SIAM J. NUMER. ANAL.
Vol. 40, No. 6, pp. 2219–2239
© 2003 Society for Industrial and Applied Mathematics

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

FLEXIBLE INNER-OUTER KRYLOV SUBSPACE METHODS*

VALERIA SIMONCINI† AND DANIEL B. SZYLD‡

Abstract. Flexible Krylov methods refers to a class of methods which accept preconditioning that can change from one step to the next. Given a Krylov subspace method, such as CG, GMRES, QMR, etc. for the solution of a linear system $Ax = b$, instead of having a fixed preconditioner M and the (right) preconditioned equation $AM^{-1}y = b$ ($Mx = y$), one may have a different matrix, say M_k, at each step. In this paper, the case where the preconditioner itself is a Krylov subspace method is studied. There are several papers in the literature where such a situation is presented and numerical examples given. A general theory is provided encompassing many of these cases, including truncated methods. The overall space where the solution is approximated is no longer a Krylov subspace but a subspace of a larger Krylov space. We show how this subspace keeps growing as the outer iteration progresses, thus providing a convergence theory for these inner-outer methods. Numerical tests illustrate some important implementation aspects that make the discussed inner-outer methods very appealing in practical circumstances.

Key words. flexible or inner-outer Krylov methods, variable preconditioning, nonsymmetric linear system, iterative solver

AMS subject classifications. 65F10, 15A06

PII. S0036142902401074

1. Introduction.

Consider the iterative solution of large sparse (symmetric or) nonsymmetric linear systems of equations of the form

$$Ax = b. \tag{1.1}$$

In recent years, several authors studied Krylov subspace methods with variable (or flexible) preconditioning, i.e., preconditioning with a different (possibly nonlinear) operator at each iteration of a Krylov subspace method. These include [1], [17], [27], [29], [36], [37], and [39]. The usual (right) preconditioning consists of replacing (1.1) by

$$AM^{-1}y = b, \quad \text{with} \quad Mx = y, \tag{1.2}$$

for a suitable preconditioner M. One of the motivations for methods with variable preconditioners is the need to solve each preconditioning equation

$$Mz = v \tag{1.3}$$

only inexactly, as is done, e.g., in [12], using multigrid or, in [40], using a two-stage preconditioner, one of which is inexact; see also [4]. This implies that we have (implicitly) a different M at the kth step of the Krylov method. One can also consider preconditioners which might improve using information from previous iterations; cf. [2], [13], [21].

*Received by the editors January 16, 2002; accepted for publication (in revised form) July 16, 2002; published electronically January 7, 2003.

http://www.siam.org/journals/sinum/40-6/40107.html

†Dipartimento di Matematica, Università di Bologna, I-40126 Bologna, Italy and Istituto di Matematica Applicata e Tecnologie Informati che del CNR, I-27100 Pavia, Italy (val@dragon.imati.cnr.it).

‡Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, PA 19122-6094 (szyld@math.temple.edu). The research of this author was supported in part by the National Science Foundation under grant DMS-0207525. Part of this paper was prepared while this author visited the Università di Bologna, within the 2001 I.N.d.A.M. project "Problemi di perturbazione singolare: tecniche adattive e di decomposizione dei dominii."

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Experiments have been reported in the literature, where the preconditioner in (1.3) is itself a Krylov subspace method. For example, some versions of GMRESR [37] fit this description [6]. We refer to [39] for a more detailed analysis specific to GMRESR versus FGMRES. In [5], [29], one has GMRES for the preconditioner, or inner iterations, and FGMRES as the (outer) flexible method. In [36], QMR is the preconditioner and FQMR is the outer method. In all these cases, we can say that at the kth outer step (or cycle) the inner iterative method is stopped after a certain number m_k of (inner) iterations. This number is either fixed a priori, such as in [5], [29], [37], or is the consequence of some stopping criteria often involving the inner residual [1], [17], [36], [38].

In this paper we analyze the convergence theory of these inner-outer methods, i.e., flexible Krylov subspace methods preconditioned by (a possibly different) Krylov subspace method. The resulting class of methods can be characterized as those choosing at the kth cycle the approximate solution to (1.1) from a particular k-dimensional subspace of a larger Krylov subspace; see sections 2–5. We show that, as the method progresses, the dimensions of this subspace keeps growing, and thus the method converges. This finite termination property is not available for restarted methods, such as GMRES(m). In other words, by restricting ourselves to preconditioners which are Krylov methods, we maintain the global iteration within a larger Krylov subspace.

An alternative view of these inner-outer methods is to consider a Krylov method with polynomial preconditioning (see, e.g., [2], [11], [15], [20], [21], [28], [30]), where a new polynomial is implicitly chosen at each cycle by a Krylov method; cf. [14], [37], [38].

Our approach is very general, and thus our analysis applies to a variety of inner-outer methods. In particular, it applies to truncated Krylov methods, such as DQGMRES [32], when preconditioned with a Krylov subspace method. Several authors have suggested strategies on how to choose which vectors to keep in the truncated basis; see, e.g., [7], [38]. Our results in section 5 provide a theoretical foundation for some of these empirical recommendations.

In section 6 we discuss the possibility of breakdown and stagnation of the inner-outer methods. Since the residual of the inner method is deflated with respect to the previous vectors in the outer basis, stagnation is much less prevalent than in restarted methods.

In section 7 we show in some particular cases why these methods behave better than the corresponding restarted counterparts and illustrate this with additional numerical experiments. In our (albeit limited) computational experience, even in cases where at intermediate steps the restarted methods have a lower residual norm, the inner-outer methods outperform the restarted methods. These experiments demonstrate how competitive these inner-outer methods are. They are a very good alternative to restarted methods such as GMRES(m). We hope that the theory presented here, assuring convergence, together with the numerical evidence, will encourage more practitioners to try these inner-outer methods.

In this paper we concentrate on nonsymmetric systems, but most of our observations are valid for symmetric systems as well, where flexible CG-type methods can be employed [17], [27]. We also note that in our descriptions the coefficient matrix is assumed to be A, as in (1.1), although everything applies to a preconditioned matrix AP^{-1} as in (1.2), or $P^{-1}A$, for a suitable fixed preconditioner P. In fact, some of our experiments in section 8 are of this type, where P corresponds to an incomplete factorization of A.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Finally, we note that exact arithmetic is assumed throughout the paper, and that numerical experiments were carried out using Matlab 5.3 with machine epsilon $\epsilon = 2.2 \cdot 10^{-16}$ [23].

2. General setup. Given an initial guess x_0 to the solution of (1.1), and the corresponding residual $r_0 = b - Ax_0$, a Krylov subspace method generates a basis $\{v_1, v_2, \ldots, v_m\}$ of the Krylov subspace $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, A^2r_0, \ldots, A^{m-1}r_0\}$. Let $V_m = [v_1, v_2, \ldots, v_m]$. An approximation to the solution of (1.1) is sought in $x_0 + \mathcal{K}_m(A, r_0)$, i.e., of the form $x_m = x_0 + V_m y_m$ for some $y_m \in \mathbb{R}^m$. The different methods arise by different choices of the basis defined by V_m and by different choices of y_m; see, e.g., [19], [30], for detailed description of these methods.

For example, in GMRES [31], the vectors in V_m are orthonormal, produced by the Arnoldi method with $v_1 = r_0/\beta$, $\beta = \|r_0\|$, the norm of the initial residual. Thus, the following relation holds:

$$AV_m = V_{m+1}H_m, \tag{2.1}$$

where H_m is upper Hessenberg of size $(m+1) \times m$. The vector y_m is chosen to minimize the norm of the residual $r_m = b - Ax_m$, i.e., find y_m which is the minimizer of

$$\min_{y \in \mathbb{R}^m} \|r_0 - AV_m y\| = \min_{y \in \mathbb{R}^m} \|V_{m+1}(\beta e_1 - H_m y)\|, \tag{2.2}$$

where e_1 is the first Euclidean vector. Our general analysis also includes truncated methods such as DQGMRES [32] as a possible outer subspace method, where only certain vectors of V_m are kept in storage. For example, one can keep the last ℓ columns of V_m (denote the $n \times \ell$ matrix containing these by $V_{m[\ell]}$). In this case, the minimization (2.2) is replaced by the minimization of

$$\min_{y \in \mathbb{R}^m} \|\beta e_1 - H_m y\|, \tag{2.3}$$

where H_m is banded with upper semibandwidth ℓ. In QMR [16], a relation like (2.1) holds, but V_m is not an orthogonal matrix (the basis is bi-orthogonal to a basis of another Krylov subspace; these bases are obtained via a two-sided Lanczos process), the minimization of (2.3) is performed, and H_m is tridiagonal.

When a standard right preconditioner is used, as in (1.2), the expression (2.1) becomes

$$AM^{-1}V_m = V_{m+1}H_m.$$

When flexible preconditioning is used, this relation is replaced with

$$AZ_m = V_{m+1}H_m, \tag{2.4}$$

cf. (1.3), and storage has to be allocated for both the matrices Z_m and V_m, i.e., one needs approximately twice the storage of the standard preconditioned case; see, e.g., [29], [36].

In order to analyze the inner-outer method, when the outer method is a flexible Krylov subspace method and the inner is a Krylov subspace method, we consider the approximations of the solution of (1.1) at the kth (outer) cycle to be taken from the affine space $x_0 + \text{span}\{z_1, \ldots, z_k\}$, where $Z_k = [z_1, \ldots, z_k]$, i.e., of the form

$$x_k = x_0 + Z_k u_k \quad \text{for some } u_k \in \mathbb{R}^k. \tag{2.5}$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We change the notation a bit, to make a distinction with the inner method, and rewrite (2.4) as

$$AZ_k = W_{k+1}T_k, \tag{2.6}$$

i.e., $W_k = [w_1, \ldots, w_k]$ is the matrix containing a basis of the outer subspace, and the $(k+1) \times k$ matrix T_k contains the coefficients used in the orthogonalization (or deflation). The matrix T_k is either upper Hessenberg (if orthogonalization is used via the Arnoldi method), banded (if a truncated method is used), or tridiagonal (if the two-sided Lanczos method is used).

In other words, given x_0 and the corresponding residual

$$r_0 = b - Ax_0, \qquad w_1 = r_0/\beta, \qquad \beta = \|r_0\|, \tag{2.7}$$

for each cycle k, first a new vector z_k is computed, which approximates the solution of

$$Az = w_k \tag{2.8}$$

(using an inner Krylov method). Then, the vector Az_k is computed, orthogonalized with respect to the previous vectors w_i, $i \le k$ (or with respect to the ℓ vectors kept in the truncated version), and normalized (or bi-orthogonalized with respect to some other vectors) to obtain the new vector w_{k+1}; cf. (2.6). Thus, the residual at the kth cycle is

$$r_k = b - Ax_k = r_0 - AZ_k u_k = r_0 - W_{k+1}T_k u_k = W_{k+1}(\beta e_1 - T_k u_k). \tag{2.9}$$

We point out that in the case of using an Arnoldi method in the outer scheme, orthogonalizing Az_k with respect to the vectors in W_k is equivalent to deflating the inner residual $w_k - Az_k$ with respect to the vectors in W_k, i.e., with respect to all previous inner Krylov subspace starting vectors. In fact, we have

$$\begin{aligned} t_{k+1,k}w_{k+1} &= (I - W_k W_k^T)Az_k \qquad (2.10)\\ &= w_k - (w_k - Az_k) - w_k + W_k W_k^T(w_k - Az_k)\\ &= -(I - W_k W_k^T)(w_k - Az_k), \qquad (2.11) \end{aligned}$$

where W_k^T stands for the transpose of W_k.

As we said, at the kth cycle a new vector z_k is computed, which approximates the solution of (2.8), using an inner Krylov method. The corresponding (inner) Krylov subspace is $\mathcal{K}_m(A, w_k)$, with $m = m_k$, and its basis is $\{v_1^{(k)}, \ldots, v_m^{(k)}\}$, with

$$v_1^{(k)} = w_k \tag{2.12}$$

(of unit norm). The matrix $V_m^{(k)} = [v_1^{(k)}, \ldots, v_m^{(k)}]$ satisfies (2.1), i.e.,

$$AV_m^{(k)} = V_{m+1}^{(k)} H_m^{(k)}, \tag{2.13}$$

and we have

$$z_k = V_m^{(k)} y_k \quad \text{for some } \; y_k \in \mathbb{R}^m, \; m = m_k. \tag{2.14}$$

It is important to realize that when the inner system is preconditioned from the right, this can be viewed as a global preconditioning strategy. More precisely, consider

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

preconditioning the inner system (2.8) from the right with a fixed matrix P, so that the inner system transforms into $AP^{-1}\hat{z} = w_k$, with $z = P^{-1}\hat{z}$. An approximation to $\hat{z}$ is determined in $\mathcal{K}_m(AP^{-1}, w_k)$, from which an approximation to z can be readily recovered. Note that we can write $\hat{z}_k = V_m^{(k)} y_k$, where $V_m^{(k)}$ is now a basis for $\mathcal{K}_m(AP^{-1}, w_k)$. Let $Z_k = P^{-1}\hat{Z}_k = [P^{-1}\hat{z}_1, \ldots, P^{-1}\hat{z}_k]$. Relation (2.6) transforms into

$$AP^{-1}\hat{Z}_k = W_{k+1}T_k,$$

which immediately shows that inner preconditioning corresponds to applying the flexible method to the system $AP^{-1}\hat{x} = b$, with $\hat{x} = Px$. In light of these considerations, from now on we shall work with the coefficient matrix A, where A could actually stand for any preconditioned matrix AP^{-1}.

The description of flexible inner-outer methods by (2.5), (2.6), (2.13), and (2.14) is pretty general, and many of our results, including the following proposition, apply to this general setting.

PROPOSITION 2.1. *Each new vector of the outer basis w_{k+1} is a linear combination of the orthogonal projections of the columns $V_{m+1}^{(k)}$ (i.e., of the basis of the inner space $\mathcal{K}_{m+1}(A, w_k)$), onto the (bi-)orthogonal complement of $\mathcal{R}(W_k)$ (or $\mathcal{R}(W_{k[\ell]})$).*

Proof. From (2.14) and (2.13), we have $Az_k = AV_m^{(k)} y_k = V_{m+1}^{(k)}(H_m^{(k)} y_k)$, and from (2.10) the proposition follows. ∎

Note that while for each of the cycles we have the columns of $V_m^{(k)}$ being a basis of a Krylov subspace, neither $\mathcal{R}(W_k)$ nor $\mathcal{R}(Z_k)$, the range of W_k or Z_k, respectively, is a Krylov subspace. This is in contrast to the standard (nonflexible) preconditioned case. As pointed out, e.g., in [9], minimal residual or orthogonal residual methods with these bases would converge, as long as $\dim \mathcal{R}(W_k) = k$ and $\dim \mathcal{R}(Z_k) = k$, i.e., as long as the new vectors z_k and w_k are linearly independent of the previous ones so that the subspaces keep growing. As we shall see, this is the case for the inner-outer methods studied here.

We can say more; the columns of Z_k and those of W_k are bases of (different) subspaces (of dimension k) of a larger Krylov subspace generated by the initial residual r_0.

LEMMA 2.2. *Assume that Z_k and W_{k+1} are full column rank. Then, $\mathcal{R}(Z_k) \subset \mathcal{K}_{q-1}(A, r_0)$, and $\mathcal{R}(W_{k+1}) \subset \mathcal{K}_q(A, r_0)$, where $q = q_k = p_k + k$ and p_k is given by $p = p_k = \sum_{j=1}^{k} m_j$.*

Proof. We use induction on k. From (2.7), $\operatorname{span}\{w_1\} = \operatorname{span}\{r_0\}$. Therefore $z_1 \in \mathcal{K}_{m_1}(A, w_1) = \mathcal{K}_{m_1}(A, r_0)$ and $w_2 \in \operatorname{span}\{Az_1, w_1\} \subset \mathcal{K}_{m_1+1}(A, r_0)$, so that $\mathcal{R}(W_2) \subset \mathcal{K}_q(A, r_0)$ with $q = p_1 + 1 = m_1 + 1$. Assume that the assertions hold for $k-1$, and thus $w_k \in \mathcal{R}(W_k) \subset \mathcal{K}_q(A, r_0)$, with $q = q_{k-1}$. From (2.12), it follows that z_k belongs to the inner Krylov subspace $\mathcal{K}_m(A, w_k)$ $(m = m_k)$, which is a subspace of $\mathcal{K}_q(A, r_0)$ with $q = q_{k-1} + m_k = p_{k-1} + k - 1 + m_k = q_k - 1$. We then have that $w_{k+1} \in \operatorname{span}\{Az_k, W_k\} \subset \mathcal{K}_q(A, r_0)$, with $q = q_k$. ∎

This lemma also applies to truncated Krylov methods, since the vectors kept in storage are chosen from W_{k+1}, i.e., we have

$$\mathcal{R}(W_{k+1[\ell]}) \subset \mathcal{R}(W_{k+1}) \subset \mathcal{K}_q(A, r_0). \tag{2.15}$$

In the special case that all inner subspaces have the same dimension $m_j = m$, we have $p = km$ and $q = km + k = k(m+1)$; cf. [37, Lemma 4.2], where a result similar to Lemma 2.2 is shown for this special case.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We also remark that in [18] and [34] the situation is analyzed when (2.1) is replaced with $AU_{k-1} = U_k B_k$ for some matrix B_k and $\mathcal{R}(U_k)$ is some general subspace; cf. (2.6). The question there is to find an appropriate matrix E for which the columns of U_k are the basis of a Krylov subspace of $A + E$, i.e., $\mathcal{R}(U_k) = \mathcal{K}_k(A + E, r_0)$. Here we instead have subspaces of a larger Krylov subspace of A.

3. The subspaces of the inner-outer methods. In this section, we further characterize the k-dimensional subspaces $\mathcal{R}(Z_k)$ and $\mathcal{R}(W_k)$. Our characterization is of interest but does not necessarily reveal the intrinsic form of these subspaces. It does help us, though, in providing part of the setup used in section 5, where we show how these subspaces grow with each cycle.

The first simple observation is that the approximation x_k can be expressed as a particular linear combination of all the bases of the k inner Krylov subspaces. Indeed, from (2.5) and (2.14), it follows that

$$x_k = x_0 + \sum_{j=1}^{k} (u_k)_j \; z_j = x_0 + \sum_{j=1}^{k} (u_k)_j \; V_{m_j}^{(j)} \; y_j = x_0 + \sum_{j=1}^{k} \sum_{i=1}^{m_j} (u_k)_j \; (y_j)_i \; v_i^{(j)} \; .$$

Equivalently, if we define

$$\mathcal{B}_k = [V_{m_1}^{(1)}, V_{m_2}^{(2)}, \ldots, V_{m_k}^{(k)}] \in \mathbb{R}^{n \times p} \; , \quad Y_k = \begin{bmatrix} y_1 & O & O & \cdots \\ O & y_2 & O & \cdots \\ \vdots & & \ddots & \\ O & \cdots & O & y_k \end{bmatrix} \in \mathbb{R}^{p \times k} \; , \tag{3.1}$$

where O stands for a submatrix with zero entries (in this case $m \times 1$ submatrices, for different values of m), we have, from (2.14),

$$Z_k = \mathcal{B}_k Y_k, \tag{3.2}$$

and thus $x_k = x_0 + Z_k u_k = x_0 + \mathcal{B}_k (Y_k u_k)$. If we write the $n \times q$ matrix $\mathcal{B}'_k = [V_{m_1+1}^{(1)}, V_{m_2+1}^{(2)}, \ldots, V_{m_k+1}^{(k)}]$ and the $q \times p$ matrix $\mathcal{H}_k$ as the "block diagonal collection" of all $H_m^{(i)}$'s, we obtain the relation

$$A\mathcal{B}_k = \mathcal{B}'_k \mathcal{H}_k, \tag{3.3}$$

reminiscent of (2.13) or (2.1). Therefore, using (3.2) and (3.3),

$$r_k = r_0 - AZ_k u_k = r_0 - A\mathcal{B}_k Y_k u_k = r_0 - \mathcal{B}'_k (\mathcal{H}_k Y_k u_k), \tag{3.4}$$

providing an explicit representation of the residual in terms of the complete set of inner bases collected in $\mathcal{B}'_k$.

Remark 3.1. In light of Proposition 2.1 and (2.12), the columns of $\mathcal{B}'_k$ are not linearly independent. We then consider the $n \times (p+1)$ matrix

$$\mathcal{V}_k = [w_1, V_{2:m_1+1}^{(1)}, V_{2:m_2+1}^{(2)}, \ldots, V_{2:m_k+1}^{(k)}], \tag{3.5}$$

where $V_{2:m+1}^{(j)} = [v_2^{(j)}, \ldots, v_m^{(j)}]$, $m = m_j$. If the matrix $\mathcal{V}_k$ is of full rank, we have

$$\mathcal{R}(\mathcal{V}_k) = \mathcal{K}_{p+1}(A, r_0).$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In other words, the columns of the matrix $\mathcal{V}_k$ (if full rank) provides us with a basis for the Krylov subspace $\mathcal{K}_{p+1}(A, r_0)$ from where the global iterates are computed; see [33], where a similar construct is used. In general, we can conclude only that

$$\mathcal{R}(\mathcal{V}_k) \subseteq \mathcal{K}_{p+1}(A, r_0). \tag{3.6}$$

In light of Lemma 2.2, the relation (3.2) is natural, and by Remark 3.1 we can improve upon the dimension of the Krylov subspace from Lemma 2.2 to obtain

$$\mathcal{R}(Z_k) \subset \mathcal{K}_p(A, r_0), \qquad p = p_k. \tag{3.7}$$

We now want to find a similar explicit dependence of W_{k+1} on $\mathcal{B}'_k$ or $\mathcal{V}_k$. We consider two cases: when the columns of W_{k+1} are produced by an Arnoldi method or by a two-sided Lanczos algorithm.

LEMMA 3.2. *If the Arnoldi method is used to generate Z_k, W_k, the bases of the outer spaces, related by* (2.6)*, and these have full column rank, then,*

$$W_{k+1} = \mathcal{V}_k R_k, \tag{3.8}$$

with $R_k = G_k S_k^{-1}$, where the $(p+1) \times (k+1)$ matrix G_k and the $(k+1) \times (k+1)$ matrix S_k are given by

$$G_k = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ O & -\hat{g}_1 & O & O & \cdots \\ O & O & -\hat{g}_2 & O & \cdots \\ \vdots & & & \ddots & \\ O & O & O & \cdots & -\hat{g}_k \end{bmatrix}, \qquad S_k = \begin{bmatrix} 1 & 0 & t_{1,2} & t_{1,3} & \\ 0 & t_{2,1} & 0 & t_{2,3} & \ddots \\ 0 & 0 & t_{3,2} & 0 & \ddots \\ 0 & 0 & 0 & t_{4,3} & \ddots \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & t_{k+1,k} \end{bmatrix},$$

$\hat{g}_k \in \mathbb{R}^m$ is defined through $g_k = e_1 - H_m^{(k)} y_k = [\gamma_k; \hat{g}_k^T]^T$, and the $(k+1) \times k$ upper Hessenberg matrix T_k has entries t_{ij}.

Proof. Using (2.13), (2.14), and the definition of g_k,

$$Az_k = AV_m^{(k)} y_k = V_{m+1}^{(k)} H_m^{(k)} y_k = V_{m+1}^{(k)}(e_1 - g_k).$$

Then, by (2.12) and the fact that we use the Arnoldi method,

$$t_{k,k} = w_k^T A z_k = w_k^T V_{m+1}^{(k)}(e_1 - g_k) = e_1^T(e_1 - g_k) = 1 - \gamma_k. \tag{3.9}$$

We can use these relations to write w_{k+1} as a linear combination of the inner bases.

$$\begin{aligned} w_{k+1} t_{k+1,k} &= Az_k - W_k T_{1:k,k} \\ &= V_{m+1}^{(k)}(e_1 - g_k) - W_{k-1} T_{1:k-1,k} - w_k t_{k,k} \\ &= v_1^{(k)}(1 - \gamma_k) - [v_2^{(k)}, \ldots, v_{m+1}^{(k)}]\hat{g}_k - W_{k-1} T_{1:k-1,k} - w_k t_{k,k} \\ &= -[v_2^{(k)}, \ldots, v_{m+1}^{(k)}]\hat{g}_k - [w_1, w_2, \ldots, w_{k-1}] T_{1:k-1,k}, \end{aligned}$$

that is,

$$[w_1, w_2, \ldots, w_{k+1}] \begin{bmatrix} T_{1:k-1,k} \\ 0 \\ t_{k+1,k} \end{bmatrix} = -V_{2:m+1}^{(k)} \hat{g}_k.$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Collecting all terms, we obtain

$$W_{k+1}S_k = [w_1, V^{(1)}_{2:m+1}, V^{(2)}_{2:m+1}, \ldots, V^{(k)}_{2:m+1}] \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ O & -\hat{g}_1 & O & O & \cdots \\ O & O & -\hat{g}_2 & O & \cdots \\ \vdots & & & \ddots & \\ O & O & O & \cdots & -\hat{g}_k \end{bmatrix}.$$

The lemma follows from the nonsingularity of the upper triangular matrix S_k. ∎

The same result (3.8) holds in the case of two-sided Lanczos. The difference is that the matrix T_k is tridiagonal, so that the entries above the third diagonal in S_k are zero. The proof proceeds in the same manner as that of Lemma 3.2, except that in (3.9) w_k is replaced by the corresponding vector of the bi-orthogonal basis to $\mathcal{R}(W_k)$.

It follows from (3.8) and Remark 3.1 that

$$\mathcal{R}(W_{k+1[\ell]}) \subset \mathcal{R}(W_{k+1}) \subset \mathcal{K}_{p+1}(A, r_0), \qquad p = p_k;$$

cf. (2.15), Lemma 2.2, and (3.7).

To complete the picture, we want to obtain relations of the form (3.3) and (3.4), using the matrix $\mathcal{V}_k$. If full rank, its columns are a basis of the global subspace $\mathcal{K}_{p+1}(A, r_0)$.

Let P be the $q \times q$ permutation matrix that moves all vectors $v_1^{(2)}, v_1^{(3)}, \ldots, v_1^{(k)}$ in $\mathcal{B}'_k$ to the end (that is, as the last columns of the whole matrix). Then we can write $\mathcal{B}'_k P = [\mathcal{V}_k, v_1^{(2)}, v_1^{(3)}, \ldots, v_1^{(k)}] = [\mathcal{V}_k, W_{2:k}]$, where $W_{2:k} = [w_2, w_3, \ldots, w_k] = W_{k+1}[O, I_{k-1}, O]^T$, I_{k-1} is the identity of order $k-1$. Thus, using (3.8), we write

$$\mathcal{B}'_k = \mathcal{V}_k \, [I, R_k[O, I_{k-1}, O]^T] P^T, \tag{3.10}$$

where I is the identity of order $p+1$, obtaining from (3.3) the desired relation

$$A\mathcal{B}_k = \mathcal{V}_k \, [I, R_k[O, I_{k-1}, O]^T] P^T \mathcal{H}_k =: \mathcal{V}_k N_k. \tag{3.11}$$

Similarly, using the definition of N_k in (3.11), replacing (3.10) in (3.4), and using (2.7), we obtain,

$$r_k = r_0 - \mathcal{V}_k N_k Y_k u_k = \mathcal{V}_k(\beta e_1 - N_k Y_k u_k). \tag{3.12}$$

4. Minimal residuals and other methods. The discussion so far applies to any Krylov subspace method including CG, MINRES, GMRES, QMR, FOM, BiCG, etc. In principle, the outer (flexible) method can be any one of them, while the inner method can be any other. In practice, FGMRES-GMRES(m), GCR-GMRES(m), and FQMR-QMR were considered [5], [6], [29], [36], [37]. GMRES is a minimal residual method [31], while QMR [16] can be seen as one if the appropriate inner product and associated norm are considered; see [9]. In fact, the proof in [9, section 4.3] applies equally well to truncated GMRES methods such as DQGMRES; we more generally call these flexible truncated GMRES (FTGMRES). Therefore, when we consider that the inner-outer method is FGMRES-GMRES(m), our analysis will equally apply to the cases of FQMR-QMR and FTGMRES-GMRES(m) with the proviso that the norm at each cycle is a different one.

In FGMRES-GMRES(m), z_k in (2.14) is chosen as to minimize the inner residual in the kth cycle, i.e.,

$$\|w_k - Az_k\| = \|w_k - AV_m^{(k)} y_k\| = \|V_m^{(k)}(e_1 - H_m^{(k)} y_k)\| = \min_{y \in \mathbb{R}^m} \|e_1 - H_m^{(k)} y\|,$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

while u_k in (2.5) is chosen as to minimize the outer residual (2.9), i.e.,

$$\|r_k\| = \min_{u\in\mathbb{R}^k} \|W_{k+1}(\beta e_1 - T_k u)\| = \min_{u\in\mathbb{R}^k} \|\beta e_1 - T_k u\|. \tag{4.1}$$

Remark 4.1. We emphasize that the inner-outer methods we discuss do not consist of just the concatenation of the inner and the outer minimizations—there is also the orthogonalization with respect of the outer basis (see, e.g., Proposition 2.1), i.e., before the outer minimization takes place, there is also a deflation step, cf. [7], [10], where some Krylov spaces with deflation are studied; see further section 7.

In light of the discussion in section 3, the minimization (4.1) can be interpreted as being performed on a subspace of dimension k of the global Krylov space $\mathcal{K}_{p+1}(A, r_0)$. Indeed, from (3.12), now we can write

$$\|r_k\| = \min_{u\in\mathbb{R}^k} \|\mathcal{V}_k(\beta e_1 - N_k Y_k u)\|.$$

We note that the columns of $\mathcal{V}_k$ are not necessarily orthogonal.

One can explicitly say that the subspace of $\mathcal{K}_{p+1}(A, r_0)$ where the minimization takes place is spanned by the (k linearly independent) columns of $\mathcal{V}_k N_k Y_k$. Another characterization of this subspace can be obtained from (2.6) and (3.8) giving us the k columns of $AZ_k = \mathcal{V}_k R_k T_k$.

5. The growing subspaces. The main result of this section is that the subspace from where the approximations are chosen keeps growing. This provides convergence of the inner-outer methods. In exact arithmetic, these methods would then terminate in at most n steps, where n is the order of the matrix A.

We have noted that, in general, we have the inclusion (3.6). It is also well known that the matrix $[r_0, Ar_0, \ldots, A^p r_0]$ may have vectors which are almost linearly dependent. We therefore want to study what we can say about the rank of $\mathcal{B}_k$ defined in (3.1) or, equivalently, that of $\mathcal{V}_k$ defined in (3.5). We show here that as k grows, i.e., as a new cycle is computed, the rank of $\mathcal{B}_k$ is guaranteed to grow as well.

For simplicity of the exposition, we assume in this section that $m_k = m$ for all k, and thus $p = mk$. We further assume that the minimal polynomial of A has degree larger than p. (More precisely, this is assumed with respect to $v_1^{(k-1)}$; see (5.2) below. In other words, we assume that the grade of $v_1^{(k-1)}$ is less than p.) We comment at the end of the section the implications of these assumptions.

LEMMA 5.1. *Let* $\mathcal{R}(V_m^{(k-1)}) = \mathcal{K}_m(A, v_1^{(k-1)})$ *and* $\mathcal{R}(V_m^{(k)}) = \mathcal{K}_m(A, v_1^{(k)})$ *be both of dimension* m *(i.e.,* $V_m^{(k-1)}$ *and* $V_m^{(k)}$ *are of full rank), with*

$$v_1^{(k)} \in \mathcal{R}(AV_m^{(k-1)}) \tag{5.1}$$

and $v_1^{(k)}$ *and* $v_1^{(k-1)}$ *being linearly independent. Then* $\dim \mathcal{R}([V_m^{(k-1)}, V_m^{(k)}]) > m$.

Proof. Let $v = v_1^{(k-1)}$; then any element $w \in \mathcal{R}(V_m^{(k-1)}) = \mathcal{K}_m(A, v)$ can be written as $w = p(A)v$, with p a polynomial of degree $m-1$ at most. Thus, from (5.1) we have that $v_1^{(k)} = Ap(A)v = q(A)v$, with $q(z)$ a nonzero polynomial of degree at most m, such that $q(0) = 0$. In other words, $1 \le \deg q(z) \le m$. In fact, since $0 \ne v_1^{(k)} \ne \alpha v$, for any $\alpha \in \mathbb{R}$, we have $2 \le \deg q(z) \le m$. Now consider

$$A^{m-1} v_1^{(k)} = A^{m-1} q(A) v = \hat{q}(A) v, \tag{5.2}$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\deg \hat{q}(z) = (m-1) + \deg q(z) \geq m+1$; then $A^{m-1}v_1^{(k)} \notin \mathcal{K}_m(A, v_1^{(k-1)})$. Therefore,

$$\dim \mathcal{R}\left([V_m^{(k-1)}, A^{m-1}v_1^{(k)}]\right) = m+1,$$

from which the result follows. ∎

We note that in Lemma 5.1 (as well as Theorem 5.2 below) the hypothesis on the bases $V_m^{(j)}$ of $\mathcal{K}_m(A, v_1^{(j)})$ is that they be of full column rank. There is no requirement that they be orthogonal. Similarly, the vectors $v_1^{(j)}$ and $v_1^{(j-1)}$ need only be linearly independent. This implies that Lemma 5.1 and Theorem 5.2 apply to any method that produces these pairwise linearly independent vectors. In particular, they apply to the cases when either Arnoldi or two-sided Lanczos are used for the construction of the matrix W_k.

In [7] and [38, section 3] it is recommended that when using truncated Krylov methods the last vector of the previous basis, say $V_m^{(k-1)}$, be kept in the new basis. This recommendation was based on empirical evidence on symmetric matrices. Lemma 5.1 provides a theoretical basis for such a recommendation. In fact, the last column of $V_m^{(k-1)}$ has a nonzero component in the direction of (5.2) (or of (5.6) below), which is the key quantity for the subspace to grow.

To evaluate how the subspaces $V_m^{(k-1)}$ and $V_m^{(k)}$ generated during the inner step behave vis-á-vis Lemma 5.1, we shall compute for some specific examples their canonical angles [35]; see later section 7. Consider the two subspaces $\mathcal{R}(V_m^{(k-1)})$ and $\mathcal{R}(V_m^{(k)})$, and let m_* be the smallest integer such that $\dim \mathcal{R}(V_{m_*}^{(j)}) = \dim \mathcal{R}(V_{m_*+1}^{(j)})$, $j = k-1,\ k$. We define $\bar{m} = m$ if $2m \leq m_*$ or $\bar{m} = m_* - m$ if $2m > m_*$. The canonical angles between these two subspaces can be computed as the nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\bar{m}}$ of $P_m^{(k-1)}(I - P_m^{(k)})$, where $P_m^{(j)} = V_m^{(j)}(V_m^{(j)})^T$, $j = k-1, k$, are the orthogonal projectors onto $\mathcal{R}(V_m^{(j)})$, $j = k-1,\ k$, respectively.

We present now the main result of this section.

THEOREM 5.2. *Let $\mathcal{R}(V_m^{(j)}) = \mathcal{K}_m(A, v_1^{(j)})$ be of dimension m, for $j = 1, \ldots, k$. Assume that each $v_1^{(j)}$ is such that*

$$v_1^{(j)} = AV_m^{(j-1)}y_{j-1} + \mathcal{B}_{j-2}d_j, \tag{5.3}$$

with $0 \neq y_{j-1} \in \mathbb{R}^m$, $j = 2, \ldots, k$, and that

$$v_1^{(j)} \quad \text{and} \quad v_1^{(j-1)} \quad \text{are linearly independent.} \tag{5.4}$$

Then

$$\dim \mathcal{R}(\mathcal{B}_k) \geq m + k - 1, \qquad k \geq 2. \tag{5.5}$$

Proof. We use induction on k. For $k = 2$, (5.5) holds by Lemma 5.1. Assume that (5.5) holds for $k-1 > 1$. We prove it for k. That is, we show that the next basis $V_m^{(k)}$ of $\mathcal{K}_m(A, v_1^{(k)})$ contains at least a vector that is not a linear combination of elements in $\mathcal{B}_{k-1}$. Consider the starting vector $v_1^{(k)}$ for the new basis $V_m^{(k)}$. From (5.3) we have

$$v_1^{(k)} = AV_m^{(k-1)}y_{k-1} + \mathcal{B}_{k-2}d_k = A\sum_{j=0}^{m-1} A^j v_1^{(k-1)}\eta_j + \mathcal{B}_{k-2}d_k$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for some values η_j, not all of which are zero. Therefore, for $A^{m-1}v_1^{(k)} \in \mathcal{K}_m(A, v_1^{(k)})$ we have

$$A^{m-1}v_1^{(k)} = A^m \sum_{j=0}^{m-1} A^j v_1^{(k-1)}\eta_j + A^{m-1}\mathcal{B}_{k-2}d_k. \tag{5.6}$$

Since not all η_j's are zero, using the same argument as in the proof of Lemma 5.1,

$$A^m \sum_{j=0}^{m-1} A^j v_1^{(k-1)}\eta_j \notin \mathcal{K}_m(A, v_1^{(k-1)}), \tag{5.7}$$

while the second term of (5.6) lies in $\mathcal{B}_{k-1}$; thus, $A^{m-1}v_1^{(k)} \notin \mathcal{B}_{k-1}$, and the result follows. ∎

Remark 5.3. The hypothesis for Lemma 5.1 and Theorem 5.2 that the degree of the minimal polynomial of A is larger than m and mk, respectively, though seldom violated in practice, is crucial. Otherwise the Krylov subspace might be invariant and one may have, e.g., $\dim \mathcal{R}([V_m^{(k-1)}, V_m^{(k)}]) = m$. We will assume that this hypothesis holds, else one may experience some early breakdown; see section 6 for the definition of breakdown and also [6] for a similar situation.

Several comments on Theorem 5.2 are in order. First, we comment on the hypothesis (5.4). In the case of (flexible) Arnoldi, this is automatically satisfied, since the columns of W_k are orthogonal. In the case of (flexible) two-sided Lanczos, one has a vector $\hat{w} = \hat{w}_{j-1}$ in the basis bi-orthogonal to W_k such that $(v_1^{(j-1)}, \hat{w}) \neq 0$, while $(v_1^{(j)}, \hat{w}) = 0$, implying (5.4); for details about two-sided Lanczos see, e.g., [19], [30]. We mention also that (5.4) implies that these vectors are not the zero vectors, i.e., $w_k \neq 0$, $k = j-1,\ j$. As described in section 6 this is equivalent to assuming that there is no breakdown of the inner-outer method.

Second, this theorem applies to the inner-outer methods described so far, including the truncated methods. The hypothesis (5.3) expresses what we have said in Proposition 2.1 and Lemma 3.2.

Third, we want to emphasize that while the proof of Theorem 5.2 implies that $\mathcal{R}(\mathcal{B}_k)$ grows in each cycle, (5.5) really provides a lower bound on its dimension, and we expect the dimension of $\mathcal{R}(\mathcal{B}_k)$ to be higher. We comment, though, that (5.5) can be rephrased as

$$\dim \mathcal{R}(\mathcal{B}_k) \geq \dim \mathcal{R}([V_m^{(1)}, W_k]) = m + k - 1,$$

where the last equality holds when $V_m^{(1)}$ and W_k are full rank.

Finally, a careful review of the proofs of Lemma 5.1 and Theorem 5.2 indicates that, for the case where m_k changes from one cycle to the next, (5.5) holds for $m = m_1$, and that one needs $m_j \geq m_{j-1}$ to guarantee (5.7).

6. Stagnation and breakdown. In this section we discuss the possibility of breakdown in the construction of the outer basis W_{k+1}, as well as the possibility of stagnation of the inner-outer methods. Some of our observations apply to the general setting described in sections 2, 3, and 5. In some other cases the discussion relates specifically to a minimal residual method, such as FGMRES-GMRES(m).

As we shall see, there are two elements that relate to the possibility of stagnation and breakdown: one is $\bar{T}_k$, the $k \times k$ principal matrix of T_k, and the other is the

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

next (outer) vector w_{k+1}. The situation on breakdown can be summarized as follows: If $w_{k+1} \neq 0$, there is no breakdown, even if $\bar{T}_k$ is singular. If $w_{k+1} = 0$, then the singularity of $\bar{T}_k$ matters. When $\bar{T}_k$ is singular we have (real) breakdown, while if $\bar{T}_k$ is nonsingular, we have what is usually called a "lucky" breakdown, meaning that this implies that x_k is the exact solution of (1.1); see [29, Proposition 2.2] where breakdown for FGMRES (using any inner solver) is considered. Stagnation of the inner-outer method is discussed at the end of the section.

In the case of FGMRES-GMRES(m), one example of breakdown ($w_{k+1} = 0$) occurs when there is stagnation in the inner iteration, i.e., in GMRES(m). In other words, given w_k, the approximation $z_k \in \mathcal{K}_m(A, w_k)$ is such that $w_k - Az_k = w_k$, cf. (2.11), implying $z_k = 0$. Unless the exact solution was found at the previous outer step, inner stagnation yields incurable breakdown, as the outer iteration no longer proceeds. Prevention of stagnation of the inner iteration has been addressed by other authors. In [37], if inner stagnation occurs, the inner step is replaced by a minimization step. At the end of section 2.2 of [29] there is a comment indicating that all that is needed to avoid stagnation is a direction of steepest descent. This is exactly what is guaranteed by the choice of the stopping criteria of the inner iterations in [17]. We also remark that in [1] a condition on the matrix A is introduced to avoid inner stagnation in the inner-outer generalized conjugate gradient method. Let $Q[w]$ denote the inner outcome. It is shown in [1] that if $AQ[\cdot]$ is coercive, that is, there exists $\delta > 0$ such that $(w, AQ[w]) \geq \delta(w, w)$ for all $w \neq 0$, then the inner iteration does not stagnate. Here (x, y) denotes the usual inner product. Note that in the symmetric positive definite case, coercivity implies that a steepest descent direction is obtained in the inner method; see, e.g., [22]. In the context of FGMRES-GMRES(m), coercivity also implies that the inner GMRES does not stagnate. Indeed, for each outer basis vector w_k and using $z_k = Q[w_k]$, we have

$$(w_k, AQ[w_k]) = (w_k, Az_k) = 1 - \|w_k - Az_k\|^2,$$

where $w_k - Az_k$ is the inner residual. Therefore, imposing $(w_k, AQ[w_k]) > 0$ implies $\|w_k - Az_k\|^2 < 1 = \|w_k\|$, which is equivalent to lack of stagnation. We should mention however that coercivity in our context is a very strong assumption. For GMRES as inner, it holds that $Q[w_k] = p_{m-1}(A)w_k$ for some polynomial p_{m-1} of degree not greater than $m-1$, where the polynomial changes at each outer cycle. Hence, $(w_k, AQ[w_k]) > 0$ is satisfied if the operator $Ap(A)$ is coercive for any polynomial p of degree at most $m-1$.

There are many examples where $\bar{T}_k$ is singular, even if T_k is of full rank. In these cases, having $t_{k+1,k} = \|w_{k+1}\| = 0$ implies real breakdown. In Lemma 5.1 and Theorem 5.2 it is assumed that there is no breakdown, see (5.4) and (2.12), and it is concluded that the space from where the solution is drawn keeps growing.

In the absence of breakdown ($w_{k+1} \neq 0$), $\bar{T}_k$ singular indicates stagnation of the outer process and thus of the overall inner-outer method. The following result is an adaptation of [3, Theorem 3.1] for FGMRES. The proof given in [3] can be used verbatim here, so we do not reproduce it.

THEOREM 6.1. *Suppose that k steps of the (outer) Arnoldi method have been taken, $w_{k+1} \neq 0$, and assume that $\bar{T}_k$ is singular. Then*

$$\min_{u \in \mathbb{R}^k} \|\beta e_1 - T_k u\| = \min_{u \in \mathbb{R}^{k-1}} \|\beta e_1 - T_{k-1} u\|. \tag{6.1}$$

If we denote by u_j the minimizer in (4.1) with j replacing k, for $j = k$ or $k-1$, then $u_k = ((u_{k-1})^T, 0)^T$, and it follows that $x_k = x_{k-1}$. Conversely, suppose that k

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

steps of the (outer) Arnoldi method have been taken and that (6.1) *holds; then* $\bar{T}_k$ *is singular.*

Note, however that since we assume that there is no breakdown, this stagnation can only be "temporary," since by Theorem 5.2 the subspace keeps growing, and therefore the inner-outer method converges [9].

7. Comparison with restarted methods. Let us consider restarted GMRES, i.e., GMRES(m). This method generates an approximate solution as sum of approximations obtained at each restart, that is,

$$x_k = x_0 + x^{(1)} + x^{(2)} + \cdots + x^{(k)}. \tag{7.1}$$

The single approximations are obtained in the following subspaces:

$$\begin{aligned} &K_m(A, r_0), && r_0 = b - Ax_0, \\ &K_m(A, r_0^{(1)}), && r_0^{(1)} = r_0 - Ax^{(1)} \equiv b - Ax_1, \\ &K_m(A, r_0^{(2)}), && r_0^{(2)} = r_0^{(1)} - Ax^{(2)} \equiv b - Ax_2, \\ &\quad \vdots \end{aligned}$$

where each starting vector $r_0^{(j)}$ is the residual in the previous Krylov subspace.

Intuitively, one can think of improving upon GMRES(m) by considering a linear combination (or weighted average) of the single approximations, say

$$\tilde{x}_k(\alpha) = x_0 + \alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_k x^{(k)}, \tag{7.2}$$

instead of (7.1); cf. [41, section 3]. One could require the parameters $\alpha \in \mathbb{R}^k$ to be constructed, for example, so as to minimize the norm of the residual, in which case we have

$$\|\tilde{r}_k\| = \min_{\alpha \in \mathbb{R}^k} \|b - A\tilde{x}_k(\alpha)\| \leq \|b - Ax_k\|,$$

where the last inequality follows from considering $\alpha = (1, \ldots, 1)^T$. In other words, such a method cannot be worse than restarted GMRES in terms of residual norms.

The inner-outer methods we study, such as FGMRES-GMRES(m), do more than just implicitly choose the k parameters in (7.2). It follows from Proposition 2.1 that one obtains an iteration of the form (7.2), but where the residual obtained for each single (inner) Krylov method is in turn deflated with respect to the previous (outer) vectors; cf. [13], [26]. We thus expect that the overall inner-outer method will perform at least as well as the restarted counterpart.

We prove this explicitly in some particular cases below ($k = 1$ or $m = 1$). For other cases, we show examples where at particular points in the computations, the inner-outer method may have a residual norm which is larger than the corresponding restarted one. Nevertheless, for these, as well as for all other numerical examples we ran, FGMRES-GMRES(m) always converges using fewer matrix-vector multiplications than GMRES(m).

In the following proposition we show that at the very first outer cycle the residual computed by FGMRES-GMRES(m) is the same as that of GMRES(m). Therefore, possible improvements versus the restarted approach are expected starting with the second outer cycle of the flexible method.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

PROPOSITION 7.1. *Let r_k^F, r_k^G be the FGMRES-GMRES(m) and GMRES(m) residuals after k cycles, respectively. For $k = 1$ we have $r_1^F = r_1^G$.*

Proof. After m iterations of the first cycle of GMRES(m), we solve the problem (2.2) where V_m spans $\mathcal{K}_m(A, r_0)$, and we obtain $y^G = (H_m^T H_m)^{-1} H_m^T e_1 \beta$, so that $r_1^G = b - AV_m(H_m^T H_m)^{-1} H_m^T e_1 \beta$.

On the other hand, let $w_1 = r_0 \beta^{-1}$ be the first outer basis vector in FGMRES-GMRES(m). After m iterations of the first inner step we solve the problem

$$\min_{y \in \mathbb{R}^m} \|w_1 - AV_m y\| = \min_{y \in \mathbb{R}^m} \|e_1 - H_m y\|,$$

where V_m spans $\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A, w_1)$. We obtain $y^F = (H_m^T H_m)^{-1} H_m^T e_1$ so that $z_1 = V_m y^F$. At the end of the first outer cycle we thus solve the problem

$$\min_{u \in \mathbb{R}} \|b - Az_1 u\| = \beta \min_{\hat{u} \in \mathbb{R}} \|w_1 - Az_1 \hat{u}\|, \qquad \hat{u} = \frac{1}{\beta} u.$$

We obtain $\hat{u} = ((Az_1)^T (Az_1))^{-1} (Az_1)^T w_1$. Explicitly writing $Az_1 = V_{m+1} H_m y^F$ and substituting into the previous expression, we derive $\hat{u} = ((y^F)^T H_m^T H_m y^F)^{-1} (y^F)^T H_m^T e_1 = 1$, from which it follows that $u = \beta \hat{u} = \beta$. Therefore, $r_1^F = b - Az_1 u = b - AV_m(H_m^T H_m)^{-1} H_m^T e_1 \beta = r_1^G$. ∎

Later in this section we provide an example that shows that Proposition 7.1 cannot be generalized to $k > 1$ for general m. Nonetheless, we can prove for $m = 1$ and for any $k > 0$ that the FGMRES-GMRES(m) iterates coincide with those of full GMRES.

PROPOSITION 7.2. *Let x_k^F be the approximate solution obtained after k outer cycles of FGMRES-GMRES(1), and assume that GMRES(1) does not stagnate. Let x_k^G be the approximate solution after k iterations of full GMRES. Then $x_k^F = x_k^G$.*

Proof. At each outer cycle k of FGMRES, the inner solver GMRES approximately solves $Az = w_k$. If GMRES(1) is used, then $z_k = \delta_k w_k$ for some scalar $\delta_k \neq 0$. Setting $D_k = \text{diag}(\delta_1, \ldots, \delta_k)$, we can write $Z_k = W_k D_k$ so that (2.6) becomes

$$AW_k = W_{k+1} T_k D_k^{-1},$$

and the latter is an Arnoldi relation so that $\mathcal{R}(W_k) = \mathcal{K}_k(A, w_1)$. Let $AV_k = V_{k+1} H_k$ be the Arnoldi relation associated with full GMRES. Since $\mathcal{R}(W_k) = \mathcal{R}(V_k)$, there exists an orthogonal $k \times k$ matrix R_k such that $W_k = V_k R_k$ and $H_k = R_{k+1} T_k D_k^{-1} R_k^{-1}$. Since $w_1 = v_1$, $(R_k)_{1,1} = 1$. Using $(H_k^T H_k)^{-1} H_k^T = R_k D_k (T_k^T T_k)^{-1} T_k^T R_{k+1}^T$, the solution $x_k^G = V_k (H_k^T H_k)^{-1} H_k^T e_1 \beta$ becomes

$$x_k^G = V_k R_k D_k (T_k^T T_k)^{-1} T_k^T R_{k+1}^T e_1 \beta = Z_k (T_k^T T_k)^{-1} T_k^T e_1 \beta = x_k^F. \qquad ∎$$

Unfortunately, the result of Proposition 7.2 and the implication that $\|r_k^F\| \leq \|r_k^G\|$ do not carry over to larger values of m, as experimentally shown in the next example.

Example 7.3. We consider the linear system $Ax = b$ of size $n = 100$, where $A = \text{bidiag}(d, 1)$, with $d = [0.01, 0.02, 0.03, 0.04, 10, 11, \ldots, 105] \in \mathbb{R}^n$. In Table 7.1 we report the residual norms for both FGMRES-GMRES(m) (F/G(m) for short) and GMRES(m) (GM(m) for short), with $m = 10$, when the right-hand side is $b_1 = (1, 2, 1, 2, \ldots)^T$ (left) and $b_2 = (1, -2, 1, -2, \ldots)^T$ (right). Both vectors were normalized so that $\|b_1\| = \|b_2\| = 1$. The reported results clearly show that $\|r_k^F\|$ is larger than $\|r_k^G\|$ at an early stage of the iterative process, highlighted in italics. Nonetheless, convergence is achieved after a few more iterations in the flexible method, whereas the restarted approach stagnates.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 7.1
FGMRES/GMRES(10) vs. GMRES(10) on bidiagonal matrix, with right-hand sides b_1 and b_2.

b_1			b_2		
k	F/G(10)	GM(10)	k	F/G(10)	GM(10)
1	0.197966	0.197966	1	0.168170	168170
2	*0.175303*	*0.170102*	2	0.153462	0.153675
3	0.145040	0.150841	3	*0.139839*	*0.138271*
4	0.145022	0.147060	4	*0.139510*	*0.137050*
⋮			⋮		
12	$6.4 \cdot 10^{-5}$	0.144093	13	0.00029268	0.136947

We recall here that each new vector is deflated before it is used as the initial vector for the inner iteration; see (2.11) or Proposition 2.1. As discussed in section 3, this deflation helps in providing inner subspaces $V_m^{(k)}$, which have a larger angle between them, justifying the good overall performance of the flexible method. This is illustrated in Table 7.2 below, where we considered the matrix in Example 7.3 and, as right-hand side b, the left singular vector corresponding to the smallest singular value of A; see [33]. Note that full GMRES reaches a residual norm less than 10^{-16} in 21 iterations; therefore in our tests we considered that an invariant subspace of A was found for $m_* = 20$.

For both FGMRES-GMRES(m) and GMRES(m), in Table 7.2 we report the smallest sine value of the canonical angles between the subspaces spanned by $V_m^{(k-1)}$ and $V_m^{(k)}$ for $k = 2, 3$; see the discussion in section 3. When using GM(12), we collected the value of the $\bar{m}$th singular value, with $\bar{m} = m_* - m = 20 - 12 = 8$.

TABLE 7.2
Smallest nonzero singular value of $V_m^{(k-1)}(V_m^{(k-1)})^T(I - V_m^{(k)}(V_m^{(k)})^T)$, measuring the sines of canonical angles between $\mathcal{R}(V_m^{(k-1)})$ and $\mathcal{R}(V_m^{(k)})$.

k		F/G(5)	F/G(10)	GM(5)	GM(10)	GMRES(12)
2	σ_{min}	$7 \cdot 10^{-5}$	$4 \cdot 10^{-8}$	$2 \cdot 10^{-16}$	$1 \cdot 10^{-16}$	$5 \cdot 10^{-6}$
3	σ_{min}	$1 \cdot 10^{-2}$	$9 \cdot 10^{-4}$	$4 \cdot 10^{-17}$	$3 \cdot 10^{-16}$	$3 \cdot 10^{-4}$

Table 7.2 shows that, for restarted GMRES, the distance between $V_m^{(1)}$ and $V_m^{(2)}$ is around machine precision for $m = 5, 10$, implying that the Krylov subspace generated after one restart is very close to the previous one. The same happens at the next outer cycle, for $V_m^{(2)}$ and $V_m^{(3)}$. Only for $m = 12$ do the subspaces generated after the first restart of GMRES provide new information. This is confirmed by the convergence history of the method, shown in the right plot of Figure 1.

Table 7.2 and Figure 1 also report results for the flexible method. In particular, by only deflating the starting vector of the first inner Krylov subspace, F/G(10) seems to be capable of capturing the information missed by restarted GMRES after one cycle, resulting in faster convergence. On the other hand, we notice that F/G(5) achieves its final approximation to the exact solution with a large delay.

8. Computational considerations. Computational efficiency leads us to consider truncated and restarted versions of optimal algorithms. This is the case both for the original GMRES method as well as for FGMRES [31], [32], [37]. In the case of a flexible method, Theorem 5.2 ensures that in exact arithmetic the method converges as long as condition (5.4) is satisfied. Therefore, if, for example,

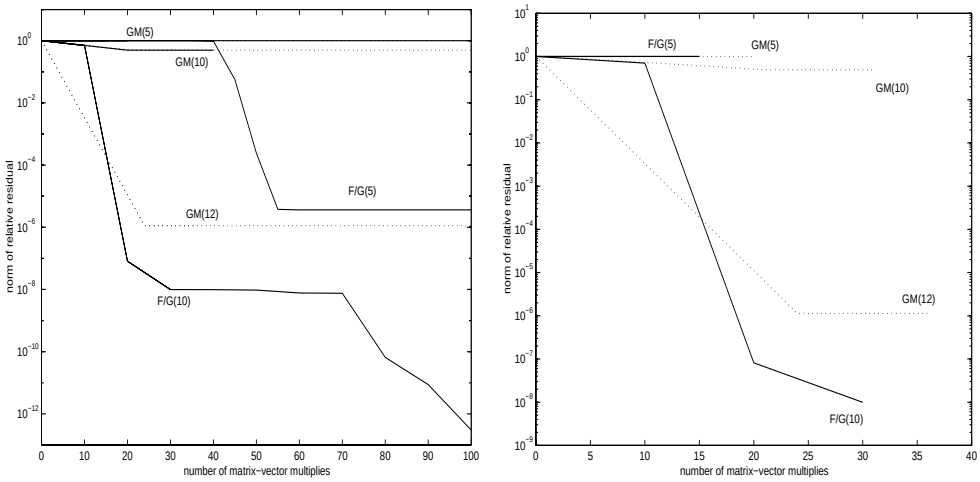

FIG. 1. *Convergence history of flexible and restarted methods for various values of m. Left: Entire picture. Right: Initial phase. See also Table* 7.2.

GMRES is used as the outer method, finite termination is obtained by orthogonalizing only the next outer basis vector w_{k+1} with respect to the previous vector w_k. By orthogonalizing w_{k+1} with respect to more basis vectors, i.e., to more columns of W_k, we expect the convergence to be faster.

In order to make fair comparisons, we recall that k cycles of FGMRES require m vectors for the inner space plus $2k$ vectors for the outer bases. These should be compared with the m vectors required by GMRES(m). A truncated version of FGMRES will instead require $2k_t$ outer vectors, where k_t is fixed beforehand. In our experiments, when using truncated flexible GMRES, we used m_t inner vectors and k_t outer vectors so that $m_t + 2k_t = m$, where m is the number of vectors used in restarted GMRES. We remark that such memory constraint on m_t, k_t forces us to work with a much smaller inner space than in restarted GMRES if outer orthogonality is maintained with respect to several vectors. Different selections of these parameters were analyzed and some of them are reported here. In addition, we mention that the larger k_t, the more expensive the reorthogonalization step. Clearly, the worst case scenario in this context is given by FGMRES, where all outer vectors are kept orthogonal.

We also remark that the matrix-vector multiplication required at each outer step of the flexible method can be avoided by exploiting the available inner residual. Therefore, at least for a low number of outer cycles, restarted and (truncated) flexible methods can be compared in terms of number of matrix-vector products, which in our experiments represents the highest computational cost. It is also customary to further precondition the flexible approach with a matrix P by applying the flexible method to the preconditioned matrix AP^{-1}. Due to the implementation consideration just mentioned, preconditioning the inner-outer method amounts to simply preconditioning the inner solver; see also the discussion in section 2.

Example 8.1. We consider the 900×900 matrix originating from the centered finite difference discretization of the operator

$$L(u) = -\Delta u + \mu x u_x, \qquad \text{for} \quad \mu = 100, 1000, \tag{8.1}$$

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

on $[0,1] \times [0,1]$ with zero Dirichlet boundary conditions. As a right-hand side we selected $b = Ae$, where e is a vector of all ones.

Figure 2 reports the convergence history of restarted GMRES(20), FGMRES-GMRES(20), and its truncated variant, using FTGMRES-GMRES(m_t, k_t) (FT/G(m-k) for short) as discussed above. In the left plot we displayed the results for $\mu = 100$, in the right plot for $\mu = 1000$. We observe that for $\mu = 100$ truncation is not harmful, especially when the generated inner space is sufficiently large, yielding similar convergence for the flexible method and its truncated variant. For $\mu = 1000$ the picture changes considerably and the convergence of the truncated methods reflects the influence of the two parameters m_t, k_t. Restarted GMRES clearly shows lack of information that is eventually recovered after many more iterations.

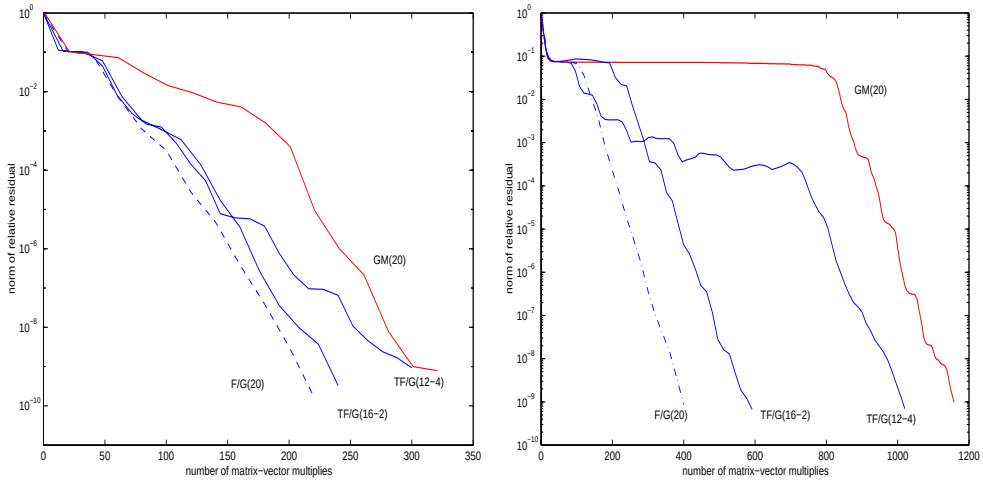

FIG. 2. *Convergence history for operator $L(u)$ in* (8.1). *Left:* $\mu = 100$. *Right:* $\mu = 1000$.

Example 8.2. We consider the operator

$$L(u) = -1000\Delta u + 2e^{4(x^2+y^2)}u_x - 2e^{4(x^2+y^2)}u_y \tag{8.2}$$

on $[0,1] \times [0,1]$, which was studied in [7, Problem 2]. Here we assume zero Dirichlet boundary conditions and a centered finite difference discretization, yielding a coefficient matrix A of size $n = 40000$. The right-hand side is determined as $b = Ae$, where e is the vector of all ones, and then normalized so as to have unit norm. The convergence history of restarted GMRES with $m = 30$ is reported in Figure 3 (left). The figure also shows the curves of the flexible variant and its truncated versions for two different values of the truncation parameter. On this problem, truncation is particularly effective. The results seem to suggest that on this problem, an inner subspace of small dimension suffices for the flexible method to converge rapidly. In the right plot of Figure 3 we show the convergence history of all methods for a larger Krylov subspace dimension, $m = 50$. Restarted GMRES considerably improves its performance; cf. [7]. On the other hand, flexible schemes do not seem to necessitate of a larger inner dimension, implying that information gathering in the outer process is very effective. Our findings corroborate similar experimental evidence in [7]. We also notice that for $m = 50$, the truncated variants converged in less than 20 outer iterations. Therefore, when large values of k_t are selected, truncation takes place only

very late in the convergence stage. As a consequence, the curves in the right plots of Figure 3 associated with the truncated methods FT/G(30-10) and FT/G(20-15) closely resemble the curves one would obtain with FGMRES(m) with $m = 30$ and $m = 20$; cf. the left plot of Figure 3 for $m = 30$.

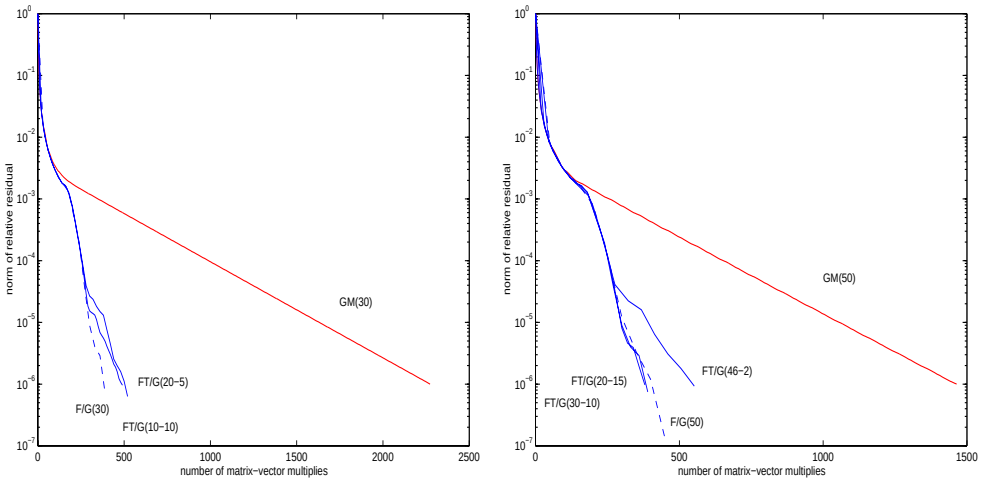

FIG. 3. *Convergence history for operator $L(u)$ in* (8.2). *Left: $m = 30$. Right: $m = 50$.*

Example 8.3. Our last set of experiments involves three matrices from the "Matrix Market" [8], [24]. For the first two matrices, the right-hand side was chosen to be the vector of all ones. Additional (fixed) incomplete LU preconditioning was applied [25], [30]. In our experiments, we used the Matlab function `luinc` with tolerance *tol* to build the preconditioning matrix P. As mentioned earlier in this section, in the flexible algorithm this amounts to run the inner solver with the preconditioned matrix AP^{-1}.

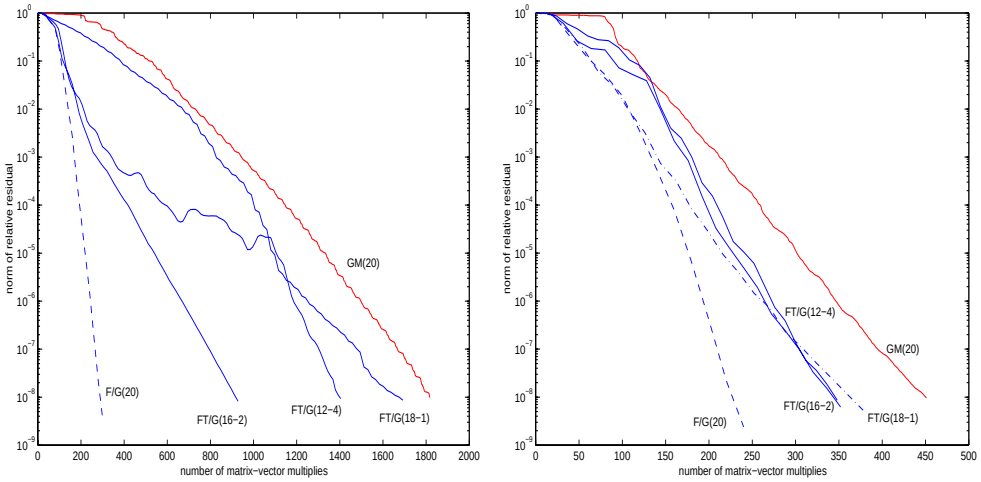

FIG. 4. *Convergence history for flexible and restarted methods on matrix Sherman5. Left: no preconditioning. Right: preconditioning using incomplete LU with tolerance tol $= 10^{-2}$.*

In Figure 4 we report experiments with the matrix SHERMAN5 from the Harwell–

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Boeing set, a nonsymmetric 3312×3312 matrix stemming from a fully implicit black oil simulation [8]. The matrix was scaled using the absolute values of its diagonal entries. The fixed ILU preconditioner was built using $tol = 10^{-2}$.

The plots show that the behavior of the methods is more homogeneous after fixed preconditioning, while when no preconditioning is applied the convergence of the truncated flexible method is less predictable.

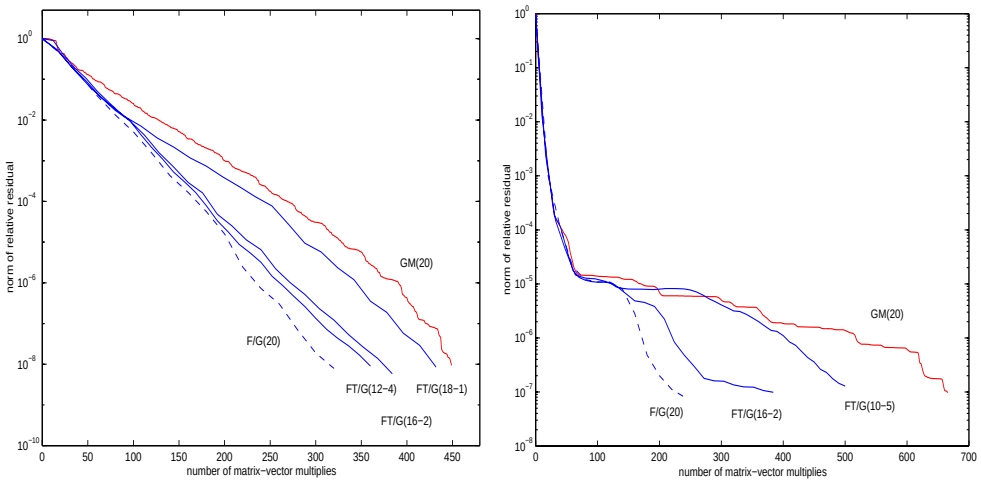

FIG. 5. *Convergence history for flexible and restarted methods on Preconditioned Oilgen (left) and on Scaled Fidap_ex*11 *(right).*

The second matrix in the set is the OILGEN 2205×2205 matrix from the same matrix set, originating from a three-dimensional oil reservoir simulation [8]. The fixed ILU preconditioning was applied with tolerance $tol = 0.5$. Convergence histories are reported in the left plot of Figure 5. The difference in the behavior of the methods is less pronounced.

Finally, we consider a larger matrix from the FIDAP group in the University of Florida collection [8], namely matrix EX11 of size $n = 16614$, which stems from a fully coupled Navier–Stokes problem. We set $b = Ae$, normalized so as to have unit norm, and we use diagonal preconditioning. The results reported in the right plot of Figure 5 show the dramatic improvements of the flexible methods over restarted GMRES with $m = 20$. These results are consistent with the other experiments on smaller matrices reported earlier.

We end this section with some comments on the behavior of the truncated schemes. By comparing the results in Figures 2, 4, and 5, we see that, except for Example 8.2, the curves of the flexible truncated variants quickly abandon the optimal curve of the flexible method, confirming that orthogonalization with respect to the previous inner starting vectors is crucial to obtain fast convergence. Among the choices we have analyzed, however, we see that maintaining orthogonality with respect to the two previous starting vectors ($k_t = 2$) seems to provide the closest to optimal convergence curve. Although in exact arithmetic $k_t = 1$ is sufficient to ensure termination, a larger value of k_t seems to pay off in our tests at the cost of a smaller inner subspace dimension (m_t). Not surprisingly, however, the performance for $k_t = 4$ indicates that a value of m_t that is too small may slow down convergence.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

9. Conclusion. We have analyzed a class of flexible inner-outer iterative Krylov subspace methods in an unified manner. These are Krylov methods where the preconditioner is itself a (preconditioned) Krylov method. We have shown convergence of this class of methods by showing that the subspace from where the approximation is chosen keeps growing. This convergence is guaranteed as long as there is no stagnation in the inner iterations.

We have shown experimentally (and in some cases theoretically) that these methods can compete favorably with the standard restarted methods such as GMRES(m).

In the case of truncated methods, our theory indicates one way in which to choose the vectors to keep in order to guarantee convergence. Our experimental evidence confirms the effectiveness of this choice. These truncated methods appear to perform better than the standard restarted versions using the same amount of storage, and in some cases they are almost as good as the untruncated flexible method (which requires more storage).

Further analysis is needed to determine which vectors in the outer basis are the important ones to keep, allowing us to use the storage for more vectors in the inner basis; cf. [7].

Acknowledgments. We are indebted to Eric de Sturler for valuable comments on an earlier version of the manuscript. In particular, they led to Remark 5.3 and to a better proof of Proposition 7.2. We thank the editor, Martin Gutknecht, for his superb job in handling the paper. His questions and observations helped improve our presentation. We also benefited from comments and suggestions from T. Faraj and J. Mas and from the referees.

REFERENCES

[1] O. Axelsson and P. S. Vassilevski, *A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 625–644.

[2] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel, *Adaptively preconditioned GMRES algorithms*, SIAM J. Sci. Comput., 20 (1998), pp. 243–269.

[3] P. N. Brown, *A theoretical comparison of the Arnoldi and the GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.

[4] B. Carpentieri, I. S. Duff, and L. Giraud, *Multilevel Preconditioning Techniques for the Solution of Large Dense Linear Systems in Electromagnetism*, Technical report, CERFACS, Toulouse, France, in preparation.

[5] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1998), pp. 43–66.

[6] E. de Sturler, *Nested Krylov methods based on GCR*, J. Comput. Appl. Math., 67 (1996), pp. 15–41.

[7] E. de Sturler, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp. 864–889.

[8] I. S. Duff, R. G. Grimes, and J. G. Lewis, *Sparse matrix test problems*, ACM Trans. Math. Software, 15 (1989), pp. 1–14.

[9] M. Eiermann and O. Ernst, *Geometric aspects in the theory of Krylov subspace methods*, Acta Numer., 2001, pp. 251–312.

[10] M. Eiermann, O. Ernst, and O. Schneider, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 261–292.

[11] S. C. Eisenstat, J. M. Ortega, and C. T. Vaughan, *Efficient polynomial preconditioning for the conjugate gradient method*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 859–872.

[12] H. C. Elman, O. G. Ernst, and D. P. O'Leary, *A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations*, SIAM J. Sci. Comput., 23 (2001), pp. 1291–1315.

[13] J. Erhel, K. Burrage, and B. Pohl, *Restarted GMRES preconditioned by deflation*, J. Comput. Appl. Math., 69 (1996), pp. 303–318.

[14] V. Faber, W. Joubert, E. Knill, and T. Manteuffel, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729.
[15] R. W. Freund, *On conjugate gradient type methods and polynomial preconditioners for a class of complex non-Hermitian matrices*, Numer. Math., 57 (1990), pp. 285–312.
[16] R. W. Freund and N. M. Nachtigal, *QMR: A quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.
[17] G. H. Golub and Q. Ye, *Inexact preconditioned conjugate gradient method with inner-outer iteration*, SIAM J. Sci. Comput., 21 (1999), pp. 1305–1320.
[18] A. Greenbaum, *Comparison of QMR-type methods with GMRES*, presented at Mathematical Journey through Analysis, Matrix Theory and Scientific Computation: A Conference on the Occasion of Richard S. Varga's 70th Birthday, Kent State University, Kent, Ohio, March 24–26, 1999.
[19] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, Frontiers Appl. Math. 17, SIAM, Philadelphia, 1997.
[20] O. J. Johnson, C. A. Micchelli, and G. Paul, *Polynomial preconditioners for conjugate gradient calculations*, SIAM J. Numer. Anal., 20 (1983), pp. 362–376.
[21] W. Joubert, *A robust GMRES-based adaptive polynomial preconditioning algorithm for non-symmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439.
[22] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Addison–Wesley, Reading, MA, 1984.
[23] *MATLAB User's Guide*, The MathWorks, Inc., Natick, MA, 1998.
[24] *Matrix Market*, http://math.nist.gov/MatrixMarket (15 July 2002).
[25] J. A. Meijerink and H. van der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., 31 (1977), pp. 148–162.
[26] R. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.
[27] Y. Notay, *Flexible conjugate gradient*, SIAM J. Sci. Comput., 22 (2000), pp. 1444–1460.
[28] Y. Saad, *Practical use of polynomial preconditionings for the conjugate gradient method*, SIAM J. Sci. Stat. Comput., 6 (1985), pp. 865–881.
[29] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.
[30] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.
[31] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
[32] Y. Saad and K. Wu, *DQGMRES: A direct quasi-minimal residual algorithm based on incomplete orthogonalization*, Numer. Linear Algebra Appl., 3 (1996), pp. 329–343.
[33] V. Simoncini, *On the convergence of restarted Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 430–452.
[34] G. W. Stewart, *Backward error bounds for approximate Krylov subspaces*, Linear Algebra Appl., 340 (2002), pp. 81–86.
[35] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
[36] D. B. Szyld and J. A. Vogel, *FQMR: A flexible quasi-minimal residual method with inexact preconditioning*, SIAM J. Sci. Comput., 23 (2001), pp. 363–380.
[37] H. A. Van der Vorst and C. Vuik, *GMRESR: A family of nested GMRES methods*, Numer. Linear Algebra Appl., 1 (1994), pp. 369–386.
[38] C. Vuik, *Further experiences with GMRESR*, Supercomputer, 55 (1993), pp. 13–27.
[39] C. Vuik, *New insights in GMRES-like methods with variable preconditioners*, J. Comput. Appl. Math., 61 (1995), pp. 189–204.
[40] J. S. Warsa, M. Benzi, T. A. Wareing, and J. E. Morel, *Preconditioning a mixed discontinuous finite element method for radiation diffusion*, Numer. Linear Algebra Appl., to appear.
[41] L. Zhou and H. F. Walker, *Residual smoothing techniques for iterative methods*, SIAM J. Sci. Comput., 15 (1994), pp. 297–312.

Downloaded 12/30/12 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php