IDR VERSUS OTHER KRYLOV SUBSPACE SOLVERS*

JENS-PETER M. ZEMKE†

Abstract. We compare members of the IDR family for the solution of linear systems and eigenvalue problems with traditional Krylov subspace solvers. This comparison is based on a description of IDR as a means to construct generalized Hessenberg decompositions, whereas traditional Krylov methods construct Hessenberg decompositions.

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

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

1. Introduction.

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. [6]. In 1940 the first modern Krylov subspace method was developed [5]. The best known Krylov subspace methods are based on Lanczos's [7, 8] and Arnoldi's [1] method. The first IDR method is [24]; the IDR(s) methods [17, 23] are relatively new. The generalization to use larger shadow spaces of dimension $s \in \mathbb{N}$ offers advantages: these appear to be more stable than the original IDR variant, BiCGStab [21, 20], and most of its relatives.

We simply term all methods form the IDR family, e.g., original IDR, BiCGStab, IDR(s) and IDRStab [18, 16, 19] amongst others, as *IDR methods* or *Sonneveld methods*. Sonneveld methods are linked to Lanczos processes termed Lanczos($s, 1$). This process 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 subleties associated with Sonneveld methods.

2. Classical Krylov subspace methods.

Essentials of classical Krylov subspace methods can be captured by a so-called *Hessenberg decomposition* [5, 4]

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

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

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

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 $\mathcal{K}_k(\mathbf{A}, \mathbf{q})$.

3. Sonneveld methods.

Sonneveld methods are based on the IDR Theorem. IDR spaces, a special case of Sonneveld subspaces [16, Definition 2.2, p. 2690], are defined as follows. Define $\mathcal{G}_0$ by

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

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

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

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

Theorem 3.1 (IDR Theorem [17]). *Under mild conditions on the matrices* $\mathbf{A}$ *and the space* $\mathcal{S}$*,*

---

*Version of October 4, 2011, 11:19.**

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

(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 [17, 15].

The relation of IDR to Krylov subspaces is given in [3, 14, 4, 15]. The latter includes an alternate description of Sonneveld spaces [15, Theorem 11, p. 1104] based on left block Krylov subspaces. Implementations of the recursion (3.2) are given in [17, 23, 22]:

Initialization: compute $s+1$ basis vectors $\mathbf{g}_i$, $1 \leqslant i \leqslant s+1$, in $\mathcal{K}_{s+1} \subset \mathcal{G}_0$.
Recursion: for $j > 0$ until convergence perform the following:
- **Intersection:** compute a linear combination $\mathbf{v}_i$ of vectors in $\mathcal{G}_{j-1} \cap \mathcal{S}$.
- **Update:** if constructing the first vector in a new space $\mathcal{G}_j$, chose a new linear polynomial g_j of exact degree 1.
- **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 $\mathcal{G}_j$.

Sonneveld Krylov subspace methods can be described by a so-called *generalized Hessenberg decomposition* [4]

$$\mathbf{A}\mathbf{V}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \quad k \in \mathbb{N}, \quad k < n, \tag{3.3}$$

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).

The small change from the Hessenberg decomposition (2.1) to the *generalized* Hessenberg decomposition (3.3) is the main change in devising new algorithms or applying well-known techniques from the pool of existing Krylov subspace method techniques.

4. Some classical techniques and remarks on results. We very briefly sketch the application of some classical Krylov subspace techniques to Sonneveld methods. The Ritz approach is based on the Sonneveld pencil $(\mathbf{H}_k, \mathbf{U}_k)$ [4]

$$\mathbf{H}_k\mathbf{s}_j = \theta_j\mathbf{U}_k\mathbf{s}_j \tag{4.1}$$

and gives Ritz pairs $(\theta_j, \mathbf{y}_j := \mathbf{V}_k\mathbf{s}_j = \mathbf{G}_k\mathbf{U}_k\mathbf{s}_j)$ [4, 12, 13]. The harmonic Ritz approach [9, 10, 2, 11]

$$\mathbf{I}_k\underline{\mathbf{s}}_j = \underline{\theta}_j\underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{U}}_k\underline{\mathbf{s}}_j \tag{4.2}$$

gives harmonic Ritz pairs $(\underline{\theta}_j, \underline{\mathbf{y}}_j := \mathbf{V}_k\underline{\mathbf{s}}_j = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{s}}_j)$. The Orthogonal Residual (OR) approach: The kth OR solution is given by

$$\mathbf{H}_k\mathbf{z}_k := \mathbf{e}_1\|\mathbf{r}_0\|, \quad \text{e.g., mostly} \quad \mathbf{z}_k := \mathbf{H}_k^{-1}\mathbf{e}_1\|\mathbf{r}_0\|, \tag{4.3}$$

the kth OR iterate by

$$\mathbf{x}_k := \mathbf{V}_k\mathbf{z}_k = \mathbf{G}_k\mathbf{U}_k\mathbf{z}_k. \tag{4.4}$$

The Minimal Residual (MR) approach: The kth MR solution is given by

$$\rho_k := \left\|\underline{\mathbf{H}}_k\underline{\mathbf{z}}_k - \underline{\mathbf{e}}_1\|\mathbf{r}_0\|\right\| = \min, \quad \text{i.e.,} \quad \underline{\mathbf{z}}_k := \underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{e}}_1\|\mathbf{r}_0\|, \tag{4.5}$$

the kth MR iterate by

$$\underline{\mathbf{x}}_k := \mathbf{V}_k\underline{\mathbf{z}}_k = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{z}}_k. \tag{4.6}$$

Other flavors like OrthoRes, OrthoMin, and OrthoDir, and techniques like flexible, multi-shift, and inexact variants can now be developed based on (3.3). Methods like (flexible or multi-shift) QMRIDR [22] provide a smooth transition between the methods of Lanczos and Arnoldi and do not rely on the transposed matrix, but are a little less stable. Some further details and examples are part of the slides.

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.

[2] R. 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.

[3] M. H. Gutknecht, *IDR explained*, Electron. Trans. Numer. Anal., 36 (2009/10), pp. 126–148.

[4] M. H. Gutknecht and J.-P. 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/.

[5] K. 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.

[6] 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).

[7] C. 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.

[8] ———, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.

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

[10] R. B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra and Its Applications, 154–156 (1991), pp. 289–309.

[11] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numerical Linear Algebra with Applications, 2 (1995), pp. 115–133.

[12] O. Rendel, *Aspects of eigenvalue computations using Induced Dimension Reduction (IDR)*, bachelor's thesis, Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, 2010.

[13] A. Rizvanolli, *Eigenwertberechnung mittels IDRStab*, Studienarbeit, Institut für Numerische Simulation, Technische Universität Hamburg-Harburg, 2011.

[14] V. Simoncini and D. B. Szyld, *Interpreting IDR as a Petrov-Galerkin method*, SIAM J. Sci. Comput., 32 (2010), pp. 1898–1912.

[15] G. L. Sleijpen, P. Sonneveld, and M. B. van Gijzen, *Bi-CGSTAB as an induced dimension reduction method*, Appl. Numer. Math., 60 (2010), pp. 1100–1114.

[16] G. L. G. Sleijpen and M. 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.

[17] P. Sonneveld and M. 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.

[18] M. Tanio and M. 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).

[19] ———, *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.

[20] 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.

[21] 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.

[22] M. B. van Gijzen, G. L. Sleijpen, and J.-P. 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/.

[23] M. B. van Gijzen and P. 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).

[24] 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.