NORTH-HOLLAND

Convergence Properties of Block GMRES and Matrix Polynomials

V. Simoncini*
IMGA-CNR
Modena, Italy

and

E. Gallopoulos*
Department of Computer Engineering and Informatics
University of Patras
26500 Patras
Greece

Submitted by Paul Van Dooren

ABSTRACT

This paper studies convergence properties of the block GMRES algorithm when applied to nonsymmetric systems with multiple right-hand sides. A convergence theory is developed based on a representation of the method using matrix-valued polynomials. Relations between the roots of the residual polynomial for block GMRES and the matrix ε-pseudospectrum are derived, and illustrated with numerical experiments. The role of invariant subspaces in the effectiveness of block methods is also discussed.

1. INTRODUCTION AND SUMMARY

Block iterative methods have been proposed as an attractive approach for handling eigenvalue problems and linear systems [10, 21, 39]. They promise

* Work performed at the Center for Supercomputing Research and Development, University of Illinois at Urbana-Champaign, with support from fellowship 2043597 from the Consiglio Nazionale delle Ricerche, Italy. (Simoncini), and from grant NSF CCR-912010 from the US National Science Foundation (Gallopoulos).

LINEAR ALGEBRA AND ITS APPLICATIONS 247:97–119 (1996)

0024-3795/96/$15.00
SSDI 0024-3795(95)00093-7

favorable convergence properties and effective exploitation of parallel computer architectures [21, 22]. Block methods are natural candidates when systems with multiple right-hand sides have to be solved, an important problem in areas such as chemistry, electromagnetics, structures, and control; see applications described in [5, 24] and references in [34, 36]. In this paper we study convergence properties of a block Krylov solver, that is, the block version of the GMRES method of Saad and Schultz [29], hereafter denoted by BGMRES, when applied to solve the system

$$AX = B, \tag{1.1}$$

where A is a large, nonsymmetric matrix of order n, and $X = [x_1, \dots, x_s]$ and $B = [b_1, \dots, b_s]$ are rectangular matrices of dimension $n \times s$ with $s \leqslant n$.

An essential component of BGMRES is the block Arnoldi procedure; see [1]. To apply BGMRES we are given an initial guess $X^{(0)}$. Since it is possible to deflate otherwise, we will assume that the initial block residual $R^{(0)} := B - AX^{(0)}$ is of full rank. BGMRES generates an approximate solution $X^{(m)}$ over the block Krylov subspace $\mathbb{K}_m(A, R^{(0)}) = \operatorname{span}\{R^{(0)}, AR^{(0)}, \dots, A^{m-1}R^{(0)}\}$. $\mathbb{K}_m$ is the space generated by the matrix $R^{(0)}$, where K_m is the space generated by a vector $r^{(0)}$. The approximate BGMRES solution is $X^{(m)} = X^{(0)} + Z_m$, where Z_m solves the minimization problem

$$\min_{Z \in \mathbb{K}_m} \|R^{(0)} - AZ\|_F, \tag{1.2}$$

with $\|\cdot\|_F$ the Frobenius norm. We refer to [40] for a detailed description of the BGMRES algorithm, and to [11] for an example of use of BGMRES.

Under certain conditions on the starting residuals, each iteration of block Arnoldi generates s linearly independent vectors, leading to finite termination in at most $\lceil n/s \rceil$ iterations. In general, however, n is very large relative to s, and the method will be considered effective only if a good approximation to the exact solution is determined in a number of iterations much smaller than $\lceil n/s \rceil$. Therefore, a convergence analysis of BGMRES is essential. Our starting point is the work of O'Leary [21], who analyzed the convergence behavior of block CG for real symmetric matrices. The first part of our discussion uses extensively the apparatus developed therein. We then apply matrix-valued polynomials in order to further study the convergence of BGMRES. Kent, in [14], used matrix-valued polynomials to describe some block Krylov methods and hence provides useful background. We also refer to work of Vital on block GMRES for nonsymmetric A [40], Sadkane [30], and Simon and Yeremin [33] on block Arnoldi. Another interesting contribution is that of Jia, who discussed the use of block Lanczos-type algorithms, including block Arnoldi, for nonsymmetric eigenvalue problems [12], and thus generalized earlier results [10, 28] and provided useful implementation discussions; and a recent effort by Broyden toward a comprehensive theory for block methods [3].

We remark that the effectiveness of block methods applied to nonsymmetric systems is still under investigation. For example, it was shown in [36] that in the absence of preconditioning and parallel processing a nonblock hybrid method based on GMRES was superior to several other methods including BGMRES. A more recent comparison [34] between block and non-block methods showed that block methods become more attractive when the matrix A is relatively dense, whereas extremely good performance on the Cray YMP/C90 with a preconditioned method based on BGMRES was reported in [15]. See also implementation discussions in [1, 12, 20, 31, 33]. In this paper we do not discuss implementation issues. It is worth noting, however, that the developed theory is useful for understanding the behavior of other block Krylov methods (see [34]) as well as for designing techniques that improve the performance of BGMRES [see [35]).

The remainder of this paper is organized as follows. We start Section 2 with a result from [40], which shows that for a fixed number of iterations, BGMRES achieves a residual at most as large as if GMRES were to be applied independently to each right-hand side. We then prove that under certain conditions, asymptotic convergence depends on the distribution of $n-s$ clustered eigenvalues of A (Theorem 2.2). In Section 3 we introduce a representation of BGMRES based on matrix-valued polynomials. This representation is used to obtain bounds for the norm of the block residual (Theorems 3.1 and 3.2). We then study some properties of these polynomials. The roots of the BGMRES polynomial are shown to be eigenvalues of a rank-s modification of the orthogonal section of A on a block Krylov subspace (Theorem 3.3). This fact leads to interesting new relations between the roots and the ε-pseudospectrum of the matrix. In Section 4 we discuss briefly the issue of finite termination and its dependence on invariant subspaces of A, and provide concluding remarks.

The following notation is used. Whenever A is diagonalizable, (u_j, λ_j) $(j = 1, \dots, n)$ denotes the eigenvector-eigenvalue pairs, U the matrix of eigenvectors, and Λ the set of eigenvalues of A. Assuming that the block Krylov subspace $\mathbb{K}_m$ is of full rank, we denote by m the block dimension of the bases $\mathscr{V}_m = [V_1, \dots, V_m]$ of $\mathbb{K}_m$ with $V_i \in \mathbb{R}^{n \times s}$; $\mathbb{P}_m$ denotes the space of scalar-valued polynomials of degree at most m, and $\overline{\mathbb{P}}_m$ the subspace of $\mathbb{P}_m$ with polynomials p_k satisfying $p_k(0) = 1$. The inner product of $x, y \in \mathbb{R}^n$ is denoted by $(x, y) = x^T y$. Unless specified otherwise, the Frobenius norm is used for matrices and vectors; therefore, we omit the subscript from $\|\cdot\|_F$. The condition number of a square matrix A is denoted by $\kappa(A)$. For rectangular matrices we also use the norm $\|\cdot\|_\psi$ defined as $\|R\|_\psi := \max_{j=1,\ldots,s}(\|r_j\|)$ with $R = [r_1, \ldots, r_s]$. I_s and 0_s denote the identity and zero matrices of order s. Finally, superscript $*$ denotes the complex conjugate (transpose, when applied to vectors or matrices).

2. RESIDUAL EVALUATION

From Equation (1.2) we see that given m, BGMRES computes the approximate solution $X^{(m)}$ such that the $\|R^{(0)} - AZ\|$ is minimized over $Z \in \mathbb{K}_m$. If the minimum is achieved for $Z_m \in \mathbb{K}_m$, we can write $R^{(m)} = R^{(0)} - AZ_m$. From the definition of the Frobenius norm, BGMRES minimizes $\mathrm{tr}[(R^{(m)})^T R^{(m)}]$ with $R^{(m)} = [r_1^{(m)}, \ldots, r_s^{(m)}]$ and tr denoting the trace.

Vital shows that $\|R^{(m)}\|_\psi$ is always smaller than the largest residual norm obtained after solving each system separately using GMRES with the same Krylov subspace dimension m [40].

THEOREM 2.1 [40]. *Let $\mathbb{K}_m(A, R^{(0)})$ be the Krylov subspace generated by $R^{(0)}$, and $K_m(A, r_j^{(0)})$ generated by $r_j^{(0)}$. Then*

$$\min_{Z \in \mathbb{K}_m} \|R^{(0)} - AZ\|_\psi \leqslant \max_{j=1,\ldots,s} \min_{z_j \in K_m} \|r_j^{(0)} - Az_j\|. \tag{2.1}$$

We will provide sharper estimates of residual behavior by assuming that A is diagonalizable, with all eigenvalues in $\mathbb{C}^+$. In particular we assume that $0 < \Re(\lambda_1) \leqslant \cdots \leqslant \Re(\lambda_n)$. Unless stated otherwise, for the remainder of this paper, we will assume that the rank of $\mathbb{K}_m$ is ms. We will be referring to this as the *full-rank assumption* and note that it is appropriate for convergence studies, though any implementation of the algorithm must take account of the possible rank loss [12, 28, 33]. The full-rank assumption is guaranteed to hold if $\sum_{j=1}^s p_j(A) r_j^{(0)} \neq 0$ for all $p_j \in P_{m-1}$ [4].

Since A is real, its eigenvalues are symmetric with respect to the real axis. We can thus construct an ellipse $F_s(c, e, a)$ which contains $\lambda_s, \ldots, \lambda_n$ and has center c on the real axis, foci $c \pm e$, and major semiaxis a [16]. Furthermore e^2 is real, and the ellipse is symmetric with respect to the real axis.

Let T_m denote the Chebyshev polynomial of the first kind and degree m. The following lemmas will be useful for the main result of the section. For a proof of the first one see for example [26].

LEMMA 2.1. *Let F denote the ellipse $F(0,1,a)$, including its interior, centered at the origin with major semiaxis a and eccentricity one. Then*

$$\max_{\zeta \in F} |T_m(\zeta)| = T_m(a).$$

The next lemma follows easily after manipulating algebraic expressions for T_m.

LEMMA 2.2. *Let a, e be both real (imaginary) with positive real (imaginary) part; let also $c \in \mathbb{R}^+$ with $|e| < c$. Then*

$$\frac{T_m(a/e)}{|T_m(c/e)|} \leqslant 2\left(\frac{a/e + \sqrt{(a/e)^2 - 1}}{|c/e| + \sqrt{(c/e)^2 - 1}}\right)^m.$$

LEMMA 2.3. *Let ∂F be the boundary of the ellipse F as defined in Lemma 2.1, with radius $\rho > 1$, so that $a = \frac{1}{2}(\rho + 1/\rho)$, Then, for $m > 1$,*

$$\left|\frac{T_{m+1}(\zeta)}{T_m(\zeta)}\right| \leqslant \frac{\rho^2 + 1}{\rho - 1} \qquad \text{for} \quad \zeta \in \partial F. \tag{2.2}$$

Proof. For $\zeta \in \partial F$ we have

$$|T_m(\zeta)| \leqslant \max_{y \in \partial F} |T_m(y)| = \frac{1}{2}\left(\rho^m + \frac{1}{\rho^m}\right)$$

and

$$|T_m(\zeta)| \geqslant \min_{y \in \partial F} |T_m(y)| = \frac{1}{2}\left(\rho^m - \frac{1}{\rho^m}\right).$$

Then

$$\left|\frac{T_{m+1}(\zeta)}{T_m(\zeta)}\right| \leqslant \frac{\rho^{m+1} + 1/\rho^{m+1}}{\rho^m - 1/\rho^m} \leqslant \frac{\rho^{m+1} + 1}{\rho^m - 1}.$$

Consider the real function $\Theta(t) = (\rho^{t+1} + 1)/(\rho^t - 1)$ with $t \in [1, \infty)$. It is easy to shows that when $\rho > 1$, Θ is a monotonically decreasing function of t, so that $\Theta(t) \leqslant \Theta(1)$ for $t \in [1, \infty)$, which proves (2.2). ∎

The next lemma is a straightforward extension of [21, Lemma 5, part (a)]. Let $F_s(c, e, a)$ be as above, and define $\tilde{R}$ to be $R^{(0)}$ with the sth column omitted.

LEMMA 2.4. *Let U be the matrix of linearly independent eigenvectors of A, and let $\{\lambda_i\}_{i=1,\dots,n}$ denote the corresponding eigenvalues. Let $F_s(c, e, a)$ contain $\lambda_s, \dots, \lambda_n$, and define*

$$L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

to be an $n \times (s-1)$ matrix such that UL is an orthogonal basis for $\operatorname{span}\{A\tilde{R}\}$, with L_1 of order $s-1$ and nonsingular. Define $\tilde{L} = (\tilde{L}_{i,j})$ as $\tilde{L} = L_2 L_1^{-1}$. Then for each $j = 1, \dots, s-1$ there is a vector $g_j \in \operatorname{span}\{A\tilde{R}, \dots, A^m\tilde{R}\}$ such that

$$g_j = u_j + \sum_{i=s}^{n} \delta_{i,j} u_i$$

with $\delta_{i,j} = \tilde{L}_{i,j} T_{m-1}(\zeta_i)/T_{m-1}(\zeta_j)$, and $\zeta_l = (c - \lambda_l)/e$.

Proof. We first observe that $ULL_1^{-1} \in \operatorname{span}\{A\tilde{R}\}$. For $j = 1, \dots, s-1$, define the vector $g_j = \mathcal{T}_j(A)U(ULL_1^{-1})_{:,j}$, where

$$\mathcal{T}_j(\zeta) := T_{m-1}(\zeta)/T_{m-1}(\zeta_j),$$

and the subscript $:, j$ indicates the jth column of the matrix. Since $ULL_1^{-1} = U[I_{s-1}, \tilde{L}^T]^T$ and $\mathcal{T}_j(\zeta_j) = 1$, it follows that $g_j = u_j + \sum_{i=s}^{n} \mathcal{T}_j(\zeta_i) u_i \tilde{L}_{i,j}$. ∎

We next characterize the norm of the individual residual vectors after approximating the solution from $\mathbb{K}_m(A, R^{(0)})$. As in the symmetric case analyzed in [21], we will provide a bound in terms of Chebyshev polynomials. To simplify notation we derive the result for the sth residual vector. The result also holds for other residuals, provided the subscript s is changed accordingly.

THEOREM 2.2. *Let A be a diagonalizable matrix of order n with its eigenvalues in $\mathbb{C}^+$. Let the ellipse $F_s(c, e, a)$ contain $\lambda_s, \dots, \lambda_n$ and be such that $c \in \mathbb{R}^+$ is its center, e^2 is real, and the origin is exterior to the ellipse. Let $r_s^{(m)} = b_s - Ax_s^{(m)}$ be the residual of the sth system of (1.1) when approximated using the* BGMRES *algorithm with starting block $R^{(0)}$. Then the sth residual satisfies*

$$\left(r_s^{(m)}, r_s^{(m)}\right) \leqslant S \left(\frac{a/e + \sqrt{(a/e)^2 - 1}}{|c/e| + \sqrt{(c/e)^2 - 1}} \right)^{2m}, \tag{2.3}$$

where S does not depend on m.

Proof. BGMRES computes $X^{(m)}$ so that the trace $\operatorname{tr}[(R^{(m)})^T R^{(m)}]$ is minimized. This is achieved by minimizing $\|r_j^{(m)}\|$ $(1 \leqslant j \leqslant s)$. The residual $r_s^{(m)}$ can be written as

$$r_s^{(m)} = r_s^{(0)} + \sum_{j=1}^{s} t_{m,j}(A) A r_j^{(0)}, \tag{2.4}$$

where $t_{m,j} \in \mathbb{P}_{m-1}$. The theorem will be proved by selecting polynomials $t_{m,j}$ which provide an appropriate upper bound for $r_s^{(m)}$.

Let $r_s^{(0)} = U\xi$, with $\xi = [\xi_1 \ \cdots \ \xi_n]^T$, be the expansion of the initial residual for the sth system in eigenvectors of A. Let also the polynomial $q_{m,s}$ be defined by $q_{m,s}(\lambda) = 1 + t_{m,s}(\lambda)\lambda$, so that

$$q_{m,s}(\lambda) = \frac{T_m\left(\dfrac{c - \lambda}{e}\right)}{T_m\left(\dfrac{c}{e}\right)} \tag{2.5}$$

with c, e parameters of the ellipse $F_s(c, e, a)$. The remaining $s - 1$ polynomials are constructed by combining vectors $\{g_1, \dots, g_{s-1}\}$ (cf. Lemma 2.4) as follows:

$$\sum_{j=1}^{s-1} t_{m,j}(A) A r_j^{(0)} = -\sum_{j=1}^{s-1} q_{m,s}(\lambda_j) \xi_j g_j;$$

hence $r_s^{(m)} = q_{m,s}(A)U\xi - \sum_{j=1}^{s-1} q_{m,s}(\lambda_j)\xi_j g_j$. To simplify notation, we will omit the subscripts in $q_{m,s}$. From the definition of g_j and after some algebra, we obtain

$$r_s^{(m)} = \sum_{i=s}^{n} \left[q(\lambda_i)\xi_i - \sum_{j=1}^{s-1} q(\lambda_j)\xi_j \delta_{i,j} \right] u_i. \tag{2.6}$$

Hence the residual to be minimized is expressed as a linear combination of $\{u_s, \dots, u_n\}$. It follows from (2.6) that

$$\left(r_s^{(m)}, r_s^{(m)}\right) = \left(\sum_{i=s}^{n} q(\lambda_i)\xi_i u_i, \sum_{k=s}^{n} q(\lambda_k)\xi_k u_k \right) \tag{2.7}$$

$$+ \left(\sum_{k=s}^{n} \sum_{l=1}^{s-1} q(\lambda_l)\xi_l \delta_{k,l} u_k, \sum_{i=s}^{n} \sum_{j=1}^{s-1} q(\lambda_j)\xi_j \delta_{i,j} u_i \right) \tag{2.8}$$

$$-2\Re\left(\sum_{i=s}^{n} q(\lambda_i)\xi_i u_i, \sum_{k=s}^{n} \sum_{l=1}^{s-1} q(\lambda_l)\xi_l \delta_{k,l} u_k \right). \tag{2.9}$$

We next bound each of the terms (2.7), (2.8), and (2.9).

(2.7): From (2.5) and Lemma 2.1 we obtain

$$\left| \left(\sum_{i=s}^{n} q(\lambda_i)\xi_i u_i, \sum_{k=s}^{n} q(\lambda_k)\xi_k u_k \right) \right| \leqslant \frac{T_m^2(a/e)}{|T_m^2(c/e)|} \sum_{k,i=s}^{n} |(\xi_k u_k, \xi_i u_i)|$$

$$= S_1 \frac{T_m^2(a/e)}{|T_m^2(c/e)|},$$

where S_1 is the value obtained from the summation term.

(2.8): From (2.5) and the definition of $\delta_{i,j}$ it follows that

$$\left(\sum_{k=s}^{n} \sum_{l=1}^{s-1} q(\lambda_l)\xi_l \delta_{k,l} u_k, \sum_{i=s}^{n} \sum_{j=1}^{s-1} q(\lambda_j)\xi_j \delta_{i,j} u_i \right)$$

$$= \sum_{k=s}^{n} \sum_{l=1}^{s-1} \frac{T_m^*\left(\dfrac{c-\lambda_l}{e}\right)}{T_m^*\left(\dfrac{c}{e}\right)} \tilde{L}_{k,l}^* \frac{T_{m-1}^*\left(\dfrac{c-\lambda_k}{e}\right)}{T_{m-1}^*\left(\dfrac{c-\lambda_l}{e}\right)} \xi_l^*$$

$$\times \sum_{i=s}^{n} \sum_{j=1}^{s-1} \frac{T_m\left(\dfrac{c-\lambda_j}{e}\right)}{T_m\left(\dfrac{c}{e}\right)} \tilde{L}_{i,j} \frac{T_{m-1}\left(\dfrac{c-\lambda_i}{e}\right)}{T_{m-1}\left(\dfrac{c-\lambda_j}{e}\right)} \xi_j (u_k, u_i).$$

For each $\lambda_j, j = 1, \dots, s-1$, consider the ellipse of radius ρ_j, centered at the origin and passing through $(c - \lambda_j)/e$. Let $\omega := \max_j(\rho_j^2 + 1)/(\rho_j - 1)$, so that for each λ_j and $m > 1$ it holds that

$$\left| \frac{T_m(\zeta_j)}{T_{m-1}(\zeta_j)} \right| \leqslant \omega, \qquad \zeta_j = \frac{c - \lambda_j}{e}$$

(cf. Lemma 2.3). Using the inequality $T_{m-1}(a/e) \leqslant 2(a/e)T_m(a/e)$ and Lemma 2.1, it follows that

$$\left| \left(\sum_{k=s}^{n} \sum_{l=1}^{s-1} q(\lambda_l)\xi_l \delta_{k,l} u_k, \sum_{i=s}^{n} \sum_{j=1}^{s-1} q(\lambda_j)\xi_j \delta_{i,j} u_i \right) \right|$$

$$\leqslant \frac{4\omega^2 (a/e)^2 T_m^2(a/e)}{|T_m^2(c/e)|} \sum_{k=s}^{n} \sum_{j=1}^{s-1} |\tilde{L}_{k,l}|\,|\xi_l| \sum_{i=s}^{n} \sum_{j=1}^{s-1} |\tilde{L}_{i,j}|\,|\xi_j|\,|(u_k, u_i)|.$$

Hence

$$\left| \left(\sum_{k=s}^{n} \sum_{i=1}^{s-1} q(\lambda_l)\xi_l \delta_{k,l} u_k, \sum_{i=s}^{n} \sum_{j=1}^{s-1} q(\lambda_j)\xi_j \delta_{i,j} u_i \right) \right|$$

$$\leqslant 4\omega^2 \left(\frac{a}{e}\right)^2 S_2 \frac{T_m^2(a/e)}{|T_m^2(c/e)|}.$$

(2.9): From (2.5) and using the definition of $\delta_{i,j}$ it follows that

$$
\begin{aligned}
&2\left|\Re \sum_{k,i=s}^{n}\left(\sum_{l=1}^{s-1} q(\lambda_l)\xi_l\delta_{k,l}u_k, q(\lambda_i)\xi_i u_i\right)\right| \\
&\quad \leqslant 4\omega\left(\frac{a}{e}\right)\frac{T_m^2(a/e)}{\left|T_m^2(c/e)\right|}\sum_{k,i=s}^{n}\sum_{l=1}^{s-1}|\tilde{L}_{k,l}||\xi_l||(u_k, \xi_i u_i)| \\
&\quad \leqslant 4\omega\left(\frac{a}{e}\right)S_3\frac{T_m^2(a/e)}{\left|T_m^2(c/e)\right|}.
\end{aligned}
$$

Combining the above bounds, it follows that

$$
\left(r_s^{(m)}, r_s^{(m)}\right) \leqslant \left[S_1 + 4\omega^2\left(\frac{a}{e}\right)^2 S_2 + 4\omega\left(\frac{a}{e}\right)S_3\right]\frac{T_m^2(a/e)}{\left|T_m^2(c/e)\right|}, \tag{2.10}
$$

where the S_i's do not depend on m. Equation (2.3) follows from Lemma 2.2. ∎

We observe that when there is only a single right-hand side, the theorem reduces to a well-known result [26]. Theorem 2.2 extends to BGMRES and real nonsymmetric matrices the convergence result shown for block CG in [21, Theorem 5]. It shows that when the matrix A is diagonalizable, after applying BGMRES, the norm of each residual is bounded by normalized Chebyshev polynomials defined on an ellipse generated by $n - s$ eigenvalues of A.

The expression (2.10) should be used with care. Recent results of Fischer and Freund have disproved some long-held perceptions about Chebyshev polynomials in $\mathbb{C}$ [7, 9]. In particular, Chebyshev polynomials may not be optimal, so that the left-hand side of the inequality (2.10) can be significantly smaller than the upper bound; in that case Theorem 2.2 will be useless as a characterization of the residual. Fortunately, as discussed in [9], if m is large enough and certain additional conditions hold, then the Chebyshev polynomials are very close to being optimal, in which case the result of Theorem 2.2 provides a reasonable estimate of the error. One example is when the domain enclosed by $F_s(c, e, a)$ is far enough from the origin and e is real. Lower bounds that the value of $\lceil c/e \rceil$ must satisfy for Chebyshev polynomials to be optimal are obtained in [9, Theorem 2].

Observe that the distribution of the excluded eigenvalues affects both the value of ω and the ellipse F_s. For example, the exclusion of the eigenvalues with small real part and large imaginary part will prevent one from having to deal with minimization over ellipses which are too close to the origin and have large radii. This will improve the behavior of the method and could make (2.3) a better bound for the residual behavior.

3. MATRIX-VALUED POLYNOMIALS AND BGMRES

In this section we represent BGMRES using matrix-valued polynomials. We obtain convergence results for the method and spectral information for the matrix.

In order to generate the basis of $\mathbb{K}_m(A, R^{(0)})$, block Arnoldi carries out the following process:

$$V_1 = R^{(0)}\chi_{0,0} \qquad V_{i+1}\,\chi_{i+1,i} = AV_i - \sum_{k=1}^{i} V_k\,\chi_{k,i} \quad i = 1,\dots,m-1, \tag{3.1}$$

where $V_i \in \mathbb{R}^{n\times s}$, $\chi_{k,i} \in \mathbb{R}^{s\times s}$. Here $\chi_{0,0}$ is computed so that the columns of V_1 are orthonormal. We define $\mathscr{H}_i$ to be the upper block Hessenberg matrix $\mathscr{H}_i = \{\chi_{j,k}\}_{j=1,\dots,i+1}^{k=1,\dots,i}$ of dimension $s(i+1)\times si$ [27, 40]. Let also $\tilde{\mathscr{H}}_i := [I_{si}, 0_{si,s}]\mathscr{H}_i$, and let ϕ_k be a matrix-valued polynomial of degree k, defined by $\phi_k(\lambda) = \sum_{i=0}^{k}\lambda^i\xi_i$, with $\xi_i \in \mathbb{R}^{s\times s}$. The solution of (1.1) is approximated by $X^{(m)} = X^{(0)} + \mathscr{V}_m y$, where $y = [y_1^T,\dots,y_m^T]^T$, with $y_i \in \mathbb{R}^{s\times s}$ $(1 \leqslant i \leqslant m)$, chosen appropriately. As shown in [14, Chapter 4], we can use the matrix-valued polynomials ϕ_k to write the relation $A\mathscr{V}_m = \mathscr{V}_{m+1}\mathscr{H}_m$ as

$$\lambda P_{m-1} = P_{m-1}\tilde{\mathscr{H}}_m + \phi_m(\lambda)\,\chi_{m+1,m}E_m, \tag{3.2}$$

where $P_{m-1} := [\phi_0(\lambda), \phi_1(\lambda), \dots, \phi_{m-1}(\lambda)]$ and $E_m = [0_s,\dots,0_s, I_s] \in \mathbb{R}^{s\times ms}$.

It follows that $R^{(m)} = R^{(0)} - A\mathscr{V}_m y$ can be written as

$$\begin{aligned} R^{(m)} &= R^{(0)} - A\sum_{i=0}^{m-1}\phi_i(A)\,R^{(0)}y_{i+1} \\ &= R^{(0)} - \sum_{i=0}^{m-1} A^{i+1}R^{(0)}a_i, \end{aligned}$$

with $a_i \in \mathbb{R}^{s\times s}$ $(1 \leqslant i \leqslant m-1)$.

Let $\overline{\mathbb{P}}_{m,s}$ be the space of matrix-valued polynomials Θ_m of degree not greater than m and order s, such that $\Theta_m(0) = I_s$. In particular, $\Phi_m(\lambda) :=$

$I_s - \sum_{i=0}^{m-1} \lambda^{i+1}\alpha_i$ is a matrix-valued polynomial (of full degree) in $\overline{\mathbb{P}}_{m,s}$. We can write[1]

$$R^{(m)} = \Phi_m(A) \circ R^{(0)} \equiv R^{(0)} - \sum_{i=0}^{m-1} A^{i+1}R^{(0)}a_i.$$

If $\Phi_m(\lambda) \in \overline{\mathbb{P}}_{m,s}$ is the matrix-valued residual polynomial for BGMRES, Φ_m solves the minimization problem

$$\min_{\Theta_m \in \overline{\mathbb{P}}_{m,s}} \|\Theta_m(A) \circ R^{(0)}\|. \tag{3.3}$$

We first consider the case of A diagonalizable. For standard GMRES, the inequality $\|r^{(m)}\| \leqslant \kappa_2(U)\epsilon_m \|r^{(0)}\|$ holds, where $\epsilon_m := \min_{p_m \in \bar{P}_m} \|p_m\|_\Lambda$, $\|p_m\|_\Lambda := \sup_{\lambda \in \Lambda} |p_m(\lambda)|$, and $\kappa_2(U)$ denotes the condition number associated with the spectral norm of U [29]. We shall derive an analogous result for the convergence of BGMRES. We first show that the norm of the residual $R^{(m)} = \Phi_m(A) \circ R^{(0)}$ of any block Krylov method is bounded by $\|\Phi_m(\lambda)\|$ with λ belonging to a certain region of $\mathbb{C}$.

For $\Phi_m \in \overline{\mathbb{P}}_{m,s}$ define $\|\Phi_m\|_\Lambda := \sup_{\lambda \in \Lambda} \|\Phi_m(\lambda)\|$. The following lemma will be needed

LEMMA 3.1. *Let $A = U^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n) U$ be a matrix of order n. Let also $\Psi_m(\lambda) := (q_{j,k}(\lambda))_{j,k=1,\ldots,s}$ with $q_{j,k} \in \mathbb{P}_m$. Then*

$$\|\Psi_m(A)\| \leqslant \kappa(U)\sqrt{n}\, \|\Psi_m\|_{\Lambda,F}. \tag{3.4}$$

Proof. We have

$$\begin{aligned}
\|\Psi_m(A)\|^2 &= \sum_{j,k=1}^{s} \|q_{j,k}(A)\|^2, \\
&\leqslant \kappa^2(U) \sum_{j,k=1}^{s} \|q_{j,k}(\Lambda)\|^2 \\
&= \kappa^2(U) \sum_{j,k=1}^{s} \sum_{i=1}^{n} |q_{j,k}(\lambda_i)|^2 = \kappa^2(U) \sum_{i=1}^{n} \|\Psi_m(\lambda_i)\|^2 \\
&\leqslant \kappa^2(U)\, n \sup_{\lambda \in \Lambda} \|\Psi_m(\lambda)\|^2. \qquad \blacksquare
\end{aligned}$$

[1] The notation $\circ$ is attributed to Gragg [14].

THEOREM 3.1. *Let* $R^{(m)} = \Phi_m(A) \circ R^{(0)}$ *with* $R^{(0)}, R^{(m)} \in \mathbb{R}^{n \times s}$, $A \in \mathbb{R}^{n \times n}$, $A = U^{-1} \operatorname{diag}(\lambda_1, \dots, \lambda_n) U$, *and* $\Phi_m \in \overline{\mathbb{P}}_{m,s}$. *Then*

$$\|R^{(m)}\| \leqslant \kappa(U)\sqrt{n}\, \|\Phi_m\|_\Lambda \, \|R^{(0)}\|. \tag{3.5}$$

Proof. Rewrite Φ_m as $\Phi_m(\lambda) = \sum_{i=0}^m \lambda^i \omega_i^T$, where $\omega_i := (w_{j,k}^{(i)})$ for $i = 0, \dots, m$, and $\Phi_m(0) = I_s$. Hence the (j,k)th element of $\Phi_m(\lambda)$ is the polynomial $q_{j,k}(\lambda) := \sum_{i=0}^m \lambda^i w_{k,j}^{(i)}$. Then

$$R^{(m)} = \left[\sum_{i=0}^{m} A^i \sum_{j=1}^{s} w_{1,j}^{(i)} r_j^{(0)}, \dots, \sum_{i=0}^{m} A^i \sum_{j=1}^{s} w_{s,j}^{(i)} r_j^{(0)} \right],$$

that is, $R^{(m)} = [\sum_{j=1}^s q_{j,1}(A) r_j^{(0)}, \dots, \sum_{j=1}^s q_{j,s}(A) r_j^{(0)}]$. Denote by $\operatorname{vec}(R^{(m)})$ the ns-vector having as elements the columns $r_1^{(m)}, \dots, r_s^{(m)}$ of $R^{(m)}$. We can thus write

$$\operatorname{vec}(R^{(m)}) = \Psi_m(A) \operatorname{vec}(R^{(0)}),$$

where $\Psi_m(A)$ is a matrix of order ns, with components $q_{k,j}(A)$. From Lemma 3.1, $\|\operatorname{vec}(R^{(m)})\| \leqslant \kappa(U)\sqrt{n}\, \|\Psi_m\|_\Lambda \, \|\operatorname{vec}(R^{(0)})\|$. Furthermore, $\Psi_m(\lambda)$ coincides with $\Phi_m(\lambda)$ and $\|\operatorname{vec}(R^{(m)})\|^2 = \operatorname{tr}[(R^{(m)})^T R^{(m)}]$, so that (3.5) follows. ∎

For the remainder of this section we no longer assume that A is diagonalizable. For highly nonnormal matrices, $\kappa(U)$ can be very large. For such cases, an attractive alternative to Theorem 3.1 uses the notion of ε-pseudospectrum of A [18]. This is defined as $\Lambda_\varepsilon(A) := \{\zeta \in \mathbb{C} : \|(\zeta I - A)^{-1}\| \geqslant \varepsilon^{-1}\}$ with $\|\cdot\|$ any norm. Let L_ε be the arc length of the boundary of $\Lambda_\varepsilon(A)$.

LEMMA 3.2. *Let* $\Psi_m(\lambda) = (q_{j,k}(\lambda))_{j,k=1,\dots,s}$ *with* $q_{j,k} \in \mathbb{P}_m$ *and* $A \in \mathbb{R}^{n \times n}$. *Then*

$$\|\Psi_m(A)\| \leqslant \frac{L_\varepsilon}{2\pi\varepsilon} \|\Psi_m\|_{\Lambda_\varepsilon, F}. \tag{3.6}$$

Proof. Let Γ be the contour or the union of contours enclosing $\Lambda_\varepsilon(A)$. For each $j, k = 1, \dots, s$, consider the integral representation $q_{j,k}(A) = (1/2\pi\imath)\int_\Gamma q_{j,k}(\lambda)(\lambda I - A)^{-1} \, d\lambda$. Then we can write

$$\Psi_m(A) = \frac{1}{2\pi\imath} \int_\Gamma \Psi_m(\lambda) \otimes (\lambda I - A)^{-1} \, d\lambda,$$

where $\otimes$ indicates the Kronecker product. From $\|\Psi_m(\lambda) \otimes (\lambda I - A)^{-1}\| = s^{-1}\|\Psi_m(\lambda)\|$ for $\lambda \in \Gamma$, it follows that

$$\|\Psi_m(A)\| \leqslant \frac{1}{2\pi\varepsilon} \|\Psi_m\|_{\Lambda_\varepsilon} \int_\Gamma |d\lambda|.$$

Using the definition of L_ε the result follows. ∎

The next result generalizes Theorem 3.1 to nondiagonalizable matrices. The proof is similar except that it uses the bound (3.6) instead of (3.4).

THEOREM 3.2. *Let* $R^{(m)} = \Phi_m(A) \circ R^{(0)}$ *with* $R^{(0)}, R^{(m)} \in \mathbb{R}^{n \times s}$, $A \in \mathbb{R}^{n \times n}$, *and* $\Phi_m \in \overline{\mathbb{P}}_{m,s}$. *Then*

$$\|R^{(m)}\| \leqslant \frac{L_\varepsilon}{2\pi\varepsilon} \|\Phi_m\|_{\Lambda_\varepsilon} \|R^{(0)}\|. \tag{3.7}$$

For the case of BGMRES, it follows from (3.3) that

$$\|R^{(m)}\| \leqslant \left(\frac{L_\varepsilon}{2\pi\varepsilon} \|R^{(0)}\| \right) \min_{\Theta_m \in \overline{\mathbb{P}}_{m,s}} \sup_{\lambda \in \Lambda_\varepsilon} \|\Theta_m(\lambda)\|.$$

We illustrate some of our points using the tridiagonal Toeplitz matrix of order $n = 200$ given by [19]

$$A = \begin{pmatrix} 5.1 & 3 & 0 & 0 & \cdots & 0 \\ 2 & 5.1 & 3 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 2 & 5.1 & 3 \\ 0 & \cdots & 0 & 0 & 2 & 5.1 \end{pmatrix}. \tag{3.8}$$

Pseudospectra of Toeplitz matrices have been studied in detail in [25], while the particular example can be found in [19]. For small ε, the ε-pseudospectrum of A is enclosed by the ellipse formed by mapping the unit circle using the symbol function $f(z) = 2z^{-1} + 5.1 + 3z$. It is worth noting that the ellipse is contained in the domain $[0.1, 10.1] \times [-1, 1]$.

Figure 1 depicts the values $\|\Phi_{25}(\lambda)\|$ in a region containing the ε-pseudospectrum $\Lambda_\varepsilon(A)$ of the matrix above. The picture shows that the

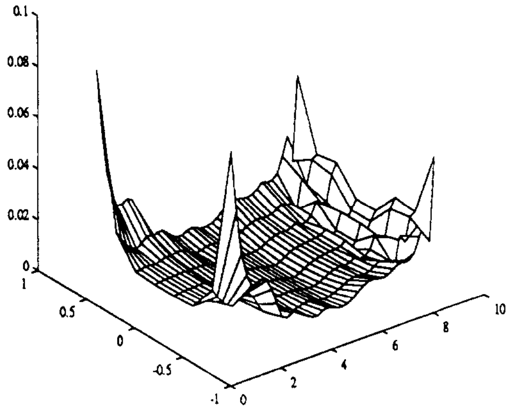

FIG. 1. Norm of BGMRES matrix-valued polynomial $\Phi_{25}(\lambda)$ for λ taking values in the ε-pseudospectrum $\Lambda_\varepsilon(A)$. $n = 200$. Random starting vectors $[r_1^{(0)}, r_2^{(0)}]$.

polynomial has small norm values on $\Lambda_\varepsilon(A)$; however, the inequality (3.5) is not sharp, and $\|R^{(m)}\|$ may be small even for large values of the norm of Φ_m.

Figure 2(a) shows the level curves $\|\Phi_{25}(\lambda)\| = \tau$, for $\tau = 0.02, 0.04$, of the polynomial used in Figure 1. Furthermore, the roots of Φ_{25} (generated with BGMRES applied to $[r_1^{(0)}, r_2^{(0)}]$) are shown in Figure 2(b). It thus seems that the roots and lemniscates of the BGMRES polynomial provide a feeling for the ε-pseudospectrum. These issues are discussed in greater detail next.

3.1. *Roots of the* BGMRES *Polynomial and ε-Pseudospectrum*

From the full-rank assumption, $\det(\Phi_m(\lambda))$ has ms zeros. From now on, we will be referring to these (latent roots) simply as roots of Φ_m.

We first note that $\tilde{\mathscr{H}}_m$ is the matrix representation of the section of A in $\mathbb{K}_m(A, R^{(0)})$ relative to the (orthogonal) basis $\mathscr{V}_m$. Under certain conditions, the eigenvalues of $\tilde{\mathscr{H}}_m$ (also called Ritz values of A), approximate some eigenvalues of A [12, 30]. On the other hand, the roots of the GMRES polynomial coincide with the eigenvalues of a rank-1 modification of the orthogonal section of A with respect to the basis of the Krylov subspace generated by Arnoldi [17, 23, 32, 37].

Therefore, we first show that the roots of the BGMRES polynomial Φ_m coincide with the eigenvalues of a modification of the block Hessenberg matrix $\tilde{\mathscr{H}}_m$. We then show relations between the roots of the BGMRES polynomial and the ε-pseudospectrum. Our results also corroborate, for the general ($s \geqslant 1$) case, experimental observations made in [19, Example 3] for GMRES.

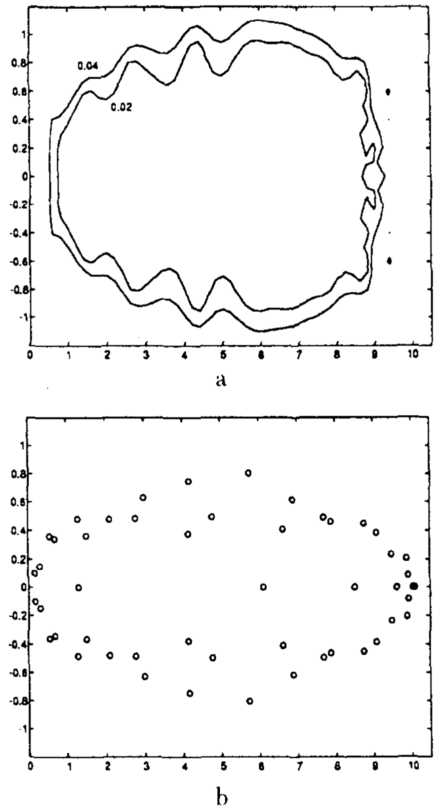

FIG. 2. Matrix (3.8) of size $n = 200$ and random starting vectors $[r_1^{(0)}, r_2^{(0)}]$: (a) level curves $\|\Phi_{25}(\lambda)\| = \tau$ for $\tau = 0.02, 0.04$; (b) roots of BGMRES polynomial Φ_{25}.

Denote by $h_{m+1} := [0_s, \ldots, 0_s, \chi_{m+1,m}]$ the last s rows of $\mathscr{H}_m$, with $\chi_{m+1,m} \in \mathbb{R}^{s\times s}$. First note that y is chosen to minimize $\|E_1^T\chi_{0,0} - \mathscr{H}_m y\|$. Let $S_{m+1} := \mathscr{H}_m y - E_1^T\chi_{0,0}$, and write $S_{m+1} = [S_m^T, \sigma_{m+1}^T]^T$ with $\sigma_{m+1} \in \mathbb{R}^{s\times s}$, $\sigma_{m+1} := \chi_{m+1,m} y_m$. Then from $\mathscr{H}_m^T S_{m+1} = 0$ it follows that

$$\tilde{\mathscr{H}}_m^T S_m + h_{m+1}^T h_{m+1} y = 0. \tag{3.9}$$

THEOREM 3.3. *The roots of the* BGMRES *matrix-valued residual polynomial $\Phi_m(\lambda)$ coincide with the eigenvalues of the matrix $\tilde{\mathscr{H}}_m + \mathscr{L}_m$, where $\tilde{\mathscr{H}}_m = \mathscr{V}_m^T A \mathscr{V}_m$, $\mathscr{V}_m$ is the orthogonal basis for the block Krylov subspace $\mathbb{K}_m(A, R^{(0)})$ generated with block Arnoldi, $\mathscr{L}_m := \tilde{\mathscr{H}}_m^{-T} h_{m+1}^T h_{m+1}$, and h_{m+1} are the last s rows of the matrix $\mathscr{H}_m = \mathscr{V}_{m+1}^T A \mathscr{V}_m$.*

Proof. The residual polynomial Φ_m is determined by $\Phi_m(\lambda) = P_m S_{m+1}$. Recalling the definition of S_{m+1},

$$\phi_m(\lambda) = \left[\Phi_m(\lambda) - P_{m-1} S_m\right] \sigma_{m+1}^{-1}.$$

Note that σ_{m+1} is nonsingular, since $\chi_{m+1,m}$ is nonsingular and y_m is the product of Givens rotations. We can thus rewrite the recursion (3.2) as

$$\lambda P_{m-1} = P_{m-1}\left(\tilde{\mathscr{H}}_m - S_m \sigma_{m+1}^{-1} \chi_{m+1,m} E_m\right)$$
$$+ \Phi_m(\lambda) \sigma_{m+1}^{-1} \chi_{m+1,m} E_m. \tag{3.10}$$

Using the definition of σ_{m+1} and relation (3.9), we see that

$$\begin{aligned} -S_m \sigma_{m+1}^{-1} \chi_{m+1,m} E_m &= -S_m y_m^{-1} E_m \\ &= \tilde{\mathscr{H}}_m^{-T} h_{m+1}^T h_{m+1} y y_m^{-1} E_m \\ &= \tilde{\mathscr{H}}_m^{-T} h_{m+1}^T \chi_{m+1,m} y_m y_m^{-1} E_m = \mathscr{L}_m. \end{aligned}$$

We replace (3.10) with $\lambda P_{m-1} = P_{m-1}(\tilde{\mathscr{H}}_m + \mathscr{L}_m) + \Phi_m(\lambda) y_m^{-1} E_m$. The result follows after noticing that $\det[\lambda I - (\tilde{\mathscr{H}}_m + \mathscr{L}_m)] = 0$ if and only if $\det(\Phi_m(\lambda) = 0$. ∎

Since from Theorem 3.3 the roots of $\Phi_m(\lambda) = 0$ satisfy

$$\left(\tilde{\mathscr{H}}_m + \mathscr{L}_m\right) z = \lambda z, \tag{3.11}$$

they also satisfy $\mathscr{H}_m^T \mathscr{H}_m z = \lambda \tilde{\mathscr{H}}_m^T z$. It is worth noting that Freund used this generalized eigenvalue problem to compute the roots of scalar residual polynomials [8]; see [23] for related results.

The rank of $\mathscr{L}_m$ is s; hence the roots of the BGMRES residual polynomial are Ritz values of a rank-s modification of A.

COROLLARY 3.1. *Let $L_m := \mathscr{V}_m \mathscr{L}_m \mathscr{V}_m^T$, let $\{\lambda, z\}$ be a solution of* (3.11), *and let $f := \mathscr{V}_m z$ with $z = [z_1^T, \ldots, z_m^T]^T \in \mathbb{R}^{ms}$. Then*

$$\|(A + L_m) f - \lambda f\| = \|\chi_{m+1,m} z_m\|. \tag{3.12}$$

Proof. We first note that

$$(A + L_m)\mathscr{V}_m = \mathscr{V}_m\left(\tilde{\mathscr{H}}_m + \mathscr{L}_m\right) + V_{m+1}\chi_{m+1,m}E_m^T.$$

Hence

$$(A + L_m)f = \mathscr{V}_m\left(\tilde{\mathscr{H}}_m + \mathscr{L}_m\right)z + V_{m+1}\chi_{m+1,m}z_m$$

$$= \lambda f + V_{m+1}\chi_{m+1,m}z_m.$$

The result follows from the orthogonality of the columns of V_{m+1}. ∎

From (3.11) and the definition of ε-approximation eigenpairs [38] we obtain:

COROLLARY 3.2. *Let* $\varepsilon = \|\mathscr{L}_m\|$. *Then any solution* $\{\lambda, z\}$ *of* (3.11) *is an* ε*-approximation eigenpair of* $\tilde{\mathscr{H}}_m$.

Since we assumed that $\mathbb{K}_m(A, R^{(0)})$ maintains full rank, the parameter $\varepsilon = \|\mathscr{L}_m\|$ becomes zero when $m = \lceil n/s \rceil$. Though one can construct examples where until termination the residual is reduced only very slowly or not at all [2, 18], in general one expects the roots of Φ_m to provide a feeling for the ε-pseudospectrum.

From the structure of $\mathscr{L}_m$, we can obtain additional information on the location of the roots of Φ_m relative to the Ritz values of A. In particular, all except the last s columns of $\mathscr{L}_m$ are zero, so that $\mathscr{L}_m = \xi E_m$, where $\mathscr{L}_m \in \mathbb{R}^{ms \times ms}$, with

$$\xi = \left[\,\xi_1^T, \dots, \xi_s^T\,\right]^T \in \mathbb{R}^{ms\times s}, \qquad \xi = \tilde{\mathscr{H}}_m^{-T} E_m^T \chi_{m+1,m}^T \chi_{m+1,m}.$$

We use a generalization of the Gerschgorin circle theorem, due to Feingold and Varga [6]; adapted to our notation and Hessenberg matrices, the theorem reads as follows.

THEOREM 3.4 [6, THEOREM 2]. *For the matrix $\tilde{\mathscr{H}}_m$ of order ms, partitioned into blocks $\chi_{i,j}$ $(1 \leqslant i, j \leqslant m)$ of order s, each of its eigenvalues λ satisfies*

$$\left\|(\chi_{j,j} - \lambda I_s)^{-1}\right\|^{-1} \leqslant \sum_{\substack{i=1 \\ i \neq j}}^{j+1} \|\chi_{i,j}\|, \qquad 1 \leqslant j \leqslant m-1, \tag{3.13}$$

$$\left\|(\chi_{m,m} - \lambda I_s)^{-1}\right\|^{-1} \leqslant \sum_{i=1}^{m-1} \|\chi_{i,m}\|. \tag{3.14}$$

For each j $(1 \leqslant j \leqslant m)$, the set of complex numbers satisfying Equations (3.13) and (3.14) is called a Gerschgorin set, and is denoted by $\mathscr{G}_j$. Hence the eigenvalues of $\tilde{\mathscr{H}}_m$ must lie in $\bigcup_{j=1}^{m} \mathscr{G}_j$. The question is: where do the roots of Φ_m lie relative to $\mathscr{G}_j$?

COROLLARY 3.3. *Each root λ of the* BGMRES *residual polynomial Φ_m satisfies at least one of the inequalities below:*

$$\left\|(\chi_{j,j} - \lambda I_s)^{-1}\right\|^{-1} \leqslant \sum_{\substack{i=1 \\ i \neq j}}^{j+1} \|\chi_{i,j}\|, \qquad 1 \leqslant j \leqslant m-1,$$

$$\left\|(\chi_{m,m} + \xi_m - \lambda I_s)^{-1}\right\|^{-1} \leqslant \sum_{i=1}^{m-1} \|\chi_{i,m}\| + \sum_{i=1}^{m-1} \|\xi_i\|.$$

Hence, for both the eigenvalues of $\mathscr{H}_m + \mathscr{L}_m$ (roots of Φ_m) and the eigenvalues of $\tilde{\mathscr{H}}_m$ (Ritz values of A), all but the last Gerschgorin sets are the same. Comparing the sets from Theorem 3.4 with those from Corollary 3.3, we see that a large $\|\xi\|$ will cause the last Gerschgorin sets to become quite different, thus radically changing the regions in which the roots are distributed relative to those of the Ritz values.

4. BLOCK METHODS AND INVARIANT SUBSPACES

Block iterative methods give rise to several interesting theoretical and practical questions. Our analysis suggests that matrix-valued polynomials can be useful tools in studying the convergence of block iterative schemes.

Another interesting issue is the relation between the type and number of right-hand sides and termination properties of the method, assuming exact arithmetic.

In particular, the block solution $X = A^{-1}B$ satisfies $X \in X^{(0)} + \mathbb{K}_m(A, R^{(0)})$ for the smallest m for which $\mathbb{K}_m(A, R^{(0)})$ is an invariant subspace of A. Dropping for notational convenience the superscript from $R^{(0)}$ for each $j = 1, \dots, s$ let $\mathscr{M}_j$ denote the smallest invariant subspace of A containing the starting vector r_j, with $m_j := \det \mathscr{M}_j$. In exact arithmetic, Arnoldi will generate an invariant subspace for r_j in at most m_j iterations. This is also true of block Arnoldi, but since each step of the block algorithm takes at least s matrix-vector multiplications, it is important to be able to form the invariant subspace as efficiently as possible. To see that even in exact arithmetic, and with full rank for R, the block method can be ineffective, let $s = 2$, and choose $r_2 = Ar_1$. Then the block subspace corresponds to the set $\{r_1, Ar_1, Ar_1, A^2r_1, \dots\}$, so that r_2 plays no role, and $m_1 - 1$ iterations of block Arnoldi are needed. In the other extreme, let all $\mathscr{M}_j$ be identical, and let the full-rank assumption hold. Then, after m_1/s iterations, block Arnoldi constructs a basis of $\mathscr{M}_1$. So, depending on the relation between the $\mathscr{M}_j$'s, a large s may or may not help improve convergence. This issue becomes even more complex when restarting has to be used. Recent work of Joubert [13] may provide insights into this problem for the case of BGMRES. Overall, a study which takes into account these concerns would be very desirable.

We thank Professor P. Van Dooren for a discussion on matrix-valued polynomials, Professor E. Lanconelli for suggestions on Lemma 2.3, and the referees for their comments and corrections and for bringing Reference [23] to our attention. We also thank Ms. Merle Levy for several editorial suggestions. An earlier version of this work appeared as CSRD Tech. Report 1287, June 1993.

REFERENCES

1 D. L. Boley and G. H. Golub, The Lanczos-Arnoldi algorithm and controllability, *Systems Control Lett.* 4:317–324 (1984).

2 P. N. Brown, A theoretical comparison of the Arnoldi and GMRES algorithms, *SIAM J. Sci. Statist. Comput.* 12(1):58–78 (Jan. 1991).

3 C. G. Broyden, Block conjugate gradient methods, *Optim. Methods and Software* 2:1–17 (1993).

4 F. Chatelin, *Spectral Approximation of Linear Operators*, Academic, New York, 1983.

5 D. C. Chatfield, M. S. Reeves, D. G. Truhlar, C. Duneczky, and D. W. Schwenke, Complex Generalized Minimal Residual Algorithm for Iterative Solution of Quantum Mechanical Reactive Scattering Equations, Technical report, Univ. of Minnesota Supercomputer Inst., Minneapolis, MN 55415, Aug. 1992.

6 D. G. Feingold and R. S. Varga, Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem, *Pacific J. Math.* 12:1241–1250 (1962).

7 B. Fischer and R. W. Freund, Optimal Chebyshev polynomials on ellipses in the complex plane, in *Approximation Theory VI* (C. K. Chui, L. L. Schumaker, and J. D. Ward, Eds.), Academic, New York, 1989, pp. 251–254.

8 R. W. Freund, Quasi-kernel polynomials and their use in non-Hermitian matrix iterations, *J. Comput. Appl. Math.* 43:135–158 (1992).

9 R. W. Freund and B. Fischer, Chebyshev polynomials are not always optimal, *J. Approx. Theory* 65:261–272 (1991).

10 G. H. Golub and R. Underwood, The block Lanczos method for computing eigenvalues, in *Mathematical Software III* (J. R. Rice, Ed.), Academic, New York, 1977, pp. 364–377.

11 I. M. Jaimoukha and E. M. Kasenally, Krylov subspace methods for solving large Lyapunov equations, *SIAM J. Numer. Anal.* 31(1):227–251 (Feb. 1994).

12 Z. Jia, Generalized block Lanczos methods for large unsymmetric eigenproblems, submitted for publication.

13 W. D. Joubert, On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems, *J. Numer. Linear Algebra Appl.*, submitted for publication.

14 M. D. Kent, Chebyshev, Krylov, Lanczos: Matrix Relationships and Computations, Ph.D. Thesis, Dept. of Computer Science, Stanford Univ., 1989.

15 S. Kharchenko, P. Kolesnikov, A. Nikishin, A. Yeremin, M. Heroux, and Q. Seikh, Iterative solution methods on the Cray YMP/C90. Part II : Dense linear systems, presented at 1993 Simulation Conference : High Performance Computing Symposium, Washington, D.C.

16 T. A. Manteuffel, The Tchebychev iteration for nonsymmetric linear systems, *Numer. Math.* 28:307–327 (1977).

17 T. A. Manteuffel and J. Otto, On the roots of the orthogonal polynomials and residual polynomials associated with a conjugate gradient method, *J. Numer. Linear Algebra*, to appear.

18 N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.* 13(3):778–795 (July 1992).

19 N. M. Nachtigal, L. Reichel, and L. N. Trefethen, A hybrid GMRES algorithm for nonsymmetric matrix iterations, *SIAM J. Matrix Anal. Appl.* 13(3):796–825 (July 1992).

20 A. A. Nikishin and A. Y. Yeremin, Variable Block CG Algorithms for Solving Large Sparse Symmetric Positive Definite Systems on Parallel Computers, I : General Iterative Scheme, Technical Report EE-RR 1/92, Elegant Mathematics, Inc., 1992, Rev. Jan. 1993.

21 D. P. O'Leary, The block conjugate gradient algorithm and related methods, *Linear Algebra Appl.* 29:293–322 (1980).

22 D. P. O'Leary, Parallel implementation of the block conjugate gradient algorithm, *Parallel Comput.* 5:127–139 (1987).

23 C. C. Paige, B. N. Parlett, and H. A. van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces, *Numer. Linear Algebra Appl.*, to appear.

24 M. Papadrakakis and S. Smerou, A new implementation of the Lanczos method in linear problems, *Internat. J. Numer. Methods Engrg.* 29:141–159 (1990).

25 L. Reichel and L. N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, *Linear Algebra Appl.* 162–164:153–185 (1992).

26 Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, *Math. Comp.* 37(155):105–125 (1981).

27 Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Halstead Press, New York, 1992.

28 Y. Saad, On the rates of convergence of the Lanczos and the block-Lanczos methods, *SIAM J. Numer. Anal.* 17(5):687–706 (Oct. 1980).

29 Y. Saad and M. H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Statist. Comput.* 7(3):856–869 (July 1986).

30 M. Sadkane, Block Arnoldi and Davidson methods for nonsymmetric large eigenvalue problems, *Numer. Math.* 64:195–211, (1993).

31 M. Sadkane and B. Vital, Davidson's Method for Linear Systems of Equations. Implementation of a Block Algorithm on a Multi-processor, Technical Report TR/PA 91/60, CERFACS, Toulouse, Sept. 1991.

32 P. E. Saylor and D. C. Smolarski, Implementation of an adaptive algorithm, for Richardson's method, *Linear Algebra Appl.* 154–156:615–646 (1991).

33 H. D. Simon and A. Yeremin, A new approach to construction of efficient iterative schemes for massively parallel algorithms: Variable block CG and BiCG methods and variable block Arnoldi procedure, in *Proceedings of the Sixth SIAM Conference on Parallel Processing for Scientific Computing*, SIAM, (R. F. Sincovec, D. E. Keyes, M. R. Leuze, L. R. Petzold, and D. A. Reed, Eds.), Philadelphia, Mar. 1993, pp. 57–60.

34 V. Simoncini and E. Gallopoulos, Iterative Methods for Complex Symmetric Systems with Multiple Right-Hand Sides, Technical Report 1322, Center for Supercomputing Research and Development, Univ. of Illinois at Urbana-Champaign, Dec. 1993.

35 V. Simoncini and E. Gallopoulos, A Hybrid Block GMRES Method for Nonsymmetric Systems with Multiple Right-Hand Sides, *J. Comput. Appl. Math.*, 66:457–469 (1996).

36 V. Simoncini and E. Gallopoulos, An iterative method for nonsymmetric systems with multiple right-hand sides, *SIAM J. Sci. Statist. Comput.*, 16(4):917–933 (1995).

37 D. C. Smolarski and P. E. Saylor, Computing the roots of complex orthogonal and kernel polynomials, *SIAM J. Sci. Statist. Comput.* 9(1):1–13 (1988).

38 L. N. Trefethen, Approximation theory and numerical linear algebra, in *Algorithms for Approximation II* (J. C. Mason and M. G. Cox, Eds.), Chapman and Hall, 1990, pp. 336–360.

39 R. Underwood, An Iterative Block Lanczos Method for the Solution of Large Sparse Symmetric Eigenproblems, Technical Report STAN-CS-75-496, Computer Science, Stanford Univ., 1975.
40 B. Vital, Etude de quelques Méthodes de Résolution de Problèmes Linéaires de Grande Taille sur Multiprocesseur, Ph.D. Thesis, Univ. de Rennes I, Rennes, Nov. 1990.

Received 28 April 1994; final manuscript accepted 15 January 1995