

TUNING IDR TO FIT YOUR APPLICATIONS*

OLAF RENDEL† AND JENS-PETER M. ZEMKE‡

Abstract. We focus on IDR methods for the computation of approximate solutions to linear systems and the partial eigenvalue problem. IDR methods offer a variety of parameters that can be tuned to the application. Most important among these seem to be the choices of the shadow residuals and the roots of the stabilizing/damping polynomials. We report on some findings for a few members of the large IDR family of methods, namely, some QMRIDR variants and methods based on IDRStab.

Key words. IDR; IDR(s); eigenvalues; Krylov subspace methods.

AMS subject classifications. 65F15 (primary); 65F10; 65F50

1. Introduction & History. Krylov subspace methods are named after the Russian naval engineer Алексей Николаевич Крылов (Aleksei Nikolaevich Krylov), who in 1931 wrote a paper on a method to compute the coefficients of the characteristic polynomial of a matrix, cf. [25]. This was based on certain subspaces nowadays known as *Krylov subspaces*. In 1940 the German engineer and mathematician Karl Hessenberg developed the first modern Krylov subspace method [20]. Yet, the best known Krylov subspace methods are those based on the methods of Lanczos [26, 27] and Arnoldi [1]. The first IDR method [54], in contrast, was developed at the end of the 70s of the last century; the more recent IDR(s) methods, starting with [45], soon superseded by [51] have been delevoped a few years ago. The original IDR variant was a predecessor to both CGS [43] and BiCGStab [49, 48], which are both also members of the vast IDR family. These methods were soon generalized to other so-called transpose free, hybrid BiCG, or Lanczos-type product methods, which are also related to the IDR family. We list as most important generalizations BiCGStab2 [17], BiCGStab(ℓ) [40], GCGS [10] (in particular CGS2 and shifted CGS), GPBiCG [59], BiCGSafe [16], TFQMR [13], QMRCGStab [3], TFIQMR and TFILanczos [2], and last but not least, ML(k)BiCGStab [55]. All these methods rely on the Lanczos method and can break down for several reasons. The composite step methods [4, 5] avoid one type of breakdown. Modifications like multi-shift TFQMR [12], multi-shift BiCGStab(ℓ) [15], and flexible BiCGStab [52] (the latter no longer spans Krylov subspaces) have also been developed. The development of IDR by Peter Sonneveld in 1976–1980 had a major impact on the subsequent development of Krylov subspace methods.

The recent development of IDR family members of IDR based on larger shadow spaces, i.e., with dimension $s \in \mathbb{N}$, e.g., IDR(s) [45, 51], offers advantages: these appear to be more stable than the original IDR variant *and* BiCGStab *and* most of its relatives. IDR(s) is closely related to ML(k)BiCGStab; BiCGStab-like implementations of IDR(s) are possible, cf. [41]. To every Krylov-based linear system solver a Krylov-based eigenvalue solver corresponds, this correspondence was worked out in [19] for the prototype IDR(s) [45]; the eigenvalue counterpart of the enhanced version [51] has been considered in the bachelor's thesis [32]. Incorporation of the basic principles in IDR(s) and BiCGStab(ℓ) has been considered independently in GIDR(s, L) [46] and GBiCGStab(s, L) [47], and IDR(s)Stab(ℓ) [42], we group these methods under the descriptive name IDRStab. The eigenvalue counterpart of IDRStab is considered in [33]. Some of the techniques that have been applied to the original IDR and derived methods have already been applied in the IDR(s) context, we mention

*Version of October 11, 2011, 11:53.

†Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, D-21073 Hamburg, Germany (olaf.rendel@tu-harburg.de).

‡Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, D-21073 Hamburg, Germany (zemke@tu-harburg.de).

QMRIDR [8] and flexible and multi-shift QMRIDR [50]. The impact of IDR(s) on the subsequent development of new methods seems to be almost as large as was the case for the original IDR, even though the development of new variants takes place more rapidly, which we intend to explain in this note.

We simply term all methods form the IDR family, e.g., original IDR, BiCGStab, IDR(s) and IDRStab amongst others, as *IDR methods*, and, for reasons explained below, *Sonneveld methods*. Most Sonneveld methods developed thus far are Krylov subspace methods. These are all linked to Lanczos processes, those based on larger shadow spaces to a process we term Lanczos($s, 1$), which is based on a left block Krylov subspace and a simple right Krylov subspace. Even though this link explains some of the details, it does not account for all the subtleties associated with Sonneveld methods.

In this note we sketch some aspects of how to chose the free parameters in Sonneveld methods. We focus mainly on the shadow vectors and the roots of the stabilizing polynomials.

1.1. Motivation. The IDR family of methods, e.g., when the focus is on those methods based on a larger shadow space, appears to be a quite recent development. In this note we recall that there are intimate connections to the other well-known Krylov subspace methods and showcase a general principle to transfer Krylov subspace method techniques to the IDR setting. Almost all IDR methods have been derived by, or based on, work by Peter Sonneveld, for this reason from now on we will term them *Sonneveld methods*.

1.2. Outline. We present a short introduction to Sonneveld methods and highlight the similarities with and differences to the methods of Lanczos [26, 27] and Arnoldi [1] (which has some similarities with the method by Hessenberg [20]). The similarities form the basis to present the application of different techniques to Sonneveld methods and to extend some results, which are well-known for the classical methods, but have only partially been presented for Sonneveld methods. We state an observation by Sonneveld on the convergence of Sonneveld methods for linear systems [44] and extend this observation to the eigenvalue case. This gives a hint how to choose the shadow vectors. The connection to Lanczos methods is used to present some ideas to choose the stabilizing polynomials.

1.3. Notation. We use standard notation. The identity matrix of size $n \times n$ is denoted by $\mathbf{I} = \mathbf{I}_n$, its column vectors by $\mathbf{e}_j$ and its elements by the Kronecker delta δ_{ij}. The vector of the sums of all columns, i.e., the vector of all ones, is denoted by $\mathbf{e}$. The matrix $\mathbf{O} = \mathbf{O}_n$ denotes the zero matrix of size $n \times n$, the zero column vector of length n is denoted by $\mathbf{o} = \mathbf{o}_n$. The sizes are omitted if easily deducible from the context. We are interested in the properties of a square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, e.g., its inverse and/or some of its eigenvalues. Unreduced Hessenberg matrices are denoted by letter $\mathbf{H}_k \in \mathbb{C}^{k\times k}$, upper triangular matrices by letter $\mathbf{U}_k \in \mathbb{C}^{k\times k}$. Extended counterparts of $\mathbf{I}_k$, $\mathbf{H}_k$, and $\mathbf{U}_k$ exist, which are denoted by $\underline{\mathbf{I}}_k$, $\underline{\mathbf{H}}_k$, and $\underline{\mathbf{U}}_k$, respectively. The rectangular matrices $\underline{\mathbf{I}}_k \in \mathbb{C}^{(k+1)\times k}$ and $\underline{\mathbf{U}}_k \in \mathbb{C}^{(k+1)\times k}$ are obtained by appending a row of zeros at the bottom, $\underline{\mathbf{H}}_k$ is an unreduced extended Hessenberg matrix that has $\mathbf{H}_k$ as leading square part. The columns of $\underline{\mathbf{I}}_k$ are denoted by $\underline{\mathbf{e}}_j$. The vector of all ones of length $k+1$ is denoted by $\underline{\mathbf{e}} \in \mathbb{C}^{k+1}$. The inverse, transpose, complex conjugate transpose, and pseudo-inverse (or Moore-Penrose inverse) is denoted by appending $^{-1}$, $^{\mathsf{T}}$, $^{\mathsf{H}}$, and †, respectively.

We remark that like in [19] we use a simplified way to denote Krylov subspace methods, e.g., we write GMRes in place of GMRES as is done in the original publication, since the acronym stands for the phrase Generalized Minimal RESidual.

2. Classical Krylov subspace methods. Starting with Hessenberg's original method [20], most essentials of the classical Krylov subspace methods for $\mathbf{A} \in \mathbb{C}^{n\times n}$

can be captured by a so-called *Hessenberg decomposition* [19]

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k, \quad k \in \mathbb{N}, \quad k < n, \tag{2.1}$$

where $\mathbf{Q}_{k+1} = \big(\mathbf{Q}_k, \mathbf{q}_{k+1}\big) = \big(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{k+1}\big) \in \mathbb{C}^{n\times(k+1)}$ accounts for the basis vectors $\mathbf{q}_j$, $1 \leqslant j \leqslant k+1$, produced to span the $(k+1)$st Krylov subspace ($\mathbf{q} := \mathbf{q}_1$)

$$\mathcal{K}_{k+1} := \mathcal{K}_{k+1}(\mathbf{A}, \mathbf{q}) := \mathsf{span}\,\{\mathbf{q}, \mathbf{A}\mathbf{q}, \mathbf{A}^2\mathbf{q}, \ldots, \mathbf{A}^k\mathbf{q}\} = \mathsf{span}\,\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{k+1}\}, \tag{2.2}$$

and an unreduced extended Hessenberg matrix $\underline{\mathbf{H}}_k \in \mathbb{C}^{(k+1)\times k}$ that in some manner collects information about the action of the operator $\mathbf{A}$ on the Krylov subspace $\mathcal{K}_k(\mathbf{A}, \mathbf{q})$. More precisely, as we do assume that $\mathsf{rank}\,(\mathbf{Q}_{k+1}) = k+1$, $\underline{\mathbf{H}}_k$ can be thought of as a rectangular projection of $\mathbf{A}$,

$$\mathbf{Q}_{k+1}^{\dagger}\mathbf{A}\mathbf{Q}_k = \underline{\mathbf{H}}_k. \tag{2.3}$$

2.1. Lanczos's method. Like Hessenberg's method [20], Lanczos's method [26, 27] is based on bi-orthogonality. In contrast to Hessenberg, who uses the standard unit vectors $\mathbf{e}_j$ as left vectors, Lanczos does not use a fixed set of vectors, but instead for this purpose computes a basis of a left Krylov subspace

$$\widehat{\mathcal{K}}_{k+1} := \mathcal{K}_{k+1}(\widehat{\mathbf{A}}, \widehat{\mathbf{q}}) = \mathsf{span}\,\{\widehat{\mathbf{q}}, \widehat{\mathbf{A}}\widehat{\mathbf{q}}, \widehat{\mathbf{A}}^2\widehat{\mathbf{q}}, \ldots, \widehat{\mathbf{A}}^k\widehat{\mathbf{q}}\} = \mathsf{span}\,\{\widehat{\mathbf{q}}_1, \widehat{\mathbf{q}}_2, \ldots, \widehat{\mathbf{q}}_{k+1}\}, \tag{2.4}$$

where $\widehat{\mathbf{A}}$ is the adjoint in some bilinear or sesquilinear form. For ease of presentation, we think of $\widehat{\mathbf{A}}$ being the Hermitean adjoint, i.e., $\widehat{\mathbf{A}} = \mathbf{A}^{\mathsf{H}}$, and the form being the usual inner product in $\mathbb{C}^n$. The method of Lanczos computes (formally) bi-orthogonal bases of $\widehat{\mathcal{K}}_k$ and $\mathcal{K}_k$ via some two-sided Gram-Schmidt process. As all vectors in the Krylov subspace $\mathcal{K}_k$ correspond to polynomials in $\mathbf{A}$,

$$\mathbf{q}_k \in \mathcal{K}_k \quad \Rightarrow \quad \mathbf{q}_k = \sum_{j=1}^{k} \mathbf{A}^{j-1}\mathbf{q}\alpha_j = p_{k-1}(\mathbf{A})\mathbf{q}, \tag{2.5}$$

the products of left vectors $\sum_{j=1}^{k} \widehat{\mathbf{A}}^{j-1}\widehat{\mathbf{q}}\widehat{\alpha}_j$ and right vectors $\sum_{j=1}^{k} \mathbf{A}^{j-1}\mathbf{q}\alpha_j$ can be expressed using solely the so-called moments of $\mathbf{A}$, compare with [26, § V., Eqn. (34) p. 258]:

$$\langle \widehat{\mathbf{A}}^i\widehat{\mathbf{q}}, \mathbf{A}^j\mathbf{q}\rangle = \langle \widehat{\mathbf{q}}, \mathbf{A}^{i+j}\mathbf{q}\rangle = c_{i+j}, \quad 0 \leqslant i, j \leqslant n. \tag{2.6}$$

Lanczos derives a three-term recurrence for the solutions $\boldsymbol{\eta}_k = \big(\eta_{0,k}, \ldots, \eta_{k-1,k}\big)^{\mathsf{T}}$ to the Hankel systems

$$\begin{pmatrix} c_0 & c_1 & \cdots & c_k \\ c_1 & c_2 & \cdots & c_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_k & c_{k+1} & \cdots & c_{2k} \end{pmatrix} \begin{pmatrix} \eta_{0,k} \\ \vdots \\ \eta_{k-1,k} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ h_k \end{pmatrix} \tag{2.7}$$

with a certain $h_k \in \mathbb{C}$ that eventually becomes zero, see [26, § VI., Eqn. (50), p. 262]. This "progressive" form of his algorithm is the basis for his method of "minimized iterations" [26, § VII., pp. 265–268], which is the well-known reduction to tridiagonal form by means of a two-sided Gram-Schmidt process. The quantities constructed in this algorithm can be captured with two Hessenberg decompositions

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k, \quad \widehat{\mathbf{A}}\widehat{\mathbf{Q}}_k = \widehat{\mathbf{Q}}_{k+1}\underline{\widehat{\mathbf{T}}}_k, \quad \text{where} \quad \widehat{\mathbf{T}}_k = \mathbf{T}_k^{\mathsf{H}}. \tag{2.8}$$

Here, $\mathbf{T}_k \in \mathbb{C}^{k\times k}$ and $\widehat{\mathbf{T}}_k \in \mathbb{C}^{k\times k}$ denote the leading square parts of the unreduced extended tridiagonal (i.e., Hessenberg) matrices $\underline{\mathbf{T}}_k$ and $\underline{\widehat{\mathbf{T}}}_k$, respectively. More interesting to us is the anticipatory remark to be found as footnote 11 on page 263 of [26]:

> "Instead of iterating with $\mathbf{A}$ and $\widehat{\mathbf{A}}$ n times, we can also iterate with $\mathbf{A}$ alone $2n$ times. Any of the columns of the iteration matrix can now be chosen as c_i numbers since these columns correspond to a dotting of the iteration matrix with $\widehat{\mathbf{q}}_1 = 1, 0, 0, \ldots$, respectively $0, 1, 0, 0, \ldots$; $0, 0, 1, 0, 0, \ldots$; and so on. The transposed matrix is not used here at all. E. C. Bower of the Douglas Aircraft Co. points out to the author that from the machine viewpoint a uniform iteration scheme of $2n$ iterations is preferable to a divided scheme of $n + n$ iterations. The divided scheme has the advantage of less accumulation of rounding errors and more powerful checks on the successive iterations. The uniform scheme has the advantage that more than one column is at our disposal. Accidental deficiencies of the $\widehat{\mathbf{q}}_1$ vector can thus be eliminated, by repeating the algorithm with a different column. (For this purpose it is of advantage to start with the trial vector $\mathbf{q}_1 = 1, 1, 1, \ldots, 1.$) In the case of a *symmetric* matrix it is evident that after n iterations the basic scalars should be formed, instead of continuing with n more iterations."

The mentioned "uniform scheme" is possible because of the property of the inner products defining the moments as sketched in Eqn. (2.6).

This quote is more or less a precise description of what is to be expected from Sonneveld methods, which are introduced in the next section. One point of view on these methods is that they are based on the trick to rewrite all inner products such that the polynomials are strictly confined to the right-hand side, as already utilized in Eqn. (2.6). Interestingly, Sonneveld methods have not been derived like that, even though this technique was cleverly used to derive CGS [43] from BiCG [27, 9] and forms the basis of most known hybrid BiCG methods, most prominent amongst them BiCGStab [49, 48].

We remark for later use that there also exist variants of Lanczos's method for more than one left and more than one right starting vector, e.g., so-called block Lanczos's methods, when the number of left and right starting vectors is equal. In the analysis of IDR we are especially interested in Lanczos's methods where the number of left-hand starting vectors is given by any $s \in \mathbb{N}$ and there is only *one* right-hand starting vector. This latter family of Lanczos's method will be referred to as Lanczos$(s, 1)$.

3. Sonneveld methods. Sonneveld, or IDR, methods are based on the so-called IDR Theorem. This theorem is about a nested sequence of spaces, where in contrast to the case in ordinary Krylov subspaces, where the dimensions only can increase with increasing indices, the dimensions of consecutive spaces are shrinking. These IDR spaces, a special case of Sonneveld subspaces [42, Definition 2.2, p. 2690], are defined as follows. Let the subspace $\mathcal{G}_0$ be the full Krylov subspace $\mathcal{K}(\mathbf{A}, \mathbf{q}) = \mathcal{K}_n(\mathbf{A}, \mathbf{q})$:

$$\mathcal{G}_0 = \mathcal{K}(\mathbf{A}, \mathbf{q}) = \mathcal{K}_n(\mathbf{A}, \mathbf{q}) = \mathsf{span}\,\{\mathbf{q}, \mathbf{A}\mathbf{q}, \ldots, \mathbf{A}^{n-1}\mathbf{q}\} \subset \mathbb{C}^n. \tag{3.1}$$

In case of non-derogatory $\mathbf{A} \in \mathbb{C}^{n\times n}$ and a generic starting vector $\mathbf{q} \in \mathbb{C}^n$, $\mathcal{G}_0 = \mathbb{C}^n$. Cases exist where $\mathcal{G}_0 \subsetneq \mathbb{C}^n$. Starting from $\mathcal{G}_0$, the Sonneveld spaces $\mathcal{G}_j$ are recursively defined by

$$\mathcal{G}_j = g_j(\mathbf{A})(\mathcal{G}_{j-1} \cap \mathcal{S}), \quad g_j(z) = \eta_j z + \mu_j, \quad \eta_j, \mu_j \in \mathbb{C},\ \eta_j \neq 0, \quad j = 1, 2, \ldots. \tag{3.2}$$

Here, $\mathcal{S}$ is a space of codimension $s \in \mathbb{N}$. The IDR Theorem is given as follows:

THEOREM 3.1 (IDR Theorem [45]). *Under mild conditions on the matrices* $\mathbf{A}$ *and the space* $\mathcal{S}$*,*

(i) $\mathcal{G}_j \subsetneq \mathcal{G}_{j-1}$ *for all* $\mathcal{G}_{j-1} \neq \{\mathbf{o}_n\}$, $j > 0$.
(ii) $\mathcal{G}_j = \{\mathbf{o}_n\}$ *for some* $j \leqslant n$.

For the proof we refer to [45, 38].

At first glance IDR seems to be detached from Lanczos's methods, let alone from Lanczos$(s, 1)$. The analysis contained in [18, 37, 19] and especially in [38] clarifies

the connections to Krylov subspace methods. We state the alternate description of Sonneveld spaces from [38], similarly to [50]. Let the columns of $\widetilde{\mathbf{R}}_0$ form a basis of $\mathcal{S}^{\perp}$. We can characterize the IDR Sonneveld spaces in terms of the orthogonal complement of left block Krylov subspaces

$$\mathcal{K}_j(\mathbf{A}^{\mathsf{H}}, \widetilde{\mathbf{R}}_0) = \Big\{ \sum_{i=0}^{j-1} (\mathbf{A}^{\mathsf{H}})^i \widetilde{\mathbf{R}}_0 \mathbf{c}_i \mid \mathbf{c}_i \in \mathbb{C}^s \Big\} \tag{3.3}$$

as

$$\mathcal{G}_j = \{ G_j(\mathbf{A})\mathbf{v} \mid \mathbf{v} \perp \mathcal{K}_j(\mathbf{A}^{\mathsf{H}}, \widetilde{\mathbf{R}}_0) \}, \quad \text{where} \quad G_j(z) = \prod_{i=1}^{j} g_i(z), \tag{3.4}$$

see [38, Theorem 11, p. 1104]. We already noted in [50] that numerical experiments indicate that the "local closeness" of this Lanczos process to an unperturbed one is the driving force behind IDR based methods. This is anticipated in [39, p. 204]:

> "[..], we expect to recover the convergence behavior of the incorporated Bi-CG process (in the BiCGstab methods) if we compute the iteration coefficients as accurately as possible. Therefore, we want to avoid all additional perturbations that might be introduced by an unfortunate choice of the polynomial process that is carried out on top of the Bi-CG process."

Implementations of Sonneveld methods are usually based on as little vectors as possible, e.g., we only pick sufficiently many vectors in each space $\mathcal{G}_j$ to ensure the next ones to be in the smaller subspace $\mathcal{G}_{j+1}$. Several implementations, each adding new ideas, have been used thus far, see, e.g., [45, 51, 50]. The latter is one of the most advanced schemes, but still not all freedom is exploited. The "best" scheme to compute a stable basis using the recursion (3.2) has yet to be discovered; this is an active area of research. The recursion (3.2) translates in the generic case into a vector recurrence as follows:

Initialization: compute $s+1$ basis vectors $\mathbf{g}_i$, $1 \leqslant i \leqslant s+1$, in $\mathcal{K}_{s+1} \subset \mathcal{G}_0$ using your favorite Krylov subspace method. We advocate the use of Arnoldi/GMRes.

Recursion: for $j > 0$ until convergence perform the following:

- **Intersection:** compute a linear combination $\mathbf{v}_i$ of vectors in $\mathcal{G}_{j-1} \cap \mathcal{S}$. Typically, $s+1$ vectors are used in this stage, mostly the newest vectors, as is done in [45, 51, 47, 50], or a fixed set of s vectors for several steps and one that changes, as is done in [42]. Numerically, using *all* vectors available is more robust (but more costly); this was observed in several experiments.
- **Update:** if constructing the first vector in a new space $\mathcal{G}_j$, chose a new linear polynomial g_j of exact degree 1. Here, the remarks on the accuracy of the Lanczos coefficients apply, the techniques from [39] find here applications. Alternatives to minimization include the use of eigenvalue information, either using another Krylov subspace method [37] or the purified Sonneveld pencil of the Sonneveld method [19].
- **Map:** compute the new vector $g_j(\mathbf{A})\mathbf{v}_i$ in $\mathcal{G}_j$ and compute a new basis vector as linear combination of $g_j(\mathbf{A})\mathbf{v}_i$ with other vectors in $\mathcal{G}_j$. The first vector is essentially unique up to scaling; experiments show that computing linear combinations increases the numerical stability significantly.

Sonneveld Krylov subspace methods can be described by a so-called *generalized Hessenberg decomposition* [19]

$$\mathbf{A}\mathbf{V}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \quad k \in \mathbb{N}, \quad k < n, \tag{3.5}$$

where $\mathbf{V}_k = \mathbf{G}_k\mathbf{U}_k$, with $\mathbf{U}_k \in \mathbb{C}^{k\times k}$ upper triangular, and all other matrices are defined like in Eqn. (2.1). We only changed the letter for the matrix capturing the

basis vectors from $\mathbf{Q}$ to $\mathbf{G}$ to reflect that these columns are selected vectors $\mathbf{g}_i$ from the spaces $\mathcal{G}_j$, $j \geqslant 0$.

The matrix $\mathbf{V}_k$ captures all the linear combinations of $\mathbf{g}_i$ vectors in the pre-image of $g_j(\mathbf{A})$ times η_j, the coefficients of these linear combinations can be found in the columns of the upper triangular and banded $\mathbf{U}_k$ and the unreduced extended Hessenberg and banded $\underline{\mathbf{H}}_k$, possibly scaled differently ($\eta_j + \mu_j \neq 0$) and mingled in the latter with the coefficients of the linear combinations in the $\mathcal{G}_j$ spaces. We remark that changes in the polynomial structurally shift (parts of) block columns in $\mathbf{U}_k$ and $\underline{\mathbf{H}}_k$. This feature is of interest in MR Sonneveld methods, which are considered in the next section, as the understanding of this behavior helps to prevent incurable stagnation.

Again we can interpret the matrices $\mathbf{U}_k$ and $\underline{\mathbf{H}}_k$ of coefficients as a kind of projection if $\mathsf{rank}\,(\mathbf{G}_{k+1}) = k+1$, this time of the pencil $(\mathbf{A}, \mathbf{I}_n)$ to the rectangular pencil $(\underline{\mathbf{H}}_k, \underline{\mathbf{U}}_k)$, where $\underline{\mathbf{U}}_k \in \mathbb{C}^{(k+1)\times k}$ denotes $\mathbf{U}_k$ with one additional zero row appended at the bottom:

$$\begin{aligned}\mathbf{G}_{k+1}^\dagger(\mathbf{A},\mathbf{I}_n)\mathbf{V}_k &= \mathbf{G}_{k+1}^\dagger(\mathbf{A},\mathbf{I}_n)\mathbf{G}_{k+1}\underline{\mathbf{U}}_k \\ &= \mathbf{G}_{k+1}^\dagger(\mathbf{A}\mathbf{G}_k\mathbf{U}_k, \mathbf{G}_{k+1}\underline{\mathbf{U}}_k) = \mathbf{G}_{k+1}^\dagger(\mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \mathbf{G}_{k+1}\underline{\mathbf{U}}_k) = (\underline{\mathbf{H}}_k, \underline{\mathbf{U}}_k). \quad (3.6)\end{aligned}$$

The derivation of Sonneveld methods gives them an unfamiliar appeal to people from the Krylov subspace community. The theoretical investigations show that Sonneveld methods are just another branch of Krylov subspace methods. We show in this note that the small change from the Hessenberg decomposition (2.1) to the *generalized* Hessenberg decomposition (3.5), i.e., the introduction of an upper triangular $\mathbf{U}_k$ in place of the simple $\mathbf{I}_k$ in classical Krylov subspace methods, is the main change in devising new algorithms or applying well-known techniques from the pool of existing Krylov subspace method techniques.

We start with very simple descriptions of some basic Krylov subspace flavors; we obtain the "classical" technique upon setting $\mathbf{U}_k$ to the identity matrix $\mathbf{I}_k$. This amounts to almost trivial changes when a "classical" Krylov subspace technique is applied to Sonneveld methods.

4. Ritz, harmonic Ritz, and the like. There are many different techniques available to extract approximations to some eigenpairs using Krylov subspace techniques, we sketch very briefly some of them. Best known is the Ritz approach. Here, the leading square part $(\mathbf{H}_k, \mathbf{U}_k)$ of the rectangular Sonneveld pencil [19] $(\underline{\mathbf{H}}_k, \underline{\mathbf{U}}_k)$ is used to compute approximate eigenvalues θ_j, the eigenvectors $\mathbf{s}_j$,

$$\mathbf{H}_k\mathbf{s}_j = \theta_j\mathbf{U}_k\mathbf{s}_j, \quad (4.1)$$

are prolonged to give the Ritz vectors $\mathbf{y}_j := \mathbf{V}_k\mathbf{s}_j = \mathbf{G}_k\mathbf{U}_k\mathbf{s}_j$. This is discussed for the prototype IDR(s) [45] in [19], for the enhanced version [51] in [32], and for IDRStab [42] in [33]. As some of these eigenvalues correspond to roots of the polynomials G_i from Eqn. (3.4), in all cases a purified and/or deflated pencil is used, for details we refer to [19, 32, 33].

On the same grounds we can compute harmonic Ritz values associated with the MR approach [29, 11]. The motivation for the name harmonic Ritz and the connection to Lehmann's optimal eigenvalue inclusions [28] can be found in [30]. In the context of Sonneveld methods we compute the eigenvalues $\underline{\theta}_j$ and the eigenvectors $\underline{\mathbf{s}}_j$ of one of the equivalent pairs $(\underline{\mathbf{H}}_k^{\mathsf{H}}\underline{\mathbf{H}}_k, \underline{\mathbf{H}}_k^{\mathsf{H}}\underline{\mathbf{U}}_k)$ or $(\mathbf{I}_k, \underline{\mathbf{H}}_k^\dagger\underline{\mathbf{U}}_k)$, e.g.,

$$\mathbf{I}_k\underline{\mathbf{s}}_j = \underline{\theta}_j\underline{\mathbf{H}}_k^\dagger\underline{\mathbf{U}}_k\underline{\mathbf{s}}_j, \quad (4.2)$$

to obtain the harmonic Ritz pairs $(\underline{\theta}_j, \underline{\mathbf{y}}_j)$, $\underline{\mathbf{y}}_j := \mathbf{V}_k\underline{\mathbf{s}}_j = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{s}}_j$. Again, it is preferable to apply this scheme to a purified and/or deflated pencil, as in this case we otherwise include information about the roots of some G_i in the process and smear

out this information, such that we can no longer separate information about the eigenvalues of $\mathbf{A}$ and these (unwanted) roots.

Similarly, more general eigenpair extraction schemes are possible, e.g., refinement [22, 23, 24] and quasi-optimal extraction [58]. The details will be given in a separate report.

5. Polynomials and perturbations. Perturbed methods will satisfy an approximate generalized Hessenberg decomposition

$$\mathbf{A}\mathbf{V}_k + \mathbf{F}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k + \mathbf{F}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \quad k \in \mathbb{N}, \tag{5.1}$$

where $\mathbf{F}_k \in \mathbb{C}^{n\times k}$ accounts for perturbations, e.g., due to finite precision and/or inexact matrix-vector multiplies. The impact of the perturbations is very different in Sonneveld and Lanczos methods, Lanczos methods tend to compute multiple Ritz copies, Sonneveld methods tend to compute what we call *ghost polynomial roots*. As an example, some of the main differences between BiCG perturbed by finite precision and as underlying method of BiCGStab are depicted in Figure 5.1. This figure and the corresponding caption is an excerpt of an experiment from a forthcoming report.

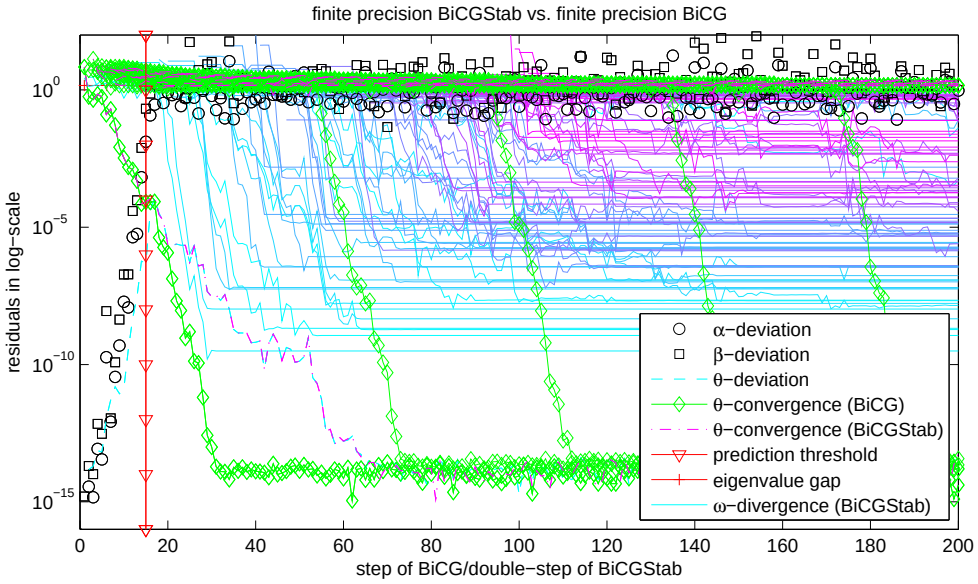

FIG. 5.1. *Approximation of the eigenvalue of maximal magnitude of a well-conditioned matrix* $\mathbf{A} \in \mathbb{C}^{200\times 200}$ *by Ritz values using* BiCG *and* BiCGStab *in the naïve implementation in ANSI/IEEE 754 arithmetic. The differences of the constructed* α_j *and* β_j *are marked with circles and squares, respectively. Any forward error analysis could predict the behaviour up to the total loss of common digits, depicted as prediction threshold. The convergence of the* BiCG*-Ritz values (green lines w. diamonds; multiple copies) and the* BiCGStab*-Ritz values (magenta dash-dotted lines) towards the eigenvalue of largest magnitude deviates after this point.* BiCG *returns multiple copies of Ritz values;* BiCGStab *computes spurious Ritz values close to the values* $1/\omega_j$*. To illustrate the latter the convergence of* BiCGStab*-Ritz values towards the first* 102 *values* $1/\omega_j$*,* $2 \leqslant j \leqslant 103$ *is depicted by colored straight lines (obtained using* MATLAB*'s color scheme "cold").*

A first step to analyze this startling behaviour is the observation given in the report [6] that in Sonneveld methods we implicitly work with a modified matrix that has these roots as eigenvalues. Another tool of trade may be the obvious generalization stated for the unperturbed case in [19, §2] of the polynomial expressions obtained in [57] with the aid of [56]. This is currently an active area of research.

6. OR and MR. Krylov subspace methods for the solution of linear systems

$$\mathbf{A}\mathbf{x} = \mathbf{r}_0, \quad \mathbf{A} \in \mathbb{C}^{n\times n}, \quad \mathbf{r}_0 \in \mathbb{C}^n \tag{6.1}$$

can be mainly distinguished into two classes: those of type OR and those of type MR. We use the classification given in [57]. OR is an abbreviation for Orthogonal

Residuals, MR is an abbreviation for Minimal Residuals. This classification easily extends to Sonneveld methods. In all cases we assume that $\mathbf{g}_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.

Orthogonal Residual (OR) approaches, the well-known ones are CG [21] and FOM [34, 35], are based on the solution of Hessenberg linear systems with the leading square part $\mathbf{H}_k \in \mathbb{C}^{k\times k}$ of $\underline{\mathbf{H}}_k \in \mathbb{C}^{(k+1)\times k}$. In the case of Sonneveld methods we define the kth OR solution $\mathbf{z}_k$ to be any solution of

$$\mathbf{H}_k\mathbf{z}_k := \mathbf{e}_1\|\mathbf{r}_0\|, \quad \text{e.g., mostly} \quad \mathbf{z}_k := \mathbf{H}_k^{-1}\mathbf{e}_1\|\mathbf{r}_0\|, \tag{6.2}$$

and set the kth OR iterate $\mathbf{x}_k$ to

$$\mathbf{x}_k := \mathbf{V}_k\mathbf{z}_k = \mathbf{G}_k\mathbf{U}_k\mathbf{z}_k. \tag{6.3}$$

The norm of the kth residual $\mathbf{r}_k := \mathbf{r}_0 - \mathbf{A}\mathbf{x}_k$ can be estimated by

$$\begin{aligned}\|\mathbf{r}_k\| &= \|\mathbf{r}_0 - \mathbf{A}\mathbf{x}_k\| = \|\mathbf{r}_0 - \mathbf{A}\mathbf{G}_k\mathbf{U}_k\mathbf{z}_k\| \\ &= \|\mathbf{G}_k(\mathbf{e}_1\|\mathbf{r}_0\| - \mathbf{H}_k\mathbf{z}_k) - \mathbf{g}_{k+1}h_{k+1,k}\mathbf{e}_k^{\mathsf{T}}\mathbf{z}_k\| = \|\mathbf{g}_{k+1}\| \cdot |h_{k+1,k}\mathbf{e}_k^{\mathsf{T}}\mathbf{z}_k|,\end{aligned} \tag{6.4}$$

i.e., using solely quantities available in the algorithm. OR methods are mostly simple to implement, yet they can break down whenever $\mathbf{H}_k$ becomes singular. In some of these methods, the numerical stability suffers in case of near-breakdown. Most IDR methods developed thus far are of type OR, e.g., the methods given in [45, 51, 42] are all of type OR.

Minimal Residual (MR) approaches, the well-known ones are MinRes [31], GMRes [35], and QMR [14], compute an approximation based on the solution of Hessenberg least-squares problems. In the case of Sonneveld methods we define the kth MR solution $\underline{\mathbf{z}}_k$ to be the unique solution of

$$\rho_k := \left\|\underline{\mathbf{H}}_k\underline{\mathbf{z}}_k - \underline{\mathbf{e}}_1\|\mathbf{r}_0\|\right\| = \min, \quad \text{i.e.,} \quad \underline{\mathbf{z}}_k := \underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{e}}_1\|\mathbf{r}_0\|, \tag{6.5}$$

and set the kth MR iterate $\underline{\mathbf{x}}_k$ to

$$\underline{\mathbf{x}}_k := \mathbf{V}_k\underline{\mathbf{z}}_k = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{z}}_k. \tag{6.6}$$

The norm of the kth residual $\underline{\mathbf{r}}_k := \mathbf{r}_0 - \mathbf{A}\underline{\mathbf{x}}_k$ can be estimated by

$$\begin{aligned}\|\underline{\mathbf{r}}_k\| &= \|\mathbf{r}_0 - \mathbf{A}\underline{\mathbf{x}}_k\| = \|\mathbf{r}_0 - \mathbf{A}\mathbf{G}_k\mathbf{U}_k\underline{\mathbf{z}}_k\| \\ &= \|\mathbf{G}_{k+1}(\underline{\mathbf{e}}_1\|\mathbf{r}_0\| - \underline{\mathbf{H}}_k\underline{\mathbf{z}}_k)\| \leqslant \|\mathbf{G}_{k+1}\| \cdot \|(\underline{\mathbf{e}}_1\|\mathbf{r}_0\| - \underline{\mathbf{H}}_k\underline{\mathbf{z}}_k)\| = \|\mathbf{G}_{k+1}\| \cdot \rho_k,\end{aligned} \tag{6.7}$$

i.e., apart from the norm of $\mathbf{G}_{k+1}$ using solely a quantity available in the algorithm. The norm of $\mathbf{G}_{k+1}$ can be bounded by $\sqrt{k+1}$ in case all columns, e.g., all basis vectors, are normalized to unit length. Using a basis that is block-wise orthonormal ensures a lower bound on $\|\mathbf{G}_{k+1}\|$, this is used in the MR implementation of IDR sketched in [50]. Other MR implementations of IDR are given in [3, 8].

It is possible to obtain the MR quantities from the OR quantities by a technique known as residual smoothing, developed by Hestenes and Stiefel [21, §7, p. 418–419], see also Schönauer and Weiß [36, 53]. This is incorporated in the code of [51].

The relations between OR and MR methods, c.f. [7], are still valid in case of OR and MR Sonneveld methods. Whenever the MR version stagnates, the OR version has a peak or breaks down. This is the so-called peak-plateau phenomenon. In Sonneveld methods OR approaches break down whenever some polynomial g_j has zero as a root, the corresponding MR method in this case *stagnates forever*, which we call an incurable stagnation in [50]. As the numerical properties of OR and MR approaches depend on the constructed Hessenberg matrices, we might try to use the freedom in the construction of the additional polynomials g_j, especially in the IDRStab approach, to ensure a better conditioned Hessenberg matrix. This is an active area of research.

7. General comments. From the previous comments it is obvious that Sonneveld methods should best be started with a variant of Arnoldi's method, as these can be used to obtain optimal results. Apart from the additional polynomials g_j introduced, Sonneveld methods perform Lanczos$(s, 1)$ in the background. As already mentioned in the previous sections, the quantities of the underlying Lanczos's process should be locally as close as possible to an exact process. MR processes are a little bit more costly, but these are simpler to depict. For this reason we use QMRIDR in our experiments for linear systems. In the eigenvalue part we use methods developed in [33] based on Sleijpen's MATLAB implementation of IDRSTAB [42].

8. Choosing the shadow space. In [44] Sonneveld presented examples that support the rule of thumb the convergence curves of IDR(s) [51] are getting close to the optimal convergence curve of GMRES, when we omit the additional multiplications, s tends to infinity (i.e., is large enough) and *we use a set of orthonormalized random vectors* as shadow space. Numerical experiments indicate that the same holds true for the eigenvalue solvers based on IDR, see Figure 8.1.

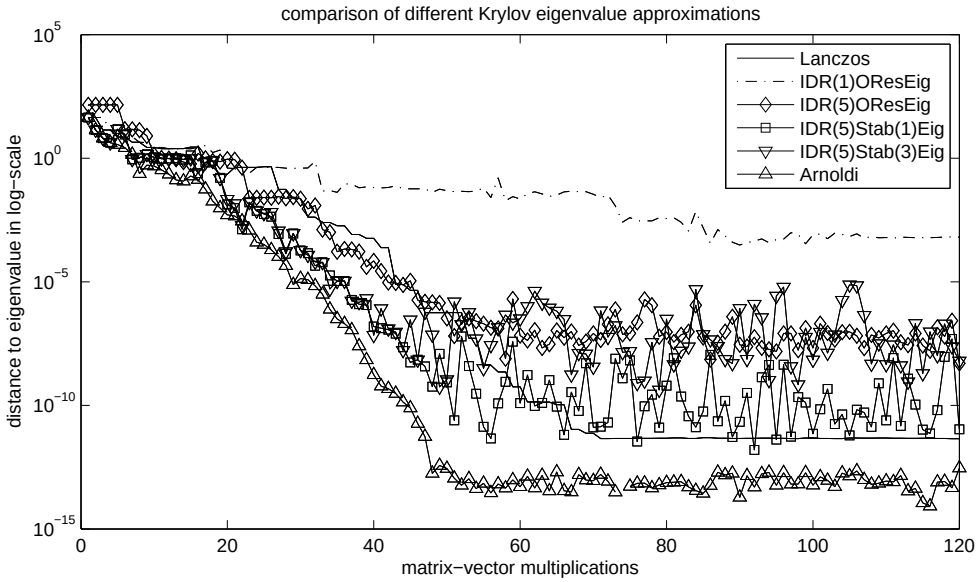

FIG. 8.1. *Comparison of Lanczos's, Arnoldi's, and four different Sonneveld methods to approximate the eigenvalue closest to* $4 + 45i$ *of the matrix* `e05r0500` *from matrix market. Increasing* s*, here the values* $s = 1$ *and* $s = 5$ *have been used as examples, gives curves that are closer to the convergence curve of Arnoldi's method, the shadow vectors have been chosen as orthonormalized random vectors. For IDR*(1) *we used the same shadow vector than in Lanczos's method, the resulting approximations should be identical but the bad numerical stability of* IDR(s)ORES *results in a large deviation from the convergence curve of Lanczos's method. The eigenvalue approximations obtained by* IDRSTAB*, using the method sketched in [33], tend to fail to reach a low level of attainable accuracy when the degree of the stabilizing polynomial is too large.*

We tried different various other choices for the shadow vectors. It turns out that using (approximate) information about (left) eigenvectors results in unpredictable behaviour. This is an active area of research. Currently we advocate the use of orthonormalized random vectors.

9. Choosing the stabilizing polynomials. The roots of the stabilizing polynomials sometimes lie far outside the field of values, which results in a severe loss of convergence properties of the eigenvalues solvers based on IDR. This is partly explained by the connection to Lanczos's methods and the rule of thumb to "stay as close as possible" to the underlying Lanczos's process. In Figure 8.1 we can observe that a larger degree of the stabilizing polynomial, which helps to increase the convergence properties of IDRSTAB as a linear system solver, has a negative effect on the eigenvalue convergence. As in real arithmetic complex conjugate roots of the stabiliz-

ing polynomials are needed to increase stability, we advocate to use only stabilizing polynomials of degree at most 2. In our experiments, the occurrence of "ghost polynomial roots" had negative effects once these roots lie far from the field of values of $\mathbf{A}$. A classical remedy is given in [39], this technique is part of Sleijpen's implementation of IDRStab and was used in the experiments. As the residual polynomials should give "small" residuals when we dampen the spectrum of (mildly non-normal) $\mathbf{A}$, and typically the convergence of the Sonneveld Ritz values starts at the outer eigenvalues, we used three theoretical experiments to investigate the stability of other choices:

- Setting the roots of the stabilizing polynomials to inner eigenvalues,
- Setting the roots of the stabilizing polynomials to the leading part of a Hessenberg matrix similar to $\mathbf{A}$.
- Setting the roots of the stabilizing polynomials to the trailing part of a Hessenberg matrix similar to $\mathbf{A}$.

The latter choice captures the first approximations of a "flipped" process, i.e., one with starting vectors given by the last vectors of the forward process. These last vectors are in some sense "richer" in the direction of eigenvectors corresponding to inner eigenvalues. Of course, the first and last method lead to impractical methods. These choices perform similar to the choices to minimize the residual or to apply the technique from [39], referred to as "vanilla". As example we refer to Figure 9.1. In general, application of the technique from [39] increases the stability compared to the former. We currently investigate mixed procedures based on the ideas in [39] and eigenvalue information based on the approaches mentioned in Section 4.

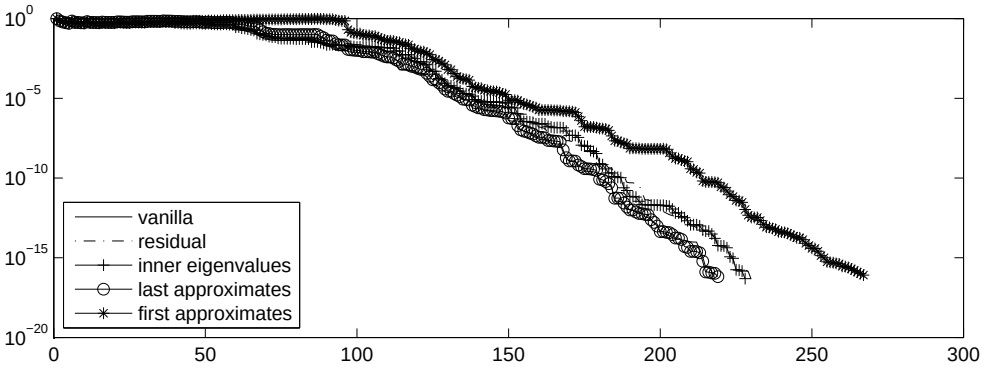

FIG. 9.1. *Comparison of two impractical different choices of stabilizing polynomials for* QMRIDR(4) *with the choices used in [45], [51, 39], and [37].*

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29. [See page 1, 2]

[2] Tony F. Chan, Lisette de Pillis, and Henk van der Vorst, *Transpose-free formulations of Lanczos-type methods for nonsymmetric linear systems*, Numerical Algorithms, 17 (1998), pp. 51–66. [See page 1]

[3] T. F. Chan, E. Gallopoulos, V. Simoncini, T. Szeto, and C. H. Tong, *A quasi-minimal residual variant of the Bi-CGSTAB algorithm for nonsymmetric systems*, SIAM J. Sci. Comput., 15 (1994), pp. 338–347. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992). [See page 1, 8]

[4] Tony F. Chan and Tedd Szeto, *A composite step conjugate gradients squared algorithm for solving nonsymmetric linear systems*, Numer. Algorithms, 7 (1994), pp. 17–32. Received Aug. 12, 1993. [See page 1]

[5] ———, *Composite step product methods for solving nonsymmetric linear systems*, SIAM J. Sci. Comput., 17 (1996), pp. 1491–1508. Received Jun. 6, 1994. [See page 1]

[6] T.P. Collignon, G.L.G. Sleijpen, and M.B. van Gijzen, *Interpreting IDR(s) as a deflation method*, Technical Report 10-21, Department of Applied Mathematical Analysis, Delft University of Technology, Delft, 2010. [See page 7]

[7] Jane Cullum and Anne Greenbaum, *Relations between Galerkin and norm-minimizing iter-*

ative methods for solving linear systems, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 223–247. [See page 8]

[8] Lei Du, Tomohiro Sogabe, and Shao-Liang Zhang, *Quasi-minimal residual smoothing technique for the IDR(s) method*, JSIAM Letters, 3 (2011), pp. 13–16. [See page 2, 8]

[9] R. Fletcher, *Conjugate gradient methods for indefinite systems*, in Numerical analysis (Proc 6th Biennial Dundee Conf., Univ. Dundee, Dundee, 1975), G. A. Watson, ed., vol. 506 of Lecture Notes in Math., Berlin, 1976, Springer, pp. 73–89. [See page 4]

[10] Diederik R. Fokkema, Gerard L. G. Sleijpen, and Henk A. van der Vorst, *Generalized conjugate gradient squared*, J. Comput. Appl. Math., 71 (1996), pp. 125–146. Received Mar. 2, 1995. [See page 1]

[11] Roland W. Freund, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, Journal of Computational and Applied Mathematics, 43 (1992), pp. 135–158. Orthogonal polynomials and numerical methods. [See page 6]

[12] ———, *Solution of shifted linear systems by quasi-minimal residual iterations*, in Numerical linear algebra (Kent, OH, 1992), de Gruyter, Berlin, 1993, pp. 101–121. [See page 1]

[13] ———, *A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems*, SIAM J. Sci. Comput., 14 (1993), pp. 470–482. [See page 1]

[14] Roland W. Freund and Noël M. Nachtigal, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339. [See page 8]

[15] A. Frommer, *BiCGStab(ℓ) for families of shifted linear systems*, Computing, 70 (2003), pp. 87–109. [See page 1]

[16] Seiji Fujino, Maki Fujiwara, and Masahiro Yoshida, *BiCGSafe method based on minimization of associate residual*, Transactions of the Japan Society for Computational Engineering and Science, 8 (2006), pp. 145–152. (Japanese). [See page 1]

[17] Martin H. Gutknecht, *Variants of BiCGStab for matrices with complex spectrum*, SIAM J. Sci. Comput., 14 (1993), pp. 1020–1033. Received Sep. 9, 1991. [See page 1]

[18] ———, *IDR explained*, Electron. Trans. Numer. Anal., 36 (2009/10), pp. 126–148. [See page 4]

[19] Martin H. Gutknecht and Jens-Peter M. Zemke, *Eigenvalue computations based on IDR*, Bericht 145, TUHH, Institute of Numerical Simulation, May 2010. Online available at http://doku.b.tu-harburg.de/volltexte/2010/875/. [See page 1, 2, 3, 4, 5, 6, 7]

[20] Karl Hessenberg, *Behandlung linearer Eigenwertaufgaben mit Hilfe der Hamilton-Cayleyschen Gleichung*, Numerische Verfahren, Bericht 1, Institut für Praktische Mathematik (IPM), Technische Hochschule Darmstadt, July 1940. Scanned report and biographical sketch of Karl Hessenberg's life online available at http://www.hessenberg.de/karl1.html. [See page 1, 2, 3]

[21] Magnus R. Hestenes and Eduard Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953). [See page 8]

[22] Zhongxiao Jia, *Some Numerical Methods for Large Unsymmetric Eigenproblems*, dissertation, Universität Bielefeld, Februar 1994. [See page 7]

[23] ———, *Refined iterative algorithms based on Arnoldi's process for large unsymmetric eigenproblems*, Linear Algebra and Its Applications, 259 (1997), pp. 1–23. [See page 7]

[24] ———, *The convergence of harmonic Ritz values, harmonic Ritz vectors, and refined harmonic Ritz vectors*, Mathematics of Computation, 74 (2004), pp. 1441–1456. [See page 7]

[25] Алексей Николаевич Крылов, *О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем*, Известия Академии Наук СССР. Отделение математических и естественных наук. Ser. VII, 4 (1931), pp. 491–539. (Russian). [See page 1]

[26] Cornelius Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282. [See page 1, 2, 3]

[27] ———, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53. [See page 1, 2, 3, 4]

[28] N. J. Lehmann, *Optimale Eigenwerteinschließungen*, Numerische Mathematik, 5 (1963), pp. 246–272. [See page 6]

[29] Ronald B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra and Its Applications, 154–156 (1991), pp. 289–309. [See page 6]

[30] C. C. Paige, Beresford N. Parlett, and Henk A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numerical Linear Algebra with Applications, 2 (1995), pp. 115–133. [See page 6]

[31] C. C. Paige and M. A. Saunders, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629. [See page 8]

[32] Olaf Rendel, *Aspects of eigenvalue computations using Induced Dimension Reduction (IDR)*, bachelor's thesis, Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, 2010. [See page 1, 6]

[33] Anisa Rizvanolli, *Eigenwertberechnung mittels IDRStab*, Studienarbeit, Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, 2011. [See page 1, 6, 9]

[34] Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp., 37 (1981), pp. 105–126. [See page 8]

[35] Youcef Saad and Martin H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869. [See page 8]

[36] W. Schönauer, *Scientific Computing on Vector Computers*, Elsevier, Amsterdam, 1987. [See

page 8]
[37] Valeria Simoncini and Daniel B. Szyld, *Interpreting IDR as a Petrov-Galerkin method*, SIAM J. Sci. Comput., 32 (2010), pp. 1898–1912. [See page 4, 5, 10]
[38] Gerard L.G. Sleijpen, Peter Sonneveld, and Martin B. van Gijzen, *Bi-CGSTAB as an induced dimension reduction method*, Appl. Numer. Math., 60 (2010), pp. 1100–1114. [See page 4, 5]
[39] Gerard L.G. Sleijpen and Henk A. van der Vorst, *Maintaining convergence properties of BiCGstab methods in finite precision arithmetic*, Numer. Algorithms, 10 (1995), pp. 203–223. [See page 5, 10]
[40] Gerard L. G. Sleijpen and Diederik R. Fokkema, *BiCGstab(l) for linear equations involving unsymmetric matrices with complex spectrum*, Electron. Trans. Numer. Anal., 1 (1993), pp. 11–32. Received Mar. 24, 1993. [See page 1]
[41] Gerard L. G. Sleijpen, Peter Sonneveld, and Martin B. van Gijzen, *Bi-CGSTAB as an induced dimension reduction method*, Appl. Numer. Math., 60 (2010), pp. 1100–1114. Special Issue: 9th IMACS International Symposium on Iterative Methods in Scientific Computing (IISIMSC 2008). Received Jun. 30, 2008, electr. publ. Jul. 9, 2009. [See page 1]
[42] Gerard L. G. Sleijpen and Martin B. van Gijzen, *Exploiting BiCGstab(ℓ) strategies to induce dimension reduction*, SIAM J. Sci. Comput., 32 (2010), pp. 2687–2709. Received Mar. 11, 2009, electr. publ. Aug. 31, 2010. [See page 1, 4, 5, 6, 8, 9]
[43] Peter Sonneveld, *CGS, a fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 36–52. [See page 1, 4]
[44] ———, *On the convergence behaviour of IDR(s)*, Technical Report 10-08, Department of Applied Mathematical Analysis, Delft University of Technology, Delft, 2010. [See page 2, 9]
[45] Peter Sonneveld and Martin B. van Gijzen, *IDR(s): a family of simple and fast algorithms for solving large nonsymmetric systems of linear equations*, SIAM J. Sci. Comput., 31 (2008/09), pp. 1035–1062. [See page 1, 4, 5, 6, 8, 10]
[46] Masaaki Tanio and Masaaki Sugihara, *GIDR(s,L): generalized IDR(s)*, in The 2008 annual conference of the Japan Society for Industrial and Applied Mathematics, Chiba, Japan, September 2008, pp. 411–412. (Japanese). [See page 1]
[47] ———, *GBi-CGSTAB(s, L): IDR(s) with higher-order stabilization polynomials*, J. Comput. Appl. Math., 235 (2010), pp. 765–784. Received May 13, 2009, electr. publ. Jul. 16, 2010. [See page 1, 5]
[48] H. A. van der Vorst, *Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644. [See page 1, 4]
[49] H. A. van der Vorst and P. Sonneveld, *CGSTAB, a more smoothly converging variant of CG-S*, Report 90-50, Department of Mathematics and Informatics, Delft University of Technology, 1990. [See page 1, 4]
[50] Martin B. van Gijzen, Gerard L.G. Sleijpen, and Jens-Peter M. Zemke, *Flexible and multi-shift induced dimension reduction algorithms for solving large sparse linear systems*, Bericht 156, TUHH, Institute of Numerical Simulation, August 2011. Online available at http://doku.b.tu-harburg.de/volltexte/2011/1114/. [See page 2, 5, 8]
[51] Martin B. van Gijzen and Peter Sonneveld, *An elegant IDR(s) variant that efficiently exploits bi-orthogonality properties.*, Technical Report 10-16, Department of Applied Mathematical Analysis, Delft University of Technology, Delft, 2010. (revised version of report 08-21). [See page 1, 5, 6, 8, 9, 10]
[52] Judith A. Vogel, *Flexible BiCG and flexible Bi-CGSTAB for nonsymmetric linear systems*, Appl. Math. Comput., 188 (2007), pp. 226–233. [See page 1]
[53] Rüdiger Weiss, *Convergence Behavior of Generalized Conjugate Gradient Methods*, Dissertation, Universität Karlsruhe, 1990. [See page 8]
[54] P. Wesseling and P. Sonneveld, *Numerical experiments with a multiple grid and a preconditioned Lanczos type method*, in Approximation methods for Navier-Stokes problems (Proc. Sympos., Univ. Paderborn, Paderborn, 1979), R. Rautmann, ed., vol. 771 of Lecture Notes in Math., Berlin, Heidelberg, New York, 1980, Springer-Verlag, pp. 543–562. [See page 1]
[55] Man-Chung Yeung and Tony F. Chan, *ML(K)BiCGSTAB: a BiCGSTAB variant based on multiple Lanczos starting vectors*, SIAM J. Sci. Comput., 21 (1999), pp. 1263–1290. Received May 16, 1997, electr. publ. Dec. 15, 1999. [See page 1]
[56] Jens-Peter M. Zemke, *Hessenberg eigenvalue–eigenmatrix relations*, Linear Algebra Appl., 414 (2006), pp. 589–606. Received Nov. 9, 2004, electr. publ. Feb. 17, 2006. [See page 7]
[57] ———, *Abstract perturbed Krylov methods*, Linear Algebra Appl., 424 (2007), pp. 405–434. Received Nov. 3, 2005, electr. publ. Feb. 21, 2007. [See page 7]
[58] ———, *Quasi-minimal residual eigenpairs*, in Special Issue: GAMM Annual Meeting 2008 — Bremen, vol. 8 of PAMM, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2009, pp. 10835–10836. [See page 7]
[59] Shao-Liang Zhang, *GPBi-CG: generalized product-type methods based on Bi-CG for solving nonsymmetric linear systems*, SIAM J. Sci. Comput., 18 (1997), pp. 537–551. Received Aug. 19, 1995. [See page 1]