those of A; see also the discussion in Freund [124], Freund, Golub, and Nachtigal [132].

An alternative to the use of complex arithmetic is to work with the real form of the problem $Ax = b$. Let $A = R + \mathrm{i}S$, with $R = (A + A^*)/2$ and $S = (A - A^*)/(2\mathrm{i})$. Then one can write the original system as

$$\begin{bmatrix} R & -S \\ S & R \end{bmatrix} \begin{bmatrix} \mathrm{Re}\, x \\ \mathrm{Im}\, x \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\, b \\ \mathrm{Im}\, b \end{bmatrix}. \tag{54}$$

Other mathematically equivalent forms may be devised. This one is particularly interesting because the coefficient matrix is real and has eigenvalues which are symmetric with respect to the real axis. Moreover, if R is positive definite, the eigenvalues lie on a half-plane, a property that is well suited for Krylov subspace methods. For general R, however, the system in (54) may not be easier to solve than the original system. When the original matrix A has some convenient structure, say, e.g. banded form, then one may consider reordering the rows and columns of the coefficient matrix

DOI: 10.1002/nla

in (54) so as to maintain that structure. We refer to [80] for a thorough discussion on different real formulations and on their use within preconditioned Krylov subspace methods. The choice between the complex and real formulations is thus problem dependent, with the availability of a good preconditioner also playing a role; see, e.g. Campobasso [60], Natarajan [242], and Sarkar, Yang, and Arvas [280] for some numerical experience on application problems. Note that in some of these cases the coefficient matrix is in fact complex symmetric, for which specific strategies may be adopted; see Section 13.3.

13. SYSTEMS WITH SPECIAL PROPERTIES I: EXPLOITING OTHER FORMS OF SYMMETRY

13.1. The role of an inner product

Inner (scalar) products have a fundamental role in devising effective Krylov subspace methods. The optimal procedures discussed in previous sections all rely on minimizing a norm in the underlying inner product, such as the energy norm (A-norm) for the Conjugate Gradients, or the Euclidean norm for MINRES and GMRES. Clearly, minimization strategies may be devised to obtain solutions that are optimal with respect to other inner products, such as $\langle x, y\rangle = x^{\mathrm{T}} M y$, where M is symmetric positive definite and may be either fixed or variable throughout the iteration [27, 108, 168, 176, 281]. This latter case occurs, for instance, in QMR, where the matrix M varies as the iteration proceeds [92]; other typical examples include $M = A^{\mathrm{T}}A$ or $M = AA^{\mathrm{T}}$, when A is non-symmetric. In [108], a diagonal matrix M is used in the context of restarted methods (see Section 8); the vector of diagonal entries is chosen proportional to the magnitude of residual entries at restart. Suitable choices of M allow one to devise methods that minimize the error even in the non-symmetric case, by possibly employing Krylov subspaces other than $\mathscr{K}_m(A, r_0)$; see, e.g. Section 5. Interestingly, by playing with various inner products it is possible to restate the given minimization problem in a seemingly more accessible way; see, e.g. Gutknecht and Rozložník [176] and references therein. As a simple example, the minimization in the A-norm of the error corresponds to the minimization in the A^{-1}-norm of the residual.

A thorough description of various algorithms based on specific M-inner products can be found in Broyden [11], Ashby, Manteuffel, and Saylor [52], where the more general class of M-normal matrices is discussed; see also Section 13.2. In this class also fall M-symmetric (or M-Hermitian) matrices, i.e. M-self-adjoint, defined as the (possibly non-symmetric) matrices A such that $\langle x, Ay\rangle = \langle Ax, y\rangle$, that is, $A^{\mathrm{T}}M = MA$ [145]; see, e.g. Freund and Nachtigal [137] for a particularly convenient exploitation of this property in Krylov subspace methods, and Gutknecht [168] for relations among different methods derived by using different inner products. This viewpoint has found great application in the context of generalized eigenproblems, see, e.g. the early references by Grimes, Lewis, and Simon [166], and Nour-Omid *et al.* [246].

The inner product framework has found an appropriate role in the analysis of preconditioned iterations, where M is the preconditioner, or a quantity related to it; see, e.g. the discussion in Axelsson [13, Section 11.1.2]. In this case, often it can be shown that the employed preconditioned method, e.g. MINRES or CG, is optimal with respect to the inner product defined by the preconditioner [100, 115, 226]. As an example, the following algorithm generalizes the standard Arnoldi recurrence to compute an M^{-1}-orthogonal basis $\{v_1, \ldots, v_m\}$ of the Krylov subspace

DOI: 10.1002/nla

$\mathscr{K}_m(AM^{-1}, v_1)$:

Given v_1, A, m, and M symmetric positive definite
$w_1 = M^{-1}v_1$
$\tau = (w_1^{\mathrm{T}} v_1)^{1/2}$
$v_1 = v_1/\tau$, $w_1 = w_1/\tau$
for $i = 1, \ldots, m$
$\hat{v}_{i+1} = Aw_i$
$\hat{v}_{i+1} = \hat{v}_{i+1} - \sum_{j=1}^{i} (w_j^{\mathrm{T}} v_{i+1}) v_j$
$\hat{w}_{i+1} = M^{-1}\hat{v}_{i+1}$
$\tau = (\hat{w}_{i+1}^{\mathrm{T}} \hat{v}_{i+1})^{1/2}$
$v_{i+1} = \hat{v}_{i+1}/\tau$, $w_{i+1} = \hat{w}_{i+1}/\tau$
end

Note that this algorithm also generates the auxiliary basis $\{w_1, \ldots, w_m\}$ with $w_i = M^{-1}v_i$, $i = 1, \ldots, m$. We point out that the Arnoldi relation (9) is maintained, i.e. we have $(AM^{-1})V_m = V_{m+1}$ $H_{m+1,m}$, but now the matrices V_m and $H_{m+1,m}$ are not the same as if one uses the Euclidean inner product. Other implementations can be devised that use, for instance, left preconditioning, see, e.g. [115, 226, 281].

Particularly appealing in some applications is the case in which, given a matrix A, its symmetric part $M = (A + A^{\mathrm{T}})/2$ is positive definite. In this case, one can see that the 'preconditioned' matrix $M^{-1}A$ is equal to the identity matrix plus a skew-symmetric matrix with respect to the M-inner product. In this way, it can be readily seen that the Arnoldi process on $M^{-1}A$ in the M-inner product simplifies, resulting in a three-term recursion associated with a tridiagonal matrix H_m; this fact has been noticed some years ago by Concus and Golub [70], and Widlund [347], and it has been reconsidered using different motivations; see, e.g. Arioli, Loghin, and Wathen [7], Faber and Manteuffel [111]. We return to this property in Section 13.2. In general, if the non-symmetric matrix A is diagonalizable and has real and positive eigenvalues, then there exists a symmetric and positive definite matrix M such that A is M-symmetric or, in other words, A is similar to a symmetric and positive definite matrix. In general, such an M is difficult to determine, however, in case it can be derived (as in, e.g. [11, 52, 34, 41, 118]), then a CG-type method with an M-inner product may be employed, as discussed above, in spite of A being non-symmetric.

More recently, interest has been extended to the case of symmetric but indefinite M for the solution of linear systems as well as of eigenvalue problems [37, 191, 255, 291]. This situation occurs very often in practice, in problems that are intrinsically indefinite. In particular, it may happen that the given matrix A is non-symmetric, but it is M-symmetric with respect to a symmetric indefinite M, as in the following example:

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

In this case, the bilinear form $\langle x, y \rangle = x^{\mathrm{T}} My$ does not define a proper inner product, as it may happen that $x^{\mathrm{T}} Mx = 0$ for $x \neq 0$ (just take $x^{\mathrm{T}} = [1, 1]$ and M as in the example above). Safeguard strategies have been proposed to overcome this weakness. The structural complexity of indefinite inner products has motivated several theoretical as well as computational studies in the past

DOI: 10.1002/nla

few years, for which the book [145] provides solid foundations. From a computational stand point, exploiting the M-symmetry leads to significant advantages within the two-sided Lanczos recurrence. More precisely, let A be M-symmetric, for some symmetric M. Then, by choosing $\hat{r}_0 = Mr_0$ as auxiliary starting vector (see Section 2.1), one can show that

$$\hat{w}_j = \xi_j M w_j, \quad j = 1, 2, \ldots,$$

where the ξ_j's are suitable scaling factors. Therefore, the left Lanczos basis need not be recursively computed, leading to a simplified recurrence that does not require multiplications by A^{T}. Freund and Nachtigal [137] present a detailed implementation of this simplified procedure. This approach has found large application in preconditioning of symmetric but indefinite matrices, when the preconditioner is itself indefinite; see, e.g. [291] and references therein. In this case, preconditioned versions of CG or MINRES cannot be applied, and the *simplified* Lanczos or QMR procedures are a particularly appealing alternative to long-term recurrence schemes or to truncated methods. On the other hand, it is shown in [267] that BiCG on a symmetric saddle point problem with a structured indefinite preconditioner is equivalent to preconditioned CG, and safeguard strategies are proposed in [267] to avoid possible misconvergence of the indefinite CG iteration.

We conclude this section devoted to the use of different inner products with pointers to some papers that attempted to significantly limit the use of *any* inner product in Conjugate Gradient-type methods. These efforts aim to go beyond the use of semiiterative methods such as Richardson or Chebyshev methods, which make no use of inner products and are therefore amenable to effective parallel/vector implementations; see Saad [273]. In particular, Fischer and Freund [116] present a polynomial method that approximates the behaviour of the Conjugate Gradient residual polynomial with information generated with a few Krylov subspace iterations. The limited numerical experience shows a similarity in the convergence behaviour of the two polynomials, although, as expected, the new residual polynomial seems unable to capture the superlinear convergence of the Krylov subspace solver.

13.2. Normal and B-normal matrices

Normality and B-normality are more subtle properties that make it possible to generalize the discussion in Section 13.1 to certain matrices that have a less simple structure. Nevertheless, this structure can be exploited to derive short-term recurrences in Krylov subspace methods. The key property of normal matrices in this context is that if their eigenvalues lie on a line in the complex plane, the Arnoldi recurrence simplifies, so that an optimal three-term recursion results. This fact was amply discussed by Faber and Manteuffel [110] and further analysed by Liesen and Saylor [212].

For more general normal matrices, that is, for those normal matrices whose eigenvalues do not lie on a line, Huhtanen [186] more recently proposed an optimal (residual norm minimizing) three-term recurrence based on the Lanczos method. The strategy exploits the fact that when A is normal, the decomposition $A = R + \mathrm{i}S$ with $R = (A + A^*)/2$ and $S = (A - A^*)/(2\mathrm{i})$ satisfies $AR = RA$, that is, A and R commute. Therefore, it is suggested to replace A with the Hermitian matrix R in the generation of the Krylov subspace, and thus determine the solution as

$$\min_{x_m \in \mathcal{K}_m(R,b)} \|b - Ax_m\| = \min_{z_m \in \mathcal{K}_m(R,Ab)} \|b - z_m\|.$$

The approximate solution x_m can be recovered from z_m by a strategy that closely resembles a flexible technique; see Section 10. The algorithm proposed in [186] can be easily implemented and

DOI: 10.1002/nla

the shown numerical results confirm the efficiency of the approach on generic normal matrices, as compared to methods that do not take normality into account. It should be mentioned, however, that the algorithm might need restarting in some specific cases; restarting strategies are also discussed in [186]. Refined schemes can be obtained when further conditions are imposed to the spectrum location, as described by Fassbender and Ikramov [112].

The special case of normal matrices in the form of scaled and shifted orthogonal matrices, has received distinct attention, and specific implementations for large linear systems have been proposed; see Jagels and Reichel [189, 190] and the references therein.

Finally, a theoretically richer class is that of B-normal(s) matrices. A matrix A is said to be B-normal(s) if there exists an Hermitian and positive definite matrix B such that $B^{-1}A^*B$ is a polynomial of A of degree s. For instance, B-Hermitian matrices are B-normal(1), since from $A^*B = BA$ it follows $B^{-1}A^*B = A$. B-normal(s) matrices are interesting because they allow the construction of Krylov subspace methods with $s+2$-term recurrence. However, they are not so interesting in practice in the context of iterative methods, since it was recently shown in [212] that either the associated iterative scheme converges very quickly, or for s large, the occurrence of a short-term recurrence is lost.

13.3. Complex symmetric matrices

A square complex matrix is said to be symmetric if $A = A^{\mathrm{T}}$, that is, A is equal to its transpose, with no conjugation. This class of matrices arises in a variety of applications, such as electromagnetics and acoustics. For this reason methods have been devised specifically tailored to exploit this particular structure and they have proven to be superior to methods that disregard this property. Joly and Meurant [194] combine the normal equations with a special definite inner product. In some particular cases, their approach reduces to one of the algorithms discussed below.

A special bilinear form has become quite popular when solving complex symmetric systems, and this is given by

$$[[x, y]] := x^{\mathrm{T}}y, \quad x, y, \in \mathbb{C}^n. \tag{55}$$

This is clearly not a definite inner product, as it may happen that $[[x, x]] = 0$ for $x \neq 0$; take, for instance, the vector $x^{\mathrm{T}} = [1, \mathrm{i}]$ where i is the imaginary unit. Such non-zero vectors are called isotropic [184]. Successful implementations of short-term recurrences employing the inner product (55) have been proposed in the eigenvalue context already by Cullum and Willoughby [75]. The Conjugate Gradient method can be formally implemented with the usual Euclidean inner product replaced by (55). The short-term recurrence generates a basis for the same Krylov subspace as with the Euclidean inner product. However, the computed basis is different, since orthogonality properties hold with respect to the indefinite inner product (55). This approach is called COCG in van der Vorst and Melissen [334]; see also [74, 75]. The method may break down if some basis vector is isotropic; however, this is not the case if A has n distinct eigenvalues; see Freund [124]. It is noticed in [334] that COCG often behaves like BiCG, in terms of number of iterations, but with half the computational cost. In fact, this is due to the interesting fact that if one starts the usual (Euclidean) BiCG iteration with auxiliary vector $\hat{r}_0 = \bar{r}_0$ (i.e. with the conjugate of the starting residual), then indeed the 'right' recurrence is the same as that of COCG, while the 'left' recurrence is redundant since the constructed matrix satisfies $\widehat{W}_m = \overline{W}_m$. In terms of the two-sided Lanczos procedure, we have the following two relations:

$$AW_m = W_{m+1,m}T_{m+1,m}, \quad A^*\overline{W}_m = \overline{W}_{m+1}T^*_{m,m+1},$$

DOI: 10.1002/nla

which are clearly the conjugate of each other for $A = A^{\mathrm{T}}$. This fact was fully exploited by Freund in [124], where a QMR procedure (quasi-minimum residual norm in the Euclidean inner product) applied to this simplified Lanczos recurrence is proposed and tested. In practice, the redundant auxiliary recurrence is not constructed, and the indefinite inner product (55) is used throughout; we refer to Boyse and Seidl [40], and Simoncini and Gallopoulos [294], as well as to the references therein for a more complete discussion on using (55) with complex symmetric matrices.

A particular variant can be implemented when the complex symmetric matrix has the form $A = M + \sigma I$, with M real symmetric and $\sigma \in \mathbb{C}$. Due to the shift-invariance property of Krylov subspaces, i.e. it holds that $\mathscr{K}_m(M, b) = \mathscr{K}_m(M + \sigma I, b)$, the space generated by A is the same as that generated by M. Therefore, if the starting residual is real, a less computationally intensive real basis for the approximation space can be generated, while complex arithmetic needs be employed only to build the approximate solution; see Freund [123] for an implementation.

A way around the indefinite inner product that still exploits the symmetric form was proposed in Bunse-Gerstner and Stöver [55], where a Lanczos-type recurrence is devised by using the basis W_m and its conjugate in the same recurrence, yielding the following relation:

$$A W_m = \bar{W}_m T_m + \bar{w}_{m+1} t_{m+1,m} e_m^{\mathrm{T}}.$$

The method works with the Euclidean inner product, and is a special implementation of the coupled recurrence method proposed by Saunders, Simon, and Yip [283]. The approach relies on the general full unitary transformation $T = W^{\mathrm{T}} A W$, with $W^* W = I_n$ and T tridiagonal complex symmetric. The Lanczos-type recurrence that partially completes the unitary transformation thus generates a matrix W_m with orthonormal columns. This allows the scheme to determine an approximate solution by an (optimal) residual minimization, together with a short-term recurrence. The resulting method is called CSYM in [55]. Convergence properties are also derived in the same reference, showing that the behaviour of the method depends on the extreme *singular values* of A. In several of the experiments presented in [55], CSYM performs very well compared to QMR and to CG applied to the system of normal equations; see Section 12.

A more 'classical' approach consists of resorting to the double size real form. Writing $A = R + \mathrm{i}S$ with R, S real and symmetric, the complex symmetric system $Ax = b$ may be written, e.g. as

$$\begin{bmatrix} R & S \\ S & -R \end{bmatrix} \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad x = x_1 + \mathrm{i}x_2, \quad b = b_1 + \mathrm{i}b_2.$$

Other formulations are possible (see Section 12, and in particular (54)), by appropriately changing the block signs; see Axelsson and Kucherov [16], and Freund [123]. In the formulation above, the new coefficient matrix is doubled in size, but it is real symmetric. In [123] it is shown that in general and without preconditioning, Krylov subspace methods are not efficient when applied to this formulation. However, in [16] the authors propose the following parameter-dependent formulation

$$\begin{bmatrix} R - \alpha S & \sqrt{1+\alpha^2}\, S \\ \sqrt{1+\alpha^2}\, S & -R - \alpha S \end{bmatrix} \begin{bmatrix} x_1 \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ d \end{bmatrix},$$

where α is a real non-zero parameter, and z, d are real vectors associated with the solution and the right-hand side of the original system; see the references cited in [16], and also [206]. Here it is

DOI: 10.1002/nla

assumed that α is such that the real symmetric matrix $R+\alpha S$ is non-singular. Under the assumption that R and S are semidefinite and that one of them is positive definite, the authors show that the Schur complement system associated with the system above can be solved efficiently by using $R+\alpha S$ as preconditioner. We refer to [16] for some considerations on the practical implementation of this idea and for a detailed spectral analysis. A performance comparison with other approaches is carried out with shifted systems (i.e. $S=\sigma I$, $\sigma\in\mathbb{R}$) stemming from real applications. In Section 14.1 we discuss more specific methods for shifted linear systems.

Preconditioning complex symmetric matrices is a very active area of research and no guidelines really exist up to today. Although we do not address this important issue in this context, we would like to remark that in some cases, the selection of the iterative method depends on the effectiveness of available preconditioners. Due to the lack of a wide variety of robust preconditioners that maintain the complex symmetry of the problem, practitioners often prefer to either employ non-symmetric preconditioners, thus destroying the problem structure, or to handle the real form of the linear system. To give standard preconditioners a chance, however, we close by noticing that a mere 'complexification' of an incomplete Cholesky-type preconditioner may sometimes be very effective; see [297] for a practical application.

14. SYSTEMS WITH SPECIAL PROPERTIES II: PARAMETRIZED SYSTEMS

14.1. Shifted systems

In this section, we discuss the solution of systems that depend linearly on a parameter σ. We start with the shifted form $(A-\sigma I)x=b$ and then generalize it to $(\mathscr{A}-\sigma\mathscr{B})x=b$.

In several applications the following non-singular system needs to be solved

$$(A-\sigma I)x=b,\quad x=x(\sigma), \tag{56}$$

for many tabulated values of the parameter σ, while the matrix A and the right-hand side b remain fixed. If σ is very close to zero, then $A-\sigma I$ may be viewed as a special perturbation of A, so that we expect x not to be too far from $A^{-1}b$, depending on the sensitivity of the coefficient matrix A. In general, however, σ varies in a possibly wide range, so that (56) needs to be solved explicitly.

Krylov subspace methods are particularly appealing for these problems because of the shift-invariance property of Krylov subspaces; see Section 13.3. This allows one to obtain approximate solutions for all values of the parameter, by generating a single approximation space. The Arnoldi relation (10) can be rewritten as

$$(A-\sigma I)V_m=V_m(H_m-\sigma I)+h_{m+1,m}v_{m+1}e_m^{\mathrm{T}}, \tag{57}$$

and we emphasize that the matrices V_m and H_m are the same as in (10) and do not depend on σ. Note that if A and b are real and σ is complex, then the matrix V_{m+1} and the upper Hessenberg representation matrix $H_{m+1,m}$ are both real. Moreover, if A is real symmetric, then $A-\sigma I$ is real or complex symmetric, so that short-term recurrences apply; in particular, for $\sigma\in\mathbb{C}$, $A-\sigma I$ is a normal matrix.

Analogously, the Lanczos relation (14) becomes

$$(A-\sigma I)W_m=W_m(T_m-\sigma I)+t_{m+1,m}w_{m+1}e_m^{\mathrm{T}}. \tag{58}$$

The rectangular matrices $H_{m+1,m}$ and $T_{m+1,m}$ are defined accordingly, following (11) and (13).

DOI: 10.1002/nla

All methods discussed in the previous sections that are based on the Arnoldi or Lanczos relations can be generalized to handle the shifted forms (57) or (58). This was first proposed by Datta and Saad [78], where a minimal residual method for each shifted system was obtained by minimizing the residual $r_m^{(\sigma)} = b - (A - \sigma I)V_m y^{(\sigma)}$. For each parameter σ, this corresponds to solving (cf. also Section 2.2)

$$\min_{y \in \mathbb{R}^m} \|e_1 \beta_0 - (H_{m+1,m} - \sigma \tilde{I})y\|, \quad \tilde{I} = \begin{bmatrix} I_m \\ 0^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}.$$

Clearly, the computationally demanding step of generating the basis vectors is carried out only once for all parameters of interest, possibly leading to great computational savings as compared to solving each shifted system separately. Moreover, as a function of the number of parameters, memory requirements only depend on m.

In the case of the two-sided Lanczos process, or when A is symmetric, the banded structure of T_m can be used to derive a short-term recurrence as in the unshifted case. However, the direction vectors depend on the parameter σ in a way that we discuss next. The general approach was first derived by Freund in [123] for CG and in [126] for TFQMR. For instance, for the QMR method, let $Q_{m+1}^{(\sigma)} R_{m+1,m}^{(\sigma)}$ be the QR factorization of the shifted Lanczos matrix $T_{m+1,m} - \sigma \tilde{I}$. Since $T_{m+1,m}$ is tridiagonal, then $R_m^{(\sigma)} = [I_m, 0] R_{m+1,m}^{(\sigma)}$ is upper triangular and banded with three non-zero diagonals, i.e. semibandwidth 2. Then, assuming $x_0 = 0$, the QMR solution can be written as follows (cf. also Sections 2.2 and 4). Set $t_m = [I_m, 0](Q_{m+1}^{(\sigma)})^{\mathrm{T}} e_1 \beta_0$ so that $x_m^{(\sigma)} = V_m (R_m^{(\sigma)})^{-1} t_m$. As discussed in the previous sections, the banded structure of $R_m^{(\sigma)}$ allows one to derive a short-term recurrence for $x_m^{(\sigma)}$. On the other hand, this procedure forces to store the last few columns of $P_m^{(\sigma)} := V_m (R_m^{(\sigma)})^{-1}$, which clearly depend on σ. Therefore, as many replications of these vectors are required as the number of available parameters. Storage savings may be obtained if only some of the solution components are required. In this case, only the corresponding components in the vectors in $P_m^{(\sigma)}$ need be carried around. Depending on the number of parameters, it may be more convenient not to form $P_m^{(\sigma)}$ and store the matrix V_m instead, or some of its row components, as needed. As a generalization, if $D^{\mathrm{T}} x_m^{(\sigma)}$ is of interest for some tall matrix D, then the small matrix $D^{\mathrm{T}} P_m^{(\sigma)}$ should be stored in place of the last columns of $P_m^{(\sigma)}$.

The sensitivity of the symmetric positive definite matrix A and the role of the shift have also been analysed. In particular, Sleijpen and van den Eshof [304] study this in connection with the solution of shifted systems arising in Tikhonov regularization procedures; see, also, Frommer and Maass [141].

A short-term recurrence alternative to Lanczos-type methods in the non-symmetric case has been recently proposed by Frommer [139], where the Bi-CGstab(ℓ) method is adapted to handle several shifts simultaneously.

If a method based on the Arnoldi relation is used in the non-symmetric case, the shifted form still suffers from the known memory limitations associated with keeping the whole basis V_m. Restarting or truncating are thus required; see Section 8. To efficiently exploit Krylov subspace invariance after restarting, all residual vectors of the unconverged systems should be collinear. If this is not the case, the corresponding subspaces are not equal and the method proceeds with completely uncoupled processes for each shift after the first restart. Unfortunately, the residuals obtained with GMRES are not collinear, so that the original method cannot be efficiently restarted to solve the

DOI: 10.1002/nla

systems for all parameters simultaneously. An attempt to overcome this problem was presented by Frommer and Glassner [140]. They propose to force collinearity of the GMRES residuals by relaxing the minimization constraint for all but one system, called the seed system, corresponding to the zero shift. In fact, the seed system does not need to be associated with the zero shift, but any of the shifted systems could be considered, after an appropriate parameter translation; see Feriani, Perotti, and Simoncini [114]. Assuming that the starting residual r_0 is the same for all shifted systems, in [140] the non-seed residual $r_m = \hat{p}(A - \sigma I)r_0$ is obtained, with $\hat{p}(\lambda) = p(\lambda + \sigma)/p(\sigma)$, where p is the optimal residual polynomial associated with the seed system; a practical implementation is discussed in [140]. It is important to notice that $\hat{p}(\lambda)$ is defined only if $p(\sigma) \neq 0$, and that the method is very sensitive to the actual magnitude of $p(\sigma)$. It is shown in [140] that convergence of all shifted systems can be obtained with the restarted procedure if $A - \sigma I$ is positive real.

A way to completely overcome the collinearity problem in Arnoldi-based algorithms is to resort to the restarted version of FOM (see Section 2.5). Indeed, since in FOM the residual vector is a multiple of v_{m+1} (cf. (32)), the residuals of all shifted systems are naturally collinear. The FOM method can thus be restarted with the same new direction vector for all unconverged shifted systems. More precisely, $r_k^{(\sigma)} = \beta_k^{(\sigma)} v_{k+1}$ for some scalar $\beta_k^{(\sigma)}$. By taking $\hat{v}_1 = r_k^{(\sigma)}/\beta_k^{(\sigma)}$ $(= \pm v_{k+1})$ at restart, for one of the σ's, the new Krylov subspace $\mathscr{K}_m(A, \hat{v}_1)$ can be built. We denote by $\hat{V}_m$ the matrix whose columns span the new subspace, so that the new approximate solutions can be written as $x_m^{(\sigma)} \leftarrow x_m^{(\sigma)} + \hat{V}_m y_m^{(\sigma)}$, with

$$y_m^{(\sigma)} = (\hat{H}_m - \sigma I)^{-1} e_1 \beta_m^{(\sigma)}, \quad \hat{H}_m = \hat{V}_m^{\mathrm{T}} A \hat{V}_m.$$

Implementation details as well as numerical comparisons can be found in [290].

A similar situation arises if one wishes to solve the following parametrized system

$$(\mathscr{A} - \sigma \mathscr{B})z = f, \quad z = z(\sigma), \tag{59}$$

for several values of σ. System (59) arises in a variety of applications, and also in shift-and-invert approaches to solve eigenvalue problems [20]. If $\mathscr{A}$ or $\mathscr{B}$ is non-singular, the problem can be restated as a shifted system. For instance, assuming $\mathscr{B}$ non-singular, then

$$(\mathscr{A}\mathscr{B}^{-1} - \sigma I)\hat{z} = f, \quad z = \mathscr{B}^{-1}\hat{z}, \tag{60}$$

to which the methods described earlier in the section can be applied. In particular, if $\mathscr{A}$ and $\mathscr{B}$ are real symmetric and $\mathscr{B}$ is also positive definite, then the coefficient matrix in (60) is symmetric with respect to the $\mathscr{B}$-inner product, so that a short-term recurrence method can effectively be applied, see Section 13.1. Various aspects related to the application of the Lanczos and MINRES methods to (60) with the $\mathscr{B}$-inner product have been recently analysed by Meerbergen [226]. If $\mathscr{B}$ is not positive definite, but $\mathscr{A}\mathscr{B}^{-1}$ is J-symmetric for some matrix J, then this can be exploited as described in Section 13.2. For instance, if $\mathscr{A}$ is symmetric, then $\mathscr{A}\mathscr{B}^{-1}$ is $\mathscr{B}$-symmetric and the theory applies.

A less restrictive approach consists of applying $(\mathscr{A} - \tau\mathscr{B})^{-1}$, for a judiciously chosen τ, that is,

$$(\mathscr{A} - \tau\mathscr{B})^{-1}(\mathscr{A} - \sigma\mathscr{B})z = (\mathscr{A} - \tau\mathscr{B})^{-1} f.$$

Clearly, τ should be chosen so that $(\mathscr{A} - \tau\mathscr{B})$ is non-singular, and also so as to improve the spectral properties of the resulting matrix; in the eigenvalue literature, this is known as the Cayley

DOI: 10.1002/nla

transformation [20, 151], and it acts on the original system as a preconditioner. The case of symmetric $\mathscr{A}$ and $\mathscr{B}$ has been analysed in [226], where a comparison of Lanczos and MINRES methods is performed, together with a detailed analysis of error estimations in finite precision arithmetic. Unless other special properties of the coefficient matrices can be taken into account, the preconditioned system can be solved by one of the Krylov subspace methods mentioned in the previous sections, keeping in mind that for large problems, the matrix $(\mathscr{A} - \tau\mathscr{B})$ should not be inverted explicitly; see also Section 11.

The preconditioner just described may be viewed as a convenient strategy that allows one to still exploit the shifted structure of the problem. Similar properties characterize more general polynomial preconditioners; see Freund [123]. On the other hand, standard preconditioning approaches may be more effective on each single shifted system, but unfortunately they destroy the shifted structure so that the convenient invariance property of the Krylov subspace can no longer be employed. At this time, determining good preconditioners for shifted systems that preserve the original structure of the problem is an open area of research.

A related but different problem is the solution of linear systems of the form

$$Ax = (A - \gamma I)v, \tag{61}$$

where γ is a non-zero complex parameter. The system solution can also be obtained as $x = v - \gamma A^{-1}v$, therefore the question arises whether it is more convenient to explicitly perform the matrix–vector product in the right-hand side of (61), or to solve $Ad = v$ first and then update x. For the symmetric case, in [296] a theoretical analysis is presented that shows the advantages of solving $Ad = v$ first. Numerical results where the right-hand side is a second degree polynomial in A are also reported.

14.2. Systems depending quadratically on a parameter

A natural generalization of the shifted problem arises in the situation where the coefficient matrix depends non-linearly on the parameter σ, and the system needs to be solved for several values of the parameter. The problem has been addressed in the literature in the case of quadratic dependence, mostly focusing on the symmetric case, whereas the problem is far from being solved in the more general case; we comment on this at the end of this section.

We consider the following linear system,

$$(\sigma^2 A + \sigma B + C)x = b, \quad x = x(\sigma). \tag{62}$$

We assume $B \neq 0$, otherwise the problem can be dealt with as in (59) by first renaming σ^2; we also assume that either A or C is non-singular. The coefficient matrix is a special case of *matrix polynomial* in σ, and more precisely a second degree matrix polynomial; we refer to [146] for a thorough analysis of its spectral properties. We assume throughout that σ is such that the coefficient matrix is non-singular. In the following we shall review two general strategies for solving system (62), which have also been employed in the eigenvalue context; see, e.g. Tisseur and Meerbergen [324]. A natural way to proceed consists of linearizing the equation in terms of σ. There are several different ways to realize this, leading to differently structured problems. For instance, (62) is equivalent to a larger system $(\mathscr{A} - \sigma\mathscr{B})z = f$, given by

$$\left(\begin{bmatrix} B & C \\ M & 0 \end{bmatrix} - \sigma \begin{bmatrix} -A & 0 \\ 0 & M \end{bmatrix} \right) \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \tag{63}$$

DOI: 10.1002/nla

for any non-singular matrix M. Note that $\sigma x = y$ and that the dependence on σ in $(\mathscr{A} - \sigma\mathscr{B})z = f$ is now linear. If B is symmetric, whenever C is non-singular and symmetric, the choice $M = C$ makes the matrix $\mathscr{A}$ symmetric. In the more general case of singular C, however, the linearized form above can still be employed in some cases. Assume that C is symmetric and positive semidefinite, as is the case in certain structural dynamics applications, and let $C = C_1 C_1^{\mathrm{T}}$, with C_1 full column-rank. Then we can write

$$\left(\begin{bmatrix} B & C_1 C_1^{\mathrm{T}} \\ C_1 C_1^{\mathrm{T}} & 0 \end{bmatrix} - \sigma \begin{bmatrix} -A & 0 \\ 0 & C_1 C_1^{\mathrm{T}} \end{bmatrix}\right) \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

$$\left(\begin{bmatrix} B & C_1 \\ C_1^{\mathrm{T}} & 0 \end{bmatrix} - \sigma \begin{bmatrix} -A & 0 \\ 0 & I \end{bmatrix}\right) \begin{bmatrix} y \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \quad \hat{x} = C_1^{\mathrm{T}} x,$$

and the vector $\sigma^{-1} y$ is the unique solution to the linear system (62) [287]. Analogously, if C is non-symmetric and it is possible to factorize C as $C = C_1 C_2^{\mathrm{T}}$ with C_1, C_2 full rank, then

$$\left(\begin{bmatrix} B & C_1 \\ C_2^{\mathrm{T}} & 0 \end{bmatrix} - \sigma \begin{bmatrix} -A & 0 \\ 0 & I \end{bmatrix}\right) \begin{bmatrix} y \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \quad \hat{x} = C_2^{\mathrm{T}} x,$$

and the vector $\sigma^{-1} y$ is the sought after solution.

To simplify the presentation, in the following we shall refer to (63) with C non-singular, keeping in mind that one can work with the 'reduced' form above if C is singular. The system $(\mathscr{A} - \sigma\mathscr{B})z = f$ can be solved as described in the previous section. If A and C are symmetric, by choosing $M = C$ we find that $\mathscr{A}\mathscr{B}^{-1}$ is $\mathscr{B}$-symmetric, and thus a simplified version of the Lanczos process can be derived; see Section 13.2. These aspects, together with several implementation issues associated with the special form of $\mathscr{A}$ and $\mathscr{B}$ are described in [297].

An alternative approach that is being currently investigated, mostly in the eigenvalue setting, attempts to avoid linearization. The reason for this is that linearization doubles the problem size, and thus the dimension of all recurrence vectors. The general idea consists of projecting the original problem onto a subspace of much smaller dimension, say m, imposing a Galerkin condition on the residual; see similar strategies in the eigenvalue setting in [20]. Let the columns of Y_m span such subspace. Then an approximation $x_m = Y_m y_m$ to the solution x is found by solving the reduced problem

$$Y_m^{\mathrm{T}}(\sigma^2 A + \sigma B + C) Y_m y_m = Y_m^{\mathrm{T}} b,$$

or equivalently, setting $b_m = Y_m^{\mathrm{T}} b$, $A_m = Y_m^{\mathrm{T}} A Y_m$, $B_m = Y_m^{\mathrm{T}} B Y_m$, and $C_m = Y_m^{\mathrm{T}} C Y_m$,

$$(\sigma^2 A_m + \sigma B_m + C_m) y_m = b_m.$$

A major issue is the choice of the approximation subspace. Bai and Su [22] present an approach where Y_m is obtained by efficiently compressing the relevant information of the Krylov subspace obtained with the linearized problem. The approach seems to work well in the context of model reduction of second-order dynamical systems, where the transfer function $h(\sigma) = d^{\mathrm{T}}(\sigma^2 A + \sigma B + C)^{-1} b$ needs to be approximated, for values of σ close to a fixed target. The authors report better numerical performance, in terms of accuracy and memory requirements, than

DOI: 10.1002/nla

with the linearized case in the general non-symmetric setting. Several issues are still open, such as restarting, to make the approach appealing on large size problems; see, also, Freund [129, 130].

15. STOPPING CRITERIA

In the discussion of the previous sections little has been said about when to stop the iterations. A good account of various stopping criteria used in practice may be found in [26], among which are

1. $\|r_m\| \leqslant \varepsilon(\|A\|\,\|x_m\| + \|b\|)$,
2. $\|r_m\| \leqslant \varepsilon\|b\|$,
3. $\|r_m\| \leqslant \varepsilon\|r_0\|$.

The quantity ε is commonly a user supplied parameter. All three tests involve the residual, which is usually available during the iteration, by means of a vector recurrence or by direct computation. The ideal quantity to be monitored, the error $x_* - x_m$ is clearly not available, since the exact solution x_* is unknown; however, see below for *estimates* of the error norm. If information is available on $\|A^{-1}\|$, say its order of magnitude, then a good estimate of the error norm may be obtained as $\|A^{-1}\|\,\|r_m\|$; alternatively, one can directly measure the $A^{\mathrm{T}}A$-norm of the error by means of the (computable) residual norm as $\|x_* - x_m\|^2_{A^{\mathrm{T}}A} = \|r_m\|^2$. We refer to [10, 11, 52], for further discussion on this equivalence and for estimates of $\|A^{-1}\|$ within the Krylov subspace iteration.

While the third criterion is commonly employed, it may be considered the less reliable one, as it may dramatically depend on the initial approximation x_0. The first criterion takes into account possible ill-conditioning of the problem, manifesting itself with the bound $\|A\|\,\|x_*\| \geqslant \|b\|$, and it is therefore less strict than the second criterion, which assumes $\|A\|\,\|x_*\| \approx \|b\|$. In fact, this latter criterion may be hard to satisfy in case of ill-conditioned problems. Moreover, the first criterion is the one closest to the actual convergence properties of methods such as GMRES. Indeed, it was shown by Dřkosová *et al.* [90] that a robust implementation of GMRES, using for instance Householder Gram–Schmidt orthogonalization, is backward stable in the sense that given the final approximation x_n, the backward error $\|b - Ax_n\|/(\|A\|\,\|x_n\| + \|b\|)$ is at worst $O(n^{5/2})\varepsilon$, where ε is the machine precision. A similar result was very recently proven for the *modified Gram–Schmidt* implementation of GMRES by Paige, Rozložník, and Strakoš [249], thus solving a longstanding open problem. The usefulness of the stopping criteria above in the finite precision setting is discussed by Strakoš and Tichý [320, 321] for the Conjugate Gradients algorithm and its preconditioned form.

If information on the single entries of A are available, a componentwise stopping criterion may be considered, to exploit the problem sparsity; see Arioli, Duff, and Ruiz [6]. We refer to [26] for a more complete discussion of these and other criteria, which have found great interest during the developments of new Krylov subspace methods.

An alternative, very classical stopping criterion relates to the number of iterations. More precisely, in some cases it is possible to give sufficient conditions on the minimum number of iterations needed to achieve an error or residual norm that is less than a prescribed tolerance, see, e.g. the discussion by Axelsson [13, Chapter 13]. Although this type of bounds provides a good tool when available, at least in exact arithmetic, the given estimate is often very loose, predicting a much

DOI: 10.1002/nla

larger number of iterations than is actually necessary. Therefore, such estimates should be always accompanied by *a posteriori* estimates such as those described above. Coupled automatic stopping criteria may be particularly appropriate when the iterative solver is embedded in another (possibly non-linear) iterative procedure; see, e.g. Axelsson and Kaporin [15] for such an analysis.

In the rest of this section, we describe some recent developments on

(a) Monitoring/estimating quantities that are possibly minimized by the method;
(b) Monitoring/estimating the quantity of interest to the discretized problem.

As for point (a), methods that (quasi-)minimize the residual norm, such as GMRES and the smoothing techniques of Section 4, naturally employ the residual norm in their stopping criterion, following one of the tests cited above. The case of CG for A symmetric and positive definite has received a lot of attention, because of its wide applicability, especially in problems associated with the discretization of partial differential equations. As we already discussed, CG minimizes the A-norm of the error at each iteration, therefore it would be important to at least estimate this quantity. It turns out that several authors have tried to address this problem. Identities relating the error A-norm and 2-norm with quantities available in the algorithm, were already present in the original paper by Hestenes and Stiefel [180], but apparently they have been reconsidered only recently by Strakoš and Tichý [320].

In the meantime, several authors have analysed the problem of approximating the error norm for A symmetric (and positive definite), by using several possibly related approaches, such as Gauss quadrature-type formulas, see, e.g. [59, 77, 147, 150, 227], together with interval arithmetic techniques [142], and purely algebraic approaches [15, 117]; we refer to [320] for a more detailed account of the literature on the subject. In particular, the authors of [320] derive computable bounds from the original results in [180], that also work in finite precision computations. More precisely, if x_m, x_{m+d} are two approximate solution iterates, with $d>0$, then the following relation is exploited,

$$\|x_* - x_m\|_A^2 = \sum_{i=m}^{m+d-1} \gamma_i \|r_i\|^2 + \|x_* - x_{m+d}\|_A^2,$$

where x_* is the exact solution, r_i the computed residual after i iterations, and $\gamma_i = r_i^{\mathrm{T}} r_i / (p_i^{\mathrm{T}} A p_i)$ is a recurrence coefficient computed at each iteration of the standard Conjugate Gradient implementation [151]. If d is an integer such that $\|x_* - x_m\|_A^2 \gg \|x_* - x_{m+d}\|_A^2$, then it is shown in [320] that the square root of

$$\sum_{i=m}^{m+d-1} \gamma_i \|r_i\|^2$$

is a good easily computable estimate, both in exact as well as in finite precision arithmetic, of the error A-norm at step m. The sharpness of the bound clearly depends on the 'delay' index d, since one can estimate the error at step m only after d more iterations are carried out. If used as stopping criterion strategy, this forces to perform d additional iterations than the stopping criterion would suggest. However, one may be willing to pay this price if some confidence in the actual error is obtained. The value of d is chosen *a priori*, and some tuning is clearly necessary.

Point (b) has recently received more attention, possibly due to the described advances in error estimates, and it actually emphasizes one of the major advantages of using iterative compared to direct methods. Arioli [5] shows that the Conjugate Gradient error A-norm and the estimates

DOI: 10.1002/nla

analysed in [320] can naturally and fruitfully be employed in stopping criteria for linear systems stemming from the finite element discretization of self-adjoint elliptic partial differential equations. In particular, in [5] it is shown how to conveniently tune the parameter η in a stopping criterion of the type

$$\|x_* - x_m\|_A \leqslant \eta \|x_*\|_A$$

so as to comply with the accuracy obtained by the finite element approximation, which can usually be expressed in terms of the mesh fineness. More precisely, η should be related to the mesh parameter, with a value that is in practice much less stringent than ε in the residual bounds at the beginning of this section. We also refer to Axelsson and Barker [14, Section 7.5] for a general discussion on the relation between η and the discretization error.

The numerical experiments reported in [5] demonstrate that the algebraic solution obtained with the suggested stopping criterion is capable of achieving the same level of accuracy as in the chosen problem discretization, even when dealing with operators having highly discontinuous coefficients. This property is explored by comparing the algebraic approximate solution with the *exact* solution to the considered elliptic equation. It turns out that residual-based stopping criteria may force the method to perform a large number of useless iterations, during which the approximation error with respect to the exact (continuous) solution does not show any further decrease. In summary, whenever the linear system to be solved stems from a suitable partial differential equation, an appropriately chosen stopping criterion may allow to terminate the computation much earlier than a generic strategy would predict. Moreover, the discussion above substantiates the fact that direct methods determine a machine precision accurate solution to the 'wrong' problem, whose distance from the continuous problem depends on the formulation used and on the approximation of the discretized problem.

More recently, Arioli, Loghin, and Wathen [7] extend the results in [5] to the non-self-adjoint case, by significantly modifying the approach by Arioli, Noulard, and Russo [8]. The authors show that a natural generalization of the A-norm of the error, or equivalently, of the A^{-1}-norm of the residual, is given by the H^{-1}-norm of the residual, where H^{-1} can be either the inverse of the symmetric part of A, or the symmetric part of A^{-1}. These appear to be the *natural* norms for the chosen problem formulation. The authors also derive numerical estimates for these norms, with which good estimates of the final attainable accuracy of the algebraic approximate solution are obtained.

Finally, in [100, section 6.2.1] the authors analyse preconditioning strategies for the Stokes problem in fluid dynamics that allow minimal residual methods to minimize the residual in the appropriate natural matrix norm. The authors show that a good norm should be consistent with the different norms used for measuring the discretization error in the spaces associated with the two different variables, velocity and pressure. In this setting, the stopping criterion can be naturally linked to the magnitude of the employed discretization error.

16. CONCLUSIONS

We have described many ideas used in the development and extensions of Krylov subspace methods. In this final section we make some concluding remarks, and in particular on the current trends in this area.

DOI: 10.1002/nla

There is a good understanding of the convergence of methods for symmetric matrices, such as CG in the positive definite case. In the case of non-symmetric matrices, for certain Arnoldi-based methods, such as full GMRES, there is an increased comprehension of their behaviour. In contrast, there is very little or no theory for the two methods most commonly used by many scientists and engineers: Bi-CGStab and *restarted* GMRES. We stress that the latter does not inherit the convergence properties of the full method, possibly leading to stagnation due to the repeated generation of the same subspace after each restart; see Section 8. This is commonly overlooked by practitioners, who too often completely ignore the fact that other methods may be more effective than *restarted* GMRES, with similar computational costs.

An area of current active research is the development of new and more effective preconditioners, especially for specific types of applications. It is often the situation that when a computational scientist is dissatisfied with the performance of a Krylov subspace solver, a better preconditioner is sought. This situation provides the motivation for the development of more advanced solvers, the deepening of their understanding and reliability, in order to considerably spread their use, and thus encourage the user to try a better (or more appropriate) method first.

Several variants, including flexible, inexact, and augmented methods have shown to be viable alternatives in many applications, and their potential has yet to be widely appreciated. All these methods have the common thread that the subspace being used is not in fact a standard Krylov subspace; see Sections 9–11. In other words, they have lost certain optimality properties, but in exchange they gain in efficiency. The same can be said about methods with indefinite inner products; see Section 13.

Frequently, people in applications do not wait for a complete convergence analysis, or for methods guaranteed to work in a variety of problems, and are ready to take the risk of (often unlikely) breakdown to satisfy their need to solve large-scale problems. Efforts in better understanding these uncommon frameworks are bound to give additional confidence in using a variety of Krylov subspace methods such as those described in this paper, and those yet to be proposed. Scientists who accept the challenge of further analysing the existing methods, enhancing their comprehension and reliability, or proposing new alternatives, will be rewarded by the applied community.

ACKNOWLEDGEMENTS

We appreciate very much the constructive comments and questions of Michael Eiermann, Serge Gratton, Martin Gutknecht, Héctor Klíe, Jörg Liesen, Ronald Morgan, Zdeněk Strakoš, and Petr Tichý, who carefully read earlier versions of this survey. We also thank Yuri Kuznetsov for his help in finding the author's first name in [138]. Work on this paper was supported in part by the U.S. National Science Foundation under grant DMS-0207525, and by the U.S. Department of Energy under grant DE-FG02-05ER25672. The first author was also supported by the PRIN 2004 project of Italian MIUR.

REFERENCES

1. Moneer M. Abu-Elnaga, Mona A. El-Kady, and Raymond D. Findlay. Sparse formulation of the transient energy function method for applications to large-scale power systems. *IEEE Transactions on Power Systems* 1988; **3**:1648–1654.
2. Guillaume Alléon, Bruno Carpentieri, Iain S. Duff, Luc Giraud, Julien Langou, Émeric Martin, and Guillaume Sylvand. Efficient parallel iterative solvers for the solution of large dense linear systems arising from the boundary element method in electromagnetism. *Proceedings of the International Conference on Supercomputing in Nuclear Application* (*SNA*), Paris, September 2003.

DOI: 10.1002/nla

3. Rajeev Alur and George J. Pappas, editors. Hybrid systems: computation and control. *Proceedings of the 7th International Workshop*, *HSCC 2004*, Philadelphia, PA, U.S.A., 25–27 March 2004. Lecture Notes in Computer Science, vol. 2993. Springer: New York, 2004.
4. Athanasios C. Antoulas, Danny C. Sorensen, and Serkan Gugercin. A survey of model reduction methods for large linear systems. *Contemporary Mathematics* 2001; **280**:193–219.
5. Mario Arioli. A stopping criterion for the Conjugate Gradient algorithm in a finite element method framework. *Numerische Mathematik* 2004; **97**:1–24.
6. Mario Arioli, Iain S. Duff, and Daniel Ruiz. Stopping criteria for iterative solvers. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:138–144.
7. Mario Arioli, Daniel Loghin, and Andrew J. Wathen. Stopping criteria in finite element problems. *Numerische Mathematik* 2005; **99**:381–410.
8. Mario Arioli, Eric Noulard, and Alessandro Russo. Stopping criteria for iterative methods: applications to PDEs. *Calcolo* 2001; **38**:97–112.
9. Walter E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics* 1951; **9**:17–29.
10. Steven F. Ashby, Michael J. Holst, Thomas A. Manteuffel, and Paul E. Saylor. The role of the inner product in stopping criteria for Conjugate Gradient iterations. *BIT Numerical Mathematics* 2001; **41**:26–52.
11. Steven F. Ashby, Thomas A. Manteuffel, and Paul E. Saylor. A taxonomy for Conjugate Gradient methods. *SIAM Journal on Numerical Analysis* 1990; **27**:1542–1568.
12. Owe Axelsson. A generalized Conjugate Gradient, least squares method. *Numerische Mathematik* 1987; **51**:209–227.
13. Owe Axelsson. *Iterative Solution Methods*. Cambridge University Press: Cambridge, New York, 1994.
14. Owe Axelsson and Vincent A. Barker. *Finite Element Solution of Boundary Value Problems*. Classics in Applied Mathematics, vol. 35. Academic Press: Orlando, 1984. Reprinted by SIAM: Philadelphia, PA, 2001.
15. Owe Axelsson and Igor Kaporin. Error norm estimation and stopping criteria in preconditioned Conjugate Gradient iterations. *Numerical Linear Algebra with Applications* 2001; **8**:265–286.
16. Owe Axelsson and Andrey Kucherov. Real valued iterative methods for solving complex symmetric linear systems. *Numerical Linear Algebra with Applications* 2000; **7**:197–218.
17. Owe Axelsson and Gunhild Lindskog. On the rate of convergence of the preconditioned Conjugate Gradient method. *Numerische Mathematik* 1986; **48**:499–523.
18. Owe Axelsson and Panayot S. Vassilevski. A black box generalized Conjugate Gradient solver with inner iterations and variable-step preconditioning. *SIAM Journal on Matrix Analysis and Applications* 1991; **12**:625–644.
19. James Baglama, Daniela Calvetti, Gene H. Golub, and Lothar Reichel. Adaptively preconditioned GMRES algorithms. *SIAM Journal on Scientific Computing* 1998; **20**:243–269.
20. Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk A. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems*: *a Practical Guide*. SIAM: Philadelphia, PA, 2000.
21. Zhaojun Bai, Patrick M. Dewilde, and Roland W. Freund. Reduced-order modeling. In *Handbook of Numerical Analysis. Numerical Methods in Electromagnetics*, Schilder WHA, ter Maten JW (eds), vol. XIII. Elsevier: Amsterdam, 2005; 825–895.
22. Zhaojun Bai and Yangfeng Su. Dimension reduction of second-order dynamical systems via a second-order Arnoldi method. *SIAM Journal on Scientific Computing* 2005; **26**:1692–1709.
23. Allison H. Baker, Elizabeth R. Jessup, and Thomas Manteuffel. A technique for accelerating the convergence of restarted GMRES. *SIAM Journal on Matrix Analysis and Applications* 2005; **26**:962–984.
24. Randolph E. Bank and Tony F. Chan. An analysis of the composite step biconjugate gradient method. *Numerische Mathematik* 1993; **66**:293–319.
25. Randolph E. Bank and Tony F. Chan. A composite step biconjugate gradient algorithm for nonsymmetric linear systems. *Numerical Algorithms* 1994; **7**:1–16.
26. Richard Barrett, Michael W. Berry, Tony F. Chan, James Demmel, June Donato, Jack Dongarra, Victor Eijkhout, Roldan Pozo, Charles Romine, and Henk A. van der Vorst. *Templates for the Solution of Linear Systems*: *Building Blocks for Iterative Methods*. SIAM: Philadelphia, PA, 1993.
27. Teri Barth and Thomas Manteuffel. Variable metric Conjugate Gradient methods. In *Advances in Numerical Methods for Large Sparse Sets of Linear Equations*, Matrix Analysis and Parallel Computing, PCG 94, Natori M, Nodera T (eds), vol. 10. Keio University: Yokohama, 1994; 165–188.
28. Bernhard Beckermann, Sergei A. Goreinov, and Eugene E. Tyrtyshnikov. Some remarks on the Elman estimate for GMRES. *SIAM Journal on Matrix Analysis and Applications* 2006; **27**:772–778.

DOI: 10.1002/nla

29. Bernhard Beckermann and Arno B.J. Kuijlaars. Superlinear convergence of Conjugate Gradients. *SIAM Journal on Numerical Analysis* 2001; **39**:300–329.
30. Bernhard Beckermann and Arno B.J. Kuijlaars. Superlinear CG convergence for special right-hand sides. *Electronic Transactions on Numerical Analysis* 2002; **14**:1–19.
31. Peter Benner, Gene H. Golub, Volker Mehrmann, and Danny C. Sorensen, editors. *Dimension Reduction of Large-Scale Systems*. Lecture Notes in Computational Science and Engineering, vol. 25. Springer: Berlin, Heidelberg, 2005.
32. Michele Benzi. Preconditioning techniques for large linear systems: a survey. *Journal of Computational Physics* 2002; **182**:418–477.
33. Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical solution of saddle point problems. *Acta Numerica* 2005; **14**:1–137.
34. Michele Benzi and Valeria Simoncini. On the eigenvalues of a class of saddle point matrices. *Numerische Mathematik* 2006; **103**:173–196.
35. George Biros and Omar Ghattas. Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver. *SIAM Journal on Scientific Computing* 2005; **27**:687–713.
36. Åke Björck. *Numerical Methods for Least Squares Problems*. SIAM: Philadelphia, PA, 1996.
37. Amina Bouras and Valérie Frayssé. A relaxation strategy for inexact matrix–vector products for Krylov methods. *Technical Report 15*, CERFACS, Toulouse, France, 2000.
38. Amina Bouras and Valérie Frayssé. Inexact matrix–vector products in Krylov methods for solving linear systems: a relaxation strategy. *SIAM Journal on Matrix Analysis and Applications* 2005; **26**:660–678.
39. Amina Bouras, Valérie Frayssé, and Luc Giraud. A relaxation strategy for inner–outer linear solvers in domain decomposition methods. *Technical Report 17*, CERFACS, Toulouse, France, 2000.
40. William E. Boyse and Andrew A. Seidl. A block QMR method for computing multiple simultaneous solutions to complex symmetric systems. *SIAM Journal on Scientific Computing* 1996; **17**:263–274.
41. James H. Bramble and Joseph E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Mathematics of Computation* 1988; **50**:1–17.
42. Claude Brezinski. *Projection Methods for Systems of Equations*. North-Holland: Amsterdam, 1997.
43. Claude Brezinski and Michela Redivo Zaglia. Hybrid procedures for solving linear systems. *Numerische Mathematik* 1994; **67**:1–19.
44. Claude Brezinski and Michela Redivo Zaglia. Look-ahead in Bi-CGSTAB and other product methods for linear systems. *BIT* 1995; **35**:169–201.
45. Claude Brezinski and Michela Redivo Zaglia. Transpose-free Lanczos-type algorithm for nonsymmetric linear systems. *Numerical Algorithms* 1998; **17**:67–103.
46. Claude Brezinski, Michela Redivo Zaglia, and Hassane Sadok. Avoiding breakdown and near-breakdown in Lanczos type algorithms. *Numerical Algorithms* 1991; **1**:261–284.
47. Claude Brezinski, Michela Redivo Zaglia, and Hassane Sadok. A breakdown-free Lanczos type algorithm for solving linear systems. *Numerische Mathematik* 1992; **63**:29–38.
48. Claude Brezinski, Michela Redivo Zaglia, and Hassane Sadok. Breakdowns in the implementation of the Lanczos method for solving linear systems. *Computers and Mathematics with Applications* 1997; **33**:31–44.
49. Claude Brezinski, Michela Redivo Zaglia, and Hassane Sadok. New look-ahead Lanczos-type algorithms for linear systems. *Numerische Mathematik* 1999; **83**:53–85.
50. Peter N. Brown. A theoretical comparison of the Arnoldi and the GMRES algorithms. *SIAM Journal on Scientific and Statistical Computing* 1991; **12**:58–78.
51. Charles G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation* 1965; **19**:577–593.
52. Charles G. Broyden. A new taxonomy of Conjugate Gradient methods. *Computers and Mathematics with Applications* 1996; **31**:7–17.
53. Charles G. Broyden and Maria T. Vespucci. *Krylov Solvers for Linear Algebraic Systems*. Studies in Computational Mathematics. vol. 11. Elsevier: Amsterdam, 2004.
54. Are Magnus Bruaset. *A Survey of Preconditioned Iterative Methods*, Pitman Research Notes in Mathematics Series, vol. 328. Lonman Scientific and Technical: Harlow, Esssex, England, 1995.
55. Angelika Bunse-Gerstner and Ronald Stöver. On a Conjugate Gradient-type method for solving complex symmetric linear systems. *Linear Algebra and its Applications* 1999; **287**:105–123.
56. Xiao-Chuan Cai, William D. Gropp, David E. Keyes, Robin G. Melvin, and David P. Young. Parallel Newton–Krylov–Schwarz algorithms for the transonic full potential equation. *SIAM Journal on Scientific Computing* 1998; **19**:246–265.

DOI: 10.1002/nla

57. Xiao-Chuan Cai and Olof B. Widlund. Multiplicative Schwarz algorithms for some nonsymmetric and indefinite problems. *SIAM Journal on Numerical Analysis* 1993; **30**:936–952.
58. Daniela Calvetti, Bryan W. Lewis, and Lothar Reichel. On the regularizing properties of the GMRES method. *Numerische Mathematik* 2002; **91**:605–625.
59. Daniela Calvetti, Serena Morigi, Lothar Reichel, and Fiorella Sgallari. Computable error bounds and estimates for the Conjugate Gradient Method. *Numerical Algorithms* 2000; **25**:79–88.
60. Sergio Campobasso. Effects of flow instabilities on the linear Harmonic analysis of unsteady flow in turbomachinery. *Ph.D. Thesis*, Oxford University, Trinity Term, 2004.
61. Zhi-Hao Cao. On the QMR approach for iterative methods including coupled three-term recurrences for solving nonsymmetric linear systems. *Applied Numerical Mathematics* 1998; **27**:123–140.
62. Bruno Carpentieri, Iain S. Duff, and Luc Giraud. A class of spectral two-level preconditioners. *SIAM Journal on Scientific Computing* 2003; **25**:749–765.
63. Bruno Carpentieri, Iain S. Duff, Luc Giraud, and Guillaume Sylvand. Combining fast multipole techniques and an approximate inverse preconditioner for large electromagnetism calculations. *SIAM Journal on Scientific Computing* 2005; **27**:774–792.
64. Tony F. Chan, Efstratios Gallopoulos, Valeria Simoncini, Tedd Szeto, and Charles H. Tong. A quasi-minimal residual variant of the Bi-CGSTAB algorithm for nonsymmetric systems. *SIAM Journal on Scientific Computing* 1994; **15**:338–347.
65. Tony F. Chan and Tedd Szeto. Composite step product methods for solving nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 1996; **17**:1491–1508.
66. Andrew Chapman and Yousef Saad. Deflated and augmented Krylov subspace techniques. *Numerical Linear Algebra with Applications* 1997; **4**:43–66.
67. David C. Chatfield, Melissa S. Reeves, Donald G. Truhlar, Csilla Duneczky, and David W. Schwenke. Complex generalized minimal residual algorithm for iterative solution of quantum mechanical reactive scattering equations. *Journal of Chemical Physics* 1992; **97**:8322–8333.
68. Ke Chen. *Matrix Preconditioning Techniques and Applications*. Cambridge Monograph on Applied and Computational Mathematics, vol. 19. Cambridge University Press: Cambridge, U.K., 2005.
69. Anthony T. Chronopoulos. s-step iterative methods for (non)symmetric (in)definite linear systems. *SIAM Journal on Numerical Analysis* 1991; **28**:1776–1789.
70. Paul Concus and Gene H. Golub. A generalized Conjugate Gradient method for nonsymmetric linear equations. In *Computing Methods in Applied Science and Engineering*, Glowinski R, Lions J-L (eds). Lecture Notes in Economics and Mathematical Systems, vol. 134. Springer: New York, 1976; 56–65.
71. Jane Cullum. Peaks, plateaux, and numerical stabilities in a Galerkin/minimal residual pair of methods for solving $Ax = b$. *Applied Numerical Mathematics* 1995; **19**:255–278.
72. Jane Cullum. Iterative methods for solving $Ax = b$, GMRES/FOM versus QMR/BiCG. *Advances in Computational Mathematics* 1996; **6**:1–24.
73. Jane Cullum and Anne Greenbaum. Relations between Galerkin and norm-minimizing iterative methods for solving linear systems. *SIAM Journal on Matrix Analysis and Applications* 1996; **17**:223–247.
74. Jane Cullum, Wolfgang Kerner, and Ralph A. Willoughby. A generalized nonsymmetric Lanczos procedure. *Computer Physics Communications* 1989; **53**:19–48.
75. Jane Cullum and Ralph A. Willoughby. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*. Classics in Applied Mathematics, vol. 41. Birkhaüser: Basel, 1985. vol. 1. Theory, vol. 2. Programs. Vol. 1 reprinted by SIAM: Philadelphia, PA, 2002.
76. Nigel Cundy, Jasper van den Eshof, Andreas Frommer, Thomas Lippert, Stefan Krieg, and Katrin Schäfer. Numerical methods for the QCD overlap operator: III. Nested iterations. *Computer Physics Communications* 2005; **165**:221–242.
77. Germund Dahlquist, Gene H. Golub, and Stephen G. Nash. Bounds for the error in linear systems. In *Workshop on Semi-Infinite Programming*, Hettich R (ed.). Springer: Berlin, 1978; 154–172.
78. Biswa N. Datta and Yousef Saad. Arnoldi methods for large Sylvester-like observer matrix equations, and an associated algorithm for partial spectrum assignment. *Linear Algebra and its Applications* 1991; **154–156**:225–244.
79. Philip J. Davis. *Interpolation and Approximation*. Blaisdell: New York, Toronto, London, 1963. Reprinted by Dover: New York, 1975.
80. David Day and Michael A. Heroux. Solving complex-valued linear systems via equivalent real formulations. *SIAM Journal on Scientific Computing* 2001; **23**(2):480–498.

DOI: 10.1002/nla

81. Eric de Sturler. Nested Krylov methods based on GCR. *Journal of Computational and Applied Mathematics* 1996; **67**:15–41.
82. Eric de Sturler. A performance model for Krylov subspace methods on mesh-based parallel computers. *Parallel Computing* 1996; **22**:57–74.
83. Eric de Sturler. Truncation strategies for optimal Krylov subspace methods. *SIAM Journal on Numerical Analysis* 1999; **36**:864–889.
84. Victorita Dolean and Stéphane Lanteri. Parallel multigrid methods for the calculation of unsteady flows on unstructured grids: algorithmic aspects and parallel performance on clusters of PCs. *Parallel Computing* 2004; **30**:503–525.
85. Tobin A. Driscol, Kim-Chuan Toh, and Lloyd N. Trefethen. From potential theory to matrix iterations in six steps. *SIAM Review* 1998; **40**:547–578.
86. Iain S. Duff. Combining direct and iterative methods for the solution of large linear systems in different application areas. *Technical Report RAL-TR-2004-033*, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, U.K., 2004. Also *TR/PA/04/128*, CERFACS, Toulouse, France.
87. Iain S. Duff. MA57—a code for the solution of sparse symmetric indefinite systems. *ACM Transactions on Mathematical Software* 2004; **30**:118–144.
88. Iain S. Duff, Albert M. Erisman, and John K. Reid. *Direct Methods for Sparse Matrices*. Clarendon Press: Oxford, 1986. (2nd edn. in preparation).
89. Iain S. Duff and Jennifer A. Scott. A parallel direct solver for large sparse highly unsymmetric linear systems. *ACM Transactions on Mathematical Software* 2004; **30**:95–117.
90. Jitka Dřkosová, Anne Greenbaum, Miroslav Rozložník, and Zdeněk Strakoš. Numerical stability of the GMRES method. *BIT Numerical Mathematics* 1995; **35**:308–330.
91. Michael Eiermann. Field of values and iterative methods. *Linear Algebra and its Applications* 1993; **180**:167–197.
92. Michael Eiermann and Oliver G. Ernst. Geometric aspects in the theory of Krylov subspace methods. *Acta Numerica* 2001; **10**:251–312.
93. Michael Eiermann and Oliver G. Ernst. A restarted Krylov subspace method for the evaluation of matrix functions. *Technical Report*, Institut für Numerische Mathematik und Optimierung, Techniche Universität Bergakademie Freiberg, June 2005; *SIAM Journal on Numerical Analysis*, in press.
94. Michael Eiermann, Oliver G. Ernst, and Olaf Schneider. Analysis of acceleration strategies for restarted minimal residual methods. *Journal of Computational and Applied Mathematics* 2000; **123**:261–292.
95. Timo Eirola and Olavi Nevanlinna. Accelerating with rank-one updates. *Linear Algebra and its Applications* 1989; **121**:511–520.
96. Stanley C. Eisenstat, Howard C. Elman, and Martin H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis* 1983; **20**:345–357.
97. Stanley C. Eisenstat, James M. Ortega, and Courtenay T. Vaughan. Efficient polynomial preconditioning for the Conjugate Gradient method. *SIAM Journal on Scientific and Statistical Computing* 1990; **11**:859–872.
98. Howard C. Elman. Iterative methods for large sparse nonsymmetric systems of linear equations. *Ph.D. Thesis*, Yale University: New Haven, CT, 1982.
99. Howard C. Elman, Oliver G. Ernst, and Dianne P. O'Leary. A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations. *SIAM Journal on Scientific Computing* 2001; **23**:1290–1314.
100. Howard C. Elman, David J. Silvester, and Andrew J. Wathen. *Finite Elements and Fast Iterative Solvers*, *with Applications in Incompressible Fluid Dynamics*, *Numerical Mathematics and Scientific Computation*, vol. 21. Oxford University Press: Oxford, New York, 2005.
101. Mark Embree. Convergence of Krylov subspace methods for non-normal matrices. *Ph.D. Thesis*, Oxford University Computing Laboratory, Numerical Analysis Group, Michaelmas Term, 1999.
102. Mark Embree. The tortoise and the hare restart GMRES. *SIAM Review* 2003; **45**:259–266.
103. Shlomo Engelberg and Aaron E. Naiman. A note on Conjugate Gradient convergence. Part III. *Numerische Mathematik* 2000; **85**:685–696.
104. Jocelyne Erhel. A parallel GMRES version for general sparse matrices. *Electronic Transactions on Numerical Analysis* 1995; **3**:160–176.
105. Jocelyne Erhel, Kevin Burrage, and Bert Pohl. Restarted GMRES preconditioned by deflation. *Journal of Computational and Applied Mathematics* 1996; **69**:303–318.
106. Jocelyne Erhel and Frédéric Guyomarc'h. An augmented Conjugate Gradient method for solving consecutive symmetric positive definite linear systems. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1279–1299.

DOI: 10.1002/nla

107. Oliver G. Ernst. Residual-minimizing Krylov subspace methods for stabilized discretizations of convection–diffusion equations. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1079–1101.
108. Azeddine Essai. Weighted FOM and GMRES for solving nonsymmetric linear systems. *Numerical Algorithms* 1998; **18**:277–292.
109. Vance Faber, Wayne D. Joubert, Emanuel Knill, and Thomas A. Manteuffel. Minimal residual method stronger than polynomial preconditioning. *SIAM Journal on Matrix Analysis and Applications* 1996; **17**:707–729.
110. Vance Faber and Thomas A. Manteuffel. Necessary and sufficient conditions for the existence of a Conjugate Gradient method. *SIAM Journal on Numerical Analysis* 1984; **21**:352–362.
111. Vance Faber and Thomas A. Manteuffel. Orthogonal error methods. *SIAM Journal on Numerical Analysis* 1987; **24**:170–187.
112. Heike Fassbender and Khakim D. Ikramov. SYMMLQ-like procedure of $Ax = b$ where A is a special normal matrix. *Calcolo* 2006; **43**:17–37.
113. Peter Feldmann, Roland W. Freund, and Emrah Acar. Power grid analysis using a flexible conjugate gradient algorithm with sparsification. *Technical Report*, Department of Mathematics, University of California, Davis, June 2006.
114. Anna Feriani, Federico Perotti, and Valeria Simoncini. Iterative system solvers for the frequency analysis of linear mechanical systems. *Computer Methods in Applied Mechanics and Engineering* 2000; **190**:1719–1739.
115. Bernd Fischer. *Polynomial Based Iteration Methods for Symmetric Linear Systems*. Wiley/Teubner: Chichester, West Essex, England, Stuttgart, 1996.
116. Bernd Fischer and Roland W. Freund. An inner product-free Conjugate Gradient-like algorithm for Hermitian positive definite systems. In *Proceedings of the Cornelius Lanczos International Centenary Conference*, David Brown J, Chu MT, Ellison DC, Plemmons RJ (eds). SIAM: Philadelphia, PA, 1994; 288–290.
117. Bernd Fischer and Gene H. Golub. On the error computation for polynomial based iteration methods. In *Recent Advances in Iterative Methods*, Golub GH, Greenbaum A, Luskin M (eds). Springer: New York, 1994; 59–67.
118. Bernd Fischer, Alison Ramage, David J. Silvester, and Andrew J. Wathen. Minimum residual methods for augmented systems. *BIT Numerical Mathematics* 1998; **38**:527–543.
119. Matthias Fischer and Lothar Gaul. Fast BEM–FEM mortar coupling for acoustic–structure interaction. *International Journal for Numerical Methods in Engineering* 2005; **62**:1677–1690.
120. Roger Fletcher. Conjugate Gradient methods for indefinite systems. In *Numerical Analysis—Dundee 1975*, Alistair Watson G (ed.). Lecture Notes in Mathematics, vol. 506. Springer: Heidelberg, 1976; 73–89.
121. Diederik R. Fokkema, Gerard L.G. Sleijpen, and Henk A. van der Vorst. Generalized conjugate gradient squared. *Journal of Computational and Applied Mathematics* 1996; **71**:125–146.
122. Jason Frank and C. (Kees) Vuik. On the construction of deflation-based preconditioners. *SIAM Journal on Scientific Computing* 2001; **23**:442–462.
123. Roland W. Freund. On Conjugate Gradient type methods and polynomial preconditioners for a class of complex non-Hermitian matrices. *Numerische Mathematik* 1990; **57**:285–312.
124. Roland W. Freund. Conjugate Gradient-type methods for linear systems with complex symmetric coefficient matrices. *SIAM Journal on Scientific and Statistical Computing* 1992; **13**:425–448.
125. Roland W. Freund. Quasi-kernel polynomials and convergence results for quasi-minimal residual iterations. In *Numerical Methods in Approximation Theory*, vol. 9. Braess D, Schumaker LL (eds), Birkhäuser: Basel, 1992; 77–95.
126. Roland W. Freund. Solution of shifted linear systems by quasi-minimal residual iterations. In *Numerical Linear Algebra*, Reichel L, Ruttan A, Varga RS (eds). W. de Gruyter: Berlin, 1993; 101–121.
127. Roland W. Freund. Transpose-free quasi-minimal residual methods for non-Hermitian linear systems. In *Recent Advances in Iterative Methods*, vol. 60. Golub GH, Greenbaum A, Luskin M (eds). IMA Volumes in Mathematics and its Applications, Springer: New York, 1994; 69–94.
128. Roland W. Freund. Model reduction methods based on Krylov subspaces. *Acta Numerica* 2003; **12**:267–319.
129. Roland W. Freund. Krylov subspaces associated with higher-order linear dynamical systems. *BIT Numerical Mathematics* 2005; **45**:495–516.
130. Roland W. Freund. Padé-type model reduction of second-order and higher-order linear dynamical systems. In *Dimension Reduction of Large-Scale Systems*, Benner P, Mehrmann V, Sorensen DC (eds), Lecture Notes in Computational Science and Engineering, vol. 45; *Proceedings of a Workshop held in Oberwolfach*, Germany, 19–25 October 2003. Springer: Berlin, Heidelberg, 2005; 191–223.
131. Roland W. Freund, Gene H. Golub, and Nöel M. Nachtigal. Iterative solution of linear systems. *Acta Numerica* 1992; **1**:57–100.

DOI: 10.1002/nla

132. Roland W. Freund, Gene H. Golub, and Nöel M. Nachtigal. Recent advances in Lanczos-based iterative methods for nonsymmetric linear systems. In *Algorithmic Trends in Computational Fluid Dynamics*, Hussaini MY, Kumar A, Salas MD (eds). Springer: Berlin, 1993; 137–162.
133. Roland W. Freund, Martin H. Gutknecht, and Nöel M. Nachtigal. An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices. *SIAM Journal on Scientific Computing* 1993; **14**:137–158.
134. Roland W. Freund and Marlis Hochbruck. On the use of two QMR algorithms for solving singular systems and applications in Markov Chain modelling. *Numerical Linear Algebra with Applications* 1994; **1**:403–420.
135. Roland W. Freund and Nöel M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik* 1991; **60**:315–339.
136. Roland W. Freund and Nöel M. Nachtigal. An implementation of the QMR method based on coupled two-term recurrences. *SIAM Journal on Scientific Computing* 1994; **15**:313–337.
137. Roland W. Freund and Nöel M. Nachtigal. Software for simplified Lanczos and QMR algorithms. *Applied Numerical Mathematics* 1995; **19**:319–341.
138. Vladimir M. Fridman. The method of minimal iterations with minimum errors for a system of linear algebraic equations with symmetric matrix. *Zhurnal Vychislitelc nou Matematiki i Matematichskou i Fiziki* 1962; **2**:341–342 (in Russian). English translation in *USSR Computational Mathematics and Mathematical Physics* 1963; **2**:362–363.
139. Andreas Frommer. BiCGSTAB(ℓ) for families of shifted linear systems. *Computing* 2003; **70**:87–109.
140. Andreas Frommer and Uwe Glässner. Restarted GMRES for shifted linear systems. *SIAM Journal on Scientific Computing* 1998; **19**:15–26.
141. Andreas Frommer and Peter Maass. Fast CG-based methods for Tikhonov–Phillips regularization. *SIAM Journal on Scientific Computing* 1999; **20**:1831–1850.
142. Andreas Frommer and Andre Weinberg. Verified error bounds for linear systems through the Lanczos process. *Reliable Computing* 1999; **5**:255–267.
143. Luc Giraud, Serge Gratton, and Julien Langou. Convergence in backward error of relaxed GMRES. *Technical Report TR/PA/06/08*, CERFACS, Toulouse, France, 2006.
144. Luc Giraud, Serge Gratton, and Émeric Martin. Incremental spectral preconditioners for sequences of linear systems. *Technical Report TR/PA/05/17*, CERFACS, Toulouse, France, 2005; *Applied Numerical Mathematics*, in press.
145. Israel Gohberg, Peter Lancaster, and Leiba Rodman. *Matrices and Indefinite Scalar Products*, *Operator Theory*: *Advances and Applications*, vol. 8. Birkhäuser: Basel, Boston, 1983.
146. Israel Gohberg, Leiba Rodman, and Peter Lancaster. *Matrix Polynomials*. Academic Press: New York, 1982.
147. Gene H. Golub and Gérard A. Meurant. Matrices, moments and quadrature II; how to compute the norm of the error in iterative methods. *BIT* 1997; **37**:687–705.
148. Gene H. Golub and Dianne P. O'Leary. Some history of the Conjugate Gradient and Lanczos algorithms: 1948–1976. *SIAM Review* 1989; **31**:50–102.
149. Gene H. Golub and Michael L. Overton. The convergence of inexact Chebyshev and Richardson iterative methods for solving linear systems. *Numerische Mathematik* 1988; **53**:571–593.
150. Gene H. Golub and Zdenek Strakoš. Estimates in quadratic formulas. *Numerical Algorithms* 1994; **8**:241–268.
151. Gene H. Golub and Charles F. Van Loan. *Matrix Computations* (3rd edn). The John Hopkins University Press: Baltimore, 1996.
152. Gene H. Golub and Qiang Ye. Inexact inverse iterations for the generalized eigenvalue problems. *BIT Numerical Mathematics* 2000; **40**:671–684.
153. Gene H. Golub and Qiang Ye. Inexact preconditioned Conjugate Gradient method with inner–outer iteration. *SIAM Journal on Scientific Computing* 2001; **21**:1305–1320.
154. Gene H. Golub, Zhenyue Zhang, and Hongyuan Zha. Large sparse symmetric eigenvalue problems with homogeneous linear constraints: the Lanczos process with inner–outer iterations. *Linear Algebra and its Applications* 2000; **309**:289–306.
155. Serge Goossens and Dirk Roose. Ritz and harmonic Ritz values and the convergence of FOM and GMRES. *Numerical Linear Algebra with Applications* 1999; **6**:281–293.
156. Serge Gratton. Private communication, 2005.
157. Anne Greenbaum. Estimating the attainable accuracy of recursively computed residual methods. *SIAM Journal on Matrix Analysis and Applications* 1997; **18**:535–551.
158. Anne Greenbaum. *Iterative Methods for Solving Linear Systems*, *Frontiers in Applied Mathematics*, vol. 17. SIAM: Philadelphia, PA, 1997.

DOI: 10.1002/nla

159. Anne Greenbaum. Some theoretical results derived from polynomial numerical hulls of Jordan blocks. *Electronic Transactions on Numerical Analysis* 2004; **18**:81–90.
160. Anne Greenbaum and Leonid Gurvits. Max–min properties of matrix factor forms. *SIAM Journal on Scientific Computing* 1994; **15**:348–358.
161. Anne Greenbaum and Leonid Gurvits. A further note on max-min properties of matrix factor forms. *SIAM Journal on Scientific Computing* 1995; **16**:496–499.
162. Anne Greenbaum, Vlastimil Pták, and Zdeněk Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM Journal on Matrix Analysis and Applications* 1996; **17**:95–118.
163. Anne Greenbaum, Miroslav Rozložnik, and Zdeněk Strakoš. Numerical behaviour of the modified Gram–Schmidt GMRES implementation. *BIT Numerical Mathematics* 1997; **37**:706–719.
164. Anne Greenbaum and Zdeněk Strakoš. Matrices that generate the same Krylov residual spaces. In *Recent Advances in Iterative Methods*, Golub GH, Greenbaum A, Luskin M (eds). IMA Volumes in Mathematics and its Applications, vol. 60. Springer: New York, 1994; 95–118.
165. Anne Greenbaum and Lloyd N. Trefethen. GMRES/CR and Arnoldi/Lanczos as matrix approximation problems. *SIAM Journal on Scientific Computing* 1994; **15**:359–368.
166. Roger G. Grimes, John G. Lewis, and Horst D. Simon. A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. *SIAM Journal on Matrix Analysis and Applications* 1994; **15**:228–272.
167. Martin H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms. Part I. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:594–639.
168. Martin H. Gutknecht. Changing the norm in Conjugate Gradient type algorithms. *SIAM Journal on Numerical Analysis* 1993; **30**:40–56.
169. Martin H. Gutknecht. Variants of BiCGStab for matrices with complex spectrum. *SIAM Journal on Scientific Computing* 1993; **14**:1020–1033.
170. Martin H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms. Part II. *SIAM Journal on Matrix Analysis and Applications* 1994; **15**:15–58.
171. Martin H. Gutknecht. Lanczos-type solvers for nonsymmetric linear systems of equations. *Acta Numerica* 1997; **6**:271–397.
172. Martin H. Gutknecht. Block Krylov space methods for linear systems with multiple right-hand sides: an introduction. In *Modern Mathematical Models*, *Methods and Algorithms for Real World Systems*, Siddiqi AH, Duff IS, Christensen O (eds). Anamaya Publishers: New Delhi, 2006, in press.
173. Martin H. Gutknecht and Klaus J. Ressel. Look-ahead procedures for Lanczos-type product methods based on three-term Lanczos recurrences. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1051–1078.
174. Martin H. Gutknecht and Miroslav Rozložník. By how much can residual minimization accelerate the convergence of orthogonal residual methods? *Numerical Algorithms* 2001; **27**:189–213.
175. Martin H. Gutknecht and Miroslav Rozložník. Residual smoothing techniques: do they improve the limiting accuracy of iterative solvers? *BIT Numerical Mathematics* 2001; **41**:86–114.
176. Martin H. Gutknecht and Miroslav Rozložník. A framework for generalized conjugate gradient methods—with special emphasis on contributions by Rüdiger Weiss. *Applied Numerical Mathematics* 2002; **41**:7–22.
177. Martin H. Gutknecht and Zdeněk Strakoš. Accuracy of two three-term and three two-term recurrences for Krylov space solvers. *SIAM Journal on Matrix Analysis and Applications* 2000; **22**:213–229.
178. Martin Hanke, Marlis Hochbruck, and Wilhelm Niethammer. Experiments with Krylov subspace methods on a massively parallel computer. *Applications of Mathematics* 1993; **38**:440–451.
179. Per Christian Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*: *Numerical Aspects of Linear Inversion*. SIAM Monographs on Mathematical Modeling and Computation, vol. 4. SIAM: Philadelphia, PA, 1997.
180. Magnus R. Hestenes and Eduard Stiefel. Methods of Conjugate Gradients for solving linear systems. *Journal of Research of the National Bureau of Standards* 1952; **49**:409–436.
181. Marlis Hochbruck and Christian Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis* 1997; **34**:1911–1925.
182. Marlis Hochbruck and Christian Lubich. Error analysis of Krylov methods in a nutshell. *SIAM Journal on Scientific Computing* 1998; **19**:695–701.
183. Marlis Hochbruck and Gerhard Starke. Preconditioned Krylov subspace methods for Lyapunov matrix equations. *SIAM Journal on Matrix Analysis and Applications* 1995; **16**:156–171.
184. Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press: Cambridge, 1985.
185. Dan Y. Hu and Lothar Reichel. Krylov-subspace methods for the Sylvester equation. *Linear Algebra and its Applications* 1992; **172**:283–313.

DOI: 10.1002/nla

186. Marko Huhtanen. A Hermitian Lanczos method for normal matrices. *SIAM Journal on Matrix Analysis and Applications* 2002; **23**:1092–1108.
187. Ilse Ipsen. Expressions and bounds for the GMRES residual. *BIT Numerical Mathematics* 2000; **40**:524–535.
188. David A. H. Jacobs. A generalization of the Conjugate-Gradient method to solve complex systems. *IMA Journal on Numerical Analysis* 1986; **6**:447–452.
189. Carl Jagels and Lothar Reichel. The isometric Arnoldi process and an application to iterative solution of large linear systems. In *Iterative Methods in Linear Algebra*, Beauwens R, Pieter PN de Groen (eds). Elsevier Science/North-Holland: Amsterdam, 1992; 361–369.
190. Carl Jagels and Lothar Reichel. A fast minimal residual algorithm for shifted unitary matrices. *Numerical Linear Algebra with Applications* 1994; **1**:361–369.
191. Kang C. Jea and David M. Young. On the simplification of generalized Conjugate-Gradient methods for nonsymmetrizable linear systems. *Linear Algebra and its Applications* 1983; **52–53**:399–417.
192. Zhongxiao Jia. On IOM(q): the incomplete orthogonalization method for large unsymmetric linear systems. *Numerical Linear Algebra with Applications* 1996; **3**:491–512.
193. Olin J. Johnson, Charles A. Micchelli, and George Paul. Polynomial preconditioners for Conjugate Gradient calculations. *SIAM Journal on Numerical Analysis* 1983; **20**:362–376.
194. Pascal Joly and Gérard Meurant. Complex Conjugate Gradient methods. *Numerical Algorithms* 1993; **4**:379–406.
195. Wayne D. Joubert. Lanczos methods for the solution of nonsymmetric systems of linear equations. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:926–943.
196. Wayne D. Joubert. A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 1994; **15**:427–439.
197. Wayne D. Joubert. On the convergence behaviour of the restarted GMRES algorithm for solving nonsymmetric linear systems. *Numerical Linear Algebra with Applications* 1994; **1**:427–447.
198. Erik F. Kaasschieter. Preconditioned conjugate gradients for solving singular systems. *Journal of Computational and Applied Mathematics* 1988; **24**:265–275.
199. Igor Kaporin. Superlinear convergence in minimum residual iterations. *Numerical Linear Algebra with Applications* 2005; **12**:453–470.
200. Ebrahim M. Kasenally. GMBACK: a generalised minimum backward error algorithm for nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 1995; **16**:698–719.
201. Ebrahim M. Kasenally and Valeria Simoncini. Analysis of a minimum perturbation algorithm for nonsymmetric linear systems. *SIAM Journal on Numerical Analysis* 1997; **34**:48–66.
202. C. T. (Tim) Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM: Philadelphia, PA, 1995.
203. C. T. (Tim) Kelley and Zhaoqin Q. Xue. GMRES and integral operators. *SIAM Journal on Scientific Computing* 1996; **17**:217–226.
204. Serge A. Kharchenko and Alex Yu. Yeremin. Eigenvalue translation based preconditioners for the GMRES(k) method. *Numerical Linear Algebra with Applications* 1995; **2**:51–77.
205. Dana A. Knoll and David E. Keyes. Jacobian-free Newton–Krylov methods: a survey of approaches and applications. *Journal of Computational Physics* 2004; **193**:357–397.
206. Andrey Kucherov. A fast iterative method for solving in real arithmetic a system of linear equations with a complex symmetric matrix. *Sovietsky Mathematischeskoe Doklady* 1991; **43**:377–379.
207. Arno B.J. Kuijlaars. Convergence analysis of Krylov subspace iterations with methods from potential theory. *SIAM Review* 2006; **48**:3–40.
208. Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards* 1950; **45**:255–282.
209. Cornelius Lanczos. Solution of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards* 1952; **49**:33–53.
210. Jörg Liesen. Computable convergence bounds for GMRES. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:882–903.
211. Jörg Liesen, Miroslav Rozložník, and Zdeněk Strakoš. Least squares residuals and minimal residual methods. *SIAM Journal on Scientific Computing* 2002; **23**:1503–1525.
212. Jörg Liesen and Paul E. Saylor. Orthogonal Hessenberg reduction and orthogonal Krylov subspace bases. *SIAM Journal on Numerical Analysis* 2005; **42**:2148–2158.
213. Jörg Liesen and Zdeněk Strakoš. Convergence of GMRES for tridiagonal Toeplitz matrices. *SIAM Journal on Matrix Analysis and Applications* 2004; **26**:233–251.
214. Jörg Liesen and Zdeněk Strakoš. GMRES convergence analysis for a convection–diffusion model problem. *SIAM Journal on Scientific Computing* 2005; **26**:1989–2009.

DOI: 10.1002/nla

215. Jörg Liesen and Zdeněk Strakoš. On numerical stability in large scale linear algebraic computations. *Zeitschrift für Angewandte Mathematik und Mechanik* 2005; **85**:307–325.
216. Jörg Liesen and Petr Tichý. Convergence analysis of Krylov subspace methods. *GAMM-Mitteilungen* 2004; **27**:153–173.
217. Jörg Liesen and Petr Tichý. The worst-case GMRES for normal matrices. *BIT Numerical Mathematics* 2004; **44**:409–435.
218. Jörg Liesen and Petr Tichý. On the worst-case convergence of MR and CG for symmetric positive definite tridiagonal Toeplitz matrices. *Electronic Transactions on Numerical Analysis* 2005; **20**:180–197.
219. Luciano Lopez and Valeria Simoncini. Analysis of projection methods for rational function approximation to the matrix exponential operator. *SIAM Journal on Numerical Analysis* 2006; **44**:613–635.
220. David G. Luenberger. *Linear and Nonlinear Programming* (2nd edn). Addison-Wesley: Reading, MA, 1984.
221. Jan Mandel. On block diagonal and Schur complement preconditioning. *Numerische Mathematik* 1990; **58**:79–93.
222. Lois Mansfield. On the use of deflation to improve the convergence of Conjugate Gradient iteration. *Communications in Applied Numerical Methods* 1988; **4**:151–156.
223. Thomas A. Manteuffel and James Otto. On the roots of the orthogonal polynomials and residual polynomials associated with a Conjugate Gradient method. *Numerical Linear Algebra with Applications* 1994; **1**:449–475.
224. Geoff Markham. Conjugate Gradient type methods for indefinite, asymmetric, and complex systems. *IMA Journal of Numerical Analysis* 1990; **10**:155–170.
225. Jiri Maryška, Miroslav Rozložník, and Miroslav Tůma. Schur complement systems in the mixed-hybrid finite element approximation of the potential fluid flow problem. *SIAM Journal on Scientific Computing* 2000; **22**:704–723.
226. Karl Meerbergen. The solution of parametrized symmetric linear systems. *SIAM Journal on Matrix Analysis and Applications* 2003; **24**:1038–1059.
227. Gérard A. Meurant. The computation of bounds for the norm of the error in the Conjugate Gradient algorithm. *Numerical Algorithms* 1994; **16**:77–87.
228. Gérard A. Meurant. *Computer Solution of Large Linear Systems*, *Studies in Mathematics and its Applications*, vol. 28. North-Holland: Amsterdam, 1999.
229. Gérard A. Meurant and Zdeněk Strakoš. The Lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numerica* 2006; **15**:471–542.
230. István Montvay and Gernot Münster. *Quantum Fields on the Lattice*. Cambridge Monographs on Mathematical Physics. Cambridge University Press: Cambridge, New York, Melbourne, 1994.
231. Ronald B. Morgan. Computing interior eigenvalues of large matrices. *Linear Algebra and its Applications* 1991; **154–156**:289–309.
232. Ronald B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM Journal on Matrix Analysis and Applications* 1995; **16**:1154–1171.
233. Ronald B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1112–1135.
234. Ronald B. Morgan. GMRES with deflated restarting. *SIAM Journal on Scientific Computing* 2002; **24**:20–37.
235. Reinhard Nabben and C. (Kees) Vuik. A comparison of deflation and coarse grid correction applied to porous media flow. *SIAM Journal on Numerical Analysis* 2004; **42**:1631–1647.
236. Reinhard Nabben and C. (Kees) Vuik. A comparison of deflation and the balancing preconditioner. *SIAM Journal on Scientific Computing* 2006; **27**:1742–1759.
237. Nöel M. Nachtigal. A look-ahead variant of the Lanczos algorithm and its application to the quasi-minimal residual method for non-Hermitian linear systems. *Ph.D. Thesis*, Massachussets Institute of Technology, Cambridge, MA, 1991.
238. Nöel M. Nachtigal, Satish C. Reddy, and Lloyd N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:778–795.
239. Nöel M. Nachtigal, Lothar Reichel, and Lloyd N. Trefethen. A hybrid GMRES algorithm for nonsymmetric matrix iterations. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:796–825.
240. Aaron E. Naiman, Ivo M. Babuška, and Howard C. Elman. A note on Conjugate Gradient convergence. *Numerische Mathematik* 1997; **76**:209–230.
241. Aaron E. Naiman and Shlomo Engelberg. A note on Conjugate Gradient convergence, Part II. *Numerische Mathematik* 2000; **85**:665–683.
242. Ramesh Natarajan. An iterative scheme for dense, complex-symmetric, linear systems in acoustics boundary-element computations. *SIAM Journal on Scientific Computing* 1998; **19**:1450–1470.

DOI: 10.1002/nla

243. Olavi Nevanlinna. *Convergence of Iterations for Linear Equations*. Birkhäuser: Basel, 1993.
244. Roy A. Nicolaides. Deflation of Conjugate Gradients with applications to boundary value problems. *SIAM Journal on Numerical Analysis* 1987; **24**:355–365.
245. Yvan Notay. Flexible Conjugate Gradient. *SIAM Journal on Scientific Computing* 2000; **22**:1444–1460.
246. Bahram Nour-Omid, Beresford N. Parlett, Thomas Ericsson, and Paul S. Jensen. How to implement the spectral transformation. *Mathematics of Computation* 1987; **48**:663–673.
247. Alexander Padiy, Owe Axelsson, and Ben Polman. Generalized augmented matrix preconditioning approach and its application to iterative solution of ill-conditioned algebraic systems. *SIAM Journal on Matrix Analysis and Applications* 2000; **22**:793–818.
248. Christopher C. Paige, Beresford N. Parlett, and Henk A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications* 1995; **2**:115–134.
249. Christopher C. Paige, Miroslav Rozložník, and Zdeněk Strakoš. Modified Gram–Schmidt (MGS), least squares, and backward stability of MGS-GMRES. *SIAM Journal on Matrix Analysis and Applications* 2006; **28**:264–284.
250. Christopher C. Paige and Michael A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis* 1975; **12**:617–629.
251. Christopher C. Paige and Michael A. Saunders. LSQR: An algorithm for sparse linear equations and least squares problems. *ACM Transactions on Mathematical Software* 1982; **8**:43–71.
252. Christopher C. Paige and Zdeněk Strakoš. Bounds for the least squares distance using scaled total least squares problems. *Numerische Mathematik* 2002; **91**:93–115.
253. Christopher C. Paige and Zdeněk Strakoš. Residual and backward error bounds in minimum residual Krylov subspace methods. *SIAM Journal on Scientific Computing* 2002; **23**:1899–1924.
254. Beresford N. Parlett. Reduction to tridiagonal form and minimal realizations. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:567–593.
255. Beresford N. Parlett and Hsin-Chu Chen. Use of indefinite pencils for computing damped natural modes. *Linear Algebra and its Applications* 1990; **140**:53–88.
256. Beresford N. Parlett, Derek R. Taylor, and Zhishun A. Liu. A look-ahead Lanczos algorithm for unsymmetric linear systems. *Mathematics of Computation* 1985; **44**:105–124.
257. Micol Pennacchio and Valeria Simoncini. Efficient algebraic solution of reaction–diffusion systems for the cardiac excitation process. *Journal of Computational and Applied Mathematics* 2002; **145**:49–70.
258. Federico Perotti and Valeria Simoncini. Analytical and numerical techniques in frequency domain response computation. In *Recent Research Developments in Structural Dynamics*, Luongo A (ed.). Research Signpost: Trivandrum, India, 2003; 33–54.
259. Claude Pommerell and Wolfgang Fichtner. Memory aspects and performance of iterative solvers. *SIAM Journal on Scientific Computing* 1994; **15**:460–473.
260. Ernesto E. Prudencio, Richard Byrd, and Xiao-Chuan Cai. Parallel full space SPQ Lagrange–Newton–Krylov–Schwarz algorithms for PDE-constrained optimization problems. *SIAM Journal on Scientific Computing* 2006; **27**:1305–1328.
261. Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, Clarendon Press: Oxford, 1999.
262. Lothar Reichel and Qiang Ye. Breakdown-free GMRES for singular systems. *SIAM Journal on Matrix Analysis and Applications* 2005; **26**:1001–1021.
263. Klaus J. Ressel and Martin H. Gutknecht. QMR-smoothing for Lanczos-type product methods based on three-term recurrences. *SIAM Journal on Scientific Computing* 1998; **19**:55–73.
264. Mickael Robbé and Miloud Sadkane. A convergence analysis of GMRES and FOM methods for Sylvester equations. *Numerical Algorithms* 2002; **30**:71–89.
265. Stefan Röllin and Martin H. Gutknecht. Variations of Zhang's Lanczos-type product method. *Applied Numerical Mathematics* 2002; **41**:119–133.
266. Miroslav Rozložník and Rüdiger Weiss. On the stable implementation of the generalized minimal error method. *Journal of Computational and Applied Mathematics* 1998; **98**:49–62.
267. Miroslav Rozložník and Valeria Simoncini. Krylov subspace methods for saddle point problems with indefinite preconditioning. *SIAM Journal on Matrix Analysis and Applications* 2002; **24**:368–391.
268. Jurgen J. Rusch. Using a complex version of GMRES for solving optical scattering problems. In *Iterative Methods in Linear Algebra*, Beauwens R, de Groen PN (eds). Elsevier Science (North-Holland): Amsterdam, 1992; 459–468.
269. Yousef Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Mathematics of Computation* 1981; **37**:105–126.

DOI: 10.1002/nla

270. Yousef Saad. Practical use of some Krylov subspace methods for solving indefinite and unsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1984; **5**:203–228.
271. Yousef Saad. Practical use of polynomial preconditioning for the Conjugate Gradient method. *SIAM Journal on Scientific and Statistical Computing* 1985; **6**:865–881.
272. Yousef Saad. A flexible inner–outer preconditioned GMRES. *SIAM Journal on Scientific Computing* 1993; **14**:461–469.
273. Yousef Saad. *Iterative Methods for Sparse Linear Systems* (2nd edn). The PWS Publishing Company: Boston, 1996. SIAM: Philadelphia, PA, 2003.
274. Yousef Saad. Analysis of augmented Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications* 1997; **18**:435–449.
275. Yousef Saad and Martin H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
276. Yousef Saad and Henk A. van der Vorst. Iterative solution of linear systems in the 20th century. *Journal of Computational and Applied Mathematics* 2000; **123**:1–33.
277. Yousef Saad and Kesheng Wu. DQGMRES: A direct quasi-minimal residual algorithm based on incomplete orthogonalization. *Numerical Linear Algebra with Applications* 1996; **3**:329–343.
278. Yousef Saad, Man-Chung Yeung, Jocelyne Erhel, and Frédéric Guyomarc'h. A deflated version of the Conjugate Gradient algorithm. *SIAM Journal on Scientific Computing* 2000; **21**:1909–1926.
279. Tapan K. Sarkar. On the application of the generalized biconjugate gradient method. *Journal of Electromagnetic Waves and Applications* 1987; **1**:223–242.
280. Tapan K. Sarkar, Xiaopu Yang, and Ercument Arvas. A limited survey of various Conjugate Gradient methods for solving complex matrix equations arising in electromagnetic wave interactions. *Wave Motion* 1988; **10**:527–546.
281. Marcus Sarkis and Daniel B. Szyld. Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. *Computer Methods in Applied Mechanics and Engineering*, 2006, in press.
282. Marcus Sarkis and Daniel B. Szyld. A proposal for a dynamically adaptive inexact additive additive Schwarz preconditioner. In *Proceedings of the Sixteenth Conference on Domain Decomposition Methods*, *2005*, Widlund O, Keyes D (eds), vol. 55. Lecture Notes in Computational Science and Engineering. Springer: Berlin, 2006, in press.
283. Michael A. Saunders, Horst D. Simon, and Elizabeth L. Yip. Two Conjugate-Gradient-type methods for unsymmetric linear equations. *SIAM Journal on Numerical Analysis* 1988; **25**:927–940.
284. Paul E. Saylor and Dennis C. Smolarski. Computing the roots of complex orthogonal and kernel polynomials. *SIAM Journal on Scientific Computing* 1988; **9**:1–13.
285. Paul E. Saylor and Dennis C. Smolarski. Implementation of an adaptive algorithm for Richardson's method. *Linear Algebra and its Applications* 1991; **154–156**:615–646.
286. Valeria Simoncini. On the numerical solution of $AX - XB = C$. *BIT Numerical Mathematics* 1996; **36**:182–198.
287. Valeria Simoncini. Linear systems with a quadratic parameter and application to structural dynamics. In *Iterative Methods in Scientific Computation IV*, Kincaid DR, Elster AC (eds). IMACS Series in Computational and Applied Mathematics, vol. 5. IMACS: New Brunswick, NJ, 1999; 451–461.
288. Valeria Simoncini. A new variant of restarted GMRES. *Numerical Linear Algebra with Applications* 1999; **6**:61–77.
289. Valeria Simoncini. On the convergence of restarted Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications* 2000; **22**:430–452.
290. Valeria Simoncini. Restarted full orthogonalization method for shifted linear systems. *BIT Numerical Mathematics* 2003; **43**:459–466.
291. Valeria Simoncini. Block triangular preconditioners for symmetric saddle-point problems. *Applied Numerical Mathematics* 2004; **49**:63–80.
292. Valeria Simoncini. Variable accuracy of matrix–vector products in projection methods for eigencomputation. *SIAM Journal on Numerical Analysis* 2005; **43**:1155–1174.
293. Valeria Simoncini and Lars Eldén. Inexact Rayleigh quotient-type methods for eigenvalue computations. *BIT Numerical Mathematics* 2002; **42**:159–182.
294. Valeria Simoncini and Efstratios Gallopoulos. Iterative methods for complex symmetric systems with multiple right-hand sides. *Technical Report 1322*, Center for Supercomputing Research and Development, December 1993. Available at http://www.dm.unibo.it/~simoncin.
295. Valeria Simoncini and Efstratios Gallopoulos. Convergence properties of block GMRES and matrix polynomials. *Linear Algebra and its Applications* 1996; **247**:97–120.

DOI: 10.1002/nla

296. Valeria Simoncini and Micol Pennacchio. The behaviour of symmetric Krylov subspace methods for solving $Mx = (M - \gamma I)v$. *Linear Algebra and its Applications* 2004; **380**:53–71.
297. Valeria Simoncini and Federico Perotti. On the numerical solution of $(\lambda^2 A + \lambda B + C)x = b$ and application to structural dynamics. *SIAM Journal on Scientific Computing* 2002; **23**:1876–1898.
298. Valeria Simoncini and Daniel B. Szyld. Flexible inner–outer Krylov subspace methods. *SIAM Journal on Numerical Analysis* 2003; **40**:2219–2239.
299. Valeria Simoncini and Daniel B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing* 2003; **25**:454–477.
300. Valeria Simoncini and Daniel B. Szyld. The effect of non-optimal bases on the convergence of Krylov subspace methods. *Numerische Mathematik* 2005; **100**:711–733.
301. Valeria Simoncini and Daniel B. Szyld. On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods. *SIAM Review* 2005; **47**:247–272.
302. Valeria Simoncini and Daniel B. Szyld. Relaxed Krylov subspace approximation. *PAMM*: *Proceedings of Applied Mathematics and Mechanics* 2005; **5**:797–800.
303. Gerard L. G. Sleijpen and Diederik R. Fokkema. BiCGStab(ℓ) for linear equations involving unsymmetric matrices with complex spectrum. *Electronic Transactions on Numerical Analysis* 1993; **1**:11–32.
304. Gerard L. G. Sleijpen and Jasper van den Eshof. Accurate Conjugate Gradient methods for families of shifted systems. *Applied Numerical Mathematics* 2004; **49**:17–37.
305. Gerard L. G. Sleijpen and Jasper van den Eshof. Inexact Krylov subspace methods for linear systems. *SIAM Journal on Matrix Analysis and Applications* 2004; **26**:125–153.
306. Gerard L. G. Sleijpen, Jasper van den Eshof, and Martin B. van Gijzen. Iterative linear solvers with approximate matrix–vector products. In *QCD and Numerical Analysis III*, *Proceedings of the Third International Workshop on Numerical Analysis and Lattice QCD*, Edinburgh, June–July 2003, Boriçi A, Frommer A, Joó B, Kennedy A, Pendleton B (eds). Lecture Notes in Computational Science and Engineering, vol. 47. Springer: Berlin, Heidelberg, 2005; 133–141.
307. Gerard L. G. Sleijpen, Jasper van den Eshof, and Martin B. van Gijzen. Relaxation strategies for nested Krylov methods. *Journal of Computational and Applied Mathematics* 2005; **177**:347–365.
308. Gerard L. G. Sleijpen, Jasper van den Eshof, and Martin B. van Gijzen. Restarted GMRES with inexact matrix–vector products. In *Numerical Analysis and Its Applications*: *Third International Conference*, *NAA 2004*, Rousse, Bulgaria, 29 June–3 July 2004, Li Z, Vulkov L, Waźniewski J (eds). Lecture Notes in Computer Science, vol. 3401. Springer: Heidelberg, 2005; 494–502.
309. Gerard L. G. Sleijpen and Henk A. van der Vorst. Maintaining convergence properties of BiCGstab methods in finite precision arithmetic. *Numerical Algorithms* 1995; **10**:203–223.
310. Gerard L. G. Sleijpen, Henk A. van der Vorst, and Diederik R. Fokkema. BiCGStab(ℓ) and other hybrid Bi-CG methods. *Numerical Algorithms* 1994; **7**:75–109.
311. Gerard L. G. Sleijpen, Henk A. van der Vorst, and Jan Modersitzki. Differences in the effects of rounding errors in Krylov solvers for symmetric indefinite linear systems. *SIAM Journal on Matrix Analysis and Applications* 2000; **22**:726–751.
312. Barry F. Smith, Petter E. Bjørstad, and William D. Gropp. *Domain Decomposition*: *Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press: Cambridge, New York, Melbourne, 1996.
313. Peter Sonneveld. CGS: a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1989; **10**:36–52.
314. Danny C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:357–385.
315. Gerhard Starke. Field-of-values analysis of preconditioned iterative methods for nonsymmetric elliptic problems. *Numerische Mathematik* 1997; **78**:103–117.
316. Gerhard Starke and Richard S. Varga. A hybrid Arnoldi–Faber iterative method for nonsymmetric systems of linear equations. *Numerische Mathematik* 1993; **64**:213–240.
317. G. W. (Pete) Stewart. Backward error bounds for approximate Krylov subspaces. *Linear Algebra and its Applications* 2002; **340**:81–86.
318. Josef Stoer. Solution of large linear systems of equations by Conjugate Gradient type methods. In *Mathematical Programming The State of the Art*. *Proceedings of the 11th International Symposium on Mathematical Programming*, University of Bonn, Bonn, 23–27 August 1982, Bachem A, Gröstchel M, Korte B (eds). Springer: Berlin, New York, 1983; 540–565.

DOI: 10.1002/nla

319. Josef Stoer and Roland W. Freund. On the solution of large linear systems of equations by Conjugate Gradient algorithms. In *Computer Methods in Applied Science and Engineering—V*, Glowinski R, Lions J-L (eds). North-Holland: Amsterdam, 1982; 35–53.
320. Zdeněk Strakoš and Petr Tichý. On error estimation in the Conjugate Gradient method and why it works in finite precision computations. *Electronic Transactions on Numerical Analysis* 2002; **13**:56–80.
321. Zdeněk Strakoš and Petr Tichý. Error estimation in preconditioned Conjugate Gradients. *BIT Numerical Mathematics* 2005; **45**:789–817.
322. Daniel B. Szyld and Judith A. Vogel. A flexible quasi-minimal residual method with inexact preconditioning. *SIAM Journal on Scientific Computing* 2001; **23**:363–380.
323. Daniel B. Szyld and Olof B. Widlund. Variational analysis of some Conjugate Gradient methods. *East–West Journal of Numerical Mathematics* 1993; **1**:51–74.
324. Françoise Tisseur and Karl Meerbergen. The quadratic eigenvalue problem. *SIAM Review* 2001; **43**:235–286.
325. Kim-Chuan Toh. GMRES vs. ideal GMRES. *SIAM Journal on Matrix Analysis and Applications* 1997; **18**:30–36.
326. Andrea Toselli and Olof B. Widlund. *Domain Decomposition Methods—Algorithms and Theory*. Springer Series in Computational Mathematics, vol. 34. Springer: Berlin, Heidelberg, 2005.
327. Lloyd N. Trefethen. Approximation theory and numerical linear algebra. In *Algorithms for Approximation II*. Papers from the *Second International Conference held at the Royal Military College of Science*, Shrivenham, 12–15 July 1988, Mason JC, Cox MG (eds). Chapman & Hall: London, 1990; 336–360.
328. Lloyd N. Trefethen and Mark Embree. *Spectra and Pseudospectra*. Princeton University Press: Princeton, Oxford, 2005.
329. Ulrich Trottenberg, Cornelis W. Oosterlee, and Anton A. Schüller. *Multigrid*. Academic Press: San Diego, London, 2001.
330. Kathryn Turner and Homer F. Walker. Efficient high accuracy solutions with GMRES(m). *SIAM Journal on Scientific and Statistical Computing* 1992; **13**:815–825.
331. Abraham van der Sluis and Henk A. van der Vorst. The rate of convergence of Conjugate Gradients. *Numerische Mathematik* 1986; **48**:543–560.
332. Henk A. van der Vorst. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 1992; **13**:631–644.
333. Henk A. van der Vorst. *Iterative Krylov Methods for Large Linear Systems*. Cambridge University Press: Cambridge, New York, Melbourne, 2003.
334. Henk A. van der Vorst and Jan B. M. Melissen. A Petrov–Galerkin type method for solving $Ax = b$, where A is symmetric complex. *IEEE Transaction on Magnetics* 1990; **26**:706–708.
335. Henk A. van der Vorst and C. (Kees) Vuik. GMRESR: a family of nested GMRES methods. *Numerical Linear Algebra with Applications* 1994; **1**:369–386.
336. Sabine Van Huffel and Joos Vandewalle. *The Total Least Squares Problem*, *Computational Aspects and Analysis*. Frontiers in Applied Mathematics, vol. 9. SIAM: Philadelphia, PA, 1991.
337. Judith A. Vogel. Flexible BiCG and flexible Bi-CGSTAB for nonsymmetric linear systems. *Applied Mathematics and Computations*, in press.
338. C. (Kees) Vuik. Further experiences with GMRESR. *Supercomputer* 1993; **55**:13–27.
339. C. (Kees) Vuik. New insights in GMRES-like methods with variable preconditioners. *Journal of Computational and Applied Mathematics* 1995; **61**:189–204.
340. C. (Kees) Vuik and Henk A. van der Vorst. A comparison of some GMRES-like methods. *Linear Algebra and its Applications* 1990; **160**:131–162.
341. Homer F. Walker. Residual smoothing and peak/plateau behaviour in Krylov subspace methods. *Applied Numerical Mathematics* 1995; **19**:279–286.
342. Homer F. Walker and Lu Zhou. A simpler GMRES. *Numerical Linear Algebra with Applications* 1994; **1**:571–581.
343. James S. Warsa, Michele Benzi, Todd A. Wareing, and Jim E. Morel. Preconditioning a mixed discontinuous finite element method for radiation diffusion. *Numerical Linear Algebra with Applications* 2004; **11**:377–389.
344. James S. Warsa, Todd A. Wareing, Jim E. Morel, John M. McGhee, and Richard B. Lehoucq. Krylov subspace iterations for deterministic k-eigenvalue calculations. *Nuclear Science and Engineering* 2004; **147**:26–42.
345. Rüdiger Weiss. Error-minimizing Krylov subspace methods. *SIAM Journal on Scientific Computing* 1994; **15**:511–527.
346. Rüdiger Weiss. *Parameter-Free Iterative Linear Solvers*. Mathematical Research Series, vol. 97. Akademie-Verlag: Berlin, 1996.

DOI: 10.1002/nla

347. Olof B. Widlund. A Lanczos method for a class of nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis* 1978; **15**:801–812.
348. Jinchao Xu and Xiao-Chuan Cai. A preconditioned GMRES method for nonsymmetric or indefinite problems. *Mathematics of Computations* 1992; **59**:311–319.
349. Ilyia Zavorin, Howard C. Elman, and Dianne P. O'Leary. Complete stagnation of GMRES. *Linear Algebra and its Applications* 2003; **367**:165–183.
350. Shao-Liang Zhang. GPBi-CG: generalized product-type methods based on Bi-CG for solving nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 1997; **18**:537–551.
351. Baojiang Zhong and Ronald B. Morgan. Complementary cycles of restarted GMRES. *Technical Report*, Department of Mathematics, Baylor University, Waco, Texas, 2004.
352. Lou Zhou and Homer F. Walker. Residual smoothing techniques for iterative methods. *SIAM Journal on Scientific Computing* 1994; **15**:297–312.

DOI: 10.1002/nla