Electronic Transactions on Numerical Analysis.
Volume 46, pp. 245–272, 2017.

ISSN 1068–9613.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

VARIANTS OF IDR WITH PARTIAL ORTHONORMALIZATION*

JENS-PETER M. ZEMKE†

Abstract. We present four variants of IDR(s) that generate vectors such that consecutive blocks of $s+1$ vectors are orthonormal. IDR methods are based on tuning parameters: an initially chosen, so-called *shadow space*, and the so-called *seed values*. We collect possible choices for the seed values. We prove that under certain conditions all four variants are mathematically equivalent and discuss possible breakdowns. We give an error analysis of all four variants and a numerical comparison in the context of the solution of linear systems and eigenvalue problems.

Key words. IDR, partial orthonormalization, minimum norm expansion, error analysis

AMS subject classifications. 65F25 (primary), 65F10, 65F15, 65F50

1. Introduction.

We present four computationally different IDR(s) variants that are based on an orthonormalization of every $s+1$ vectors computed in the recurrence. IDR(s) [16] is a recent Krylov subspace method for the solution of linear systems [16, 18, 19] or eigenvalue problems [5, 9]. IDR is an acronym for *induced dimension reduction*, a quite recent[1] technique in the setting of Krylov subspace methods. There exist several implementations of IDR(s), but the implementation in [18] is the only published one that computes vectors such that every $s+1$ consecutive in the same space are orthonormalized. We call IDR methods with this property "IDR with partial orthonormalization" and present three other IDR variants with partial orthonormalization. We prove that in the generic case, the four variants are mathematically equivalent with the exception of possible additional breakdowns of the variant in [18]. We classify breakdowns of all four variants and give a simple *a posteriori* error analysis, i.e., the recurrence error is bounded in terms of the computed quantities. IDR is related to the two-sided Lanczos process and suffers from the same possible breakdowns, making an *a priori* error analysis more or less impossible.

1.1. Motivation.

In the IDR variant in [16], several vectors in the same space are computed that are differences of residual vectors corresponding to a linear system of equations to be solved. This minimizes the amount of vectors that have to be stored at the price of additional instabilities. In [19] linear combinations of these vectors are used that simplify the algorithm and speed up the solution process of small linear systems that arise in IDR(s). Numerical evidence shows that this local basis transformation makes the method more stable, but a stability analysis is missing. As another remedy we used in [18] orthonormalization of the computed vectors in the same space. Numerical experiments suggest that the latter variant is the most stable of these three variants. At the same time we experimented in [9] with different ways of generating the new vectors in the spaces combined with the orthonormalization used in [18]. In this note we introduce the four most interesting variants we tested, prove the mathematical equivalence in the generic case, give a common rough error analysis of all four variants, and showcase with two toy examples the typical behavior of the four variants in the context of linear systems and eigenvalue problems.

1.2. Notation.

We use standard notation. Matrices are denoted by upper case bold letters $\mathbf{A} \in \mathbb{C}^{n\times n}$, its columns by lower case bold letters $\mathbf{a}_j$, $1 \leqslant j \leqslant n$, and its elements by lower case letters $a_{i,j}$, $1 \leqslant i,j \leqslant n$. The identity matrix is denoted by $\mathbf{I}_n \in \mathbb{C}^{n\times n}$, its columns by

*Received March 24, 2016. Accepted March 21, 2017. Published online on July 14, 2017. Recommended by Martin Gutknecht.

†Technische Universität Hamburg, Institut für Mathematik, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany (zemke@tu-harburg.de).

[1] Original IDR [21] dates to 1979, but the more interesting variant IDR(s) [16] dates to 2008.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$\mathbf{e}_j$, $1 \leqslant j \leqslant n$, its elements by the Kronecker delta $\delta_{i,j}$, $1 \leqslant i,j \leqslant n$. $\mathbf{O}$ denotes a zero matrix, $\mathbf{o}$ a zero vector. $\overline{\mathbf{A}}$ is the (elementwise) complex conjugate of $\mathbf{A}$. A lower bar appended to a matrix or vector such as $\underline{\mathbf{H}}_m \in \mathbb{C}^{(m+1)\times m}$ or $\underline{\mathbf{e}}_1 \in \mathbb{C}^{m+1}$ indicates one extra row appended at the bottom of $\mathbf{H}_m \in \mathbb{C}^{m\times m}$, $\mathbf{e}_1 \in \mathbb{C}^m$, with the exception of $\underline{\mathbf{z}}_m \in \mathbb{C}^m$, $\underline{\mathbf{x}}_m \in \mathbb{C}^n$, and $\underline{\mathbf{r}}_m \in \mathbb{C}^n$, which are quantities related to $\underline{\mathbf{H}}_m \in \mathbb{C}^{(m+1)\times m}$ and $\underline{\mathbf{e}}_1 \in \mathbb{C}^{m+1}$. The inverse, Moore-Penrose pseudo-inverse, transpose, and complex conjugate transpose are denoted by the superscripts $^{-1}$, †, $^\mathsf{T}$, and $^\mathsf{H}$, respectively. Spaces are denoted by upper case calligraphic letters (e.g., $\mathcal{G}$), vectors from this spaces are usually denoted by the same lower case bold letter (e.g., $\mathbf{g}$). For $x \in \mathbb{R}$, $\lfloor x \rfloor \in \mathbb{Z}$ is the largest integer with $\lfloor x \rfloor \leqslant x$. Similarly, $\lceil x \rceil \in \mathbb{Z}$ is the smallest integer with $x \leqslant \lceil x \rceil$. Inclusion of sets is denoted by $\subseteq$, strict inclusion is denoted by $\subset$.

1.3. Outline. In Section 2 we gather basic definitions and present the IDR theorem, the core of all IDR methods. Section 3 contains a generic IDR algorithm and the four IDR algorithms with partial orthonormalization. We introduce the concept of the so-called *generalized Hessenberg decomposition* that describes the computed quantities in theory and give a brief sketch how to apply IDR in the context of linear systems and eigenvalue problems. Section 4 is devoted to the choice of the so-called seed values. In Section 5 the mathematical equivalence of the four algorithms is analyzed and different types of breakdowns are classified. Section 6 is devoted to an error analysis of all four IDR algorithms with partial orthonormalization. In Section 7 we present two numerical examples, one for a linear system and one for an eigenvalue problem. We conclude in Section 8 with a discussion how to select the appropriate variant.

2. Basics. IDR methods comprise a class of Sonneveld methods; Sonneveld methods comprise a class of Krylov subspace methods. Our definition of Krylov subspaces is tailored to define a class of Sonneveld methods that includes the prototype IDR(s) of [16]:

DEFINITION 2.1 (Krylov subspaces). *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *and* $\mathbf{q} \in \mathbb{C}^n$. *We define the* right Krylov subspaces

$$\mathcal{K}_j := \mathcal{K}_j(\mathbf{A},\mathbf{q}) := \mathsf{span}\{\mathbf{q}, \mathbf{A}\mathbf{q}, \ldots, \mathbf{A}^{j-1}\mathbf{q}\} = \{p(\mathbf{A})\mathbf{q} \mid \deg(p) < j\}, \quad j \geqslant 1,$$
$$\mathcal{K}_0 := \mathcal{K}_0(\mathbf{A},\mathbf{q}) := \{\mathbf{o}\}, \quad \mathcal{K} := \mathcal{K}(\mathbf{A},\mathbf{q}) := \mathcal{K}_n(\mathbf{A},\mathbf{q}).$$

Let additionally $\widehat{\mathbf{Q}} = \left[\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_s\right] \in \mathbb{C}^{n\times s}$ *with full rank* s, *typically* $s \ll n$. *We define the* left block Krylov subspaces

$$\widehat{\mathcal{K}}_0 := \mathcal{K}_0(\mathbf{A}^\mathsf{H}, \widehat{\mathbf{Q}}) := \{\mathbf{o}\},$$
$$\widehat{\mathcal{K}}_j := \mathcal{K}_j(\mathbf{A}^\mathsf{H}, \widehat{\mathbf{Q}}) := \left\{\sum_{i=0}^{j-1} (\mathbf{A}^\mathsf{H})^i \widehat{\mathbf{Q}}\mathbf{c}_i \mid \mathbf{c}_i \in \mathbb{C}^s\right\} = \sum_{i=1}^{s} \mathcal{K}_j(\mathbf{A}^\mathsf{H}, \widehat{\mathbf{q}}_i), \quad j \geqslant 1.$$

Just like Krylov subspace methods are based on Krylov subspaces, Sonneveld methods are based on Sonneveld spaces [14, Definition 2.2, p. 2690]:

DEFINITION 2.2 (Sonneveld/IDR spaces; seed polynomials/values). *Let* $p \in \mathbb{C}[z]$, $\mathbf{A} \in \mathbb{C}^{n\times n}$, $\mathbf{q} \in \mathbb{C}^n$, *and* $\widehat{\mathbf{Q}} \in \mathbb{C}^{n\times s}$ *with full rank* s. *We define the* Sonneveld space

$$\mathcal{P}(p, \mathbf{A}, \mathbf{q}, \widehat{\mathbf{Q}}) := p(\mathbf{A})\big(\mathcal{K}(\mathbf{A},\mathbf{q}) \cap \mathcal{K}_{\deg(p)}(\mathbf{A}^\mathsf{H}, \widehat{\mathbf{Q}})^\perp\big).$$

In this paper we focus on the IDR spaces

$$\mathcal{G}_j := \mathcal{P}(M_j, \mathbf{A}, \mathbf{q}, \widehat{\mathbf{Q}}), \quad j \geqslant 0,$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where the seed polynomials M_j, $j \geqslant 0$, *are defined recursively by*

$$M_0(z) := 1, \quad M_j(z) := (z - \mu_j)M_{j-1}(z), \quad j \geqslant 1.$$

The roots μ_j, $j \geqslant 1$, *are called* seed values.

The following theorem states some well-known properties of IDR spaces. In particular, they are nested and can be represented recursively without referring to $\mathbf{A}^{\mathsf{H}}$:

THEOREM 2.3 (IDR Theorem). *Let* $\mathcal{S} := \{\mathbf{v} \in \mathbb{C}^n \mid \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{v} = \mathbf{o}\} = \mathcal{K}_1(\mathbf{A}^{\mathsf{H}}, \widehat{\mathbf{Q}})^{\perp}$. *Then*

$$\begin{aligned} &\mathcal{G}_0 = \mathcal{K} = \mathcal{K}(\mathbf{A}, \mathbf{q}), \\ &\mathcal{G}_j = (\mathbf{A} - \mu_j\mathbf{I})\mathcal{V}_{j-1}, \quad \textit{where} \quad \mathcal{V}_{j-1} := \mathcal{G}_{j-1} \cap \mathcal{S}, \quad j \geqslant 1. \end{aligned} \tag{2.1}$$

In particular, it holds that $\mathcal{G}_j \subseteq \mathcal{G}_{j-1}$, $j \geqslant 1$.

Suppose that $\mathcal{G}_0$ *and* $\mathcal{S}$ *do not share a nontrivial invariant subspace of* $\mathbf{A}$ *and that* $\mu_j \notin \mathsf{spec}(\mathbf{A})$, $j \geqslant 1$.

Then there exists a uniquely defined $j_\infty \in \mathbb{N}_0$, $j_\infty \leqslant n$, *such that the first* j_∞ *inclusions are strict,*

$$\mathcal{G}_j \subset \mathcal{G}_{j-1}, \quad 1 \leqslant j \leqslant j_\infty, \quad \textit{and} \quad \mathcal{G}_{j_\infty} = \{\mathbf{o}\}. \tag{2.2}$$

Proof. See [16], [11, Theorem 11, p. 1104, Note 2, p. 1105]. ∎

By (2.2) and (2.1) of Theorem 2.3 it follows that

$$0 < \mathsf{dim}(\mathcal{G}_{j-1}) - \mathsf{dim}(\mathcal{G}_j) \leqslant \mathsf{codim}(\mathcal{S}) = s, \quad 1 \leqslant j \leqslant j_\infty.$$

In [16, p. 1043] it is stated without proof that if $\mathcal{S}$ is a random space, then, with probability one, $\mathsf{dim}(\mathcal{G}_{j-1}) - \mathsf{dim}(\mathcal{G}_j) = s$, $1 \leqslant j \leqslant \lfloor n/s \rfloor = j_\infty - 1$. This is referred to as the *generic case* in [16]. The proof can be found in the appendix of the report [15].

Sonneveld methods and *IDR methods* are methods that compute approximations (e.g., to eigenvectors or to the solution of a linear system) that are linear combinations of vectors in a Sonneveld space and in an IDR space, respectively. IDR methods are Sonneveld methods but not *vice versa*.

The approximations computed by a Sonneveld method take the form $\mathbf{G}_m\mathbf{c}_m$ for $\mathbf{c}_m \in \mathbb{C}^m$ and $\mathbf{G}_m = \big[\mathbf{g}_1, \ldots, \mathbf{g}_m\big]$ with columns $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_m \in \mathcal{K}_m$, $m \geqslant 1$. In contrast to Krylov subspace methods such as the Arnoldi method [1] and the Lanczos method [7, 8], we do not enforce $\mathsf{rank}(\mathbf{G}_m) = m$ in a Sonneveld method. In an IDR method, we typically have $m - \lfloor m/(s+1) \rfloor \leqslant \mathsf{rank}(\mathbf{G}_m) \leqslant m$, compare with the structure of $\mathbf{G}_{m+1}$ in (2.3).

In the generic case, it holds that $\mathsf{dim}(\mathcal{G}_{j-1}/\mathcal{G}_j) = s$, $1 \leqslant j < j_\infty$. The known IDR algorithms compute $s+1$ linearly independent vectors $\mathbf{g}_{(j-1)(s+1)+1}, \ldots, \mathbf{g}_{j(s+1)}$ that lie in the set $\mathcal{G}_{j-1} \setminus \mathcal{G}_j$, $1 \leqslant j < j_\infty$; the first s vectors comprise the representatives of a basis of the quotient space $\mathcal{G}_{j-1}/\mathcal{G}_j$, the last vector is an auxiliary vector to guarantee that the intersection $\mathsf{span}\{\mathbf{g}_{(j-1)(s+1)+1}, \ldots, \mathbf{g}_{j(s+1)}\} \cap \mathcal{S}$ contains a non-trivial vector. To ease the presentation of the algorithms in the next section, we define the *local IDR matrices*

$$\begin{aligned} \mathbf{G}_s^{(j-1)} &:= \big[\mathbf{g}_{(j-1)(s+1)+1}, \ldots, \mathbf{g}_{j(s+1)-1}\big], \\ \mathbf{G}_{s+1}^{(j-1)} &:= \big[\mathbf{G}_s^{(j-1)}, \mathbf{g}_{j(s+1)}\big], \end{aligned} \quad j \geqslant 1,$$

and the *local IDR vectors*

$$\mathbf{g}_k^{(j-1)} := \mathbf{g}_{(j-1)(s+1)+k}, \quad 1 \leqslant k \leqslant s+1, \quad j \geqslant 1.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The matrix $\mathbf{G}_{s+1}^{(j-1)}$ contains all $s+1$ vectors in $\mathcal{G}_{j-1} \setminus \mathcal{G}_j$, and the matrix $\mathbf{G}_s^{(j-1)}$ only the representatives of the basis of $\mathcal{G}_{j-1}/\mathcal{G}_j$. The *global* matrix $\mathbf{G}_{m+1}$ is given in terms of local matrices and vectors by

$$\begin{aligned} \mathbf{G}_{m+1} &= \left[\mathbf{G}_{s+1}^{(0)}, \mathbf{G}_{s+1}^{(1)}, \ldots, \mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}\right], \\ j &:= \left\lfloor \frac{m+1}{s+1} \right\rfloor, \quad k := (m+1) - j(s+1), \quad m \geqslant 0. \end{aligned} \tag{2.3}$$

3. Algorithms. In this section we describe algorithms that compute unique vectors

$$\mathbf{g}_k \in \mathcal{K}_k \setminus \mathcal{K}_{k-1}, \quad \mathbf{g}_k \in \mathcal{G}_j, \quad j = \left\lfloor \frac{k-1}{s+1} \right\rfloor, \quad 1 \leqslant k \leqslant m+1, \tag{3.1}$$

based on the assumption that the vectors constructed in two consecutive $\mathcal{G}_j$ spaces are linearly independent except possibly the last one and that no linear combinations of the first s vectors constructed in each $\mathcal{G}_j$ are in the kernel of $\widehat{\mathbf{Q}}^{\mathsf{H}}$,

$$\mathsf{rank}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_s^{(j)}] = 2s+1, \qquad 1 \leqslant j < \left\lfloor \frac{m+1}{s+1} \right\rfloor, \tag{3.2a}$$

$$\mathsf{rank}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{m-j(s+1)}^{(j)}] = s+1+(m-j(s+1)), \qquad j = \left\lfloor \frac{m+1}{s+1} \right\rfloor, \tag{3.2b}$$

$$\mathsf{rank}(\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j)}) = s, \qquad 0 \leqslant j < \left\lfloor \frac{m+1}{s+1} \right\rfloor, \tag{3.2c}$$

this we term the *generic case*. First we present a *generic IDR algorithm* to compute the vectors $\mathbf{g}_1, \ldots, \mathbf{g}_{m+1}$ under this assumption. We derive four computationally different variants of it, which are named srIDR, fmIDR, mnIDR, and ovIDR. Afterwards we specialize the generic IDR algorithm to an IDR algorithm that has the property that all $\mathbf{G}_{s+1}^{(j)}$ are orthonormal,

$$\left(\mathbf{G}_{s+1}^{(j)}\right)^{\mathsf{H}} \mathbf{G}_{s+1}^{(j)} = \mathbf{I}_{s+1}, \quad j \geqslant 0. \tag{3.3}$$

We term this algorithm *IDR with partial orthonormalization*.

3.1. Generic IDR. In this section we derive our generic IDR algorithm that includes all known IDR algorithms as special cases.

Let the function $[h_{1,0}, \underline{\mathbf{H}}_m, \mathbf{G}_{m+1}] \leftarrow \mathsf{Krylov}(\mathbf{A}, \mathbf{q}, m)$ denote a generic Krylov subspace method that computes a matrix $\mathbf{G}_{m+1}$, $\mathsf{im}(\mathbf{G}_{m+1}) = \mathcal{K}_{m+1}(\mathbf{A}, \mathbf{q})$, a scalar $h_{1,0}$ such that $\mathbf{G}_{m+1}\underline{\mathbf{e}}_1 h_{1,0} = \mathbf{q}$, and an extended Hessenberg matrix $\underline{\mathbf{H}}_m$, such that the *Hessenberg decomposition*

$$\mathbf{A}\mathbf{G}_m = \mathbf{G}_{m+1}\underline{\mathbf{H}}_m$$

is satisfied. Algorithm 3.1 with a rule for the computation of the scalars $h_{i,j} \in \mathbb{C}$ results in any Krylov subspace method; these scalars might be given (e.g., the power method is obtained for $h_{i,k} = \delta_{i-1,k}$), or they might be computed in Algorithm 3.1; an example is Arnoldi's process given here in pseudocode as Algorithm 3.3.

To highlight the dependency on the basis used to define the shadow space we write $\ker(\widehat{\mathbf{Q}}^{\mathsf{H}})$ in place of $\mathcal{S}$ in the IDR algorithms. The generic IDR algorithm is Algorithm 3.2. In Algorithm 3.2, there is a lot of freedom: the choice of the starting Krylov subspace method (line 1), the computation of the seed values (line 5), the solution of the s consecutive underdetermined systems (line 9), and the choice of the scalars $h_{i,k}^{(j)}$ (line 7, line 12).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Algorithm 3.1 Krylov (generic variant)

INPUT: $\mathbf{A} \in \mathbb{C}^{n\times n}$; $\mathbf{q} \in \mathbb{C}^n$; $m \in \mathbb{N}$.
OUTPUT: $h_{1,0} \in \mathbb{C}$; $\underline{\mathbf{H}}_m \in \mathbb{C}^{(m+1)\times m}$; $\mathbf{G}_{m+1} \in \mathbb{C}^{n\times(m+1)}$.

1: $\mathbf{g}_1 \leftarrow \mathbf{q}/h_{1,0}$;
2: **for** $k = 1 : m$ **do**
3: $\quad \mathbf{g}_{k+1} \leftarrow \big(\mathbf{A}\mathbf{g}_k - \sum_{i=1}^{k} \mathbf{g}_i h_{i,k}\big)/h_{k+1,k}$;
4: **end for**

Algorithm 3.2 IDR (generic variant)

INPUT: $\mathbf{A} \in \mathbb{C}^{n\times n}$; $\mathbf{q} \in \mathbb{C}^n$; $\widehat{\mathbf{Q}} \in \mathbb{C}^{n\times s}$; $m \in \mathbb{N}$.
OUTPUT: $\mathbf{G}_{m+1} \in \mathbb{C}^{n\times(m+1)}$; ... % see (2.3)

1: $[h_{1,0}^{(0)}, \underline{\mathbf{H}}_s^{(0)}, \mathbf{G}_{s+1}^{(0)}] \leftarrow \mathsf{Krylov}(\mathbf{A}, \mathbf{q}, s)$; % $\mathsf{im}(\mathbf{G}_{s+1}^{(0)}) = \mathcal{K}_{s+1} \subset \mathcal{G}_0$
2: **for** $j = 1 \ldots$ **do**
3: $\quad \mathbf{c}_0^{(j)} \leftarrow (\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j-1)})^{-1}(\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_{s+1}^{(j-1)})$; % see (3.2)
4: $\quad \mathbf{v}_0^{(j)} \leftarrow \mathbf{g}_{s+1}^{(j-1)} - \mathbf{G}_s^{(j-1)}\mathbf{c}_0^{(j)}$; % $\mathbf{v}_0^{(j)} \in \mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) \cap \ker(\widehat{\mathbf{Q}}^{\mathsf{H}}) \subset \mathcal{V}_{j-1}$
5: $\quad$ choose μ_j % discussed in Section 4
6: $\quad \mathbf{r}_1^{(j)} \leftarrow \mathbf{A}\mathbf{v}_0^{(j)} - \mathbf{v}_0^{(j)}\mu_j$; % $\mathbf{r}_1^{(j)} \in (\mathbf{A} - \mu_j\mathbf{I})\mathcal{V}_{j-1} = \mathcal{G}_j$
7: $\quad \mathbf{g}_1^{(j)} \leftarrow \mathbf{r}_1^{(j)}/h_{1,0}^{(j)}$; % $\mathbf{g}_1^{(j)} \in \mathcal{G}_j$
8: $\quad$ **for** $k = 1 : s$ **do**
9: $\quad\quad$ solve $\widehat{\mathbf{Q}}^{\mathsf{H}}(\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)})\mathbf{c}_k^{(j)} = \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_k^{(j)}$; % see (3.6)
10: $\quad\quad \mathbf{v}_k^{(j)} \leftarrow \mathbf{g}_k^{(j)} - (\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)})\mathbf{c}_k^{(j)}$; % $\mathbf{v}_k^{(j)} \in \mathcal{V}_{j-1}$
11: $\quad\quad \mathbf{r}_{k+1}^{(j)} \leftarrow \mathbf{A}\mathbf{v}_k^{(j)} - \mathbf{v}_k^{(j)}\mu_j$; % $\mathbf{r}_{k+1}^{(j)} \in \mathcal{G}_j$
12: $\quad\quad \mathbf{g}_{k+1}^{(j)} \leftarrow \big(\mathbf{r}_{k+1}^{(j)} - \sum_{i=1}^{k} \mathbf{g}_i^{(j)} h_{i,k}^{(j)}\big)/h_{k+1,k}^{(j)}$; % $\mathbf{g}_{k+1}^{(j)} \in \mathcal{G}_j$
13: $\quad$ **end for**
14: **end for**

The choice of the scalars in the starting Krylov method and in line 7, line 12 will be used to derive our IDR algorithm with partial orthonormalization; the solution of the underdetermined systems in line 9 defines our four variants srIDR, fmIDR, mnIDR, and ovIDR. Mathematically speaking, the selection of the seed values is not very important, however, from a numerical point of view it is; the selection of appropriate seed values will be discussed in Section 4.

We assume that (3.2) is satisfied and show that for any fixed choice of the free parameters, provided that the algorithm does not break down, it generates vectors $\mathbf{g}_k$ that satisfy (3.1): In line 1 a matrix $\mathbf{G}_{s+1}^{(0)}$ is computed with $\mathsf{im}(\mathbf{G}_{s+1}^{(0)}) = \mathcal{K}_{s+1} \subseteq \mathcal{G}_0$. We recall that $\mathbf{G}_{s+1}^{(j-1)}$ satisfies assumption (3.2) and that $\mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) \subseteq \mathcal{G}_{j-1}$. By assumption (3.2c), the *Sonneveld coefficients* $\mathbf{c}_0^{(j)}$ can be computed in line 3 and determine a vector $\mathbf{v}_0^{(j)}$ in the intersection $\mathcal{V}_{j-1} = \mathcal{G}_{j-1} \cap \mathcal{S}$ in line 4. By a dimensional argument this vector is unique up to scaling if $\dim(\mathsf{im}(\mathbf{G}_{s+1}^{(j)}) + \mathcal{S}) = n$ since

$$\dim(\mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) \cap \mathcal{S}) = \underbrace{\mathsf{rank}(\mathbf{G}_{s+1}^{(j-1)})}_{s+1} + \underbrace{\dim(\mathcal{S})}_{n-s} - \underbrace{\dim(\mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) + \mathcal{S})}_{\leqslant n} \geqslant 1. \tag{3.4}$$

It is easy to prove that (3.2c) implies $\dim(\mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) + \mathcal{S}) = n$. To ensure a nonzero

ETNA
Kent State University and
Johann Radon Institute (RICAM)

component in the direction of the latest $\mathbf{g}_{s+1}^{(j)}$, we scale $\mathbf{v}_0^{(j)}$ such that this component is one,

$$\mathbf{v}_0^{(j)} = \mathbf{G}_{s+1}^{(j-1)} \begin{bmatrix} -\mathbf{c}_0^{(j)} \\ 1 \end{bmatrix} \quad \Rightarrow \quad \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{G}_s^{(j-1)} \mathbf{c}_0^{(j)} = \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{g}_{s+1}^{(j-1)},$$

which results in the linear system solved in line 3. We are free to choose a new seed value in line 5, which uniquely determines the next IDR space $\mathcal{G}_j$. Possible selection schemes are discussed in Section 4. In line 6 a first vector $\mathbf{r}_1^{(j)} \in \mathcal{G}_j$ is computed. As no more information about $\mathcal{G}_j$ is available at this step, the only possible transformation is a scaling; e.g., normalization. This is done in line 7 and results in the first $\mathbf{g}_1^{(j)} \in \mathcal{G}_j$. To repeat the whole procedure on the next level, we need s additional vectors $\mathbf{g}_{k+1}^{(j)} \in \mathcal{G}_j$, $1 \leqslant k \leqslant s$. Here we make use of the fact that $\mathcal{G}_j \subseteq \mathcal{G}_{j-1}$, which implies that $\mathsf{im}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}] \subseteq \mathcal{G}_{j-1}$. By assumption (3.2a) and (3.2b),

$$\mathsf{rank}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}] = s + k + 1,$$

so we can use a dimensional argument such as in (3.4):

$$\begin{aligned} (3.5) \quad \mathsf{dim}(\mathsf{im}([\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}]) \cap \mathcal{S}) \\ = \underbrace{\mathsf{rank}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}]}_{s+k+1} + \underbrace{\mathsf{dim}(\mathcal{S})}_{n-s} \\ - \underbrace{\mathsf{dim}(\mathsf{im}([\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}]) + \mathcal{S})}_{\leqslant n} \geqslant k + 1. \end{aligned}$$

Again, assumption (3.2c) implies $\mathsf{dim}(\mathsf{im}([\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}]) + \mathcal{S}) = n$. The vectors $\mathbf{v}_1^{(j)}, \ldots, \mathbf{v}_k^{(j)}$ in the intersection $\mathsf{im}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_k^{(j)}] \cap \mathcal{S} \subseteq \mathcal{G}_{j-1} \cap \mathcal{S} = \mathcal{V}_{j-1}$ are not uniquely defined. We are looking for a linear combination of the vectors in $\mathcal{G}_{j-1}$ that are *known* at this step and which lies in $\mathcal{S}$. To ensure that the right Krylov subspace is expanded, $\mathcal{K}_k \rightsquigarrow \mathcal{K}_{k+1}$, we scale the component of the current vector $\mathbf{g}_k^{(j)}$ to one, which results in the underdetermined linear systems in line 9. By assumption (3.2c), the $s \times (s+k)$ matrix in line 9 has full rank s for $1 \leqslant k \leqslant s$,

$$(3.6) \qquad s \geqslant \mathsf{rank}(\widehat{\mathbf{Q}}^{\mathsf{H}}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}]) \geqslant \mathsf{rank}(\widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{G}_s^{(j-1)}) = s,$$

thus the underdetermined systems are all solvable. We present four variants that result in *uniquely* defined vectors $\mathbf{v}_1^{(j)}, \ldots, \mathbf{v}_k^{(j)}$. Typically the four variants compute *different* $\mathbf{v}_1^{(j)}, \ldots, \mathbf{v}_k^{(j)}$:

- srIDR: We set the k first components of $\mathbf{c}_k^{(j)}$ to zero; this results in the *shortest recurrence* possible as we no longer need the vectors $\mathbf{g}_1^{(j-1)}, \ldots, \mathbf{g}_k^{(j-1)}$. This might not always be possible as the rank of the matrix

  $$(3.7) \qquad \mathbf{M}_{\mathsf{sr}}^{(j,k)} := \widehat{\mathbf{Q}}^{\mathsf{H}} \left[\mathbf{g}_{k+1}^{(j-1)}, \ldots, \mathbf{g}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}\right] \in \mathbb{C}^{s \times s}$$

  might be less than s. The Sonneveld coefficients of srIDR are given by

  $$(3.8) \qquad \mathbf{c}_k^{(j),\mathsf{sr}} := \begin{bmatrix} \mathbf{o}_k \\ \left(\mathbf{M}_{\mathsf{sr}}^{(j,k)}\right)^{-1} \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{g}_k^{(j)} \end{bmatrix}.$$

  This variant is used in [16, 17, 18, 19].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

- fmIDR: We set the k last components of $\mathbf{c}_k^{(j)}$ to zero; by assumption (3.2c) this results in a non-singular matrix

$$\mathbf{M}_{\mathsf{fm}}^{(j)} := \mathbf{M}_{\mathsf{fm}}^{(j,k)} := \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{G}_s^{(j-1)} \in \mathbb{C}^{s \times s}$$

that is used for $s+1$ consecutive steps; we use the *factored matrix* more than once. The Sonneveld coefficients of fmIDR are given by

$$\mathbf{c}_k^{(j),\mathsf{fm}} := \begin{bmatrix} \left(\mathbf{M}_{\mathsf{fm}}^{(j)}\right)^{-1} \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{g}_k^{(j)} \\ \mathbf{o}_k \end{bmatrix}.$$

This variant is sketched in [16, Section 4.3] and used in [14].

- mnIDR: We compute the *minimum-norm* solution of the underdetermined system and use the full-rank matrix

$$\mathbf{M}_{\mathsf{mn}}^{(j,k)} := \widehat{\mathbf{Q}}^{\mathsf{H}} [\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}] \in \mathbb{C}^{s \times (s+k)}.$$

The Sonneveld coefficients of mnIDR are given by

$$\mathbf{c}_k^{(j),\mathsf{mn}} := \left(\mathbf{M}_{\mathsf{mn}}^{(j,k)}\right)^{\dagger} \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{g}_k^{(j)}.$$

This variant has been described first in [9], and it is used in numerical examples in [23].

- ovIDR: We use the k degrees of freedom to *orthogonalize* against the computed *vectors* $\mathbf{v}_0^{(j)}, \ldots, \mathbf{v}_{k-1}^{(j)}$, which have to be stored, thereby increasing the storage. We define

$$\mathbf{V}_k^{(j)} := \left[\mathbf{v}_0^{(j)}, \ldots, \mathbf{v}_{k-1}^{(j)}\right] \in \mathbb{C}^{n \times k} \tag{3.9}$$

and

$$\mathbf{M}_{\mathsf{ov}}^{(j,k)} := [\widehat{\mathbf{Q}}, \mathbf{V}_k^{(j)}]^{\mathsf{H}} [\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}] \in \mathbb{C}^{(s+k) \times (s+k)}.$$

The Sonneveld coefficients of ovIDR are given by

$$\mathbf{c}_k^{(j),\mathsf{ov}} := \left(\mathbf{M}_{\mathsf{ov}}^{(j,k)}\right)^{-1} [\widehat{\mathbf{Q}}, \mathbf{V}_k^{(j)}]^{\mathsf{H}} \mathbf{g}_k^{(j)}.$$

This variant has not been published before. It is included for two reasons, both related to the orthonormalization: already the orthonormalization of the $\mathbf{g}$-vectors resulted in a more stable algorithm, and it is easy to detect linearly dependent $\mathbf{v}$-vectors and thus a possible breakdown in (3.2).

Regardless of the variant used, in line 10 *unique* vectors $\mathbf{v}_k^{(j)} \in \mathcal{V}_{j-1}$ in the intersection are computed. These are mapped to vectors $\mathbf{r}_{k+1}^{(j)} \in \mathcal{G}_j$ in line 11. As in step k of the inner loop already k previously computed vectors $\mathbf{g}_k^{(j)}$ exist, we can compute linear combinations with these without leaving $\mathcal{G}_j$, and we can scale the result. This is done in line 12, where $\mathbf{g}_{k+1}^{(j)} \in \mathcal{G}_j$ is computed. In this manner the algorithm computes $s+1$ vectors $\mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{s+1}^{(j)} \in \mathcal{G}_j$ and we can move to the next level.

The srIDR variant in [16] is implicitly based on scalars $h_{i,k}$ that sum column-wise to zero, $\sum_{i=1}^{k+1} h_{i,k} = 0$, the srIDR variant in [19] uses these scalars to enforce the orthogonality $\mathbf{e}_i^{\mathsf{T}} \widehat{\mathbf{Q}}^{\mathsf{H}} \mathbf{g}_{k+1}^{(j)} = 0$, $1 \leqslant i \leqslant k \leqslant s$, the fmIDR variant in [14] uses them to orthonormalize the vectors $\mathbf{v}_0^{(j)}, \ldots, \mathbf{v}_s^{(j)}$. A natural idea, first mentioned in [9] and first published for srIDR in [18], is to orthonormalize the resulting vectors $\mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{s+1}^{(j)}$; the more general algorithm is described in the next subsection.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

3.2. IDR with partial orthonormalization. Here we specialize Algorithm 3.2 to an IDR algorithm with partial orthonormalization. We replace the generic Krylov method given in Algorithm 3.1 by Arnoldi's process [1], $[h_{1,0}, \underline{\mathbf{H}}_m, \mathbf{G}_{m+1}] \leftarrow \mathsf{Arnoldi}(\mathbf{A}, \mathbf{q}, m)$; see Algorithm 3.3. This ensures that $\mathbf{G}_{s+1}^{(0)}$ is orthonormal; see (3.3).

Algorithm 3.3 Arnoldi

INPUT: $\mathbf{A} \in \mathbb{C}^{n\times n}$; $\mathbf{q} \in \mathbb{C}^n$; $m \in \mathbb{N}$.
OUTPUT: $h_{1,0} \in \mathbb{C}$; $\underline{\mathbf{H}}_m \in \mathbb{C}^{(m+1)\times m}$; $\mathbf{G}_{m+1} \in \mathbb{C}^{n\times(m+1)}$.

1: $h_{1,0} \leftarrow \|\mathbf{q}\|$;
2: $\mathbf{g}_1 \leftarrow \mathbf{q}/h_{1,0}$;
3: **for** $k = 1 : m$ **do**
4: $\quad \mathbf{r}_{k+1} \leftarrow \mathbf{A}\mathbf{g}_k$;
5: $\quad$ **for** $i = 1 : k$ **do**
6: $\quad\quad h_{i,k} \leftarrow \mathbf{g}_i^{\mathsf{H}}\mathbf{r}_{k+1}$;
7: $\quad$ **end for**
8: $\quad \mathbf{p}_{k+1} \leftarrow \mathbf{r}_{k+1} - \sum_{i=1}^{k} \mathbf{g}_i h_{i,k}$;
9: $\quad h_{k+1,k} \leftarrow \|\mathbf{p}_{k+1}\|$;
10: $\quad \mathbf{g}_{k+1} \leftarrow \mathbf{p}_{k+1}/h_{k+1,k}$;
11: **end for**

We add the orthonormalization scheme to Algorithm 3.2 to ensure (3.3) and replace the solution of the underdetermined systems in line 9 of Algorithm 3.2 by one of the variants srIDR, fmIDR, mnIDR, or ovIDR to obtain Algorithm 3.4, IDR with partial orthonormalization.

3.3. The generalized Hessenberg decomposition. In this subsection we collect the relations between the vectors constructed and the scalars used into matrix equations that will be useful later on. We define the *local matrices*

$$\mathbf{R}_{s+1}^{(j)} := \left[\mathbf{r}_1^{(j)}, \ldots, \mathbf{r}_{s+1}^{(j)}\right] \in \mathbb{C}^{n\times(s+1)}$$

collecting vectors computed in line 6 and line 11 of Algorithm 3.2 and use $\mathbf{V}_{s+1}^{(j)}$ as defined by (3.9). In the call to Krylov in Algorithm 3.2 in line 1 and in line 12 of Algorithm 3.2, GR decompositions[2] of $\left[\mathbf{q}, \mathbf{A}\mathbf{G}_s^{(0)}\right]$ and $\mathbf{R}_{s+1}^{(j)}$, $1 \leqslant j$, are computed, respectively:

$$\left[\mathbf{q}, \mathbf{A}\mathbf{G}_s^{(0)}\right] = \mathbf{G}_{s+1}^{(0)} \begin{bmatrix} \underline{\mathbf{e}}_1 h_{1,0}^{(0)} & \underline{\mathbf{H}}_s^{(0)} \end{bmatrix}, \quad \mathbf{R}_{s+1}^{(j)} = \mathbf{G}_{s+1}^{(j)} \begin{bmatrix} \underline{\mathbf{e}}_1 h_{1,0}^{(j)} & \underline{\mathbf{H}}_s^{(j)} \end{bmatrix}. \tag{3.10}$$

In Algorithm 3.3 and Algorithm 3.4 these are QR decompositions.

We define the *global matrix* $\mathbf{V}_m$ in terms of local vectors and matrices by

$$\mathbf{V}_m := \left[\mathbf{g}_1^{(0)}, \ldots, \mathbf{g}_s^{(0)}, \mathbf{V}_{s+1}^{(1)}, \ldots, \mathbf{V}_{s+1}^{(j-1)}, \mathbf{v}_0^{(j)}, \ldots, \mathbf{v}_{k-1}^{(j)}\right].$$

In line 4 and in line 10 of Algorithm 3.2 the vectors $\mathbf{v}_0^{(j)}, \ldots, \mathbf{v}_s^{(j)} \in \mathcal{V}_{j-1}$ are computed as linear combinations

$$\mathbf{V}_{s+1}^{(j)} = \begin{bmatrix} \mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)} \end{bmatrix} \begin{bmatrix} -\mathbf{c}_0^{(j)} & \ddots & -\mathbf{c}_s^{(j)} \\ 1 & \ddots & 1 \\ \mathbf{o}_{s+1} & \ddots & \mathbf{o}_1 \end{bmatrix} =: \begin{bmatrix} \mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)} \end{bmatrix} \underline{\mathbf{U}}_{s+1}^{(j)}. \tag{3.11}$$

[2] A GR decomposition is a decomposition of the form "general matrix" times "upper triangular matrix"; see [20].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Algorithm 3.4 IDR (partial orthonormalization)

INPUT: $\mathbf{A} \in \mathbb{C}^{n\times n}$; $\mathbf{q} \in \mathbb{C}^n$; $\widehat{\mathbf{Q}} \in \mathbb{C}^{n\times s}$; $m \in \mathbb{N}$.
OUTPUT: $\mathbf{G}_{m+1} \in \mathbb{C}^{n\times(m+1)}$; ... % see (2.3)

$[h_{1,0}^{(0)}, \underline{\mathbf{H}}_s^{(0)}, \mathbf{G}_{s+1}^{(0)}] \leftarrow \mathsf{Arnoldi}(\mathbf{A}, \mathbf{q}, s)$; % $\mathsf{im}(\mathbf{G}_{s+1}^{(0)}) = \mathcal{K}_{s+1} \subset \mathcal{G}_0$
for $j = 1 \ldots$ **do**
  $\mathbf{c}_0^{(j)} \leftarrow (\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j-1)})^{-1}(\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_{s+1}^{(j-1)})$; % see (3.2)
  $\mathbf{v}_0^{(j)} \leftarrow \mathbf{g}_{s+1}^{(j-1)} - \mathbf{G}_s^{(j-1)}\mathbf{c}_0^{(j)}$; % $\mathbf{v}_0^{(j)} \in \mathsf{im}(\mathbf{G}_{s+1}^{(j-1)}) \cap \ker(\widehat{\mathbf{Q}}^{\mathsf{H}}) \subset \mathcal{V}_{j-1}$
  choose μ_j % discussed in Section 4
  $\mathbf{r}_1^{(j)} \leftarrow \mathbf{A}\mathbf{v}_0^{(j)} - \mathbf{v}_0^{(j)}\mu_j$; % $\mathbf{r}_1^{(j)} \in (\mathbf{A} - \mu_j\mathbf{I})\mathcal{V}_{j-1} = \mathcal{G}_j$
  $h_{1,0}^{(j)} \leftarrow \|\mathbf{r}_1^{(j)}\|$;
  $\mathbf{g}_1^{(j)} \leftarrow \mathbf{r}_1^{(j)}/h_{1,0}^{(j)}$; % $\mathbf{g}_1^{(j)} \in \mathcal{G}_j$
  for $k = 1 : s$ **do**
    if srIDR **then**
      $\mathbf{c}_k^{(j)} \leftarrow \left[\mathbf{o}_k; \left(\widehat{\mathbf{Q}}^{\mathsf{H}}[\mathbf{g}_{k+1}^{(j-1)}, \ldots, \mathbf{g}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}]\right)^{-1}\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_k^{(j)}\right]$;
    else if fmIDR **then**
      $\mathbf{c}_k^{(j)} \leftarrow \left[(\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j-1)})^{-1}\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_k^{(j)}; \mathbf{o}_k\right]$;
    else if mnIDR **then**
      $\mathbf{c}_k^{(j)} \leftarrow \left(\widehat{\mathbf{Q}}^{\mathsf{H}}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}]\right)^{\dagger}\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_k^{(j)}$;
    else if ovIDR **then**
      $\mathbf{c}_k^{(j)} \leftarrow \left([\widehat{\mathbf{Q}}, \mathbf{V}_k^{(j)}]^{\mathsf{H}}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}]\right)^{-1}\left([\widehat{\mathbf{Q}}, \mathbf{V}_k^{(j)}]^{\mathsf{H}}\mathbf{g}_k^{(j)}\right)$;
    else
      solve $\widehat{\mathbf{Q}}^{\mathsf{H}}[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)}]\mathbf{c}_k^{(j)} = \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_k^{(j)}$; % see (3.6)
    end if
    $\mathbf{v}_k^{(j)} \leftarrow \mathbf{g}_k^{(j)} - (\mathbf{G}_{s+1}^{(j-1)}, \mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{k-1}^{(j)})\mathbf{c}_k^{(j)}$; % $\mathbf{v}_k^{(j)} \in \mathcal{V}_{j-1}$
    $\mathbf{r}_{k+1}^{(j)} \leftarrow \mathbf{A}\mathbf{v}_k^{(j)} - \mathbf{v}_k^{(j)}\mu_j$; % $\mathbf{r}_{k+1}^{(j)} \in \mathcal{G}_j$
    for $i = 1 : k$ **do**
      $h_{i,k}^{(j)} \leftarrow (\mathbf{g}_i^{(j)})^{\mathsf{H}}\mathbf{r}_{k+1}^{(j)}$;
    end for
    $\mathbf{p}_{k+1}^{(j)} \leftarrow \mathbf{r}_{k+1}^{(j)} - \sum_{i=1}^{k}\mathbf{g}_i^{(j)}h_{i,k}^{(j)}$;
    $h_{k+1,k}^{(j)} \leftarrow \|\mathbf{p}_{k+1}^{(j)}\|$;
    $\mathbf{g}_{k+1}^{(j)} \leftarrow \mathbf{p}_{k+1}^{(j)}/h_{k+1,k}^{(j)}$; % $\mathbf{g}_{k+1}^{(j)} \in \mathcal{G}_j$
  end for
end for

The local matrices $\mathbf{R}_{s+1}^{(j)}$ are given by

$$\mathbf{R}_{s+1}^{(j)} = \mathbf{A}\mathbf{V}_{s+1}^{(j)} - \mathbf{V}_{s+1}^{(j)}\operatorname{diag}(\mu_j, \ldots, \mu_j). \tag{3.12}$$

Combining equations (3.10)–(3.12), we obtain the coupling between two local blocks $\mathbf{G}_{s+1}^{(j-1)}$ and $\mathbf{G}_{s+1}^{(j)}$, $j \geqslant 1$,

$$\begin{aligned}&\mathbf{A}\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\underline{\mathbf{U}}_{s+1}^{(j)}\\ &\qquad = \begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\left(\begin{bmatrix}\mathbf{o}_{s+1} & \mathbf{O}_{s+1,s}\\ \mathbf{e}_1 h_{1,0}^{(j)} & \underline{\mathbf{H}}_s^{(j)}\end{bmatrix} + \underline{\mathbf{U}}_{s+1}^{(j)}\operatorname{diag}(\mu_j, \ldots, \mu_j)\right).\end{aligned} \tag{3.13}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Gluing these relations together and topping them with the first equation in (3.10), we obtain the *generalized Hessenberg decomposition*

$$\mathbf{A}\mathbf{G}_m\mathbf{U}_m = \mathbf{G}_{m+1}\underline{\mathbf{H}}_m^{\text{total}}, \quad \underline{\mathbf{H}}_m^{\text{total}} := \underline{\mathbf{H}}_m + \underline{\mathbf{U}}_m\mathbf{D}_m \tag{3.14}$$

of IDR that captures the recurrences of the vectors $\mathbf{g}_k$, $1 \leqslant m+1$, in both Algorithm 3.2 and Algorithm 3.4. The structure of the resulting matrices is described below.

The matrix $\underline{\mathbf{U}}_m \in \mathbb{C}^{(m+1)\times m}$ has $\mathbf{I}_s$ as leading $s \times s$ block, followed by all $\underline{\mathbf{U}}_{s+1}^{(j)}$, $j \geqslant 1$, aligned such that all ones are on the diagonal; the last block column may have less than $s+1$ columns. The matrix $\mathbf{U}_m$ results from $\underline{\mathbf{U}}_m$ by stripping of the last (zero) row. $\mathbf{U}_m$ is unit upper triangular, banded with upper bandwith $2s$, and has a staircase-like structure; see the example (3.15) taken from [23], where $\mathbf{U}_m$ and $\underline{\mathbf{H}}_m$ are depicted for $s = 2$ and $m = 9 = 3(s+1)$,

$$\mathbf{U}_9 = \left[\begin{array}{c}\circ\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\end{array}\right], \qquad \underline{\mathbf{H}}_9 = \left[\begin{array}{c}\bullet\bullet\\ \circ\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\end{array}\right]. \tag{3.15}$$

Circles in $\mathbf{U}_9$ depict the unit diagonal elements and bullets in $\mathbf{U}_9$ depict the Sonneveld coefficients $-\mathbf{c}_k^{(j)}$ defined by the IDR variant; e.g., srIDR (line 3 & line 11 of Algorithm 3.4), fmIDR (line 3 & line 13 of Algorithm 3.4), mnIDR (line 3 & line 15 of Algorithm 3.4), or ovIDR (line 3 & line 17 of Algorithm 3.4). The matrices $\mathbf{U}_m^{\text{sr}}$ and $\mathbf{U}_m^{\text{fm}}$ have additional known zero elements (e.g., the upper bandwidth of $\mathbf{U}_m^{\text{sr}}$ drops from $2s$ to s); the structure is depicted here for $s = 2$ and $m = 9 = 3(s+1)$,

$$\mathbf{U}_9^{\text{sr}} = \left[\begin{array}{c}\circ\\ \circ\bullet\bullet\\ \circ\bullet\bullet\\ \circ\bullet\bullet\\ \circ\bullet\bullet\\ \circ\bullet\bullet\\ \circ\bullet\\ \circ\end{array}\right], \qquad \mathbf{U}_9^{\text{fm}} = \left[\begin{array}{c}\circ\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\bullet\bullet\\ \circ\bullet\\ \circ\end{array}\right].$$

The matrix $\underline{\mathbf{H}}_m \in \mathbb{C}^{(m+1)\times m}$ has $\underline{\mathbf{H}}_s^{(0)}$ as leading $(s+1) \times s$ block, followed by all upper triangular basis transformation matrices

$$\begin{bmatrix}\underline{\mathbf{e}}_1 h_{1,0}^{(j)} & \underline{\mathbf{H}}_s^{(j)}\end{bmatrix} \in \mathbb{C}^{(s+1)\times(s+1)}, \quad j \geqslant 1,$$

aligned such that the nonzero scaling elements $h_{k+1,k}^{(j)}$ are on the first subdiagonal and the last block column may have less than $s+1$ columns. The band matrix $\underline{\mathbf{H}}_m$ is an unreduced extended upper Hessenberg matrix with upper bandwidth $s-1$. The example (3.15) reveals the structure of $\underline{\mathbf{H}}_9$ for $s = 2$. Circles in $\underline{\mathbf{H}}_9$ depict the nonzero scaling elements $h_{k+1,k}^{(j)}$, $0 \leqslant k \leqslant s$, $j \geqslant 0$, omitting $h_{1,0}^{(0)}$. Bullets depict the other elements $h_{i,k}^{(j)}$, $1 \leqslant i \leqslant k \leqslant s$, $j \geqslant 0$, that are used in Algorithm 3.2 and Algorithm 3.4.

The diagonal matrix $\mathbf{D}_m \in \mathbb{C}^{m\times m}$ is obtained by taking an $s \times s$ zero matrix and diagonally gluing together all diagonal matrices $\mu_j\mathbf{I}_{s+1}$ from (3.13); i.e.,

$$\mathbf{D}_m = \operatorname{diag}(\underbrace{0,\ldots,0}_{s \text{ times}},\underbrace{\mu_1,\ldots,\mu_1}_{s+1 \text{ times}},\underbrace{\mu_2,\ldots,\mu_2}_{s+1 \text{ times}},\ldots,\underbrace{\mu_j,\ldots,\mu_j}_{k \text{ times}}), \quad j = \left\lfloor \frac{m}{s+1} \right\rfloor,$$

where $k = m+1-j(s+1)$.

The generalized Hessenberg decomposition (3.14) is the basis for different algorithms to approximate solutions of linear systems and eigenvectors. These are introduced rather briefly in the next sections.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

3.3.1. Linear systems. In this section we use boldface $\mathbf{r}$ to denote residual vectors. We want to approximate the solution $\mathbf{x}$ of a given linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. Let $\mathbf{x}_0$ be an initial approximation and define the residual $\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$. Suppose that Algorithm 3.4 is invoked with $\mathbf{q} = \mathbf{r}_0$. Then by (3.14)

$$\begin{aligned} \mathbf{A}\mathbf{G}_m\mathbf{U}_m = \mathbf{G}_{m+1}\underline{\mathbf{H}}_m^{\text{total}} = \mathbf{G}_m\mathbf{H}_m^{\text{total}} + \mathbf{g}_{m+1}h_{m+1,m}\mathbf{e}_m^{\mathsf{T}}, \\ \mathbf{G}_{m+1}\underline{\mathbf{e}}_1 h_{1,0}^{(0)} = \mathbf{G}_m\mathbf{e}_1 h_{1,0}^{(0)} = \mathbf{r}_0. \end{aligned} \tag{3.16}$$

In the OR approach [9, p. 1048], we define the mth *OR solution* $\mathbf{z}_m \in \mathbb{C}^m$ and the mth *OR iterate* $\mathbf{x}_m \in \mathbb{C}^n$ by

$$\mathbf{z}_m := \left(\mathbf{H}_m^{\text{total}}\right)^{-1}\mathbf{e}_1 h_{1,0}^{(0)}, \quad \mathbf{x}_m := \mathbf{x}_0 + \mathbf{V}_m\mathbf{z}_m. \tag{3.17}$$

The mth OR solution need not exist. By (3.16), the mth *OR residual* $\mathbf{r}_m \in \mathbb{C}^n$ is given by

$$\mathbf{r}_m := \mathbf{b} - \mathbf{A}\mathbf{x}_m = -\mathbf{g}_{m+1}h_{m+1,m}\mathbf{e}_m^{\mathsf{T}}\mathbf{z}_m, \quad \|\mathbf{r}_m\| = |h_{m+1,m}\mathbf{e}_m^{\mathsf{T}}\mathbf{z}_m|,$$

thus, the mth residual is parallel to the next vector $\mathbf{g}_{m+1}$. The computation of $\mathbf{x}_m$ is possible without the need to store all vectors $\mathbf{g}_1, \ldots, \mathbf{g}_m$.

In the MR approach [9, p. 1048], we define the mth *MR solution* $\underline{\mathbf{z}}_m \in \mathbb{C}^m$ and the mth *MR iterate* $\underline{\mathbf{x}}_m \in \mathbb{C}^n$ by

$$\underline{\mathbf{z}}_m := \left(\underline{\mathbf{H}}_m^{\text{total}}\right)^{\dagger}\underline{\mathbf{e}}_1 h_{1,0}^{(0)}, \quad \underline{\mathbf{x}}_m := \mathbf{x}_0 + \mathbf{V}_m\underline{\mathbf{z}}_m. \tag{3.18}$$

The mth MR solution always exists. The mth *MR residual* $\underline{\mathbf{r}}_m \in \mathbb{C}^n$ is defined by and can be bounded using (3.16) by

$$\underline{\mathbf{r}}_m := \mathbf{b} - \mathbf{A}\underline{\mathbf{x}}_m, \quad \|\underline{\mathbf{r}}_m\| \leqslant \|\mathbf{G}_{m+1}\|\|\underline{\mathbf{H}}_m^{\text{total}}\underline{\mathbf{z}}_m - \underline{\mathbf{e}}_m h_{1,0}^{(0)}\|.$$

By [9, Lemma 4, p. 1058], $\|\mathbf{G}_{m+1}\| \leqslant \sqrt{\lceil (m+1)/(s+1) \rceil}$ in case of Algorithm 3.4. An implementation of the MR approach for the srIDR variant of IDR with partial orthonormalization is given in [18].

3.3.2. Eigenvalue problems. The seed values are eigenvalues of the *Sonneveld pencil* $(\mathbf{H}_m^{\text{total}}, \mathbf{U}_m)$; see [23]. The other eigenvalues θ_j can be used as approximations to eigenvalues of $\mathbf{A}$. The corresponding approximate eigenvectors $\mathbf{y}_j$ are given by

$$\mathbf{H}_m^{\text{total}}\mathbf{s}_j = \theta_j\mathbf{U}_m\mathbf{s}_j, \quad \mathbf{y}_j := \mathbf{V}_m\mathbf{s}_j.$$

It is possible to define other pencils based on the seed values, Sonneveld coefficients and orthonormalization coefficients to compute eigenvalues (see [23]) or to extend this *Ritz approach* to the so-called *harmonic Ritz approach* [9, p. 1047].

4. Seed values. From a mathematical point of view the selection of the seed values is not that important; the induced dimension reduction occurs independently of their selection as long as no seed value is an eigenvalue of $\mathbf{A}$. A natural idea is to use a fixed seed value, $\mu_j = \mu$, $j \geqslant 1$, e.g., $\mu = 0$. The latter choice results in a singular $\mathbf{H}_m^{\text{total}} = \mathbf{H}_m$ for all $m > s$, and the OR approach (3.17) fails for all $m > s$ while the MR approach (3.18) stagnates for all $m > s$ [9, Lemma 3, p. 1057].

A constant μ results in a Jordan block at μ in the Sonneveld pencil because the matrix $\mathbf{H}_m^{\text{total}} - \mu\mathbf{U}_m = \mathbf{H}_m + \mathbf{U}_m(\mathbf{D}_m - \mu\mathbf{I}_m)$ has the same nonzero structure as $\mathbf{H}_m$, i.e., $\mathbf{H}_m^{\text{total}} - \mu\mathbf{U}_m$ has the eigenvalue 0 with algebraic multiplicity at least $\lfloor m/(s+1) \rfloor$ and

ETNA
Kent State University and
Johann Radon Institute (RICAM)

geometric multiplicity 1; compare with the example (3.15). This might cause problems with numerical eigenvalue computations if $\mathbf{A}$ has eigenvalues close to μ.

Numerically, IDR and other Sonneveld methods deviate from their exact counterparts, and ghost eigenvalues close to the seed values are computed. Numerical experiments indicate that the best constant seed value is the mean $\mu = \mathsf{trace}(\mathbf{A})/n$ of the eigenvalues of $\mathbf{A}$.

More interesting are seed value selection schemes that take into account local information when computing μ_j, mostly the vectors $\mathbf{v}_0^{(j)}$ and $\mathbf{A}\mathbf{v}_0^{(j)}$. We present a few general schemes, divided into those designed for linear systems and those designed for eigenvalue problems. A new scheme is presented that combines the ideas underlying both approaches.

4.1. Seed values for linear systems. OR methods construct residuals parallel to the vectors $\mathbf{g}_{k+1}$. The residuals in a Krylov subspace method can always be written in the form $\mathbf{r}_k = r_k(\mathbf{A})\mathbf{r}_0$, where the *residual polynomial* r_k satisfies $r_k(0) = 1$. To minimize the residual, we think in terms of residual polynomials and replace $z - \mu_j$ by the differently scaled $1 - z\omega_j$, where $\omega_j = \mu_j^{-1}$, and minimize the scaled $\mathbf{r}_1^{(j)}$ with respect to ω_j,

$$\min_{\omega_j \in \mathbb{C}} \|\mathbf{v}_0^{(j)} - \mathbf{A}\mathbf{v}_0^{(j)}\omega_j\| \quad \Rightarrow \quad \omega_j = \frac{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}}{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}, \tag{4.1}$$

i.e., we define μ_j by

$$\mu_j := \omega_j^{-1} = \frac{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}}. \tag{4.2}$$

This results in a *harmonic Rayleigh quotient*, i.e., the resulting μ_j are inverses of elements of the field of values of the inverse of $\mathbf{A}$ [10] since

$$\mu_j = \frac{\widetilde{\mathbf{v}}^{\mathsf{H}}\widetilde{\mathbf{v}}}{\widetilde{\mathbf{v}}^{\mathsf{H}}\mathbf{A}^{-1}\widetilde{\mathbf{v}}}, \quad \text{where} \quad \widetilde{\mathbf{v}} := \mathbf{A}\mathbf{v}_0^{(j)}.$$

In [12] the resulting linear polynomials $1 - z\omega_j$ are termed MR(1)-polynomials. This approach is used in [16, 19, 21] and results in seed values that are not too close to zero. It turns out that it is unstable for $\mathbf{A}$ such that the field of values includes zero since then the seed values may become very large.

A modification known as "vanilla" technique has been developed and motivated for BiCGStab and related methods in [12, Theorem 3.1, p. 210; Eqn. (28), p. 213]: compute the minimizer (4.1) and the cosine c of the Hermitean angle between $\mathbf{A}\mathbf{v}_0^{(j)}$ and $\mathbf{v}_0^{(j)}$,

$$c := \frac{\left|\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{v}_0^{(j)}\right|}{\|\mathbf{A}\mathbf{v}_0^{(j)}\|\|\mathbf{v}_0^{(j)}\|}.$$

If $c < \kappa$ (i.e., if this angle is too large[3]) then ω_j is scaled, and the new value

$$\begin{aligned} \tilde{\omega}_j := \frac{\kappa}{c}\omega_j, \quad \mu_j := \tilde{\omega}_j^{-1} &= \kappa^{-1} \cdot \frac{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}} \cdot \frac{\left|\left(\mathbf{A}\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}\right|}{\|\mathbf{A}\mathbf{v}_0^{(j)}\|\|\mathbf{v}_0^{(j)}\|} \\ &= \kappa^{-1} \cdot \frac{\|\mathbf{A}\mathbf{v}_0^{(j)}\|}{\|\mathbf{v}_0^{(j)}\|} \cdot \mathsf{sign}\left(\frac{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}}\right) \end{aligned} \tag{4.3}$$

is used. This modification ensures that all computed seed values are only moderately outside the field of values of $\mathbf{A}$ and not too close to zero.

[3] In [12] the value $\kappa = 0.7$ is used as upper bound, which corresponds to a rounded value of the obvious choice $\sqrt{2}/2 = \cos(\pi/4)$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

4.2. Seed values for eigenvalue problems. A natural idea in eigenvalue computations is to minimize $\mathbf{r}_1^{(j)}$ with respect to μ_j. This gives the Rayleigh quotient with $\mathbf{v}_0^{(j)}$, and as a consequence, $\mathbf{r}_1^{(j)}$ is perpendicular to $\mathbf{v}_0^{(j)}$,

$$\min_{\mu_j \in \mathbb{C}} \|\mathbf{A}\mathbf{v}_0^{(j)} - \mathbf{v}_0^{(j)}\mu_j\| \quad \Rightarrow \quad \mu_j := \frac{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}} \mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}} \mathbf{v}_0^{(j)}}, \quad \mathbf{r}_1^{(j)} \perp \mathbf{v}_0^{(j)}. \tag{4.4}$$

This technique ensures that all computed seed values are in the field of values of $\mathbf{A}$. If the field of values encloses zero, a zero or small seed value may occur. This leads to problems in the OR and MR approaches.

We could use other Rayleigh quotients. We can ensure that the last diagonal element in $\mathbf{H}_{s+1}^{\text{total}}$ is the same as in Arnoldi's process if we set

$$\mu_1 = \frac{\mathbf{g}_{s+1}^{\mathsf{H}} \mathbf{A}\mathbf{g}_{s+1}}{\mathbf{g}_{s+1}^{\mathsf{H}} \mathbf{g}_{s+1}}, \quad \text{i.e., we set} \quad \mu_j := \frac{\left(\mathbf{g}_{s+1}^{(j-1)}\right)^{\mathsf{H}} \mathbf{A}\mathbf{g}_{s+1}^{(j-1)}}{\left(\mathbf{g}_{s+1}^{(j-1)}\right)^{\mathsf{H}} \mathbf{g}_{s+1}^{(j-1)}}. \tag{4.5}$$

In every cycle, IDR computes $s+1$ vectors that are a basis of a Krylov subspace of a deflated matrix $\mathbf{A}$, where the eigenvalues of interest are those of the undeflated $\mathbf{A}$. This choice ensures for $s+1$ consecutive diagonal elements that they are equal to those of Arnoldi's method, which mimics a restarted Arnoldi's method.

In [10] $\lfloor m/(s+1) \rfloor$ extra multiplications by $\mathbf{A}$ are invested to compute Ritz values using Arnoldi's method that are used as seed values.

4.3. A balanced approach to seed values. The approaches (4.2) and (4.4) minimize the norm of a multiple of $\mathbf{r}_1^{(j)}$ subject to a scaling of the vector $\mathbf{A}\mathbf{v}_0^{(j)}$ (see (4.2) and (4.1)) or subject to a scaling of the vector $\mathbf{v}_0^{(j)}$; see (4.4).

To treat both ingredients equally, we normalize both $\mathbf{A}\mathbf{v}_0^{(j)}$ and $\mathbf{v}_0^{(j)}$ to get rid of scaling issues with large or small $\mathbf{A}$, solve

$$\min_{\alpha,\beta \in \mathbb{C}} \left\| \frac{\mathbf{A}\mathbf{v}_0^{(j)}}{\|\mathbf{A}\mathbf{v}_0^{(j)}\|}\alpha - \frac{\mathbf{v}_0^{(j)}}{\|\mathbf{v}_0^{(j)}\|}\beta \right\| \quad \text{s.t.} \quad \left\| \begin{bmatrix} \alpha \\ -\beta \end{bmatrix} \right\| = 1, \tag{4.6}$$

and set $\mu_j = \beta/\alpha \cdot \|\mathbf{A}\mathbf{v}_0^{(j)}\|/\|\mathbf{v}_0^{(j)}\|$. This is a mixture of an eigenvalue-based and a linear system solver-based approach. We expect the seed values to be away from zero for non-singular $\mathbf{A}$ and not too large.

The solution of (4.6) is given by the left singular vector of the smallest singular value of the matrix

$$\mathbf{B} := \begin{bmatrix} \dfrac{\mathbf{A}\mathbf{v}_0^{(j)}}{\|\mathbf{A}\mathbf{v}_0^{(j)}\|} & \dfrac{\mathbf{v}_0^{(j)}}{\|\mathbf{v}_0^{(j)}\|} \end{bmatrix}.$$

We compute this singular vector as the eigenvector to the smallest eigenvalue of

$$\mathbf{B}^{\mathsf{H}}\mathbf{B} = \begin{bmatrix} 1 & \overline{b} \\ b & 1 \end{bmatrix}, \quad b := \frac{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}} \mathbf{A}\mathbf{v}_0^{(j)}}{\|\mathbf{A}\mathbf{v}_0^{(j)}\|\|\mathbf{v}_0^{(j)}\|}.$$

The eigenvector to the smallest eigenvalue $1 - |b|$ is given by

$$\begin{bmatrix} 1 & \overline{b} \\ b & 1 \end{bmatrix} \begin{bmatrix} |b| \\ -b \end{bmatrix} = \begin{bmatrix} |b| \\ -b \end{bmatrix} (1 - |b|),$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

which leads to a "simplified vanilla scheme" that we call "cinnamon" technique,

$$\mu_j := \frac{\|\mathbf{A}\mathbf{v}_0^{(j)}\|}{\|\mathbf{v}_0^{(j)}\|} \cdot \frac{b}{|b|} = \frac{\|\mathbf{A}\mathbf{v}_0^{(j)}\|}{\|\mathbf{v}_0^{(j)}\|} \cdot \mathsf{sign}\left(\frac{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{v}_0^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}}\right). \tag{4.7}$$

This scheme is a mixture between an eigenvalue based and an SVD based approach: we take the direction given by the Rayleigh quotient but the length given by the amplification of $\mathbf{v}_0^{(j)}$ by $\mathbf{A}$. These values will be on the annulus defined by $\{z \in \mathbb{C} \mid \sigma_n(\mathbf{A}) \leqslant |z| \leqslant \sigma_1(\mathbf{A})\}$ and in direction of the field of values of $\mathbf{A}$. This approach might cause problems: if $\mathbf{A}\mathbf{v}_0^{(j)} \perp \mathbf{v}_0^{(j)}$, then both singular values coincide and the sign in (4.7) is not defined.

4.4. Additional orthogonality. Additional orthogonality between the last $\mathbf{g}_{s+1}^{(j-1)}$ and the new $\mathbf{g}_1^{(j)}$ can be enforced by setting

$$\mu_j := \frac{\left(\mathbf{g}_{s+1}^{(j)}\right)^{\mathsf{H}}\mathbf{A}\mathbf{v}_0^{(j)}}{\left(\mathbf{g}_{s+1}^{(j)}\right)^{\mathsf{H}}\mathbf{v}_0^{(j)}} \quad \text{because then} \quad \mathbf{g}_1^{(j)} \parallel \mathbf{r}_1^{(j)} = \mathbf{A}\mathbf{v}_0^{(j)} - \mathbf{v}_0^{(j)}\mu_j \perp \mathbf{g}_{s+1}^{(j-1)}. \tag{4.8}$$

Unfortunately, numerical experiments[4] indicate that this approach is very unstable; many of the resulting values of μ_j lie far outside the field of values of $\mathbf{A}$.

5. Mathematical equivalence and classification of breakdowns. The following theorem states that the four variants of IDR with partial orthonormalization are all equivalent as long as assumption (3.2) holds true, except that the srIDR variant may break down.

THEOREM 5.1. *Suppose that $\widehat{\mathbf{Q}}$ is chosen such that assumption (3.2) holds true and that one of the following seed value selection schemes*

- *preselected seed values, e.g., constant or a list of given seed values,*
- *a local seed value selection scheme; i.e., (4.2), (4.3), (4.4), (4.5), (4.7), (4.8)*

is used.

Then the variants fmIDR, mnIDR, *and* ovIDR *of IDR with partial orthonormalization are mathematically equivalent, in particular, given the same input data, they compute the same vectors $\mathbf{g}_k$, $k \geqslant 1$.*

There exist cases where assumption (3.2) holds true and the srIDR *variant of IDR with partial orthonormalization breaks down, which we term a* pivot breakdown. *When no pivot breakdown in the* srIDR *variant occurs, it constructs the same vectors $\mathbf{g}_k$, $k \geqslant 1$, as the other three variants.*

Proof. All four variants of IDR with partial orthonormalization are completely deterministic. We first suppose that no variant breaks down and prove that the spaces $\mathcal{G}_j$ and the vectors $\mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{s+1}^{(j)} \in \mathcal{G}_j$, $j \geqslant 0$, are the same in all four variants, regardless of the choice of the seed value selection scheme listed in the theorem.

The IDR spaces $\mathcal{G}_j$, $j \geqslant 1$, are uniquely defined by the seed values μ_j, $j \geqslant 1$, which in turn are either fixed or computed based on the vector $\mathbf{v}_0^{(j)}$ and, possibly, the vector $\mathbf{g}_{s+1}^{(j-1)}$. The vector $\mathbf{v}_0^{(j)}$ is the same vector in all four variants if the $s+1$ orthonormal vectors $\mathbf{g}_1^{(j-1)}, \ldots, \mathbf{g}_{s+1}^{(j-1)} \in \mathcal{G}_{j-1}$ are the same in all four variants, which implies that when additionally all $\mathcal{G}_\ell$, $\ell < j$, are the same, then in this case the next $\mathcal{G}_j$ is the same in all four variants.

The initial vectors $\mathbf{g}_1^{(0)}, \ldots, \mathbf{g}_{s+1}^{(0)} \in \mathcal{G}_0$ are uniquely determined by Arnoldi's process and the positive signs of the nonzero scaling elements. We prove that if the previous vectors

[4] This observation was also made by Martin van Gijzen, who first came up with this idea and experimented with this kind of additional orthogonality.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$\mathbf{g}_1^{(j-1)}, \ldots, \mathbf{g}_{s+1}^{(j-1)} \in \mathcal{G}_{j-1}$ are uniquely determined, then the next $s+1$ orthonormal vectors $\mathbf{g}_1^{(j)}, \ldots, \mathbf{g}_{s+1}^{(j)} \in \mathcal{G}_j$ are uniquely determined. The next vectors are the columns of the Q-factors of the QR decompositions (3.10) of $(\mathbf{A}-\mu_j\mathbf{I})\mathbf{V}_{s+1}^{(j),\mathsf{xxIDR}}$, where xxIDR denotes the variant used. The vectors $\mathbf{v}_k^{(j),\mathsf{xxIDR}}$ mostly differ, yet the spaces $\mathsf{span}\{\mathbf{v}_0^{(j),\mathsf{xxIDR}}, \ldots, \mathbf{v}_k^{(j),\mathsf{xxIDR}}\}$ are uniquely defined and coincide for all four variants since we assume (3.2); see (3.5). The restriction $\mathbf{g}_k \in \mathcal{K}_k \setminus \mathcal{K}_{k-1}$ fixes the new vectors up to a sign. The positive sign of the diagonal elements of the R-factors of the QR decompositions (3.10) and the scaling of the component of $\mathbf{g}_k^{(j)}$ to one fixes the sign to be the same in all four variants.

It remains to give an example of a pivot breakdown of srIDR. Let $n = 10$, $\mathbf{q} = \mathbf{e}_1$, $s = 2$, $\mu_j = 1$, $j \geqslant 1$,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 1 & -1 & -3 & -2 & 0 \\ 1 & 0 & -1 & 1 & 1 & -2 & 1 & 5 & 4 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 & 0 & -2 & -2 & 1 \\ 0 & 0 & 1 & 0 & -1 & 2 & -1 & -5 & -4 & 0 \\ 0 & 0 & 0 & 1 & 2 & -2 & 1 & 5 & 4 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -2 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\widehat{\mathbf{Q}}^{\mathsf{H}} = \begin{bmatrix} 0 & -1 & -1 & 1 & 0 & -1 & 0 & 2 & 2 & 0 \\ 1 & 1 & 1 & -1 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

A computation shows that the partial orthonormal $\mathbf{g}_k$, $1 \leqslant k \leqslant 14 = 5(s+1) = m$, $5 = \lfloor n/s \rfloor$ of the three variants other than srIDR are given by $\mathbf{g}_k = \mathbf{e}_k$, $1 \leqslant k \leqslant n = 10$,

$$\begin{aligned} \mathbf{g}_{11} &= \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}}, \\ \mathbf{g}_{12} &= \frac{1}{\sqrt{22}}\begin{bmatrix} -1 & 1 & -1 & 3 & -2 & -2 & 1 & 0 & -1 & 0 \end{bmatrix}^{\mathsf{T}}, \\ \mathbf{g}_{13} &= \frac{1}{\sqrt{238}}\begin{bmatrix} 5 & -5 & 5 & -3 & 4 & 4 & -7 & 6 & -1 & -6 \end{bmatrix}^{\mathsf{T}}, \\ \mathbf{g}_{14} &= \frac{1}{\sqrt{2723}}\begin{bmatrix} -11 & 11 & -11 & -41 & 15 & 15 & 0 & -9 & 12 & 2 \end{bmatrix}^{\mathsf{T}}, \end{aligned}$$

and $\mathbf{g}_{15} = \mathbf{o}$. Assumptions (3.2a) and (3.2b) are satisfied, and all matrices $\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j)}$, $0 \leqslant j \leqslant 4$, i.e.,

$$\begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{3}{\sqrt{238}} & \frac{-50}{\sqrt{2723}} \\ \frac{5}{\sqrt{238}} & \frac{38}{\sqrt{2723}} \end{bmatrix},$$

are regular, thus also assumption (3.2c) is satisfied. Naïvely implemented, the srIDR variant breaks down at step 4 since by definition (3.7), the matrix

$$\mathbf{M}_{\mathsf{sr}}^{(1,1)} = \widehat{\mathbf{Q}}^{\mathsf{H}}\begin{bmatrix} \mathbf{g}_s^{(0)}, \mathbf{g}_{s+1}^{(0)} \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

is singular. Yet the system

$$\mathbf{M}_{\mathsf{sr}}^{(1,1)}\widehat{\mathbf{c}}_1 = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}\widehat{\mathbf{c}}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_1^{(1)}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

contained in (3.8) has a consistent right-hand side; e.g., $\widehat{\mathbf{c}}_1 = \begin{bmatrix}-.5, -.5\end{bmatrix}^{\mathsf{T}}$ is a solution. Any such solution gives, by lucky choice of $\mu_1 = 1$, the vectors $\mathbf{g}_2^{(1)} = \mathbf{g}_5 = \mathbf{e}_5$. The next system to be solved in srIDR is

$$\mathbf{M}_{\mathsf{sr}}^{(1,2)}\widehat{\mathbf{c}}_2 = \begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\widehat{\mathbf{c}} = \begin{bmatrix}0\\ 1\end{bmatrix} = \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{g}_2^{(1)},$$

with a singular $\mathbf{M}_{\mathsf{sr}}^{(1,2)}$ and an inconsistent right-hand side. At this point srIDR breaks down. □

REMARK 5.2. The variant srIDR of IDR with partial orthonormalization always breaks down at step $s+2$ if $\mathbf{q} = \widehat{\mathbf{Q}}\mathbf{e}_1$, as Arnoldi's process computes an orthonormal $\mathbf{G}_{s+1}$ with $\mathbf{g}_k \perp \mathbf{q}$, $2 \leqslant k \leqslant s+1$, and thus the first row of the matrix

$$\mathbf{M}_{\mathsf{sr}}^{(1,1)} = \widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_{2:s+1}$$

will be zero. This choice of $\widehat{\mathbf{Q}}$ is often used in applications for almost symmetric matrices $\mathbf{A}$. In contrast to srIDR, the other three variants do not necessarily break down with this choice.

We have seen that the four variants are mathematically equivalent in case they do not break down. To understand why srIDR breaks down more easily than the others, and to understand what other types of breakdown are possible, we consider the conversions between them, or, more generally, the conversion of data from any IDR with partial orthonormalization to the four variants.

COROLLARY 5.3. *Assume that assumption (3.2) holds true. Let*

$$\mathbf{A}\mathbf{G}_m\mathbf{U}_m = \mathbf{G}_{m+1}(\underline{\mathbf{H}}_m + \underline{\mathbf{U}}_m\mathbf{D}_m)$$

be any generalized Hessenberg decomposition such that the vectors $\mathbf{g}_k$, $1 \leqslant k \leqslant m+1$, are partially orthonormalized and that the unit upper triangular $\mathbf{U}_m = \underline{\mathbf{I}}_m^{\mathsf{H}}\underline{\mathbf{U}}_m$, $\underline{\mathbf{H}}_m$, and $\mathbf{D}_m$ conform with the outcome of an IDR algorithm for some $s \in \mathbb{N}$.

The conversion between and to the four variants are given by left-multiplication by unit upper triangular block-diagonal matrices. The first block is $\mathbf{I}_s$, all other blocks are $(s+1)\times(s+1)$ except possibly the last block.

We describe how to obtain the non-trivial blocks except the last:

fmIDR: *the jth block of the transformation matrix is $(\mathbf{U}_{j(s+1)+(0:s),j(s+1)+(0:s)})^{-1}$. The conversion to* fmIDR *is always possible.*

mnIDR: *the jth block of the transformation matrix is given by the inverse of the R-factor of the short QR factorization of $\mathbf{U}_{j(s+1)+(-s:s),j(s+1)+(0:s)}$, left-scaled such that the diagonal is one. The conversion to* mnIDR *is always possible.*

ovIDR: *the jth block of the transformation matrix is given by the inverse of the R-factor of the short QR factorization of $\mathbf{V}_{s+1}^{(j)}$, left-scaled such that the diagonal is one. The conversion to* ovIDR *is always possible.*

srIDR: *the jth block of the transformation matrix is given by the inverse of the transpose of the L-factor of the LU factorization of $(\mathbf{U}_{j(s+1)+(-s:0),j(s+1)+(0:s)})^{\mathsf{T}}$ without permutations. The conversion to* srIDR *is* not *always possible.*

The last block is obtained by truncating the above constructions.

REMARK 5.4. The variant srIDR breaks down when the LU decomposition without permutations does not exist. For this reason we termed such type of breakdown a *pivot breakdown*.

Proof. The left-multiplication by a unit upper triangular block-diagonal matrix with blocks as described does not change the vectors $\mathbf{g}_k$, $1 \leqslant k \leqslant m+1$, nor the structure of $\mathbf{U}_m$, $\underline{\mathbf{U}}_m$,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$\underline{\mathbf{H}}_m$, and $\mathbf{D}_m$. We only consider the case of the full unit upper triangular $(s+1)\times(s+1)$ blocks, the proof for the truncated last block is completely analogous.

The sketched transformation to fmIDR is well-defined as the unit upper triangular matrix $\mathbf{U}_{j(s+1)+(0:s),j(s+1)+(0:s)}$ has a unit upper triangular inverse. It results in a new $\mathbf{U}_m^{\mathsf{fm}}$ that has the correct structure.

The coefficients in mnIDR are defined by minimality of coefficients; e.g., by orthogonal columns in $\mathbf{U}_{j(s+1)+(-s:s),j(s+1)+(0:s)}$. The columns are linearly independent because of the upper trapezoidal structure, thus the sketched transformation is well-defined and results in the wanted orthogonal columns with the correct scaling of the last elements.

The variant ovIDR is defined by orthogonal vectors $\mathbf{v}_k^{(j)}$, which is ensured by the transformation sketched. The matrix $\mathbf{V}_{s+1}^{(j)}$ can be written as

$$\mathbf{V}_{s+1}^{(j)} = \left[\mathbf{G}_{s+1}^{(j-1)}, \mathbf{G}_s^{(j)}\right]\mathbf{U}_{j(s+1)+(-s:s),j(s+1)+(0:s)};$$

see (3.11). Assumption (3.2a) ensures that the vectors $\mathbf{g}_1^{(j-1)},\dots,\mathbf{g}_{s+1}^{(j-1)},\mathbf{g}_1^{(j)},\dots,\mathbf{g}_s^{(j)}$ are linearly independent. $\mathbf{U}_{j(s+1)+(-s:s),j(s+1)+(0:s)}$ has full rank, thus also the vectors $\mathbf{v}_0^{(j)},\dots,\mathbf{v}_s^{(j)}$ are linearly independent, which proves that the QR decomposition with regular R-factor is always possible.

The variant srIDR is defined by the banded structure of the matrix $\mathbf{U}_m$, which is obtained by carrying out the transformation sketched. If for some k, $1 \leqslant k \leqslant s$, a leading submatrix $\mathbf{U}_{j(s+1)+(-s:k-s),j(s+1)+(0:k)}$ is singular, then the LU factorization does not exist. An example is given in the proof of Theorem 5.1. ∎

We have shown that fmIDR, mnIDR, and ovIDR do not break down when we assume the generic case (3.2) but that srIDR may suffer from a pivot breakdown. In the following we look at what happens when (3.2) is violated and remark briefly on how the variants of IDR with partial orthonormalization behave.

We identify two types of breakdown:

- A *lucky breakdown* occurs if assumption (3.2a) or (3.2b) is violated, i.e., the vectors $\mathbf{g}_1^{(j-1)},\dots,\mathbf{g}_{s+1}^{(j-1)},\mathbf{g}_1^{(j)},\dots,\mathbf{g}_k^{(j)}$ become linearly dependent. In this case we have found an invariant subspace. Linear dependence among the vectors $\mathbf{g}_1^{(j)},\dots,\mathbf{g}_k^{(j)}$ can be determined when the vectors $\mathbf{v}_0^{(j)},\dots,\mathbf{v}_{k-1}^{(j)}$ become linearly dependent, as $\mathsf{im}(\mathbf{G}_k^{(j)}) = \mathsf{im}((\mathbf{A}-\mu_j\mathbf{I})\mathbf{V}_k^{(j)})$ in case μ_j is not an eigenvalue of $\mathbf{A}$.
- A *Lanczos breakdown* occurs if assumption (3.2c) is violated, i.e., if $\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j-1)}$ is rank deficient. This corresponds to the case of having found a new vector $\mathbf{v}\in\mathcal{V}_{j-1}$ "too early".

All variants compute the first vector as a solution of a system with the matrix $\widehat{\mathbf{Q}}^{\mathsf{H}}\mathbf{G}_s^{(j-1)}$, thus in case of a Lanczos breakdown all four variants break down. Here we need some form of look-ahead and/or deflation, which is not part of the paper. Nevertheless, the condition of this matrix should be monitored in any case.

6. Error analysis. In finite precision or subject to more general perturbations, the generalized Hessenberg decomposition (3.14) will no longer be satisfied, instead we need to introduce an error term $\mathbf{F}_m\in\mathbb{C}^{n\times m}$ that balances the equation and obtain the *perturbed generalized Hessenberg decomposition*

$$\mathbf{A}\mathbf{G}_m\mathbf{U}_m + \mathbf{F}_m = \mathbf{G}_{m+1}(\underline{\mathbf{H}}_m + \underline{\mathbf{U}}_m\mathbf{D}_m), \tag{6.1}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where all other quantities now denote the *computed* quantities. Suppose Arnoldi's process and the other QR decompositions (3.10) are perturbed,

$$\begin{aligned}\left[\mathbf{q}, \mathbf{A}\mathbf{G}_s^{(0)}\right] + \mathbf{F}_{s+1}^{\text{orthonormalize},0} &= \mathbf{G}_{s+1}^{(0)} \begin{bmatrix} \underline{\mathbf{e}}_1 h_{1,0}^{(0)} & \underline{\mathbf{H}}_s^{(0)} \end{bmatrix}, \\ \mathbf{R}_{s+1}^{(j)} + \mathbf{F}_{s+1}^{\text{orthonormalize},j} &= \mathbf{G}_{s+1}^{(j)} \begin{bmatrix} \underline{\mathbf{e}}_1 h_{1,0}^{(j)} & \underline{\mathbf{H}}_s^{(j)} \end{bmatrix}.\end{aligned} \tag{6.2}$$

The computation of the Sonneveld coefficients using the four variants based on oblique projections takes place in a perturbed variant of (3.11),

$$\mathbf{V}_{s+1}^{(j)} = \begin{bmatrix} \mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)} \end{bmatrix} \underline{\mathbf{U}}_{s+1}^{(j)} + \mathbf{F}_{s+1}^{\text{intersect},j}. \tag{6.3}$$

The choice of the seed values influences the size of the perturbation term that has to be added to (3.12),

$$\mathbf{R}_{s+1}^{(j)} = \mathbf{A}\mathbf{V}_{s+1}^{(j)} - \mathbf{V}_{s+1}^{(j)} \operatorname{diag}(\mu_j, \ldots, \mu_j) + \mathbf{F}_{s+1}^{\text{map},j}. \tag{6.4}$$

LEMMA 6.1. *Define the global perturbations* $\mathbf{F}_m^{\text{orthonormalize}}$, $\mathbf{F}_m^{\text{intersect}}$, *and* $\mathbf{F}_m^{\text{map}}$ *by*

$$\begin{aligned}\mathbf{F}_m^{\text{orthonormalize}} &:= \left[\mathbf{F}_{2:s+1}^{\text{orthonormalize},0}, \mathbf{F}_{s+1}^{\text{orthonormalize},1}, \mathbf{F}_{s+1}^{\text{orthonormalize},2}, \ldots\right], \\ \mathbf{F}_m^{\text{intersect}} &:= \left[\mathbf{O}_{n,s}, \mathbf{F}_{s+1}^{\text{intersect},1}, \mathbf{F}_{s+1}^{\text{intersect},2}, \mathbf{F}_{s+1}^{\text{intersect},3}, \ldots\right], \\ \mathbf{F}_m^{\text{map}} &:= \left[\mathbf{O}_{n,s}, \mathbf{F}_{s+1}^{\text{map},1}, \mathbf{F}_{s+1}^{\text{map},2}, \mathbf{F}_{s+1}^{\text{map},3}, \ldots\right].\end{aligned} \tag{6.5}$$

Then the perturbation $\mathbf{F}_m$ *in* (6.1) *is given by*

$$\mathbf{F}_m = \mathbf{F}_m^{\text{orthonormalize}} + \mathbf{A}\mathbf{F}_m^{\text{intersect}} - \mathbf{F}_m^{\text{intersect}}\mathbf{D}_m + \mathbf{F}_m^{\text{map}}. \tag{6.6}$$

Proof. Combining equations (6.2)–(6.4), the coupling between two local blocks $\mathbf{G}_{s+1}^{(j-1)}$ and $\mathbf{G}_{s+1}^{(j)}$ in the perturbed IDR algorithm, $j \geqslant 1$, looks as follows:

$$\begin{aligned}&\mathbf{A}\begin{bmatrix} \mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)} \end{bmatrix} \underline{\mathbf{U}}_{s+1}^{(j)} + \mathbf{F}_m^{\text{orthonormalize},j} + \mathbf{A}\mathbf{F}_m^{\text{intersect},j} - \mathbf{F}_m^{\text{intersect},j}\mu_j + \mathbf{F}_m^{\text{map},j} \\ &\quad = \begin{bmatrix} \mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)} \end{bmatrix} \left(\begin{bmatrix} \mathbf{o}_{s+1} & \mathbf{O}_{s+1,s} \\ \underline{\mathbf{e}}_1 h_{1,0}^{(j)} & \underline{\mathbf{H}}_s^{(j)} \end{bmatrix} + \underline{\mathbf{U}}_{s+1}^{(j)} \operatorname{diag}(\mu_j, \ldots, \mu_j) \right).\end{aligned} \tag{6.7}$$

Gluing these relations together and topping them with the first equation in (6.2), omitting the first column, we obtain (6.1) with the perturbation term (6.6). ∎

In a reasonable implementation this perturbation term can be bounded independently of the computed quantities only if we assume that no breakdown or near-breakdown occurs. We give a bound based on the computed Sonneveld coefficients and seed values, which both can be monitored.

THEOREM 6.2. *Suppose that we execute IDR with partial orthonormalization in IEEE 754 arithmetic with orthonormalization based on Householder reflections, Givens rotations, or the iterated Gram-Schmidt process [4, CGS2, p. 89]. Suppose further that the condition of* $\mathbf{A}$ *is not too large, no near-breakdown occurs. and the computed Sonneveld coefficients and seed values are not too large.*

Then all computed $\mathbf{G}_{s+1}^{(j)}$, $0 \leqslant j$, *are orthonormal up to machine precision* ϵ_M,

$$\left\| \left(\mathbf{G}_{s+1}^{(j)}\right)^{\mathsf{H}} \mathbf{G}_{s+1}^{(j)} - \mathbf{I}_{s+1} \right\|_F \leqslant C_1\, \epsilon_M, \tag{6.8}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

and the perturbation term in (6.1) is bounded by

$$\|\mathbf{F}_m\| \leqslant \|\mathbf{F}_m\|_F \leqslant C_2\,\epsilon_M \max_{j\geqslant 0}\Big(\|\mathbf{A}\| + |\mu_j|\Big)\|\underline{\mathbf{U}}_{s+1}^{(j)}\|_F, \tag{6.9}$$

where $\mu_0 = 0$, $\underline{\mathbf{U}}_{s+1}^{(0)} = \underline{\mathbf{I}}_{s+1}$. *Here,* $C_1 = C_1(n,s)$ *and* $C_2 = C_2(n,s)$ *are constants depending on the method used and on the matrix* $\mathbf{A}$.

REMARK 6.3. If $\max_{j\geqslant 0}\|\mathbf{U}_{s+1}^{(j)}\|_F$ is bounded and $\max_{j\geqslant 0}|\mu_j| \approx \|\mathbf{A}\|$, then the error bound (6.9) simplifies to the very satisfactory $\|\mathbf{F}_m\| \leqslant \widetilde{C}_2\,\epsilon_M\|\mathbf{A}\|$ for some error constant $\widetilde{C}_2 = \widetilde{C}_2(n,s)$.

Proof. It is well-known that on a computer conforming with IEEE 754

$$\begin{aligned}\|\mathbf{F}_{s+1}^{\text{orthonormalize},0}\|_F &\leqslant C_3\,\epsilon_M\sqrt{s+1}\|\mathbf{A}\|,\\ \|\mathbf{F}_{s+1}^{\text{orthonormalize},j}\|_F &\leqslant C_4\,\epsilon_M\|\mathbf{R}_{s+1}^{(j)}\|_F, \qquad 1\leqslant j,\end{aligned} \tag{6.10}$$

for small constants $C_3 = C_3(n,s+1)$ and $C_4 = C_4(n,s+1)$. This follows from bounds on the QR decomposition [6, § 3.6, p. 73 (complex arithmetic); Theorem 19.4 & p. 361 (Householder); Theorem 19.10 (Givens)] and a standard error analysis for CGS2, using the technique for Arnoldi's process described in [3, p. 314]; see also [2, Theorem 2.3, p. 311; Theorem 2.2, p. 310].

The result (6.8) on partial orthonormality can be found in [2, Theorem 2.1, p. 309–310, Page 312 (note in middle of page)] (Householder, Givens) and [4, Theorem 2] (CGS2).

Utilizing standard error analysis [6, 22], we can bound the perturbation due to finite precision in (6.4),

$$\begin{aligned}\mathbf{R}_{s+1}^{(j)} &= fl\Big(fl\big(\mathbf{A}\mathbf{V}_{s+1}^{(j)}\big) - fl\big(\mathbf{V}_{s+1}^{(j)}\operatorname{diag}(\mu_j,\ldots,\mu_j)\big)\Big)\\ &= \mathbf{A}\mathbf{V}_{s+1}^{(j)} - \mathbf{V}_{s+1}^{(j)}\operatorname{diag}(\mu_j,\ldots,\mu_j) + \mathbf{F}_{s+1}^{\text{map},j},\\ &\qquad |\mathbf{F}_{s+1}^{\text{map},j}| \leqslant \gamma_{r+1}(|\mathbf{A}| + |\mu_j\mathbf{I}|)|\mathbf{V}_{s+1}^{(j)}|,\end{aligned} \tag{6.11}$$

where r denotes the maximum number of nonzeros in a row of $\mathbf{A}$, and in (6.3),

$$\begin{aligned}\mathbf{V}_{s+1}^{(j)} = fl\left(\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\underline{\mathbf{U}}_{s+1}^{(j)}\right) &= \begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\underline{\mathbf{U}}_{s+1}^{(j)} + \mathbf{F}_{s+1}^{\text{intersect},j},\\ |\mathbf{F}_{s+1}^{\text{intersect},j}| &\leqslant \gamma_{2s+1}\left|\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\right|\left|\underline{\mathbf{U}}_{s+1}^{(j)}\right|,\end{aligned} \tag{6.12}$$

where the factor γ_{2s+1} follows since the last row of $\underline{\mathbf{U}}_{s+1}^{(j)}$ is zero. In mnIDR and ovIDR we have at most $2s+1$ nonzero elements in the columns of $\mathbf{U}_m$ and in srIDR and fmIDR at most $s+1$. In the latter case the term γ_{2s+1} can be replaced by γ_{s+1}. By (6.6), (6.5), and (6.7) we can bound $\|\mathbf{F}_m\|_F$ by

$$\sqrt{\left\lceil\frac{m+1}{s+1}\right\rceil}\max_{0\leqslant j}\|\mathbf{F}_m^{\text{orthonormalize},j} + \mathbf{A}\mathbf{F}_m^{\text{intersect},j} - \mathbf{F}_m^{\text{intersect},j}\mu_j + \mathbf{F}_m^{\text{map},j}\|_F, \tag{6.13}$$

where the undefined terms $\mathbf{F}_m^{\text{intersect},0}$ and $\mathbf{F}_m^{\text{map},0}$ are zero. We use the triangle inequality on (6.13) and look at individual terms. We express the term $\mathbf{R}_{s+1}^{(j)}$ in the normwise bound (6.10) using (6.11), (6.12), $\|\mathbf{A}\mathbf{F}\|_F \leqslant \|\mathbf{A}\|\cdot\|\mathbf{F}\|_F$, and the submultiplicativity of the Frobenius

ETNA
Kent State University and
Johann Radon Institute (RICAM)

norm,

$$\begin{aligned}(6.14)\quad \|\mathbf{F}_{s+1}^{\text{orthonormalize},j}\|_F &\leqslant C_4\,\epsilon_M\|\mathbf{A}\mathbf{V}_{s+1}^{(j)} - \mathbf{V}_{s+1}^{(j)}\operatorname{diag}(\mu_j,\ldots,\mu_j) + \mathbf{F}_{s+1}^{\text{map},j}\|_F\\ &\leqslant C_4\,\epsilon_M\left((\|\mathbf{A}\| + |\mu_j|)\|\mathbf{V}_{s+1}^{(j)}\|_F + \|\mathbf{F}_{s+1}^{\text{map},j}\|_F\right)\\ &\leqslant C_4\,\epsilon_M\left((\|\mathbf{A}\| + |\mu_j|)\left\|\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\underline{\mathbf{U}}_{s+1}^{(j)} + \mathbf{F}_{s+1}^{\text{intersect},j}\right\|_F + \|\mathbf{F}_{s+1}^{\text{map},j}\|_F\right)\\ &\leqslant C_4\,\epsilon_M\left(\sqrt{2}(1+\delta_1)(\|\mathbf{A}\| + |\mu_j|)\|\underline{\mathbf{U}}_{s+1}^{(j)}\|_F + \|\mathbf{F}_{s+1}^{\text{intersect},j}\|_F + \|\mathbf{F}_{s+1}^{\text{map},j}\|_F\right),\end{aligned}$$

where δ_1 is of order of the machine precision and defined by

$$\delta_1 := \sqrt{\frac{\|\mathbf{G}_{s+1}^{(j-1)}\|^2 + \|\mathbf{G}_s^{(j)}\|^2}{2}} - 1.$$

Krylov subspace methods are mostly used for sparse matrices $\mathbf{A}$. We rewrite the componentwise bound (6.11) using $\||\mathbf{A}|\| \leqslant \sqrt{r}\|\mathbf{A}\|$ [13, Lemma A.1] and (6.12), assuming that $r \geqslant 1$,

$$\begin{aligned}(6.15)\quad \|\mathbf{F}_{s+1}^{\text{map},j}\|_F &\leqslant \gamma_{r+1}(\sqrt{r}\|\mathbf{A}\| + |\mu_j\mathbf{I}|)\|\mathbf{V}_{s+1}^{(j)}\|_F\\ &\leqslant \gamma_{r+1}\sqrt{r}(\|\mathbf{A}\| + |\mu_j\mathbf{I}|)\left\|\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_{s+1}^{(j)}\end{bmatrix}\underline{\mathbf{U}}_{s+1}^{(j)} + \mathbf{F}_{s+1}^{\text{intersect},j}\right\|_F\\ &\leqslant \gamma_{r+1}\sqrt{r}\sqrt{2}(1+\delta_1)(\|\mathbf{A}\| + |\mu_j\mathbf{I}|)(\|\underline{\mathbf{U}}_{s+1}^{(j)}\|_F + \|\mathbf{F}_{s+1}^{\text{intersect},j}\|_F).\end{aligned}$$

Using (6.12), we obtain similarly

$$\begin{aligned}(6.16)\quad \|\mathbf{A}\mathbf{F}_m^{\text{intersect},j} - \mathbf{F}_m^{\text{intersect},j}\mu_j\|_F &\leqslant (\|\mathbf{A}\| + |\mu_j|)\|\mathbf{F}_m^{\text{intersect},j}\|_F\\ &\leqslant \gamma_{2s+1}(\|\mathbf{A}\| + |\mu_j|)\left\|\left|\begin{bmatrix}\mathbf{G}_{s+1}^{(j-1)} & \mathbf{G}_s^{(j)}\end{bmatrix}\right|\right\|_F\left\|\left|\underline{\mathbf{U}}_{s+1}^{(j)}\right|\right\|_F\\ &\leqslant \gamma_{2s+1}(1+\delta_2)\sqrt{2s+1}(\|\mathbf{A}\| + |\mu_j|)\left\|\underline{\mathbf{U}}_{s+1}^{(j)}\right\|_F,\end{aligned}$$

where δ_2 is of order of the machine precision and defined by

$$\delta_2 := \sqrt{\frac{\|\mathbf{g}_{s+1}^{(j-1)}\|_2^2 + \sum_{i=1}^{s}\left(\|\mathbf{g}_i^{(j-1)}\|_2^2 + \|\mathbf{g}_i^{(j)}\|_2^2\right)}{2s+1}} - 1.$$

We assume that all perturbations are small enough that second order terms can be incorporated into the constant. Combining the local bounds (6.14), (6.15), and (6.16) with the global bound (6.13) proves (6.9). ∎

Numerical experiments indicate that the size of the perturbation term $\mathbf{F}_m$ in (6.1) has a strong influence on the attainable accuracy of the OR and MR iterates as well as of the Ritz pairs. The seed value selection schemes in Section 4 offer control on the first part of the bound (6.9). The second part of the bound (6.9) is minimized locally by the variant mnIDR. In the next section we present two academic toy examples where we compare the accuracy that is obtained in the four variants and show that on average indeed mnIDR gives the smallest residuals of the four variants discussed.

7. Numerical examples. We present two numerical examples to analyse the four different IDR variants, one for linear systems using the MR approach and one for eigenvalue

ETNA
Kent State University and
Johann Radon Institute (RICAM)

computations using the Ritz approach. To analyse the dependence of IDR on the selection of the seed values we include a comparison for the best variant mnIDR and the worst variant ovDIR. In the applications, $\mathbf{q}$ is often chosen as initial residual $\mathbf{q} = \mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$ (solution of a linear system) or as initial eigenvector approximation or as a random vector (eigenvalue problem). We used in both cases a non-physical $\mathbf{q}$ chosen to depict the expected average behavior.

All algorithms have been implemented in MATLAB. We reorthogonalize the $\mathbf{g}$-vectors once (based on CGS2) and rebiorthogonalize the $\mathbf{v}$-vectors once against $\widehat{\mathbf{Q}}$ (based on an oblique analogue of CGS2). For the solution of the small (rectangular) systems we use known backward stable solvers; in the experiments below we used MATLAB's backslash for srIDR and ovIDR, MATLAB's built-in LU decomposition and backslash (i.e., LAPACK) for fmIDR and MATLAB's built-in pseudoinverse (i.e., LAPACK) for mnIDR. The perturbation of the generalized Hessenberg decomposition (6.1) induced by finite precision behaves as predicted by the bound (6.9).

7.1. Linear systems via the MR approach. The MR approach (3.18) is based on solving small extended Hessenberg least-squares problems

$$\min_{\underline{\mathbf{z}}_m \in \mathbb{C}^m} \left\| \underline{\mathbf{H}}_m \underline{\mathbf{z}}_m - \underline{\mathbf{e}}_1 \|\mathbf{r}_0\|_2 \right\|_2,$$

which is done by QR decomposition with a sequence of Givens rotations,

$$\left\| \underline{\mathbf{H}}_m \underline{\mathbf{z}}_m - \underline{\mathbf{e}}_1 \|\mathbf{r}_0\|_2 \right\|_2 = \left\| \mathbf{G}_{m+1,m} \cdots \mathbf{G}_{3,2}\mathbf{G}_{2,1} (\underline{\mathbf{H}}_m \underline{\mathbf{z}}_m - \underline{\mathbf{e}}_1 \|\mathbf{r}_0\|_2) \right\|_2.$$

The Givens rotation $\mathbf{G}_{j+1,j}$ rotates the plane spanned by $\mathbf{e}_j^\mathsf{T}, \mathbf{e}_{j+1}^\mathsf{T}$ to annihilate the element in position $j+1, j$. The R-factor $\mathbf{R}_m$ of the short QR decomposition,

$$\mathbf{R}_m := \underline{\mathbf{I}}_m^\mathsf{T} \mathbf{G}_{m+1,m} \cdots \mathbf{G}_{3,2}\mathbf{G}_{2,1} \underline{\mathbf{H}}_m \in \mathbb{C}^{m \times m},$$

is banded and has the structure of $\underline{\mathbf{H}}_m$ moved up by one, thus it can be stored in the same place. The update of the right-hand side only changes the last two components in every step,

$$\underline{\boldsymbol{\phi}}_m := \mathbf{G}_{m+1,m} \cdots \mathbf{G}_{3,2}\mathbf{G}_{2,1} \underline{\mathbf{e}}_1 \|\mathbf{r}_0\|_2 \in \mathbb{C}^{m+1}, \quad \boldsymbol{\phi}_m := \underline{\mathbf{I}}_m^\mathsf{T} \underline{\boldsymbol{\phi}}_m.$$

The residual of the small least-squares problem is given by $|\underline{\boldsymbol{\phi}}_m(m+1)|$, the MR solution by $\underline{\mathbf{z}}_m = \mathbf{R}_m^{-1}\boldsymbol{\phi}_m$. There are two main styles (compare with [13, equations (7), (8)]) that can be used to compute the MR iterate $\underline{\mathbf{x}}_m = \mathbf{V}_m \underline{\mathbf{z}}_m$, like GMRES or MINRES, i.e.,

$$\underline{\mathbf{x}}_m = \mathbf{V}_m(\mathbf{R}_m^{-1}\boldsymbol{\phi}_m) = (\mathbf{V}_m\mathbf{R}_m^{-1})\boldsymbol{\phi}_m. \tag{7.1}$$

The GMRES style is based on the first grouping in (7.1). After $\underline{\mathbf{z}}_m = \mathbf{R}_m^{-1}\boldsymbol{\phi}_m$ has been computed, we need to compute $\underline{\mathbf{x}}_m = \mathbf{V}_m \underline{\mathbf{z}}_m$. As we do not store the vectors in $\mathbf{V}_m$, we rerun the algorithm with known $\underline{\mathbf{z}}_m$ and compute the linear combination $\underline{\mathbf{x}}_m = \mathbf{V}_m \underline{\mathbf{z}}_m$ along with the vectors $\mathbf{v}_j$, $1 \leqslant j \leqslant m$, roughly doubling the computing time and increasing the storage by one n-vector.

The MINRES style is based on the second grouping in (7.1). We define direction vectors $\mathbf{W}_m := \mathbf{V}_m\mathbf{R}_m^{-1}$, which can be computed by a short recurrence using $\mathbf{W}_m\mathbf{R}_m = \mathbf{V}_m$, and update $\underline{\mathbf{x}}_m$ by $\underline{\mathbf{x}}_m = \underline{\mathbf{x}}_{m-1} + \mathbf{w}_m\boldsymbol{\phi}_m(m)$; see [18] for details. This roughly doubles the storage requirements and, apart from multiplications by $\mathbf{A}$, also the computing time compared to only computing the vectors $\mathbf{g}$. This is like in most OR approaches where we additionally have to update the iterates along with the residuals. If the multiplication of a vector with $\mathbf{A}$ is not $O(n)$, then this clearly is the preferred variant.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

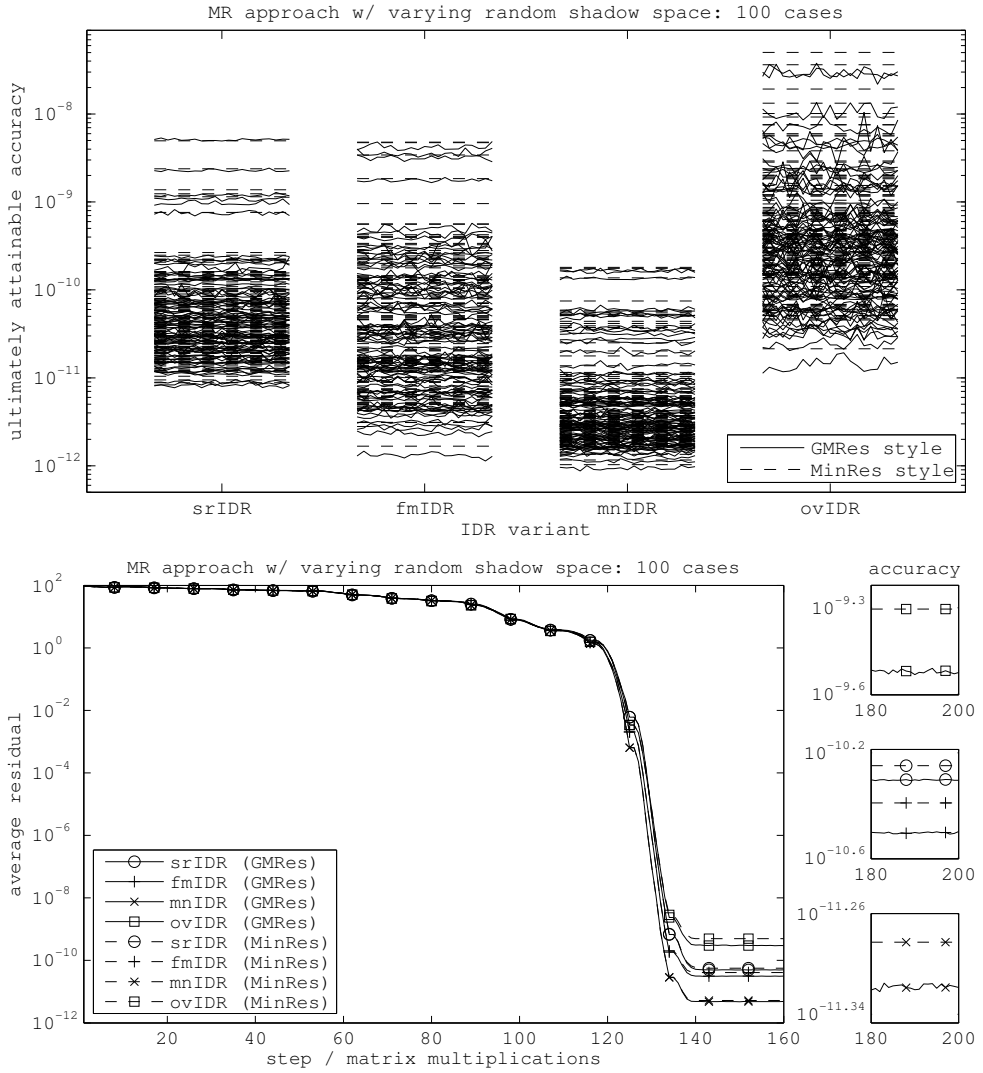

FIG. 7.1. *The ultimately attainable accuracy of the four IDR*(8) *variants with the MR approach on the toy example* $\mathbf{A} =$ `randn(100)+4*eye(100)`, $\mathbf{b} = \mathbf{A}*$`ones(100,1)`, *random* $\mathbf{x}_0$, *for the vanilla seed selection scheme and 100 random shadow spaces, both the* GMRES *and* MINRES *styles, resulting in a total of 800 lines (top). Average residual convergence for the 100 cases (bottom).*

A matrix was generated as $\mathbf{A} =$ `randn(100)+4*eye(100)`. The matrix as well as all eigenvalues are well-conditioned. By Girko's circular law all eigenvalues are approximately uniformly distributed in the circle with center 4 and radius 10. This matrix serves as an example of a "hard case" for Krylov subspace methods. The matrix is non-normal and zero is in the field of values, thus it is indefinite. Because of this, the initial speed of convergence will be small. As right-hand side we used $\mathbf{b} = \mathbf{A}*$`ones(100,1)`, as starting guess $\mathbf{x}_0 =$ `randn(100,1)`.

In the IDR algorithm we used the MR approach for the typical value of $s = 8$ and the typical choice of the vanilla technique for the seed values. We tested both styles for all four variants. We tested the eight possible mixtures of style and variant for 100 different randomly

ETNA
Kent State University and
Johann Radon Institute (RICAM)

chosen shadow spaces and computed the average behavior of the variants and styles. To give a sketch of the ultimately attainable accuracy in each variant/style-pair, we did chose the region 180–200 steps, where all pairs did converge to the ultimately attainable accuracy. The resulting 800 lines are given in the upper plot of Figure 7.1.

The lower plot in Figure 7.1 depicts the average behavior of all eight variant/style-pairs for the geometric mean. In the first 100 steps, all eight pairs behave very similar as predicted by their mathematical equivalence. The best variant is mnIDR as predicted by the bound (6.9), followed by fmIDR, closely followed by the standard variant srIDR. The worst variant by far is ovIDR. On average, the GMRES style always beats the MINRES style in terms of accuracy, which is in accordance with the observations in [13].

7.2. Eigenvalue approximation. For the eigenvalue computations we used a Grcar matrix of size 100×100. A Grcar matrix is an upper Hessenberg banded Toeplitz matrix with upper bandwith 3, the lower diagonal contains -1, and the other four nonzero diagonals contain 1. The eigenvalues come in complex conjugate pairs, and some of them have a condition number of order 10^{18}, which makes them good candidates for analyzing an eigenvalue solver. We used the starting vector $\mathbf{A}\star$`ones(100,1)`, all four variants of IDR(8) with

TABLE 7.1
Average backward error of the computed eigenvalues for 1000 randomly chosen shadow spaces.

	Sonneveld pencil	purified pencil
srIDR	$2.9945 \cdot 10^{-13}$	$2.0659 \cdot 10^{-13}$
fmIDR	$8.2185 \cdot 10^{-14}$	$7.4958 \cdot 10^{-14}$
mnIDR	$3.0661 \cdot 10^{-14}$	$4.3102 \cdot 10^{-14}$
ovIDR	$3.3671 \cdot 10^{-12}$	$1.1462 \cdot 10^{-12}$

the vanilla technique, 1000 randomly chosen shadow spaces, and stopped the algorithm at $m = 150$. We computed the Sonneveld Ritz values (i.e., the eigenvalues of the Sonneveld pencil $(\mathbf{H}_m^{\text{total}}, \mathbf{U}_m)$) that lie close to the badly conditioned eigenvalues of the Grcar matrix. For each Sonneveld Ritz value θ we computed its backward error $\sigma_{\min}(\mathbf{A} - \theta\mathbf{I})$ and the geometric mean of all backward errors for all Sonneveld Ritz values for all shadow spaces for each IDR variant. These numbers can be found in the column entitled "Sonneveld pencil" in Table 7.1. They are depicted along with a contour plot of a section of the pseudospectra of the Grcar matrix in the upper plot in Figure 7.2. The contours are plotted for the values $10^{-7}, 10^{-8}, \ldots, 10^{-16}$.

The Sonneveld pencil has the seed values as eigenvalues. In [23] we described how to construct another pencil that no longer has the seed values as eigenvalues. We used the shifted purified pencil with shift $\kappa = 7$ from [23] for each variant of IDR and all 1000 shadow spaces. We computed again the geometric mean of all backward errors. These numbers can be found in the column entitled "purified pencil" in Table 7.1. They are depicted along with a contour plot of a section of the pseudospectra of the Grcar matrix in the lower plot in Figure 7.2. The contours are again plotted for the values $10^{-7}, 10^{-8}, \ldots, 10^{-16}$.

The four different IDR variants return eigenvalue approximations that lie around the pseudospectral contour lines in Figure 7.2. To understand to what extent these overlap, we depict in Figure 7.2 the geometric mean together with the sample standard deviation for the 1000 instances for each variant/pencil-pair.

Again, mnIDR gives the best results with respect to minimal average backward error, followed by fmIDR, srIDR, and ovIDR. The best choice in this example is to use the mnIDR variant and the Sonneveld pencil.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

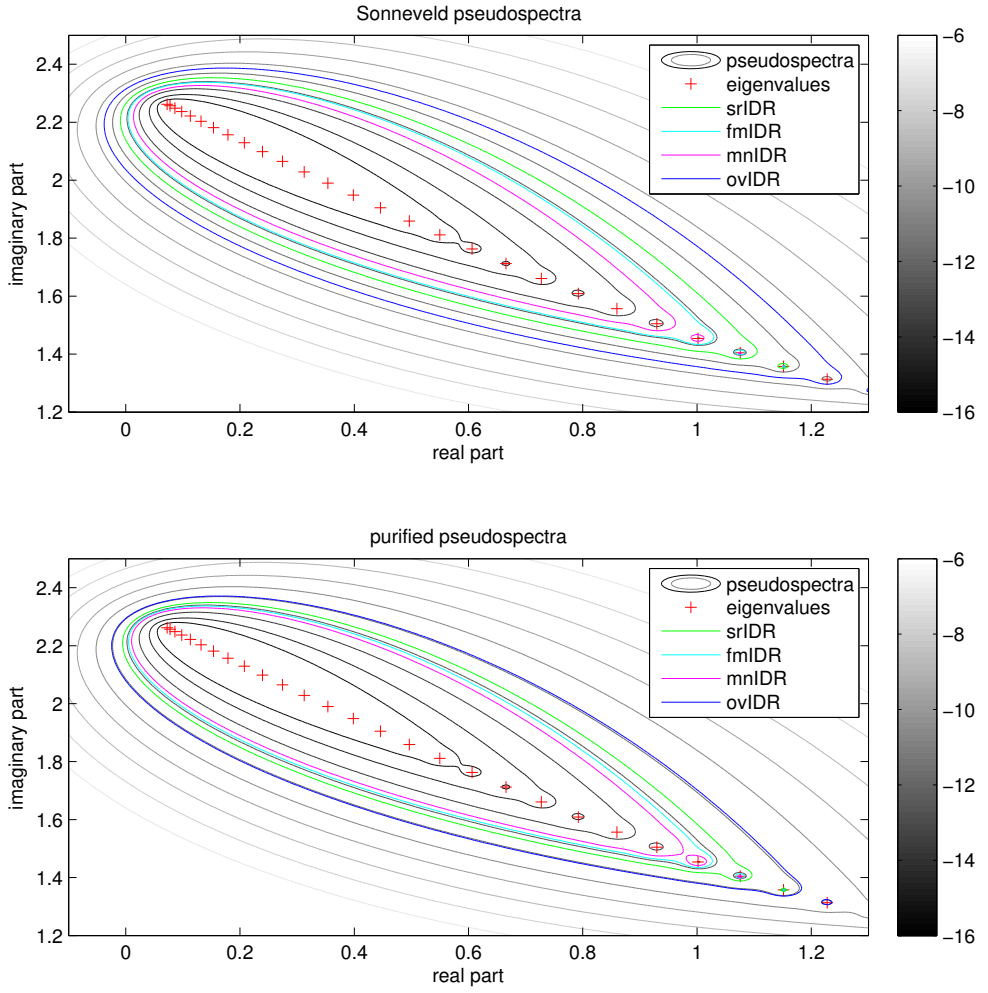

FIG. 7.2. *Pseudospectral contour plot for the eigenvalues of the* 100×100 *Grcar matrix* $\mathbf{A} =$ `gallery('grcar',100)` *that are badly conditioned. Red plusses depict the eigenvalues of the Grcar matrix. The 10 gray contour lines depict the pseudospectrum for* $10^{-7}, 10^{-8}, \ldots, 10^{-16}$. *The colored contour lines correspond to the average geometric mean of the four IDR*(8) *variants for 1000 instances using the vanilla technique for the seed values;* srIDR *in green,* fmIDR *in cyan,* mnIDR *in magenta, and* ovIDR *in blue.*

7.3. The influence of the seed values. To analyse the influence of the seed value selection scheme on the different IDR variants, we tested 11 different schemes on a Grcar matrix of size 100 for the standard value $s = 8$. We used the same shadow space and pencils for both the MR approach and the Ritz approach and computed the average behaviour over 100 runs for all four IDR variants with partial orthonormalization.

In Figure 7.4 we depict the results for the best variant mnIDR and in Figure 7.5 the results for the worst variant ovIDR. The pictures of the srIDR and fmIDR variants are omitted as they look similar to the plot for the mnIDR variant. The only difference is that they are not as accurate as mnIDR.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

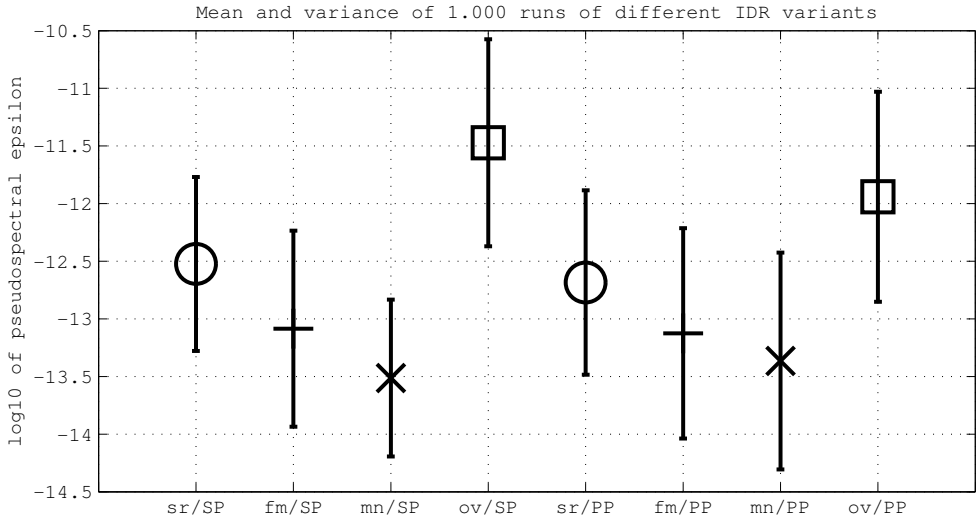

FIG. 7.3. *The average and sample standard deviation of the backward error of all four IDR variants and the Sonneveld and purified pencil.*

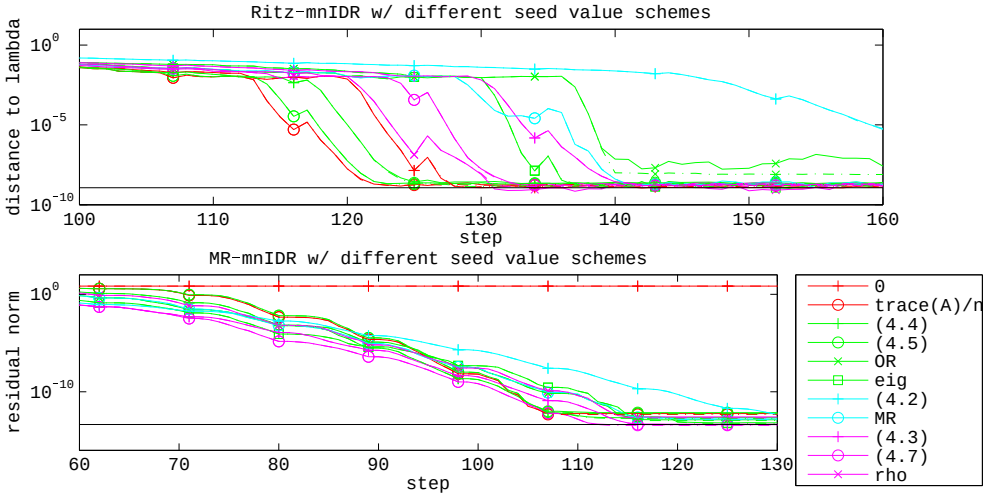

FIG. 7.4. *The average of 100 runs of mnIDR with MR approach (top) and Ritz approach (bottom) for 11 different seed value selection schemes (legend) for a Grcar matrix of size 100 and complex valued random right-hand side. Red curves correspond to constant seed value selection schemes, green curves to schemes related to Rayleigh quotients, cyan curves to local minimization, and magenta curves to mixed approaches.*

The seed value selection schemes we tested can be grouped into:

constant: we used $\mu = 0$ (thus MR stagnates, clearly visible in both lower plots) and $\mu = \mathsf{trace}(\mathbf{A})/n = 1$, which performs best in the Ritz approach.

Rayleigh: we used Rayleigh quotients with $\mathbf{v}$-vectors (4.4) and $\mathbf{g}$-vectors (4.5), precomputed OR values (Ritz values from j steps of Arnoldi applied to $\mathbf{A}$ and $\mathbf{q}$) [10], and exact eigenvalues of $\mathbf{A}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

minimizing: we used the harmonic Rayleigh quotients (4.2) and precomputed MR values (harmonic Ritz from j steps of Arnoldi applied to $\mathbf{A}$ and $\mathbf{q}$).

mixed: we used the vanilla technique (4.3), the cinnamon technique (4.7), and so-called ρ-values (Rayleigh quotients of the harmonic Ritz vectors from j steps of Arnoldi applied to $\mathbf{A}$ and $\mathbf{q}$).

There is no clear winner but a clear loser: for the given setting the original scheme (4.2) performs worst for both the Ritz and the MR approach.

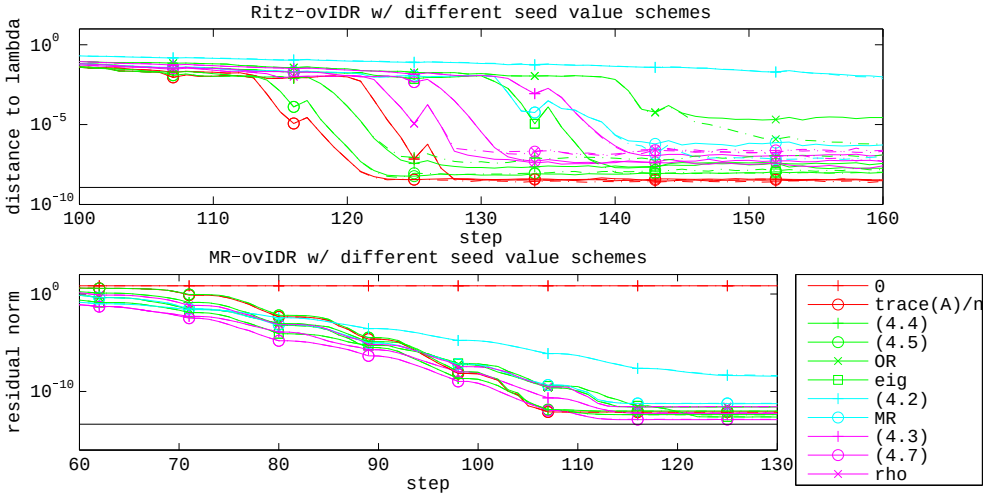

FIG. 7.5. *The average of 100 runs of ovIDR with MR approach (top) and Ritz approach (bottom) for 11 different seed value selection schemes (legend) for a Grcar matrix of size 100 and complex valued random right-hand side. Red curves correspond to constant seed value selection schemes, green curves to schemes related to Rayleigh quotients, cyan curves to local minimization, and magenta curves to mixed approaches.*

In the Ritz approach we analyzed the convergence of Ritz values to the simple eigenvalue $\lambda \approx 1.6786 + 1.1348i$ in the upper right corner of the spectrum of the Grcar matrix; see Figure 7.6. The black line in the upper plots of Figure 7.4 and Figure 7.5 depicts ten times the machine precision times the condition number of λ, which is the level of attainable accuracy of mnIDR of the Ritz approach. We plotted the approximations using the Sonneveld pencil (lines) and those of the purified pencil (dash-dotted). Ignoring the approaches based on extra multiplications with $\mathbf{A}$, the constant choice $\mu = \mathsf{trace}(\mathbf{A})/n$ is the winner, closely followed by (4.5) and (4.4). Next comes the other constant scheme $\mu = 0$, followed by the mixed approaches and those based on local residual minimization. Careful selection of seed values can result in faster convergence. There is no advantage in using more information, e.g., in the OR-, MR- and ρ-values or the exact eigenvalues of $\mathbf{A}$.

In the MR approach we computed the iterates using the MINRES (dash-dotted) and the GMRES style (lines). The black line depicts 100 times the condition number of $\mathbf{A}$ times machine precision, which is the level of attainable accuracy of mnIDR of the MR approach using (4.7) and (4.3). The differences between all seed value selection schemes, apart from the stagnating $\mu = 0$ and the scheme (4.2), are much less pronounced than in the Ritz approach. In this particular case the cinnamon technique (4.7) slightly outperforms the standard vanilla technique (4.3). In other cases the vanilla technique was better than the cinnamon technique, e.g., when zero is outside the field of values of $\mathbf{A}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

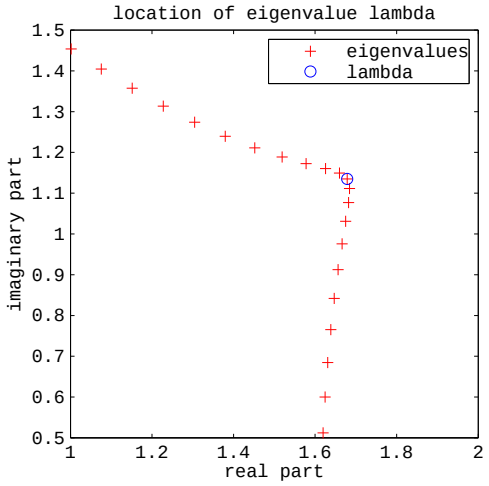

FIG. 7.6. *The location of the eigenvalue λ of the Grcar matrix of order 100.*

In the comparison of the results of using different seed value selection schemes, we observe that the accuracy of a less stable variant such as ovIDR is influenced by the choice of seeds, whereas for the more stable variant mnIDR, the attainable accuracy does not vary much with respect to the choice of the seeds.

8. Conclusion and outlook. We presented a generic IDR that we specialized to what we call IDR with partial orthonormalization. The freedom left in the generic algorithm was used to distinguish four different types of IDR. A common rough error analysis and two numerical experiments suggest that the variant mnIDR is the one that computes the quantities with smallest backward error. From a computational point of view, the variant srIDR used in [18] is more interesting, as we do not need to store as many long vectors as in mnIDR for the same value of s. The experiments indicate that on average, even though srIDR could break down when mnIDR does not, the behaviour of srIDR is not too far from that of mnIDR. If accuracy is more important, then we suggest to use mnIDR. If computational time or storage requirements are more important, then we suggest to use srIDR.

The variants fmIDR and ovIDR may play a more vital role in case of Lanczos breakdowns. Breakdowns of IDR when using random shadow spaces are very rare, thus, we do not expect that IDR variants with a look-ahead strategy are needed if we use random shadow spaces and finite precision computations.

It is easy to refine the rough error analysis in this paper for a particular variant. Bounds on the gap between the true residuals and the cheap estimates that can be computed in the algorithms are based on bounds on the perturbation such as (6.9). It remains a hard task to analyze the influence of the perturbation on the recurrence.

Acknowledgments. The author would like to thank the anonymous referee for helpful remarks that substantially increased the quality of the paper. Furthermore he wants to thank Olaf Rendel, without whom this paper never would have been written. Many of the ideas presented in this paper originated from discussions I had with Olaf during his bachelor's and master's thesis. Unfortunately he refused to be my co-author. Sabine Le Borne read an older version of this note and made many helpful comments.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

REFERENCES

[1] W. E. ARNOLDI, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[2] T. BRACONNIER, P. LANGLOIS, AND J. C. RIOUAL, *The influence of orthogonality on the Arnoldi method*, in Proceedings of the International Workshop on Accurate Solution of Eigenvalue Problems (University Park, PA, 1998), J. Barlow, B. N. Parlett, and K. Veselić, eds., Linear Algebra Appl., 309 (2000), pp. 307–323.
[3] J. DRKOŠOVÁ, A. GREENBAUM, M. ROZLOŽNÍK, AND Z. STRAKOŠ, *Numerical stability of GMRES*, BIT, 35 (1995), pp. 309–330.
[4] L. GIRAUD, J. LANGOU, M. ROZLOŽNÍK, AND J. VAN DEN ESHOF, *Rounding error analysis of the classical Gram-Schmidt orthogonalization process*, Numer. Math., 101 (2005), pp. 87–100.
[5] M. H. GUTKNECHT AND J.-P. M. ZEMKE, *Eigenvalue computations based on IDR*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 283–311.
[6] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.
[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] O. RENDEL, A. RIZVANOLLI, AND J.-P. M. ZEMKE, *IDR: a new generation of Krylov subspace methods?*, Linear Algebra Appl., 439 (2013), pp. 1040–1061.
[10] V. SIMONCINI AND D. B. SZYLD, *Interpreting IDR as a Petrov-Galerkin method*, SIAM J. Sci. Comput., 32 (2010), pp. 1898–1912.
[11] G. L. G. SLEIJPEN, P. SONNEVELD, AND M. B. VAN GIJZEN, *Bi-CGSTAB as an induced dimension reduction method*, Appl. Numer. Math., 60 (2010), pp. 1100–1114.
[12] G. L. G. SLEIJPEN AND H. A. VAN DER VORST, *Maintaining convergence properties of BiCGstab methods in finite precision arithmetic*, Numer. Algorithms, 10 (1995), pp. 203–223.
[13] G. L. G. SLEIJPEN, H. A. VAN DER VORST, AND J. MODERSITZKI, *Differences in the effects of rounding errors in Krylov solvers for symmetric indefinite linear systems*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 726–751.
[14] 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.
[15] P. SONNEVELD AND M. B. VAN GIJZEN, *IDR(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations*, Tech. Report 07-07, Department of Applied Mathematical Analysis, Delft University of Technology, Delft, 2007.
[16] ———, 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.
[17] M. TANIO AND M. SUGIHARA, *GBi-CGSTAB(s, L): IDR(s) with higher-order stabilization polynomials*, J. Comput. Appl. Math., 235 (2010), pp. 765–784.
[18] M. B. VAN GIJZEN, G. L. G. SLEIJPEN, AND J.-P. M. ZEMKE, *Flexible and multi-shift induced dimension reduction algorithms for solving large sparse linear systems*, Numer. Linear Algebra Appl., 22 (2015), pp. 1–25.
[19] M. B. VAN GIJZEN AND P. SONNEVELD, *Algorithm 913: an elegant* IDR(s) *variant that efficiently exploits biorthogonality properties*, ACM Trans. Math. Software, 38 (2011), pp. Art. 5, 19.
[20] D. S. WATKINS AND L. ELSNER, *Convergence of algorithms of decomposition type for the eigenvalue problem*, Linear Algebra Appl., 143 (1991), pp. 19–47.
[21] 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., Springer, Berlin, 1980, pp. 543–562.
[22] J. H. WILKINSON, *Rounding Errors in Algebraic Processes*, Dover, New York, 1994.
[23] J.-P. M. ZEMKE, *On structured pencils arising in Sonneveld methods*, Bericht 186, Institute of Mathematics, Technische Universität Hamburg, Hamburg, July 2014.
http://dx.doi.org/10.15480/882.1180