SAN099-1756J

ON THE CONVERGENCE OF AN IMPLICITLY RESTARTED ARNOLDI METHOD *

R. B. LEHOUCQ†

Abstract. We show that Sorensen's [35] implicitly restarted Arnoldi method (including its block extension) is simultaneous iteration with an implicit projection step to accelerate convergence to the invariant subspace of interest. By using the geometric convergence theory for simultaneous iteration due to Watkins and Elsner [43], we prove that an implicitly restarted Arnoldi method can achieve a super-linear rate of convergence to the dominant invariant subspace of a matrix. Moreover, we show how an IRAM computes a nested sequence of approximations for the partial Schur decomposition associated with the dominant invariant subspace of a matrix.

Key words. Simultaneous iteration, Arnoldi reduction, Schur decomposition, restarting, eigenvalues.

AMS subject classifications. 65F15, 65G05

1. Introduction. A classical method of solution for the large-scale eigenvalue problem is simultaneous iteration [6, 9, 26, 27, 30, 37, 40]. Simultaneous iteration was originally introduced by Bauer [7], who called the method *Treppeniteration* (staircase iteration). It is a straightforward method for computing the eigenvalues of largest modulus of a matrix $\mathbf{A}$ and is a generalization of the power method in that a subspace of size larger than one is employed. Unfortunately, the rate of convergence of simultaneous iteration is linear. For many large eigenvalue problems, this rate is prohibitive.

This article shows that an implicitly restarted Arnoldi method (IRAM) (including its block extension) is simultaneous iteration in disguise. In particular, we are concerned with the convergence of $\Phi_i(A)\mathcal{Z}$ where $\mathcal{Z}$ is a Krylov subspace and $\Phi_i(\cdot)$ is a polynomial. The relationship with simultaneous iteration is demonstrated by exploiting the well-known connection [25, 39, 42] with the QR algorithm. By appealing to the results in [43], a possible super-linear rate of convergence for an IRAM is established. Moreover, we show how an IRAM computes a nested sequence of approximations for the partial Schur decomposition associated with the dominant invariant subspace of a matrix. This practical convergence theory is possible by appealing to the results in [40] that were developed for simultaneous iteration with a projection step. Our numerical experiments will show an IRAM converges significantly faster than classical simultaneous iteration.

We feel that the strength of the relationship outlined in this article is two-fold. First, it shows that IRAMs should be considered as simultaneous iteration methods. This allows us to use the convergence theory developed for simultaneous iteration. The resulting convergence theory is in contrast to the standard theory for Arnoldi methods. Standard theory considers the degree of approximation to eigenvalues and eigenvectors as a function of increasing Arnoldi iterations (size of the Krylov subspace). Instead, our theory considers the degree of approximation to an orthonormal

*This work was supported by the Department of Energy (Contracts DE-FG0f-91ER25103 and W-31-109-Eng-38).

†Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque, NM 87185-1110 rblehou@sandia.gov. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

representation of the invariant subspace as a function of the number of restarts, or equivalently, simultaneous iterations. Second, it demonstrates why robust software can be developed for the large-scale eigenvalue problem using IRAMs. Robust software exists for simultaneous iteration because it is a backward stable algorithm. This stability is a direct result of an exclusive reliance on unitary transformations. Golub and Wilkinson [12] also examine the many practical difficulties involved when computing invariant subspaces. They conclude that working with an orthonormal basis of approximate Schur vectors is a better-behaved numerical process. Within the context of simultaneous iteration, Stewart [40] also arrives at the same conclusion.

Section 2 discusses the eigenvalue problem and Schur decompositions. Section 3 introduces the Arnoldi reduction and its computation. The relationship between an IRAM and simultaneous iteration is derived in Section 4. Section 5 describes a geometric convergence theory of an IRAM. A relationship between an IRAM and Stewart's [40] generalization of the Rayleigh–Ritz method to non-Hermitian matrices is the subject of Section 6. Section 7 discusses some practical issues associated with an IRAM. The results of some numerical experiments comparing classical simultaneous iteration with an IRAM are given in Section 8. Finally, the quality of the approximation to an invariant subspace of $\mathbf{A}$ computed by an IRAM is considered in section 9.

We conclude this section with the basic notation to be used in this article. We employ Householder notational conventions. Capital and lower-case letters denote matrices and vectors, respectively, while lower-case Greek letters denote scalars.

The transpose of a vector $\mathbf{x}$ is denoted by $\mathbf{x}^T$, and the complex conjugate of $\mathbf{x}^T$ is denoted by $\mathbf{x}^H$. The norms used are the Euclidean and Frobenius, denoted by $\|\cdot\|$ and $\|\cdot\|_F$, respectively. The range of a matrix $\mathbf{A}$ is denoted by $\mathcal{R}(\mathbf{A})$.

A matrix of lower bandwidth b will be called a banded upper Hessenberg matrix. We drop "upper" when the context is clear. Omission of the word *band* implies that the block size is one. We say that a banded Hessenberg matrix is unreduced if all the elements on the bth subdiagonal are nonzero.

2. The Eigenvalue Problem. Let $\mathbf{A}$ be a complex matrix of order n. We are interested in computing the $k \ll n$ dominant (those of largest magnitude) eigenvalues and associated invariant subspace of

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{2.1}$$

The eigenvalues and eigenvectors of $\mathbf{A}$ are denoted by λ_j and $\mathbf{x}_j$, respectively, for $j = 1, \ldots, n$. For the remainder of our article, we suppose that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.

The following decomposition proves central to the eigenvalue algorithms considered in this article. Its value is in providing us with a canonical form for which stable algorithms may be developed. For us, a stable algorithm computes the exact Schur decomposition of nearby matrix.

THEOREM 2.1. *(Schur Decomposition) If $\mathbf{A} \in \mathbf{C}^{n\times n}$, then there exists a unitary $\mathbf{Z} \in \mathbf{C}^{n\times n}$ such that*

$$\mathbf{Z}^H\mathbf{A}\mathbf{Z} = \mathbf{T}, \tag{2.2}$$

where $\mathbf{T}$ is an upper triangular matrix. The eigenvalues can appear in any order along the diagonal.

Proof. See [11, page 313]. □

Let $\mathbf{D}$ be a diagonal unitary matrix. Then $(\mathbf{Z}\mathbf{D})^H\mathbf{A}\mathbf{Z}\mathbf{D} = \mathbf{D}^H\mathbf{T}\mathbf{D}$ has diagonal blocks equal to those of $\mathbf{T}$. Thus, apart from the eigenvalues of multiplicity larger than

DISCLAIMER

one, the decomposition is essentially unique, given some ordering of the eigenvalues. Denote the leading principal matrix of order k of $\mathbf{T}$ by $\widehat{\mathbf{T}}_k$. Let $\mathbf{Z}_k$ be the corresponding columns of $\mathbf{Z}$. Then $\mathbf{A}\mathbf{Z}_k = \mathbf{Z}_k\widehat{\mathbf{T}}_k$ is a *partial* Schur decomposition of $\mathbf{A}$ of order k. When $\mathbf{A}$ is Hermitian $\mathbf{T}$ is a diagonal matrix, and hence the eigenvalues are real numbers.

The full decomposition is computed by the practical QR algorithm in the EISPACK [34] and LAPACK [1] software packages. Instead simultaneous iteration attempts to compute a partial Schur decomposition for $\mathbf{A}$ with the the dominant eigenvalues located on the diagonal of $\widehat{\mathbf{T}}_k$.

3. Partial Reduction to Band Hessenberg Form. The first step of the practical QR algorithm is to reduce $\mathbf{A}$ to upper Hessenberg form via Householder transformations. This done, as is well known, so that each step of the QR iteration performed on the Hessenberg matrix only involves order n^2 work. This is in contrast to the order n^3 work that would be required during each step of a QR iteration on $\mathbf{A}$ (assuming that $\mathbf{A}$ is a dense matrix).

Unfortunately, for large eigenvalue problems, Householder transformations cannot be used as they destroy any sparsity or structure in $\mathbf{A}$. We say an eigenvalue problem is large if the dense QR algorithm is prohibitive, in storage and/or efficiency. Instead, the Arnoldi reduction [2] only requires knowledge of $\mathbf{A}$ through a matrix-vector product. Moreover, it allows us to sequentially reduce $\mathbf{A}$ to upper Hessenberg form, producing the leading portion of the an upper Hessenberg matrix at every step. When the matrix $\mathbf{A}$ is Hermitian, the Lanczos reduction [17] is recovered.

Since our concern is in the solution of eigenvalue problems in which $\mathbf{A}$ is not only large but expensive to apply, block Arnoldi reductions [31, 32] are considered. In many instances, the cost of computing a few matrix vector products is commensurate with that of one matrix vector product. Moreover, in exact arithmetic, an unblocked Arnoldi reduction cannot detect the multiplicity of an eigenvalue. A blocksize enables the reduction to compute multiplicities less than or equal to the blocksize.

Let $b > 0$, an integer, be the block size and let $\mathbf{E}_j \equiv \left[\begin{array}{ccc} \mathrm{e}_{(j-1)b+1} & \cdots & \mathrm{e}_{jb} \end{array}\right]$ where the jth canonical basis vector is denoted by e_j. We say that

$$\mathbf{A}\mathbf{V}_j = \mathbf{V}_j\mathbf{H}_j + \mathbf{F}_j\mathbf{E}_j^T \tag{3.1}$$

is a block Arnoldi reduction of length j when $\mathbf{V}_j^H\mathbf{A}\mathbf{V}_j = \mathbf{H}_j$ is a banded upper Hessenberg matrix, $\mathbf{V}_j^H\mathbf{V}_j = \mathbf{I}_{j\cdot b}$, and $\mathbf{V}_j^H\mathbf{F}_j = \mathbf{0}$. It follows that $\mathbf{H}_j$ is the projection of $\mathbf{A}$ onto the column span of $\mathbf{V}_j$. The $j \cdot b$ columns of $\mathbf{V}_j$ are an orthonormal basis for the block Krylov subspace

$$\mathcal{K}_j^b(\mathbf{A}, \mathbf{U}_1) \equiv \{\mathbf{U}_1, \mathbf{A}\mathbf{U}_1, \cdots, \mathbf{A}^{j-1}\mathbf{U}_1\}$$

where $\mathbf{U}_1$ is a full rank matrix with b columns. Note that if $\mathbf{A} = \mathbf{A}^H$, then $\mathbf{H}_j$ is a block tridiagonal matrix.

In order to introduce notation that will be needed, we rewrite (3.1) as

$$\mathbf{A}\,\mathbf{V}_j = \left[\begin{array}{ccc} \mathbf{U}_1 & \cdots & \mathbf{U}_j \end{array}\right] \left[\begin{array}{cccc} \mathbf{G}_{1,1} & \cdots & \cdots & \mathbf{G}_{1,j} \\ \mathbf{G}_{2,1} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{G}_{j,j-1} & \mathbf{G}_{j,j} \end{array}\right]$$

$$+\mathbf{U}_{j+1}\mathbf{G}_{j+1,j}\mathbf{E}_j^T$$

where $\mathbf{U}_{j+1}\mathbf{G}_{j+1,j}$ is the QR factorization of $\mathbf{F}_j$. The reduction is advanced by one step (or its length incremented) by the following three operations:

1. $\mathbf{W} = \mathbf{A}\mathbf{U}_{j+1}$
2. $\mathbf{G}_{i,j+1} = \mathbf{U}_i^H\mathbf{W} \;\; i = 1, \ldots, j+1$
3. $\mathbf{F}_{j+1} = \mathbf{W} - \sum_{i=1}^{j+1}\mathbf{U}_i\mathbf{G}_{i,j+1}$

A direct implementation of the second and third steps will not, in general, produce an orthogonal set of Arnoldi vectors. We refer the reader to [18] for details on an efficient and robust implementation. See [13] for references and information on a block Lanczos reduction implemented via a three term block recurrence.

3.1. Computing an Approximate Partial Schur Decomposition. Suppose that $\mathbf{H}_j\mathbf{Z} = \mathbf{Z}\mathbf{T}$ is a Schur decomposition ordered so that the eigenvalues of $\mathbf{T}$ are in descending order of magnitude along the diagonal. Postmultiplying (3.1) with $\mathbf{Z}$ gives

$$\|\mathbf{A}\mathbf{V}_j\mathbf{Z} - \mathbf{V}_j\mathbf{Z}\mathbf{T}\| = \|\mathbf{F}_j\mathbf{E}_j^T\mathbf{Z}\| = \|\mathbf{G}_{j+1,j}\mathbf{E}_j^T\mathbf{Z}\|. \tag{3.2}$$

Because $\mathbf{E}_j^T\mathbf{Z}$ is a matrix with $j \cdot b$ columns consisting of the last b rows of $\mathbf{Z}$, the quality of an approximate partial Schur decomposition is determined by noting that if

$$\|\mathbf{F}_j\mathbf{E}_j^T\mathbf{Z}\begin{bmatrix} e_1 & e_2 & \cdots & e_k \end{bmatrix}\| \equiv \|\mathbf{F}_j\mathbf{E}_j^T\mathbf{Z}_k\|$$

is small then an approximate partial Schur decomposition of order k for the dominant invariant subspace is computed. In fact, if $\mathbf{H}_j\mathbf{Z}_k = \mathbf{Z}_k\widehat{\mathbf{T}}_k$ is an order k partial Schur decomposition for $\mathbf{H}_j$, then

$$(\mathbf{A} + \mathbf{M})\mathbf{V}_j\mathbf{Z}_k = \mathbf{V}_j\mathbf{Z}_k\widehat{\mathbf{T}}_k$$

where

$$\|\mathbf{M}\| = \| - \mathbf{F}_j\mathbf{E}_j^T\mathbf{Z}_k(\mathbf{V}_j\mathbf{Z}_k)^H\| = \|\mathbf{G}_{j+1,j}\mathbf{E}_j^T\mathbf{Z}_k\|$$

implying that we have computed an exact partial Schur decomposition for a matrix near $\mathbf{A}$.

If approximate eigenvectors are of interest, they can be computed from the partial Schur decomposition $\mathbf{H}_j\mathbf{Z}_k = \mathbf{Z}_k\widehat{\mathbf{T}}_k$ because

$$\mathbf{A}\mathbf{V}_j\mathbf{Z}_k\mathbf{y} - \mathbf{V}_j\mathbf{Z}_k\mathbf{y}\theta = \mathbf{F}_j\mathbf{E}_j^T\mathbf{Z}_k\mathbf{y} \tag{3.3}$$

where $\widehat{\mathbf{T}}_k\mathbf{y} = \mathbf{y}\theta$. We call $\mathbf{V}_j\mathbf{Z}_k\mathbf{y}$ a Ritz vector and θ a Ritz value. Note that the first Schur vector is always an eigenvector.

4. Connection with Simultaneous Iteration. As explained in our introduction, there is a well-known connection between simultaneous iteration and the QR algorithm. This section will show a similar high-level connection between an IRAM and simultaneous iteration. The following elementary but technical result[1] is needed for this connection. It is a generalization of the special case $\psi(\lambda) = \lambda$ shown in [20]. A similar result was proved in Lemma 1 of [24] for the Lanczos reduction.

[1] I thank Chris Beattie for suggesting this result to me during a workshop held in the spring of 1995

LEMMA 4.1. *Suppose that an integer p satisfies $1 \leq p < m$, and let $r = m - p$. Let $\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + \mathbf{F}_m\mathbf{E}_m^T$ be a length $r + p$ Arnoldi reduction, where $\mathbf{H}_m$ is an unreduced band upper Hessenberg matrix. If*

$$\psi_p(\lambda) = \prod_{i=1}^{p}(\lambda - \tau_i),$$

then

$$\psi_p(\mathbf{A})\mathbf{V}_m = \mathbf{V}_m\psi_p(\mathbf{H}_m) + \sum_{j=1}^{p}\psi_{j+1}^{p}(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{j-1}(\mathbf{H}_m), \tag{4.1}$$

where $\psi_j(\lambda) = \prod_{i=1}^{j}(\lambda - \tau_i)$ and $\psi_j^p(\lambda) = \prod_{i=j}^{p}(\lambda - \tau_i)$.

Moreover,

$$\psi_p(\mathbf{A})\mathbf{V}_r = \mathbf{V}_m\psi_p(\mathbf{H}_m)\begin{bmatrix} \mathbf{E}_1 & \cdots & \mathbf{E}_r \end{bmatrix} \tag{4.2}$$

Proof. The proof is by mathematical induction. Define $m \equiv r+p$. The subscripts are suppressed on $\mathbf{V}_m$ and $\mathbf{H}_m$ for the proof. Since $\psi_1(\mathbf{A})\mathbf{V} = \mathbf{V}\psi_1(\mathbf{H}) + \mathbf{F}_m\mathbf{E}_m^T$, where $\psi_1(\lambda) = \lambda - \tau_1$, the base case for $p = 1$ is established. Assume the lemma's truth for polynomials $\psi_j(\lambda)$ of degree $j \leq p$. Let $\psi_{p+1}(\lambda) = (\lambda - \tau_{p+1})\psi_p(\lambda)$. With the induction hypothesis, it follows that

$$\begin{aligned}
\psi_{p+1}(\mathbf{A})\mathbf{V} &= (\mathbf{A} - \tau_{p+1}\mathbf{I})\psi_p(\mathbf{A})\mathbf{V} \\
&= (\mathbf{A} - \tau_{p+1}\mathbf{I})\left\{\mathbf{V}\psi_p(\mathbf{H}) + \sum_{j=1}^{p}\psi_{j+1}^{p}(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{j-1}(\mathbf{H})\right\} \\
&= \mathbf{V}(\mathbf{H} - \tau_{p+1}\mathbf{I})\psi_p(\mathbf{H}) + \mathbf{F}_m\mathbf{E}_m^T\psi_p(\mathbf{H}) \\
&\quad + (\mathbf{A} - \tau_{p+1}\mathbf{I})\sum_{j=1}^{p}\psi_{j+1}^{p}(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{j-1}(\mathbf{H}) \\
&= \mathbf{V}\psi_{p+1}(\mathbf{H}) + \sum_{j=1}^{p+1}\psi_{j+1}^{p+1}(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{j-1}(\mathbf{H}),
\end{aligned}$$

which establishes Equation (4.1).

Since $\mathbf{H}$ is unreduced, $\psi_{j-1}(\mathbf{H})$ is a band Hessenberg matrix of lower bandwidth $(j-1)\cdot b$. Thus $\mathbf{E}_i^T\psi_{j-1}(\mathbf{H})\mathbf{E}_l = \mathbf{0}$ for $l = 1, \ldots, m-j$, and the last matrix on the right-hand side of Equation (4.1) is zero through its first $r \cdot b$ columns. Equation (4.2) is established. □

In words, Equation (4.2) shows that $\psi_p(\mathbf{A})$ applied to the first $r \cdot b$ columns of $\mathbf{V}_m$ is equivalent to the basis representation given by the first $r \cdot b$ columns of $\mathbf{V}_m\psi_p(\mathbf{H}_m)$.

Suppose that p steps of the QR algorithm are performed on $\mathbf{H}_m$ with the $p < m$ shifts $\tau_1, \ldots, \tau_p$ resulting in

$$\mathbf{H}_m\mathbf{W} = \mathbf{W}\mathbf{H}_m^+. \tag{4.3}$$

Note that $\mathbf{W}$ is a Hessenberg matrix of lower bandwidth $p \cdot b$ because it is a product of p unitary matrices each of lower bandwidth b each computed during a step of the

QR algorithm. If $\mathbf{W}_1$ denotes the initial $r \cdot b$ columns of $\mathbf{W}$, $\mathbf{W}_2$ the next b columns and $\mathbf{W}_3$ the remaining columns, equating the first $r \cdot b$ columns of (4.3) results in

$$\mathbf{H}_m \mathbf{W}_1 \equiv \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \mathbf{W}_3 \end{bmatrix} \begin{bmatrix} \mathbf{H}_r^+ \\ \mathbf{G}_{r+1,r}^+ \mathbf{E}_r^T \\ \mathbf{0} \end{bmatrix}. \tag{4.4}$$

Post-multiplying Equation (3.1) with $\mathbf{W}_1$ and using (4.4), we obtain

$$\begin{aligned} \mathbf{A}\mathbf{V}_m\mathbf{W}_1 &= \mathbf{V}_m\mathbf{H}_m\mathbf{W}_1 + \mathbf{F}_m\mathbf{E}_m^T\mathbf{W}_1, \\ &= \mathbf{V}_m\mathbf{W}_1\mathbf{H}_r^+ + \mathbf{V}_m\mathbf{W}_2\mathbf{G}_{r+1,r}^+\mathbf{E}_r^T + \mathbf{F}_m\mathbf{E}_m^T\mathbf{W}_1, \\ &\equiv \mathbf{V}_r^+\mathbf{H}_r^+ + \mathbf{F}_r^+\mathbf{E}_r^T, \end{aligned} \tag{4.5}$$

where

$$\mathbf{V}_r^+ \equiv \mathbf{V}_m\mathbf{W}_1 \quad \text{and} \quad \mathbf{F}_r^+ \equiv \mathbf{V}_m\mathbf{W}_2\mathbf{G}_{r+1,r}^+ + \mathbf{F}_m\mathbf{E}_m^T\mathbf{W}_1\mathbf{E}_r. \tag{4.6}$$

Note the use of the identity $\mathbf{E}_m^T\mathbf{W}_1 = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_m^T\mathbf{W}_1\mathbf{E}_r \end{bmatrix}$ in Equation (4.6). We remark that (4.3) through (4.6) define a restart of the Arnoldi reduction via a QR algorithm. An IRAM is a sequence of implicit restarts that are terminated when the partial Schur decomposition of interest is sufficiently well approximated. The restart is implicit because it relies upon the implicitly shifted QR algorithm. Moreover, a scheme is provided for the restart that does not require explicit application of $\mathbf{A}$.

The following theorem establishes a direct relationship between simultaneous iteration and the QR algorithm. It is a partial or truncated version of Theorem 3.1 in [39, p. 353].

THEOREM 4.2. *Assume the hypothesis of Lemma 4.1 and the notation in equation (4.4)-(4.6). If the QR algorithm computed on* $\mathbf{H}_m$ *with the* p *shifts* $\tau_1, \ldots, \tau_p$ *results in* $\mathbf{H}_m\mathbf{W} = \mathbf{W}\mathbf{H}_m^+$ *then*

$$\psi_p(\mathbf{A})\mathbf{V}_r = \mathbf{V}_r^+\widehat{\mathbf{R}}_r \tag{4.7}$$

where $\widehat{\mathbf{R}}_r$ *is an upper triangular matrix of order* $r \cdot b$.

Proof. From Lemma 4.1 it suffices to show that the QR factorization of

$$\psi_p(\mathbf{H}_m)\begin{bmatrix} \mathbf{E}_1 & \cdots & \mathbf{E}_r \end{bmatrix} = \mathbf{W}_1\widehat{\mathbf{R}}_r$$

for some upper triangular $\widehat{\mathbf{R}}_r$ of order $r \cdot b$. But this is a consequence of the link between the QR algorithm and simultaneous iteration because

$$\psi_p(\mathbf{H}_m) = \mathbf{W}\mathbf{R}$$

is a QR factorization [39, p.353]. The result follows from equating the initial $r \cdot b$ columns of this equality and letting $\widehat{\mathbf{R}}_r$ denote the leading sub-matrix of order $r \cdot b$ of $\mathbf{R}$. □

The theorem allow us to link simultaneous iteration with an IRAM. If $p = 1$, $\psi_1(\lambda) = \lambda$, $r = 1$ and $b > 1$ then (4.7) becomes $\mathbf{A}\mathbf{V}_1 = \mathbf{V}_1^+\widehat{\mathbf{R}}_1$ and classical simultaneous iteration is recovered. From (4.5) it follows that an IRAM is simultaneous iteration in disguise because $\mathbf{V}_1^+$ is computed.

We remark that if $p = m$, the previous discussion gives that implicitly restarting with $m - 1$ shifts produces

$$\mathbf{A}\mathbf{U}_1^+ = \mathbf{U}_1^+\mathbf{H}_1^+ + \mathbf{F}_1^+. \tag{4.8}$$

If $\mathbf{QR} = \mathbf{H}_1^+ - \tau_m \mathbf{I}$, then

$$(\mathbf{A} - \tau_m \mathbf{I})\mathbf{U}_1^+ = \mathbf{U}_1^+ \mathbf{QR} + \mathbf{F}_1^+. \tag{4.9}$$

The right-hand side of (4.9) defines a new starting block of vectors (after orthogonalization) for a subsequent block Arnoldi reduction. Thus we can implicitly apply polynomials of degree m in $\mathbf{A}$. This was first established in [3].

5. Convergence of an IRAM. Sorensen [35] gave some convergence results for an IRAM (blocksize of one). For non-symmetric $\mathbf{A}$, a linear rate of convergence was given for a fixed $\psi_p(\cdot)$ per restart. He also showed that for symmetric $\mathbf{A}$, using the unwanted $m - k$ eigenvalues as shifts during each implicit restart resulted in convergence to k eigenvalues (of $\mathbf{A}$).

Traditional convergence theory [14, 15, 28, 28] for Arnoldi reductions investigates the quality of $\mathcal{K}_j^b(\mathbf{A}, \mathbf{U}_1)$ to approximate eigenvectors of $\mathbf{A}$ as j increases. However, with the connection between an IRAM and simultaneous iteration in hand, a more elegant and powerful theory is possible. A comprehensive geometric convergence theory for the shifted QR algorithm is presented by Watkins and Elsner [43] within the more general framework of generic *GR algorithms*. A GR algorithm is a generalization of the QR algorithm where where the QR factorization of $\mathbf{A} - \tau\mathbf{I}$ is replaced with a GR factorization $\mathbf{G}$ is a nonsingular matrix. The convergence theory is based on the idea that a GR algorithm is a nested sequence of non-stationary simultaneous iterations.

Theorem 5.1 in [43] shows the convergence of non-stationary simultaneous iteration. We specialize this theorem to our case of simultaneous iteration. Recall the notation established in section 2 and 4. In addition, let $\phi_i(\cdot)$ denote $\psi_p(\cdot)$ (the polynomial in $\mathbf{A}$ implicitly applied) used at the i-th restart of an IRAM.

THEOREM 5.1. *Let $\mathbf{A}$ be a simple matrix of order n. Let $r \cdot b \ll n$ where b and r are positive. Define the invariant subspaces $\mathcal{Z} = Span(\mathbf{x}_1, \ldots, \mathbf{x}_{r\cdot b})$ and $\mathcal{U} = Span(\mathbf{x}_{r\cdot b+1}, \ldots, \mathbf{x}_n)$. Let $\Phi_i = \phi_i \cdots \phi_1$ and suppose that $\Phi_i(\lambda_j) \neq 0$ for $j = 1, \ldots, r\cdot b$ and let*

$$\rho_i = \frac{\max_{j=r\cdot b+1,\ldots,n} |\Phi_i(\lambda_j)|}{\min_{j=1,\ldots,r\cdot b} |\Phi_i(\lambda_j)|}. \tag{5.1}$$

If $\mathcal{V}$ is a subspace of dimension $r \cdot b$ satisfying $\mathcal{V} \cap \mathcal{U} = \{0\}$ then there exists a constant C such that

$$dist(\Phi_i(\mathbf{A})\mathcal{V}, \mathcal{Z}) \leq C\rho_i$$

for all i.

The distance between the subspaces [8, 11] $\mathcal{V}$ and $\mathcal{Z}$ may be shown to be equal to $\sqrt{1 - \|\mathbf{V}^H\mathbf{Z}\|^2}$ where the columns of $\mathbf{V}$ and $\mathbf{Z}$ provide an orthonormal basis for $\mathcal{V}$ and $\mathcal{Z}$. For increasing values of i, the approximating subspace $\Phi_i(\mathbf{A})\mathcal{V}$ aligns itself with $\mathcal{Z}$ and hence the distance between the two subspaces goes to zero.

Because of the relationship between an IRAM and simultaneous iteration established in the last section, the $\text{dist}(\Phi_i(\mathbf{A})\mathcal{V}, \mathcal{Z}) \to 0$ and the eigenvalues of leading block of $\mathbf{H}_m$ of order $r \cdot b$ computed after i restarts are approximations to $\lambda_1, \ldots, \lambda_{r\cdot b}$.

An expression for the constant C is given by Watkins and Elsner. They show that

$$C = \|\mathbf{X}\| \, \|\mathbf{X}^{-1}\| \frac{\sqrt{1 - \|\mathbf{V}^H\mathbf{Z}\|^2}}{\|\mathbf{V}^H\mathbf{Z}\|}$$

where $\mathbf{X}$ is the matrix of eigenvectors for $\mathbf{A}$. The size of C is large whenever the eigenvectors form an ill-conditioned basis and/or the starting subspace $\mathcal{V}$ is nearly orthonormal to $\mathcal{Z}$.

We remark that the hypothesis on $\mathbf{A}$ is only done for notational and elaborative purposes. The paper [43] considers the more general case of an arbitrary matrix.

For a stationary iteration ($\Phi_i(\cdot) = \phi_1(\cdot)^i$) where $|\Phi_i(\lambda_j)| \geq |\Phi_i(\lambda_{j+1})|$ for $j = 1, \ldots, n-1$ and $\rho_i < 1$, then $\Phi_i(\mathbf{A})\mathcal{V}$ converges to $\mathcal{Z}$ at the linear rate ρ_1. This is an alternate proof of Theorem 5.1 in [35].

In Example 5.3 following Theorem 5.1 in [43], a possible super-linear rate of convergence is demonstrated. Suppose that $\Phi_i(\lambda_j) \to 0$ for $j = k+1, \ldots, n$ and $\Phi_i(\lambda_j) \to d_j (\neq 0)$ for $j = 1, \ldots, k$ for all i. It follows then that for $\epsilon > 0$ there exists an i_ϵ so that $\rho_i < \epsilon$ for $i > i_\epsilon$ and hence $\rho_i \leq K\epsilon^i$ for some constant K. Because the bound on the ratio of the iterates is ϵ, $\rho_i \to 0$ super-linearly since this holds for all positive ϵ.

Finally, we remark that a that a shift strategy that leads to cubic and quadratic convergence rates of convergence of the QR algorithm cannot be adopted for an IRAM. The shift strategy requires information available only when a full Hessenberg reduction is at hand.

6. Simultaneous Iteration with a Projection Step. Stewart [40] introduced a projection step for simultaneous iteration on non-Hermitian matrices. The iteration

1. $\mathbf{AU}_1 = \mathbf{QR}$ QR factorization
2. $\mathbf{C} = \mathbf{Q}^H\mathbf{AQ}$
3. $\mathbf{CZ} = \mathbf{ZT}$ Schur decomposition
4. $\mathbf{U}_1 \leftarrow \mathbf{QZ}$

is continued until the first k columns $\mathbf{AU}_1 - \mathbf{U}_1\mathbf{T}$ are small in norm at which point a partial Schur decomposition for a nearby matrix has been computed. We note that the above algorithm iterates on the Krylov subspace $\mathcal{K}_1^b(\mathbf{A}, \mathbf{U}_1)$ and hence $r = m = 1$ and b denotes the number of columns of $\mathbf{U}_1$. The Schur decomposition computed in step 3 is ordered so that the eigenvalues appear in descending order of magnitude along the diagonal of $\mathbf{T}$. Step 4 is a Rayleigh-Ritz step via the Schur vectors $\mathbf{Z}$, and under some mild conditions, the j-th column of $\mathbf{U}_1$ converges to the j-th Schur vector associated with the j-largest (in magnitude) eigenvalue at a rate of $|\lambda_{b+1}/\lambda_j|$. We emphasize that the convergence of $\mathbf{U}_1$ is not accelerated—only ordered so that the initial columns contain the best approximations to the dominant invariant subspace.

For non-trivial block Krylov subspaces $\mathcal{K}_r^b(\mathbf{A}, \mathbf{U}_1)$ Theorem 4.2 shows that an IRAM is simultaneous iteration with a projection step. In analogy to Stewart's explicit projection computed via a Schur decomposition, an IRAM employs an implicitly shifted QR algorithm so that the columns of $\mathbf{V}_r^+$ contain increasingly better approximations to the desired partial Schur decomposition. If $|\Phi_i(\lambda_j)| \geq |\Phi_i(\lambda_{j+1})|$ for $j = 1, \ldots, r \cdot b$, all i, and $\rho_i < 1$ then Stewart's convergence theory implies that the ℓ-th ($1 \leq \ell \leq r \cdot b$) column of $\mathbf{V}_r^+$ converges at the rate

$$\frac{\max_{j=r\cdot b+1,\ldots,n} |\Phi_i(\lambda_j)|}{|\Phi_i(\lambda_\ell)|} \tag{6.1}$$

during an IRAM. Hence, an IRAM computes a nested sequence of orthonormal bases for the dominant invariant subspace.

TABLE 8.1
Computing the partial Schur decomposition corresponding to the 4 dominant eigenvalues of a convection-diffusion matrix of order 2,500.

	k	m	Matvec	Restarts	b
Subit	4	1	17,664	552	16
IRAM	4	16	253	23	1
IRAM	4	8	288	35	2
Subit	4	1	16,800	350	24
IRAM	4	24	230	12	1
IRAM	4	12	264	16	2
IRAM	4	8	330	35	3
IRAM	4	6	376	45	4
IRAM	4	200	199	0	1

7. Some Practical Issues. The question of a near-optimal shift strategy that achieves super-linear convergence is still the work of research. However, it is clear from Theorem 5.1 that a polynomial Φ_i that minimizes ρ_i for the discrete min-max problem (5.1) is required. However, there is additional flexibility, namely, changing the value of b, r and m for smaller values of ρ_i. Hence, we can consider ρ_i as a function of b, r and m to improve convergence to the partial Schur decomposition of order k.

The value b is selected so that eigenvalues up to multiplicity b can be resolved. Increasing the value of m improves the amount of information available to select the polynomial implicitly applied thus leading a decrease in ρ_i. On the other hand, increasing the value of r and/or b leads to an improved convergence rate if the numerator in (5.1) decreases.

An exact shift scheme was proposed by Sorensen. This is the default scheme used by ARPACK [19]. An exact scheme uses a number of the unwanted eigenvalues of $\mathbf{H}_m$ for the $\psi_p(\cdot)$ implicitly applied. As explained in [19, p. 71], the value of r is increased from the initial value of k as approximations to the eigenvalues of $\mathbf{A}$ satisfy the acceptance criterion. As explained in the previous paragraph, this has the effect decreasing the numerator in (5.1). The recent paper [36] discusses such an adaptive strategy for the symmetric eigenvalue problems. However, if the value of m is too small to allow r to be increased, an exact shift scheme leads to slow convergence. For symmetric eigenvalue problems, the roots of Leja polynomials [3, 5, 4] are successfully used for small m along with a deflation scheme that allows r to be increased from a value of one as satisfactory eigenvalue-eigenvector approximations emerge.

Recent papers have elucidated the virtues of other restart polynomials. These include the roots of Chebyshev polynomials [30], Harmonic Ritz values [22, 21, 24, 33], refined shifts [16].

8. Numerical Experiments. We present results for computing the 4 dominant eigenvalues and invariant subspace for the two-dimensional model convection-diffusion problem

$$-\Delta \mathbf{u}(x,y) + 4\mathbf{u}_x(x,y) + .5\mathbf{u}_y(x,y) = \lambda \mathbf{u}(x,y),$$

on the unit square $[0,1] \times [0,1]$ with zero boundary conditions. The problem is discretized by using centered finite differences where the mesh size $h = 1/(50+1)$ results in a matrix of order 2,500. Table 8.1 lists the results of some MATLAB experiments (IEEE double precision floating point arithmetic is used). Subit refers to a classical simultaneous iteration along with a projection at each restart for unscrambling the Schur vector approximations.

Each computed dominant partial Schur decomposition of order 4 gave a residual no larger than 10^{-3} and all experiments produced Ritz values that agreed to at least four significant digits. The first three lines of the table give results in terms of the number of restarts and matrix-vector products used when using $m \cdot b = 16$ vectors. The next five lines list the same output when $m \cdot b = 24$ and the final line indicates the total number of matrix vector products needed when $b = 1$ and m is run out until the partial Schur decomposition is approximated. This final experiment gives an indication of the minimum number of matrix-vector products needed. When $b = 1$, IRAM uses the strategy for increasing r from k as the eigenvalues satisfy the acceptance criterion to a maximum of 8; for $b = 2$ and $b = 4$, the value of r stayed fixed at $m - 4$ and for $b = 3$ was fixed at $r = 5$.

Some other conclusions resulting from the experiments of Table 8.1 are:

- The simultaneous iteration results always performed an extra m matrix-vector products per restart needed for the projection step. This projection is not needed at every restart. Hence the minimum number of matrix-vector products needed is half the number listed.
- Decreasing the blocksize requires a larger value of acceptance tolerance to be used so that the four dominant eigenvalues are computed to four significant digits.
- If $\mathbf{A}$ is efficiently applied to the block of vectors, then the number of matrix-vector products should be divided by the blocksize. Hence simultaneous iteration may be feasible.
- Decreasing the acceptance tolerance and/or increasing k (number of eigenvalues to computes) hinders the efficiency of simultaneous iteration. This effect is less pronounced as b decreases. IRAM with a blocksize of one has an amazing ability to compute partial Schur decompositions with small backward error.

In summary, if storage considerations and/or the cost of orthgonalizations prevent a large Arnoldi reduction from being computed, an IRAM does not require a substantial number more matrix-vector products. This was also noted by Morgan [23].

9. Approximations Drawn from an Arnoldi Reduction. In this section, we give a brief discussion of the quality of the approximation to a partial Schur decomposition produced by an IRAM. It serves to underscore the limitations of computing an approximate partial Schur decomposition with a small residual.

Suppose that $r \cdot b$ is at least as large as the size of the partial Schur decomposition needed. Since $\mathbf{A}\mathbf{V}_r = \mathbf{V}_r\mathbf{H}_r + \mathbf{F}_r\mathbf{E}_r^T$, an IRAM is attempting to drive $\|\mathbf{G}_{r+1,r}\|$ to zero so that $\mathbf{V}_r$ approaches a Schur representation of an invariant subspace. Suppose we complete this Arnoldi reduction to the full one:

$$\mathbf{A}\begin{bmatrix} \mathbf{V}_r & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_r & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{H}_r & \mathbf{M} \\ \mathbf{G}_{r+1,r}\mathbf{E}_r^T & \mathbf{C} \end{bmatrix}. \tag{9.1}$$

Using the generalization of Theorem 5.1 in [43], Watkins and Elsner show the rate of convergence of $\|\mathbf{G}_{r+1,r}\|$ is asymptotic with that of simultaneous iteration on the subspace spanned by the columns of $\mathbf{V}_r$. Thus when $\|\mathbf{G}_{r+1,r}\|$ is small, the $\mathcal{R}(\mathbf{V}_r)$

is nearly an invariant subspace for $\mathbf{A}$. However, the accuracy of $\mathbf{V}_r$ as an invariant subspace depends upon the sensitivity of $\mathbf{A}$'s invariant subspaces.

Stewart [38] considers how accurate $\mathbf{V}_r$ is to an invariant subspace of $\mathbf{A}$ for small $\|\mathbf{G}_{r+1,r}\|$. He considers whether an orthonormal matrix $\mathbf{Y}$ deviating little from $\mathbf{I}_n$ can be found so that $\mathbf{V}_r\mathbf{Y}$ is an invariant subspace for $\mathbf{A}$. Stewart chooses

$$\mathbf{Y} = \begin{bmatrix} \mathbf{I}_r & -\mathbf{P}^H \\ \mathbf{P} & \mathbf{I}_{n-r} \end{bmatrix} \begin{bmatrix} (\mathbf{I}_r + \mathbf{P}^H\mathbf{P})^{-1/2} & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{n-r} + \mathbf{P}\mathbf{P}^H)^{-1/2} \end{bmatrix},$$

where, because both $\mathbf{I}_r + \mathbf{P}^H\mathbf{P}$ and $\mathbf{I}_{n-r} + \mathbf{P}\mathbf{P}^H$ are Hermitian positive definite matrices, the square roots are uniquely defined. The answer to whether the column space of $\mathbf{V}_r$ is an accurate approximation to an invariant subspace of $\mathbf{A}$ becomes that of analyzing the interaction of $\mathbf{P}$ with $\mathbf{H}_r$, $\mathbf{M}$, $\mathbf{G}_{r+1,r}$ and $\mathbf{C}$. The following result explains the situation.

THEOREM 9.1. *Suppose that* $\mathbf{A}\mathbf{V}_r = \mathbf{V}_r\mathbf{H}_r + \mathbf{F}_r\mathbf{E}_r^T$ *is a length* r *block Arnoldi reduction. Suppose the reduction is completed to a band Hessenberg decomposition of* $\mathbf{A}$ *given by Equation (9.1), where* $\|\mathbf{G}_{r+1,r}\| = \|\mathbf{F}_r\|$. *Let*

$$\delta_r = \mathrm{sep}(\mathbf{H}_r, \mathbf{C}) \equiv \min_{\mathbf{X}\neq 0} \frac{\|\mathbf{X}\mathbf{H}_r - \mathbf{C}\mathbf{X}\|_F}{\|\mathbf{X}\|_F},$$

and denote $\beta_{r+1} \equiv \|\mathbf{G}_{r+1,r}\|$, $\gamma_r = \|\mathbf{C}\|$.

If $4\beta_{r+1}\gamma_r < \delta_r^2$, *there is a matrix* $\mathbf{P}$ *that satisfies the bound*

$$\|\mathbf{P}\| \leq 2\frac{\beta_{r+1}}{\delta_r}$$

so that the columns of $\mathbf{Z}_r = (\mathbf{V}_r + \mathbf{W}\mathbf{P})(\mathbf{I} + \mathbf{P}^H\mathbf{P})^{-1/2}$ *are an unitary basis for an invariant subspace of* $\mathbf{A}$.

Proof. The conclusion now follows directly from Theorem 4.1 of Stewart [38]. □

The size of γ_r measures the amount of coupling between the $\mathcal{R}(\mathbf{V}_r)$ and $\mathcal{R}(\mathbf{W})$. The reciprocal of δ_r measures the sensitivity of the $\mathcal{R}(\mathbf{Z}_r)$ as an invariant subspace. It may be shown that

$$\mathrm{sep}(\mathbf{H}_r, \mathbf{C}) \leq \min_{k,l} |\lambda_k(\mathbf{H}_r) - \lambda_l(\mathbf{C})|.$$

Varah [41] shows that if the matrices involved are highly nonnormal, the smallest difference between the spectrums of $\mathbf{H}_r$ and $\mathbf{C}_r$ may be an overestimate of the true separation.

Theorem 9.1 shows the dependence of β_{r+1} upon γ_r and δ_r in determining the quality of the $\mathcal{R}(\mathbf{V}_r)$ as an eigenspace of $\mathbf{A}$. Since $\mathbf{V}_r^H\mathbf{Z}_r = (\mathbf{I} + \mathbf{P}^H\mathbf{P})^{-1/2}$, Stewart [38] shows that the singular values of $\mathbf{P}$ are the tangents of the canonical, or principal, angles [8, 10, 38] between the two spaces spanned by the columns of $\mathbf{V}_r$ and $\mathbf{Z}_r$, respectively.

Traditional convergence theory for Arnoldi reductions investigates the quality of $\mathcal{K}_j^b(\mathbf{A}, \mathbf{U}_1)$ to approximate eigenvectors of $\mathbf{A}$ as j increases. Saad [28] considered the distance between a Ritz vector drawn from a block Lanczos reduction to an eigenvector. He extended his result to a $b = 1$ Arnoldi reduction in [29] with the assumption that $\mathbf{A}$ is diagonalizable. Jia [14] removed this restriction. One of Jia's main conclusions is that although a Arnoldi reduction may produce an approximation to an eigenvalue of $\mathbf{A}$, the associated eigenvector may not be well approximated by the

reduction. This situation occurs when the eigenvector is sensitive to perturbations or, in other words, is ill conditioned. The resolution of this dilemma is to instead consider the convergence of an orthonormal representation of the invariant subspace. An IRAM allows us to do this by varying b, r, and m.

10. Conclusions. We showed that an IRAM (including its block extension) is simultaneous iteration. Convergence theory and numerical experiments confirmed that $\Phi_i(A)\mathcal{K}_r^b(\mathbf{A}, \mathbf{U}_1)$ tends to the dominant invariant subspace of order $r \cdot b$ at a substantially faster rate than $\mathbf{A}^i\mathcal{K}_1^{r \cdot b}(\mathbf{A}, \mathbf{U}_1)$ if $|\lambda_{r \cdot b}| > |\lambda_{r \cdot b+1}|$. The theory also demonstrates that the convergence is not uniform; the invariant subspace of order ℓ is better approximated than the subspace of order $\ell + 1$ for $1 \leq \ell \leq r \cdot b - 1$.

Acknowledgments. I thank Steve Wright for a careful reading of the manuscript.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK Users' Guide*, SIAM, Philadelphia, second ed., 1995.

[2] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. J. Applied Mathematics, 9 (1951), pp. 17–29.

[3] J. Baglama, D. Calvetti, and L. Reichel, *Iterative methods for the computation of a few eigenvalues of a large symmetric matrix*, BIT, 36 (1996), pp. 400–440.

[4] ———, *Fast leja points*, ETNA, 7 (1998), pp. 124–140.

[5] J. Baglama, D. Calvetti, L. Reichel, and A. Ruttan, *Computation of a few close eigenvalues of a large matrix with application to liquid crystal modeling*, Journal of Computational Physics, 146 (1998), pp. 203–226.

[6] Z. Bai and G. W. Stewart, *SRRIT — A FORTRAN subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix*, ACM Trans. Mathematical Software, 23 (1997), pp. 494–513.

[7] F. L. Bauer, *Das Verfahren der Treppeniteration und verwandte Verfahren zur Losung algebraischer Eigenwertproblem*, Z. Angew. Mat. Phys., 8 (1957), pp. 214–235.

[8] F. Chatelin, *Eigenvalues of Matrices*, Wiley, 1993.

[9] I. S. Duff and J. A. Scott, *Computing selected eigenvalues of large sparse unsymmetric matrices using subspace iteration*, ACM Trans. Mathematical Software, 19 (1993), pp. 137–159.

[10] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, second ed., 1989.

[11] ———, *Matrix Computations*, Johns Hopkins University Press, Baltimore, third ed., 1996.

[12] G. H. Golub and J. H. Wilkinson, *Ill-conditioned eigensystems and the computation of the Jordan canonical form*, SIAM Review, 18 (1976), pp. 578–619.

[13] R. G. Grimes, J. G. Lewis, and H. D. Simon, *A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems*, SIAM J. Matrix Analysis and Applications, 15 (1994), pp. 228–272.

[14] Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*, SIAM J. Matrix Analysis and Applications, 16 (1995).

[15] ———, *Generalized block Lanczos methods for large unsymmetric eigenproblems*, Numerische Mathematik, 80 (1998), pp. 239–266.

[16] ———, *Polynomial characterizations of the approximate eigenvectors by the refined arnoldi method and an implicitly restarted refined arnoldi algorithm*, Linear Algebra and Its Applications, 287 (1998), pp. 191–214.

[17] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research of the National Bureau of Standards, 45 (1950), pp. 255–282. Research Paper 2133.

[18] R. B. Lehoucq and K. J. Maschhoff, *Block Arnoldi method*, in Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide, Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. V. der Vorst, eds., SIAM, To be published.

[19] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Phildelphia, PA, 1998.

[20] K. MEERBERGEN AND A. SPENCE, *Implicitly restarted Arnoldi with purification for the shift–invert transformation*, Mathematics of Computation, 218 (1997), pp. 667–689.
[21] R. MORGAN AND M. ZENG, *Harmonic projection methods for large non-symmetric eigenvalue problems*, Numerical Linear Algebra with Applications, 5 (1998), pp. 33–55.
[22] R. B. MORGAN, *Computing interior eigenvalues of large matrices*, Linear Algebra and Its Applications, 154/156 (1991), pp. 289–309.
[23] ———, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Mathematics of Computation, 65 (1996), pp. 1213–1230.
[24] C. C. PAIGE, B. N. PARLETT, AND H. A. VAN DER VORST, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numerical Linear Algebra with Applications, 2 (1995), pp. 115–134.
[25] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
[26] H. RUTIHAUSER, *Computational aspects of F.L. Bauer's simultaneous iteration method for symmetric matrices*, Numerische Mathematik, 13 (1969), pp. 4–13.
[27] ———, *Simultaneous iteration method for symmetric matrices*, Numerische Mathematik, 16 (1970), pp. 205–223.
[28] Y. SAAD, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra and Its Applications, 34 (1980), pp. 269–295.
[29] ———, *Projection methods for solving large sparse eigenvalue problems*, in Matrix Pencil Proceedings, B. Kågström and A. Ruhe, eds., vol. 973 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1982, pp. 121–144.
[30] ———, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Mathematics of Computation, 42 (1984), pp. 567–588.
[31] M. SADKANE, *A block Arnoldi–Chebyshev method for computing the leading eigenpairs of large sparse unsymmetric matrices*, Numerische Mathematik, 64 (1993), pp. 181–193.
[32] J. A. SCOTT, *An Arnoldi code for computing selected eigenvalues of sparse real unsymmetric matrices*, ACM Trans. Mathematical Software, 21 (1995), pp. 432–475.
[33] G. L. G. SLEIJPEN AND H. VAN DER VORST, *A Jacobi-Davidson iteration method for linear eigenvalue problems*, SIAM J. Matrix Analysis and Applications, 17 (1996), pp. 401–425.
[34] B. T. SMITH, J. M. BOYLE, J. J. D. B. S. GARBOW, Y. IKEBE, V. C. KLEMA, AND C. B. MOLER, *EISPACK Guide*, Springer–Verlag, Berlin, second ed., 1976. Volume 6 of *Lecture Notes in Computer Science*.
[35] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Analysis and Applications, 13 (1992), pp. 357–385.
[36] A. STATHOPOULOS, Y. SAAD, AND K. WU, *Dynamic thick restarting of the Davidson, and the implicitly restarted Arnoldi methods*, SIAM J. Scientific Computing, 19 (1998), pp. 227–245.
[37] G. W. STEWART, *Accelerating the orthogonal iteration for the eigenvectors of a Hermitian matrix*, Numerische Mathematik, 13 (1969), pp. 362–376.
[38] ———, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review, 15 (1973), pp. 727–764.
[39] ———, *Introduction to Matrix Computations*, Academic Press, San Diego, California, 1973.
[40] ———, *Simultaneous iteration for computing invariant subspaces of non-Hermitian matrices*, Numerische Mathematik, 25 (1976), pp. 123–136.
[41] J. M. VARAH, *On the separation of two matrices*, SIAM J. Numerical Analysis, 16 (1979), pp. 216–222.
[42] D. S. WATKINS, *Understanding the QR algorithm*, SIAM Review, 24 (1982), pp. 427–439.
[43] D. S. WATKINS AND L. ELSNER, *Convergence of algorithms of decomposition type for the eigenvalue problem*, Linear Algebra and Its Applications, 143 (1991), pp. 19–47.