TRUNCATED QR ALGORITHMS AND THE SOLUTION OF LARGE-SCALE EIGENVALUE PROBLEMS*

R. B. LEHOUCQ†

Abstract. The QR algorithm has emerged as the general-purpose method of choice for computing the Schur decomposition of a matrix. For most large eigenvalue problems, however, the QR algorithm cannot be used because of the explicit storage of the matrix and because often only the action of the matrix upon a vector (or group of vectors) is available. Typically, only a small number of eigenvalues and the associated invariant subspace are required. This article considers a truncated QR algorithm. We show that a truncated QR algorithm is a generalization of Sorensen's implicitly restarted Arnoldi method to block Arnoldi reductions. Moreover, implicitly restarting an Arnoldi reduction is simultaneous iteration with an implicit projection step to accelerate convergence to the invariant subspace of interest. This is a generalization of the Rayleigh–Ritz procedure on a block Krylov subspace for a non Hermitian matrix. The moral of our story is that the large scale eigenvalue problem is intimately involved with the dense one.

Key words. QR algorithm, simultaneous iteration, Arnoldi reduction, restarting, eigenvalues.

AMS subject classifications. 65F15, 65G05

1. Introduction.

The QR algorithm is a general-purpose method for computing all the eigenvalues of a matrix. The LR-iteration of Rutishauser [31], which preceded its discovery, is based on a triangular sequence of similarity transformation. The QR algorithm, developed independently by both Francis [11, 12] and Kublanovskaya [20], instead uses a sequence of unitary similarity transformations. The algorithm iteratively computes an approximation to a Schur decomposition of the matrix. The QR algorithm is implemented in the EISPACK [42] and LAPACK [1] software packages.

Unfortunately, for large-scale eigenvalue problems, the QR algorithm is not a practical method. An eigenvalue problem is considered *large* if it cannot be solved with the standard QR algorithm (as implemented in EISPACK and LAPACK). This QR algorithm relies on dense matrix similarity transformations that require explicit storage of the matrix. For most large eigenvalue problems, this requirement is prohibitive; quite often, only the action of the matrix upon a small group vectors is available. Moreover, users typically require only a small number of eigenvalues relative to the dimension of the problem. A further complication is that a representation for the associated invariant subspace is often required.

This article consider a truncated QR algorithm. We show that a truncated QR algorithm is equivalent to simultaneous iterations. This relationship allows us to exploit the well-known connection [29, 47, 50] between simultaneous iteration and the QR algorithm. In [48], Stewart presented a generalization of the Rayleigh–Ritz method to non-Hermitian matrices. This involved performing an *explicit* projection step on the matrix with orthonormal columns representing the subspace. We show that a block truncated QR algorithm performs simultaneous iteration with an *implicit* projection step. A block truncated QR algorithm is an extension of Sorensen's implicitly restarted Arnoldi method [43] to block Arnoldi reductions. The moral of our

* This work was supported in part by ARPA (U.S. Army ORA4466.01), by the U.S. Department of Energy (Contracts DE-FG0f-91ER25103 and W-31-109-Eng-38), and by the National Science Foundation (Cooperative agreement CCR-9120008).

† Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439 lehoucq@mcs.anl.gov, http://www.mcs.anl.gov/home/lehoucq/index.html.

story is that the large-scale eigenvalue problem is intimately involved with the dense one.

Since our goal is to provide a viewpoint in which to consider connecting the large-scale eigenvalue problem with the small one, this report works with general complex matrices. We will mention the standard simplifications when the matrix is Hermitian or can be computed in real arithmetic. The article is organized as follows. Notation is given in § 2 and § 3 where the eigenvalue problem introduced. The QR algorithm's connection with simultaneous iteration is the subject of § 4. The Arnoldi method, including a block formulation, is discussed in § 5. Because it will prove fundamental to the development of a QR algorithm, we discuss some of the technicalities of the implicit QR algorithm in § 6. We discuss how to compute a partial Schur decomposition from an Arnoldi reduction in § 7. The stage is finally set in § 8 for the main subject of a truncated QR algorithm. We review restarting methods for eigenvalue problems in § 9. The convergence of a truncated QR algorithm is discussed in § 10. Two important issues needed for a practical truncated QR algorithm are summarized in § 11.

2. Notation and Definitions. This section establishes 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 order of $\mathbf{A}$ will always be denoted by n. The identity matrix of order m is denoted by $\mathbf{I}_m$. The jth canonical basis vector is denoted by $\mathbf{e}_j$, the jth column of the identity matrix, and $\mathbf{E}_j \equiv \left[\begin{array}{ccc} \mathbf{e}_{(j-1)b+1} & \cdots & \mathbf{e}_{jb} \end{array} \right]$, where b is a positive integer. We will call b the block size.

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 band Hessenberg matrix is unreduced if all the elements on the bth subdiagonal are nonzero.

We now define several matrices that will prove useful. $\mathbf{H}_j$ denotes a band Hessenberg matrix of order bj of lower bandwidth b; $\mathbf{T}_j$ denotes an upper triangular matrix of order j, regardless of any block size b; and $\mathbf{F}_j$ and $\mathbf{U}_j$ denote matrices with n rows and b columns, where the subscript acts as an index. On the other hand, $\mathbf{V}_j$, $\mathbf{Z}_j$ denote matrices with n rows and bj columns. $\mathbf{U}_j$ denotes the jth block of b vectors of $\mathbf{V}_m$, and $\mathbf{G}_{i,j}$ denotes the square matrix of order b located in the i,jth block of order b of $\mathbf{H}_m$. Note that $\mathbf{G}_{j+1,j}$ is an upper triangular matrix. These matrices will define the dimensions of other matrices used in this article.

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 $\mathcal{R}(\mathbf{A})$.

3. The Eigenvalue Problem. Let $\mathbf{A}$ be a real matrix of order n. We are interested in a specified set of $k \ll n$ solutions to the matrix eigenvalue problem

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}. \tag{3.1}$$

The eigenvalues and eigenvectors of $\mathbf{A}$ are denoted by λ_j and $\mathbf{x}_j$, respectively, for $j = 1, \ldots, n$. We shall refer to these k eigenvalues as the *wanted* ones. The wanted eigenvalues of $\mathbf{A}$ requiring approximation typically are contained within some convex set of interest in the complex plane. Examples include those nearest the origin and of largest real part. An important exception might be the *dominant* eigenvalues of $\mathbf{A}$, those largest in magnitude. The following decomposition proves central to the

- Let $\mathbf{A}^{(0)} = \mathbf{A}$ and let $\{\tau_j\}_{j=1}^{p}$ be a sequence of real shifts. Set $\mathbf{Z}^{(1)} \leftarrow \mathbf{I}_n$.
- For $j = 1, \ldots, p$
  1. Compute the QR factorization $\mathbf{Q}^{(j)}\mathbf{R}^{(j)} = \mathbf{A}^{(j-1)} - \tau_{j-1}\mathbf{I}_n$.
  2. Update $\mathbf{A}^{(j)} \leftarrow \mathbf{R}^{(j)}\mathbf{Q}^{(j)} + \tau_{j-1}\mathbf{I}_n$.
  3. Update $\mathbf{Z}^{(j)} \leftarrow \mathbf{Z}^{(j)}\mathbf{Q}^{(j)}$.

FIG. 4.1. *The QR algorithm*

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 3.1. *(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*

$$\mathbf{Z}^H\mathbf{A}\mathbf{Z} = \mathbf{T}, \tag{3.2}$$

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

Proof. See [14, page 313]. □

Let $\mathbf{D}$ be a diagonal unitary matrix. Then $(\mathbf{ZD})^H\mathbf{AZD} = \mathbf{D}^H\mathbf{TD}$ has diagonal blocks equal to those of $\mathbf{T}$. Thus, apart from the eigenvalues of multiplicity larger than one, the decomposition is essentially unique, given some ordering of the eigenvalues. Denote the leading principal matrix of k blocks of $\mathbf{T}$ by $\mathbf{T}_k$. Let $\mathbf{Z}_k$ be the corresponding columns of $\mathbf{Z}$. Then $\mathbf{AZ}_k = \mathbf{Z}_k\mathbf{T}_k$ 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.

This decomposition is computed by the practical QR algorithm in the EISPACK and LAPACK software packages. A real Schur decomposition allows all computation to take place in real arithmetic; see [14, p. 341] for further details. There is also software to reorder the computed Schur decomposition.

The methods reported here attempt to compute a partial Schur decomposition for $\mathbf{A}$ with the group of the wanted eigenvalues located on the diagonal blocks of $\mathbf{T}_k$. The methods considered require $\mathcal{O}(kn)$ storage and $\mathcal{O}(kn^2)$ work. The full decomposition requires $\mathcal{O}(n^2)$ storage and $\mathcal{O}(n^3)$ work. We say an eigenvalue problem is large if the dense QR algorithm is prohibitive, in storage and/or efficiency.

4. The QR Algorithm. We quickly examine the QR algorithm and some of its fundamental properties. A wealth of excellent material exists on the QR algorithm. Thorough introductions are given by Golub and Van Loan [14], Parlett [29], Stewart [47], Watkins [50, 51] and of course Wilkinson [53]. Figure 4.1 lists the explicitly shifted QR iteration.

The following properties are consequences of the iteration. They are easily established using mathematical induction; see, for example, [47, pp. 351–354]. Assume the notation of the algorithm listed in Figure 4.

THEOREM 4.1. $\mathbf{AZ}^{(p)} = \mathbf{Z}^{(p)}\mathbf{A}^{(p)}$.

THEOREM 4.2. *Let* $\mathbf{T}^{(p)} = \mathbf{R}^{(p)}\cdots\mathbf{R}^{(1)}$*. Then* $\mathbf{Z}^{(p)}\mathbf{T}^{(p)} = \mathcal{P}(\mathbf{A})$*, where* $\mathcal{P}(\lambda) = (\lambda - \tau_1)\cdots(\lambda - \tau_p)$.

Theorem 4.1 gives that $\mathbf{A}^{(p)}$ is unitarily similar to $\mathbf{A}$. What is remarkable is that the off-diagonal elements in the last row of $\mathbf{A}^{(p)}$ approach zero with the choice of zero

shifts. The second theorem explains why. It follows that

$$\mathbf{A}^{(p)} \left[\begin{array}{ccc} \mathbf{e}_1 & \cdots & \mathbf{e}_i \end{array} \right] = \mathbf{Z}_i^{(p)} \mathbf{T}_i^{(p)}, \tag{4.1}$$

where $\mathbf{T}_i^{(p)}$ is the leading principal matrix of order i of $\mathbf{T}^{(p)}$ and $\mathbf{Z}_i^{(p)}$ contains the first i columns of $\mathbf{Z}_i^{(p)}$. Thus, the QR algorithm computes the QR factorization of the matrix on the left-hand side of (4.1). But this expression is nothing more than simultaneous iteration (with the starting subspace being the span of the first b columns of the identity matrix). Since $1 \leq b \leq n$, the QR algorithm is performing a nested sequence of simultaneous iteration. Orthonormal iteration is a more descriptive term because the columns of $\mathbf{Z}^{(p)}$ are themselves orthonormal.

We emphasize that the above two theorems imply that after the initial QR factorization of $\mathbf{A}$, orthonormal iteration can be carried out *without* explicit use of $\mathbf{A}$. It is also clear that the QR algorithm is invariant under a unitary change of basis. Together, these two observations suggest that a similarity transformation might allow us to perform orthonormal iteration more efficiently. Indeed, a truncated QR algorithm is motivated precisely by this observation.

The basic convergence of the iteration is easily established. If the eigenvalues of $\mathbf{A}$ are ordered in decreasing order of magnitude, it can be shown (under a mild condition on the starting subspace) that the off-diagonal elements in row i to left of the diagonal element converge to zero at a rate of proportional to $|\lambda_i/\lambda_{i-1}|$. Thus, the $\mathbf{Z}_i^{(p)}\mathbf{T}_i^{(p)}$ tends toward a partial Schur decomposition for $\mathbf{A}$ associated with the dominant i eigenvalues. See [14, p. 333] for details.

A comprehensive geometric convergence theory for the shifted QR iteration is presented by Watkins and Elsner [52] within the more general framework of generic *GR algorithms*. A GR algorithm is an iterative procedure in which the QR factorization is replaced with any other decomposition of the form $\mathbf{GR} = \mathbf{H} - \tau\mathbf{I}$, where $\mathbf{R}$ is upper triangular and $\mathbf{G}$ is a nonsingular matrix.

4.1. A Practical QR Algorithm. We list and briefly discuss the issues involved in a practical implementation of the QR algorithm. The remainder of the report will discuss these issues in more detail when when we wish to draw analogies between a full and truncated QR algorithm.

1. Initial reduction to upper Hessenberg form. $\mathbf{A}$ is initially reduced to upper Hessenberg form via a unitary similarity transformation. Each step of the resulting QR algorithm then becomes an $\mathcal{O}(n^2)$ process instead of an $\mathcal{O}(n^3)$ one.
2. Selection of shifts. In practice, a set of shifts are computed that lead to quadratic and cubic rates of convergence for non-Hermitian and Hermitian matrices, respectively.
3. Deflation. Since the off-diagonal elements in the last row $\mathbf{A}^{(j)}$ tend to converge to zero rapidly, they are set to zero, and the last diagonal element of $\mathbf{A}$ is an approximation to an eigenvalue. This process continues up the diagonal of $\mathbf{A}^{(j)}$. After $n-1$ such deflations, a Schur decomposition has been computed.
4. The *implicitly shifted* QR iteration. The explicit computation of the QR factorization is not carried out. Instead, the implicit Q theorem allows computation of $\mathbf{Z}^{(j)}$ and $\mathbf{A}^{(j)}$ to be interleaved.
5. Possible reordering of the Schur decomposition. If only k eigenvalues are of interest, these can be efficiently moved to the leading portion of the final

Schur matrix.

The last point gives a simple method for computing a partial Schur decomposition of interest. Its drawback is the $\mathcal{O}(n^2)$ storage required, as well as the $\mathcal{O}(n^3)$ work associated with the computation of a (full) Schur decomposition.

In the remainder of this article, we will develop the idea of a truncated QR algorithm that only requires $\mathcal{O}(kn)$ storage and $\mathcal{O}(n^2)$ work for computing a partial Schur decomposition.

5. Partial Reduction to Band Hessenberg Form. The initial step of the practical QR algorithm reduces $\mathbf{A}$ to an upper Hessenberg matrix via a sequence of elementary unitary matrices. Unfortunately, these elementary matrices require accessing the entire matrix, possibly destroying any sparsity or structure the matrix possess. The Arnoldi reduction [2], on the other hand, requires only the application of $\mathbf{A}$ with a vector. Moreover, it allows us to sequentially reduce $\mathbf{A}$ to upper Hessenberg form, producing the leading portion of the final upper Hessenberg matrix at every step. In fact, this was the motivation in Arnoldi's study. When the matrix $\mathbf{A}$ is Hermitian, the Lanczos reduction [21] 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 [39, 40] are considered. In many instances, the cost of computing a few matrix vector products is commensurate with that of one matrix vector product. There is also the issue of reliably computing clustered and/or multiple eigenvalues. See [17] for references and information on a block Lanczos reduction.

Let $b > 0$, an integer, be the block size. We say that

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + \mathbf{F}_m\mathbf{E}_m^T \tag{5.1}$$

is a block Arnoldi reduction of length m when $\mathbf{V}_m^H\mathbf{A}\mathbf{V}_m = \mathbf{H}_m$ is a banded upper Hessenberg matrix, $\mathbf{V}_m^H\mathbf{V}_m = \mathbf{I}_{mb}$, and $\mathbf{V}_m^H\mathbf{F}_m = \mathbf{0}$.

The columns of $\mathbf{V}_m$ are an orthogonal basis for the block Krylov subspace

$$\mathcal{K}_m(\mathbf{A}, \mathbf{U}_1) \equiv \{\mathbf{U}_1, \mathbf{A}\mathbf{U}_1, \cdots, \mathbf{A}^{m-1}\mathbf{U}_1\}.$$

$\mathbf{H}_m$ is the projection of $\mathbf{A}$ onto the column span of $\mathbf{V}_m$. If $m > \bar{m} \equiv \text{ceiling}(n/b)$, then $\mathbf{F}_m = \mathbf{0}$ and $\mathbf{H}_{\bar{m}}$ is the orthogonal reduction of $\mathbf{A}$ into banded upper Hessenberg form. Note that if $\mathbf{A} = \mathbf{A}^H$, then $\mathbf{H}_m$ is a block tridiagonal matrix.

Figure 5.1 lists an algorithm to compute a block Arnoldi reduction. In practical computation, two steps of orthogonalization are needed to ensure that $\mathbf{V}_{m+1}$ is orthonormal to $\mathbf{F}_{m+1}$. If $\mathbf{F}_m$ is rank deficient, then care must be taken to fill out the block with vectors that result in a $\mathbf{V}_{m+1}$ with orthonormal columns. Using the notation established in § 2, we have

$$\mathbf{A}\,\mathbf{V}_m = \begin{bmatrix} \mathbf{U}_1 & \cdots & \mathbf{U}_m \end{bmatrix} \begin{bmatrix} \mathbf{G}_{1,1} & \cdots & \cdots & \mathbf{G}_{1,m} \\ \mathbf{G}_{2,1} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{G}_{m,m-1} & \mathbf{G}_{m,m} \end{bmatrix} + \mathbf{U}_{m+1}\mathbf{G}_{m+1,m}\mathbf{E}_m^T$$

.

The following classical result explains that a block Arnoldi reduction is completely specified by the starting block.

- Let $\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + \mathbf{F}_m\mathbf{E}_m^T$ be a length-m block Arnoldi reduction
  1. Compute the orthogonal factorization $\mathbf{U}_{m+1}\mathbf{G}_{m+1,m} = \mathbf{F}_m$.
  2. $\mathbf{V}_{m+1} = \begin{bmatrix} \mathbf{V}_m & \mathbf{U}_{m+1} \end{bmatrix}$.
  3. $\mathbf{W} = \mathbf{A}\mathbf{U}_{m+1}$ and $\mathbf{G}_{m+1,m+1} = \mathbf{U}_{m+1}^H\mathbf{W}$.
  4. $\mathbf{H}_{m+1} = \begin{bmatrix} \mathbf{H}_m & \mathbf{V}_m^T\mathbf{W} \\ \mathbf{G}_{m+1,m}\mathbf{E}_m^T & \mathbf{G}_{m+1,m+1} \end{bmatrix}$.
  5. $\mathbf{F}_{m+1} = \mathbf{W} - \mathbf{V}_{m+1}\begin{bmatrix} \mathbf{V}_m^T\mathbf{W} \\ \mathbf{G}_{m+1,m+1} \end{bmatrix}$.

FIG. 5.1. *Extending a Block Arnoldi Reduction*

THEOREM 5.1. *(Implicit Q) Let two length-m block Arnoldi reductions be given by*

$$\mathbf{A}\mathbf{W}_m = \mathbf{W}_m\mathbf{B}_m + \mathbf{C}_m\mathbf{E}_m^T,$$
$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + \mathbf{F}_m\mathbf{E}_m^T,$$

where $\mathbf{W}_m$ *and* $\mathbf{V}_m$ *have orthonormal columns, and* $\mathbf{H}_m$ *and* $\mathbf{B}_m$ *are band Hessenberg matrices with positive elements on the* b*th subdiagonal. If the first* b *columns of* $\mathbf{W}_m$ *and* $\mathbf{V}_m$ *are equal and* $\mathbf{W}_m^H\mathbf{C}_m = \mathbf{0} = \mathbf{V}_m^H\mathbf{F}_m$*, then* $\mathbf{H}_m = \mathbf{B}_m$*,* $\mathbf{W}_m = \mathbf{V}_m$ *and* $\mathbf{C}_m = \mathbf{F}_m$*.*

Proof. The extension of [14, page 367] to a block formulation is straightforward. The requirement on the elements on the bth subdiagonal is equivalent to uniquely specifying the QR factorization of $\mathbf{F}_j$ for $j = 1, \ldots, m-1$. □

6. The Implicit QR Algorithm on Band Hessenberg Matrices. We now present a technical lemma that will provide useful in the remainder of the report. One of its conclusions is that a step of the QR algorithm on a band upper Hessenberg matrix remains one.

LEMMA 6.1. *Let* $\mathbf{H} - \tau\mathbf{I} = \mathbf{Q}\mathbf{R}$ *be a QR factorization, where* $\mathbf{H}$ *is an unreduced upper Hessenberg matrix of order* $m_b \equiv mb$ *with lower bandwidth* b*. Denote* $\mathbf{e}_i^T\mathbf{R}\mathbf{e}_i = \rho_i$*. Then the following properties hold:*

1. $\mathbf{Q}$ *is an upper Hessenberg matrix with lower bandwidth* b*.*
2. $\rho_i \neq 0$ *for* $i = 1, \ldots, m_b - 1$*.*
3. $\rho_{m_b} = 0$ *if and only if* τ *is an eigenvalue of* $\mathbf{H}$*.*
4. $\mathbf{e}_{m_b}^T(\mathbf{R}\mathbf{Q} + \tau\mathbf{I}) = \tau\mathbf{e}_{m_b}^T$ *if and only if* τ *is an eigenvalue of* $\mathbf{H}$*.*

Proof. A sequence of elementary unitary matrices $\mathbf{P}_i$ is easily constructed so that

$$\mathbf{P}_{m_b-1}^H \cdots \mathbf{P}_1^H(\mathbf{H} - \tau\mathbf{I})$$

is upper triangular [14, p. 233]. Each $\mathbf{P}_i$ is designed to annihilate the entries below the diagonal element of $\mathbf{P}_{i-1}^H \cdots \mathbf{P}_1^H(\mathbf{H} - \tau\mathbf{I})\mathbf{e}_i$. The product $\mathbf{P}_1 \cdots \mathbf{P}_{m_b-1}$ is band upper Hessenberg, and $\mathbf{P}_{m_b-1}^H \cdots \mathbf{P}_1^H(\mathbf{H} - \tau\mathbf{I})$ is upper triangular. Set $\mathbf{Q} = \mathbf{P}_1 \cdots \mathbf{P}_{m_b-1}$ and $\mathbf{R} = \mathbf{Q}^H(\mathbf{H} - \tau\mathbf{I})$.

A simple calculation reveals that $\mathbf{e}_{i+b}^T(\mathbf{H} - \tau\mathbf{I})\mathbf{e}_i = \mathbf{e}_{i+b}^T\mathbf{Q}\mathbf{e}_i\rho_i$. Since $\mathbf{H}$ is an unreduced band upper Hessenberg matrix, $0 < |\mathbf{e}_{i+b}^T\mathbf{H}\mathbf{e}_i| = |\mathbf{e}_{i+b}^T\mathbf{Q}\mathbf{e}_i|\ |\rho_i| \leq |\rho_i|$ for $i = 1, \ldots, m_b - 1$, establishing the second property.

The matrix $\mathbf{H} - \tau\mathbf{I}$ is singular if and only if τ is an eigenvalue of $\mathbf{H}$. The third property follows immediately, since $\det(\mathbf{H} - \tau\mathbf{I}) = \det(\mathbf{R}) = \rho_1 \cdots \rho_{m_b}$ is zero if and only if ρ_{m_b} is.

The third property gives $\rho_{m_b} = 0$ if τ is an eigenvalue of $\mathbf{H}$. Since $\mathbf{e}_{m_b}^T \mathbf{R} = \mathbf{e}_{m_b}^T \rho_{m_b}$, the final property holds. □

When $b = 1$, we may substitute plane rotations in the above lemma. Note that when $b = 1$, the QR factorization is an $\mathcal{O}(b(mb)^2)$ process. Thus, for small values of b the QR algorithm is an $\mathcal{O}((mb)^2)$ process.

Practical deflation procedures are motivated by Part 4 of Lemma 6.1. If the elements in the last row of $\mathbf{RQ} + \tau\mathbf{I} = \mathbf{Q}^H\mathbf{HQ}$ to the left of the diagonal element are small, the diagonal element is regarded as an approximation to an eigenvalue. The off-diagonal elements are set to zero, and the QR algorithm works on the leading principal matrix of order $bm - 1$ of $\mathbf{H}$. This procedure is summarized in the following result.

LEMMA 6.2. *Suppose p steps of the QR algorithm are applied on a band Hessenberg matrix of order bm. Let $\mathbf{Q}$ denote the accumulation of the unitary matrices in the QR algorithm so that $\mathbf{HQ} = \mathbf{QH}^+$.*

If the p shifts are eigenvalues of $\mathbf{H}$, then

$$\mathbf{H}^+ = \begin{bmatrix} \mathbf{H}^+_{1,1} & \mathbf{H}^+_{1,2} \\ \mathbf{0} & \mathbf{H}^+_{2,2} \end{bmatrix},$$

where $\mathbf{H}_{1,1}$ contains the $bm - p$ eigenvalues that were not used as shifts.

Proof. The proof is by induction on p. The base case $p = 1$ is just Part 4 of Lemma 6.1. Assume the lemma's truth for p. Re-apply Part 4 of Lemma 6.1 on $\mathbf{H}^+_{1,1}$ to establish the lemma's conclusion for $p + 1$ eigenvalues as shifts. □

This lemma implies that the last p columns of $\mathbf{Q}$ are an unitary basis of the left eigenspace associated with these deflated eigenvalues.

As remarked in § 4.1, practical implementations use the implicit QR algorithm. We first outline this procedure and discuss its implications. Theorems 4.1 and 4.2 imply that only the unitary matrix matters, while the Implicit Q Theorem uniquely specifies any procedure computing a partial band Hessenberg reduction. Let's put all this together.

Suppose we have the shifts $\tau_1, \ldots, \tau_p$ and define the polynomial $\mathcal{P}(\lambda) = (\lambda - \tau_1) \cdots (\lambda - \tau_p)$. The implicit QR algorithm first computes the QR factorization of

$$\mathcal{P}(\mathbf{H}_m)\mathbf{E}_1 = \mathbf{U}_1\mathbf{R}_b \equiv \begin{bmatrix} \mathbf{U}_{11} \\ \mathbf{0} \end{bmatrix} \mathbf{R}_b,$$

where $\mathbf{R}_b$ is an upper triangular matrix of order b and $\mathbf{U}_{11}$ has $(p + 1)b$ rows and b columns. Then, the similarity transformation

$$\mathbf{H}^+_m = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(m-p-1)b} \end{bmatrix}^H \mathbf{H}_m \begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(m-p-1)b} \end{bmatrix}$$

is performed. This updated matrix is returned to band Hessenberg form via a sequence of elementary unitary matrices. This is a straightforward generalization of the familiar *bulge chasing sweeps* in the standard ($b = 1$) Hessenberg QR algorithm. The crucial observation is that these bulge chasing sweeps in the band Hessenberg QR algorithm do not modify $\mathbf{U}_1$. Thus, the implicit Q theorem implies that the implicit QR algorithm on band Hessenberg matrices is equivalent to performing the explicit version of the algorithm.

Francis [12] originally proposed use of the implicit QR algorithm in order to perform the the algorithm on real matrices in real arithmetic. This allows a complex

conjugate pair of shifts to be applied by using the degree 2 polynomial $(\lambda-\tau_1)(\lambda-\bar{\tau}_1)$. The LAPACK subroutines _SHEQR also allow more than two shifts to be applied. This is based on the multi-shift QR algorithm [5] by Bai and Demmel for upper ($b = 1$) Hessenberg matrices.

7. Computing Eigenvalues. Since our interest is in partial Schur decompositions, we use employ the following convergence criterion. Suppose that $\mathbf{H}_m\mathbf{Y}_m = \mathbf{Y}_m\mathbf{T}_m$ is a Schur decomposition ordered so that the k best approximations to the eigenvalues of interest are located in the initial portion of $\mathbf{T}_m$. Thus,

$$\|\mathbf{A}\mathbf{V}_m\mathbf{Y}_k - \mathbf{V}_m\mathbf{Y}_k\mathbf{T}_k\| = \|\mathbf{F}_m\mathbf{E}_m^T\mathbf{Y}_k\| = \|\mathbf{G}_{m+1,m}\mathbf{E}_m^T\mathbf{Y}_k\| \tag{7.1}$$

where $\mathbf{H}_m\mathbf{Y}_k = \mathbf{Y}_k\mathbf{T}_k$ and the first k columns of $\mathbf{Y}_m$ are denoted by $\mathbf{Y}_k$. In words, the last b rows of $\mathbf{Y}_k$ need to be small. This Schur decomposition based criterion will always ensure that we compute a partial Schur decomposition of a nearby matrix. If approximate eigenvectors are of interest, they can be computed from $\mathbf{T}_k$. If $\mathbf{T}_k\mathbf{s}_i = \mathbf{s}_i\theta_i$, $1 \le i \le k$, then

$$\mathbf{A}\mathbf{V}_m\mathbf{Y}_k\mathbf{s}_i - \mathbf{V}_m\mathbf{Y}_k\mathbf{s}_i\theta = \mathbf{F}_m\mathbf{E}_m^T\mathbf{Y}_k\mathbf{s}_i \tag{7.2}$$

and so $\|\mathbf{A}\mathbf{V}_m\mathbf{Y}_k\mathbf{s}_i - \mathbf{V}_m\mathbf{Y}_k\mathbf{s}_i\theta_i\| = \|\mathbf{F}_m\mathbf{E}_m^T\mathbf{Y}_k\mathbf{s}_i\|$. We call $\mathbf{V}_m\mathbf{Y}_k\mathbf{s}_i$ a Ritz vector and θ_i a Ritz value. Note that $\mathbf{Y}_k\mathbf{s}_1 = \mathbf{Y}_k\mathbf{e}_1$. The first Schur vector is always an eigenvector.

For symmetric $\mathbf{A}$, Saad [34] shows that as m increases, the quality of the extremal Ritz values improves to the well-separated extremal eigenvalues of $\mathbf{A}$. For the unblocked Arnoldi reduction, he also shows [36] a similar improvement for increasing m. Unfortunately, given a large value of n, the value of m needed for a desired degree of approximation may be impractical because of storage constraints required for the Arnoldi vectors. This situation is particularly exacerbated for non-Hermitian $\mathbf{A}$.

Because of the connection with a block Krylov space, a block Arnoldi Reduction is a generalization of subspace iteration in that a sequence of subspaces are joined together. Let $\mathbf{K}_m(\mathbf{A},\mathbf{U}_1) \equiv \begin{bmatrix} \mathbf{U}_1 & \mathbf{A}\mathbf{U}_1 & \cdots & \mathbf{A}^{m-1}\mathbf{U}_1 \end{bmatrix}$. The relationship

$$\mathbf{A}\,\mathbf{K}_m(\mathbf{A},\mathbf{U}_1) = \mathbf{K}_m(\mathbf{A},\mathbf{U}_1)\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Omega}_0 \\ \mathbf{I}_b & & \vdots & \mathbf{\Omega}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_b & \mathbf{\Omega}_{m-1} \end{bmatrix} + \tilde{\mathbf{F}}_m\mathbf{E}_m^T \tag{7.3}$$

holds with each $\mathbf{\Omega}_i$ a matrix of order b. We denote the square matrix of order bm by $\mathbf{C}_m$. Equation (7.3) is equivalent to the least-squares problem

$$\min\|(\mathbf{A}^m\mathbf{U}_1) - \mathbf{K}_m(\mathbf{A},\mathbf{U}_1)\mathbf{\Omega}\| = \|\tilde{\mathbf{F}}_m\|. \tag{7.4}$$

Let $\mathbf{\Omega}^*$ be the least-squares solution partitioned conformably with the last column of the square matrix in (7.3).

Denote the QR factorization of $\mathbf{K}_m(\mathbf{A},\mathbf{U}_1) = \mathbf{W}_m\mathbf{R}_m$, where $\mathbf{R}_m$ is an upper triangular matrix of order bm. If $\mathbf{R}_m$ is invertible, then

$$\mathbf{A}\,\mathbf{W}_m = \mathbf{W}_m(\mathbf{R}_m\mathbf{C}_m\mathbf{R}_m^{-1}) + \tilde{\mathbf{F}}_m(\mathbf{E}_m^T\mathbf{R}_m^{-1}\mathbf{E}_m)\mathbf{E}_m^T,$$

where we use the identity $\mathbf{E}_m^T\mathbf{R}_m^{-1} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_m^T\mathbf{R}_m^{-1}\mathbf{E}_m \end{bmatrix}$. By the implicit Q theorem, it follows that $\mathbf{W}_m = \mathbf{V}_m$, $\tilde{\mathbf{F}}_m(\mathbf{E}_m^T\mathbf{R}_m^{-1}\mathbf{E}_m) = \mathbf{F}_m$ and $\mathbf{H}_m = \mathbf{R}_m\mathbf{C}_m\mathbf{R}_m^{-1}$,

since the matrix of the first b columns of $\mathbf{K}_m(\mathbf{A}, \mathbf{U}_1)$ is equal $\mathbf{U}_1$. Thus, the starting block characterizes the reduction. The residual matrix $\mathbf{F}_m$ is a matrix polynomial function of the initial block.

Ruhe [30] showed that the least-squares solution of (7.4) gives the coefficients associated with the monic polynomial of degree m that minimizes $\|\hat{\psi}_m(\mathbf{A})\mathbf{u}_1\|$ over all monic polynomials $\hat{\psi}_m$ of degree m. Saad [36] uses projection arguments to solve the minimization problem. For a block Arnoldi reduction, the situation is more complicated and there does not appear to be an equivalent minimization property. See [41] for a characterization of a block Arnoldi process in terms of matrix polynomials.

8. A Truncated QR Algorithm. The following elementary but technical result is needed for the connection with simultaneous iteration we desire.

LEMMA 8.1. *Suppose that an integer p satisfies $2 \le p \le m$, 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

$$\psi_p(\mathbf{A})\mathbf{V}_m = \mathbf{V}_m\psi_p(\mathbf{H}_m) + \sum_{j=1}^{p} \psi_{j+1}^{p}(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{p-j}(\mathbf{H}_m), \tag{8.1}$$

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

Moreover,

$$\psi_p(\mathbf{A})\mathbf{V}_r = \mathbf{V}_m\psi_p(\mathbf{H}_m)\begin{bmatrix} \mathbf{E}_1 & \cdots & \mathbf{E}_r \end{bmatrix} + \tilde{\mathbf{F}}_1 \tag{8.2}$$

where $\tilde{\mathbf{F}}_1 \equiv \psi_2^p(\mathbf{A})\mathbf{F}_m\mathbf{E}_m^T\psi_{p-1}(\mathbf{H}_m)\mathbf{E}_1$.

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 $j \le p$. Let $\psi_{p+1}(\lambda) = (\lambda - \tau_{p+1})\psi_p(\lambda)$. With the induction hypothesis, it follows that

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

which establishes Equation (8.1).

Since $\mathbf{H}$ is unreduced, $\psi_{p-j}(\mathbf{H})$ is a band Hessenberg matrix of lower bandwidth $(p-j)b$. Thus $\mathbf{E}_i^T \psi_{p-j}(\mathbf{H}) \mathbf{E}_l = \mathbf{0}$ for $l - p + j < i$, and the last matrix on the right-hand side of Equation (8.1) is zero through its first rb columns. Equation (8.2) is established. $\tilde{\mathbf{F}}_1$ is zero except when $p = m$. □

In plain words, Equation (8.2) shows that $\psi_p(\mathbf{A})$ applied to the first br columns of $\mathbf{V}_m$ is equivalent $\psi_p(\mathbf{H}_m)$ acting on the subspace consisting of the span of the first br columns of $\mathbf{I}_{bm}$. The unitary basis constructed by the block Arnoldi reduction provides the change of basis needed for the equivalence. The fundamental implication here is that considerable computation can be avoided by working with a significantly smaller $\mathbf{H}_m$ in the coordinate system given by $\mathbf{V}_m$. We remark that for degree one polynomials $\psi_1(\cdot)$, the block Arnoldi reduction is not truncated.

Recall from Theorems 4.1 and 4.2 that the unitary matrix $\mathbf{Z}^{(p)}$ links simultaneous iteration with the QR algorithm. For a truncated QR algorithm, the previous lemma nearly provides the crucial link.

Compute the QR factorization $\mathbf{Q}_r \mathbf{R}_r = \psi_p(\mathbf{H}_m) \begin{bmatrix} \mathbf{E}_1 & \cdots & \mathbf{E}_r \end{bmatrix}$. From (8.2), we obtain

$$\psi_p(\mathbf{A}) \mathbf{V}_r = \mathbf{V}_m \mathbf{Q}_r \mathbf{R}_r + \tilde{\mathbf{F}}_1 \equiv \mathbf{V}_r^+ \mathbf{R}_r + \tilde{\mathbf{F}}_1 .$$

This result gives rise to a truncated version of Theorem 4.1 with the starting subspace defined by the span of the columns of $\mathbf{V}_r$. This establishes the following theorem.

THEOREM 8.2. *Assume the hypothesis of Lemma 8.1 with $p < m$. Let the QR factorization $\mathbf{Q}_r \mathbf{R}_r = \psi_p(\mathbf{H}_m) \begin{bmatrix} \mathbf{E}_1 & \cdots & \mathbf{E}_r \end{bmatrix}$ be given.*

Then, the columns of $\mathbf{V}_m \mathbf{Q}_r$ provide an unitary basis for the $\mathcal{R}(\psi_p(\mathbf{A}) \mathbf{V}_r)$.

The theorem allows us to filter the eigenvalues of $\mathbf{H}_m$ in a desired order. By applying a polynomial $\psi_p(\cdot)$ that emphasizes the desired (or damps the unwanted) eigenvalues, the column space of $\mathbf{V}_m$ is ordered into $\mathbf{V}_m \mathbf{Q}_r$. This is a generalization of the acceleration technique discussed by Stewart [45, 48] for simultaneous iteration. This will be discussed further in § 8.1.

We may now exploit the connection between simultaneous iteration and a QR iteration. This will allow us to affect an ordering of the eigenvalues of $\mathbf{H}_m$ suggested by Theorem 8.2 without having to compute (or apply) the matrix polynomial $\psi_p(\mathbf{H}_m)$.

The QR algorithm on $\mathbf{H}_m$ with the $p < m$ shifts $\tau_1, \ldots, \tau_p$ gives $\mathbf{H}_m \mathbf{Q}_m = \mathbf{Q}_m \mathbf{H}_m^+$. Lemma 6.1 gives that $\mathbf{Q}_m$ is a Hessenberg matrix of lower bandwidth bp, since it is a product of p unitary matrices each of lower bandwidth b. If we equate the first rb columns of the previous matrix equality, we get

$$\mathbf{H}_m \mathbf{Q}_r \equiv \begin{bmatrix} \mathbf{Q}_r & \mathbf{W}_1 & \mathbf{W}_2 \end{bmatrix} \begin{bmatrix} \mathbf{H}_r^+ \\ \mathbf{G}_{r+1,r}^+ \mathbf{E}_r^T \\ \mathbf{0} \end{bmatrix} . \tag{8.3}$$

Post-multiplying Equation (5.1) with $\mathbf{Q}_r$ and using (8.3), we obtain

$$\begin{aligned} \mathbf{A} \mathbf{V}_m \mathbf{Q}_r &= \mathbf{V}_m \mathbf{H}_m \mathbf{Q}_r + \mathbf{F}_m \mathbf{E}_m^T \mathbf{Q}_r , \\ &= \mathbf{V}_m \mathbf{Q}_r \mathbf{H}_r^+ + \mathbf{V}_m \mathbf{W}_1 \mathbf{G}_{r+1,r}^+ \mathbf{E}_r^T + \mathbf{F}_m \mathbf{E}_m^T \mathbf{Q}_r , \\ &= \mathbf{V}_m \mathbf{Q}_r \mathbf{H}_r^+ + \mathbf{F}_r^+ \mathbf{E}_r^T , \end{aligned} \tag{8.4}$$

where

$$\mathbf{F}_r^+ \equiv \mathbf{V}_m \mathbf{W}_1 \mathbf{G}_{r+1,r}^+ + \mathbf{F}_m \mathbf{E}_m^T \mathbf{Q}_r \mathbf{E}_r^T . \tag{8.5}$$

Note the use of the identity $\mathbf{E}_m^T\mathbf{Q}_r = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_m^T\mathbf{Q}_r\mathbf{E}_r \end{bmatrix}$ in Equation (8.5). This proves the following result.

THEOREM 8.3. *Assume the hypothesis of Lemma 8.1 with $p < m$. Suppose that the QR algorithm on $\mathbf{H}_m$ with the p shifts $\tau_1, \ldots, \tau_p$ gives $\mathbf{H}_m\mathbf{Q}_m = \mathbf{Q}_m\mathbf{H}_m^+$, where $\mathbf{Q}_m$ is a band Hessenberg matrix with lower bandwidth pb. If $\mathbf{Q}_r$ is the matrix consisting of the first rb columns of $\mathbf{Q}_m$, then*

$$\mathbf{A}\mathbf{V}_m\mathbf{Q}_r = \mathbf{V}_m\mathbf{Q}_r\mathbf{H}_r^+ + \mathbf{F}_r^+\mathbf{E}_r^T, \tag{8.6}$$

where equations (8.3) and (8.5) define $\mathbf{F}_r^+$.

If $p = m$, the previous theorem gives that application of $m - 1$ shifts gives

$$\mathbf{A}\mathbf{U}_1^+ = \mathbf{U}_1^+\mathbf{H}_1^+ + \mathbf{F}_1^+. \tag{8.7}$$

If $\mathbf{QR} = \mathbf{H}_1^+ - \tau_m\mathbf{I}$, post-multiplication of (8.7) results in

$$\mathbf{A}\mathbf{U}_1^+\mathbf{Q} = \mathbf{U}_1^+(\mathbf{R}\mathbf{Q} + \tau_m\mathbf{I}) + \mathbf{F}_1^+\mathbf{Q} = \mathbf{U}_1^+\mathbf{H}_1^{++} + \mathbf{F}_1^+\mathbf{Q}. \tag{8.8}$$

The right-hand side of (8.8) defines a new starting block of vectors (after orthogonalization) for a subsequent block Arnoldi reduction.

The two theorems allow us to link simultaneous iteration with a truncated QR algorithm. They show how a QR algorithm performed on $\mathbf{H}_m$ is equivalent to a truncated QR algorithm on $\mathbf{A}$.

8.1. Subspace Iteration. A classical method of solution for the large-scale eigenvalue problem is subspace (or simultaneous) iteration [6, 10, 32, 33, 37, 45, 48]. Subspace 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 and is a generalization of the power method in that a matrix representation of a subspace of size larger than one is employed.

Suppose we have the length one block Arnoldi reduction

$$\mathbf{A}\mathbf{U}_1 = \mathbf{U}_1\mathbf{G}_{1,1} + \mathbf{F}_1. \tag{8.9}$$

Since the block size is b, $\mathbf{G}_{1,1}$ is a dense matrix of that order representing the projection of $\mathbf{A}$ onto the column span of $\mathbf{U}_1$. Equation (8.9) is nothing more than a step of simultaneous (or orthonormal) iteration in matrix form with error $\mathbf{F}_1$.

If $\mathbf{G}_{1,1}\mathbf{W} = \mathbf{W}\mathbf{T}_b$ is Schur decomposition ordered in decreasing order of magnitude, then the initial columns of $\mathbf{W}$ contain the directions associated with the dominant eigenvalues. Post-multiplying (8.9) with $\mathbf{W}$ gives

$$\mathbf{A}\mathbf{U}_1\mathbf{W} = \mathbf{U}_1\mathbf{W}(\mathbf{W}^H\mathbf{G}_{1,1}\mathbf{W}) + \mathbf{F}_1\mathbf{W} \tag{8.10}$$

as an accelerated length-one block Arnoldi reduction. Stewart [45, 48] shows how this leads to improved convergence of orthonormal iteration to the dominant invariant subspace. We emphasize that the convergence of $\mathbf{U}_1\mathbf{W}$ over $\mathbf{U}_1$ is not accelerated—only ordered so that the initial columns of $\mathbf{U}_1\mathbf{W}$ contain the best approximations to the dominant invariant subspace from among the columns of $\mathbf{U}_1$. It is the span of these columns that are accelerated toward the dominant invariant subspace. Chatelin [8, pp. 253–257] and Saad [38, pp. 156–159] provide a discussion that builds upon the work of Stewart [45, 48]. This technique is also referred to as orthonormal iteration

with projection. It is a generalization of a Rayleigh–Ritz procedure to a non-Hermitian matrix. Since it explicitly computes $\mathbf{W}$, it is an explicit projection step.

However, there is another way to affect the projection step. Suppose we are interested in the $k < b$ dominant eigenvalues. If we perform $b - k$ QR steps on $\mathbf{G}_{1,1}$ with the associated unwanted eigenvalues, Lemma 6.2 implies that the k columns of $\mathbf{U}_1\mathbf{Z}$ span the same space as those of the initial k of $\mathbf{U}_1\mathbf{W}$. That is, orthonormal iteration with projection is equivalent to implicitly restarting orthonormal iteration.

For block Arnoldi reductions, Theorems 8.2 and 8.3 explain how to perform a projection step in an implicit fashion. A truncated QR algorithm shows how to push the directions associated with the desired eigenvalues to the leading portion of the underlying Krylov subspace. This strategy not only reduces the number of applications with $\mathbf{A}$ but is also more stable than traditional methods of restarting Arnoldi reductions.

The next section reviews traditional restarting mechanisms and points to the recent work of Sorensen [43] as the impetus behind a truncated QR algorithm.

9. Restarting Arnoldi Reductions. During each step of computing a block Arnoldi reduction, a partial orthogonal reduction of $\mathbf{A}$ into a banded upper Hessenberg matrix is produced. The eigenvalues of this Hessenberg matrix are used to approximate a subset of the eigenvalues of the large matrix $\mathbf{A}$. The approximation to the eigenvalues of $\mathbf{A}$ generally improves as the order of the Hessenberg matrix increases. Unfortunately, so do the cost and storage of the reduction.

A popular alternative is to define an iteration by restarting the reduction with information in a length $m < n/b$ block Arnoldi reduction. The hope is that this restarted reduction has improved estimates to the eigenvalues of $\mathbf{A}$.

The iteration is defined by a two-stage process. First, an Arnoldi reduction of length $m < n/b$ is computed. From the information available in this reduction, a new reduction is computed. This defines the iteration and is deemed successful if improved estimates to the eigenvalues of $\mathbf{A}$ appear in the subsequent reductions.

A restarted Arnoldi iteration was introduced by Saad [35] to overcome these difficulties, based on similar ideas developed for the Lanczos process by Paige [28], Cullum and Donath [9], and Golub and Underwood [15]. Karush [19] proposes what appears to be the first example of a restarted iteration. Sadkane considers a restarted block Arnoldi method using Chebyshev polynomials [39]. Scott has produced a block Arnoldi code [40]. We call all these related schemes *explicitly* restarted Arnoldi methods because they are not truncated QR algorithms. They do not use the implicit (or explicit) QR algorithm on $\mathbf{H}_m$ as a mechanism to restart a reduction. Instead, the matrix $\mathbf{A}$ is explicitly applied to some linear combination of the columns of $\mathbf{V}_m$. Saad's original scheme used a linear combination of the wanted Ritz vectors.

A relatively recent variant was developed by Sorensen [43] as a more efficient and numerically stable way to implement restarting. This technique, the *implicitly* restarted Arnoldi method, is implemented in the `ARPACK` [25] software package. The paper [43] only considered the block size $b = 1$ and stated that the method is equivalent to a truncated QR algorithm. The results in the preceding section showed a direct connection. The remainder of this section reviews restarting schemes and then end with an example.

9.1. Explicit Polynomial Acceleration. Suppose $\mathbf{A}$ is diagonalizable with eigenpairs $(\mathbf{x}_j, \lambda_j)$ for $j = 1, \ldots, n$. If $\psi(\cdot)$ is some polynomial and we expand the

- *Start:* Build a length m block Arnoldi reduction.
- *Iteration:*
  1. Compute the eigensystem of $\mathbf{H}_m$, and determine convergence. Exit if a partial Schur decomposition of order k satisfies the approximation criterion.
  2. *Restart:* Compute a new starting block $\mathbf{U}_1^+ = \psi(\mathbf{A})\mathbf{Y}$, where $\mathcal{R}(\mathbf{Y}) \subset \mathcal{R}(\mathbf{V}_m)$.
  3. Extend the length r block Arnoldi reduction to a length m one.

FIG. 9.1. *A Polynomial Accelerated Arnoldi Iteration*

current starting vector $\mathbf{u}_1$ in terms of the basis of eigenvectors, then

$$\psi(\mathbf{A})\mathbf{u}_1 = \mathbf{x}_1\psi(\lambda_1)\zeta_1 + \cdots + \mathbf{x}_n\psi(\lambda_n)\zeta_n. \tag{9.1}$$

Assuming that the eigenpairs $(\mathbf{x}_i, \lambda_i)$ are ordered so that the wanted k ones are at the beginning of the expansion, we seek a polynomial such that

$$\max_{i=k+1,\ldots,n} |\psi(\lambda_i)| < \min_{i=1,\ldots,k} |\psi(\lambda_i)|. \tag{9.2}$$

A good polynomial $\psi(\lambda)$ acts as a *filter*. Components in the direction of unwanted eigenvectors are damped, or, equivalently, components in the direction of wanted eigenvectors are amplified.

The acceleration techniques and hybrid methods presented by Saad in Chapter 7 of [38] attempt to improve explicit restarting by approximately solving the min-max problem of equation (9.2). Motivated by Manteuffel's scheme [26], Saad first proposed the use of Chebyshev polynomials in [37]. A Chebyshev polynomial $\psi(\mathbf{A})$ on an ellipse containing the unwanted Ritz values is applied to the restart vector in an attempt to accelerate convergence of the original ERA iteration. The polynomial is applied with the use of the familiar three-term recurrence. Figure 9.1 outlines the procedure. Note that after application of the polynomial filter, the reduction must be built from scratch. The columns of the n by b matrix $\mathbf{Y}$ typically contain the Ritz or Schur vectors of interest. By the results in § 7,

$$\mathbf{Y} = \mathbf{U}_1\mathbf{\Xi}_0 + \mathbf{A}\mathbf{U}_1\mathbf{\Xi}_1 + \cdots + \mathbf{A}^{m-1}\mathbf{U}_1\mathbf{\Xi}_{m-1},$$

so that $\mathbf{Y}$ itself is a matrix polynomial in $\mathbf{U}_1$.

9.2. Implicit Restarting. Figure 9.2 lists a truncated QR algorithm. Theorems 8.2 and 8.3 give that this is mathematically equivalent to explicitly computing $(\mathbf{A}-\tau_1\mathbf{I})\cdots(\mathbf{A}-\tau_p\mathbf{I})\mathbf{U}_1$ for the next starting vector. If $p > 1$, a restart from scratch is not needed—a length r Arnoldi reduction remains.

In his paper that introduced implicit restarting, Sorensen [43] suggested using the unwanted $m-k$ eigenvalues of $\mathbf{H}_m$ as shifts in line 2. By Lemma 6.2, $\mathbf{H}_k^+$ contains the k eigenvalues of interest. As Sorensen showed in Lemma 3.10, this is mathematically equivalent to an explicit restart with a linear combination of the wanted Ritz vectors (or Schur vectors). This is called an *exact shift* strategy.

It is important to realize that implicit restarting is always performed with a basis of Schur vectors of $\mathbf{H}_m$ without explicit application of $\mathbf{A}$ or construction of the approximating Schur vectors for $\mathbf{A}$. Implicit restarting, as pointed out in § 8, is formally equivalent to a Rayleigh–Ritz step (the projection) on the current Arnoldi basis.

- *Start:* Build a length m block Arnoldi reduction.
- *Iteration:*
  1. Compute the eigensystem of $\mathbf{H}_m$ and determine convergence. Exit if a partial Schur decomposition of order k satisfies the approximation criterion.
  2. Perform $m - r = p$ steps of the QR algorithm with the p shifts τ_i resulting in $\mathbf{H}_m\mathbf{Q}_m = \mathbf{Q}_m\mathbf{H}_m^+$.
  3. *Restart:* Postmultiply the length m block Arnoldi reduction with $\mathbf{Q}_r$ (the first br columns of $\mathbf{Q}_m$) to obtain the length r block Arnoldi reduction
$$\mathbf{A}\mathbf{V}_m\mathbf{Q}_r = \mathbf{V}_m\mathbf{Q}_r\mathbf{H}_r^+ + \mathbf{F}_r^+\mathbf{E}_r^T.$$
  4. Extend the length r block Arnoldi reduction to a length m one.

FIG. 9.2. *A Truncated QR Algorithm.*

9.3. Explicit and Implicit Restarting. We present a striking example that compares the explicit and implicit restarting an (unblocked) Arnoldi reduction. Let $\mathbf{A} \in \mathbf{R}^{10\times 10}$ be zero everywhere except for diagonal elements

$$\alpha_{11} = 1, \alpha_{22} = 1, \alpha_{33} = 0, \alpha_{44} = 0, \alpha_{ii} = (5-i)\cdot 10^{-1}, \text{ for } i = 1, \ldots, 6,$$

and ones on the subdiagonal. Suppose that the vector $\mathbf{e}_1$ is used to compute an initial Arnoldi reduction. We set $k = 2$ and $m = 4$ with the interest to compute the k eigenvalues equal to one. Using the two unwanted eigenvalues as shifts for the QR iteration, an implicit restart computes the approximate partial real Schur decomposition $\mathbf{A}\mathbf{Q}_2 \approx \mathbf{Q}_2\mathbf{R}_2$, where

$$\mathbf{R}_2 \approx \begin{bmatrix} .94919 & .95789 \\ -2.6952\cdot 10^{-3} & 1.0508 \end{bmatrix},$$

with eigenvalues equal to $1 \pm i1.129168612228906\cdot 10^{-8}$. The number of restarts needed was four, for a total of ten matrix vector products.

However, explicitly restarting the Arnoldi reduction stagnates if the expansion coefficients are chosen as originally proposed by Saad. This restart chooses the linear combination $\mathbf{V}_m\mathbf{y} \leftarrow \mathbf{V}_m(\mathbf{s}_1\gamma_1 + \mathbf{s}_2\gamma_2)$, where $\gamma_i = |\mathbf{e}_m^T\mathbf{s}_i|\ \|\mathbf{f}_m\|(= \|\mathbf{A}\mathbf{V}_m\mathbf{s}_i - \mathbf{V}_m\mathbf{s}_i\theta_i\|)$. The effect is to emphasize the Ritz vectors associated with Ritz values that are not yet acceptable approximations. The resulting vector $\mathbf{V}_m\mathbf{y}$ is a linear combination of the wanted Ritz vectors. If θ_i has a nonzero imaginary part, we set $\mathbf{s}_1$ and $\mathbf{s}_2$ to be the real and imaginary portions of the complex eigenvector of $\mathbf{H}_m$ associated with θ_1.

In fact, the starting vector $\mathbf{e}_1$ is continually computed at every restart. At every restart,

$$\mathbf{H}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

is computed. The MATLAB function `EIG` computes the two eigenvectors

$$\mathbf{s}_1^T = \begin{bmatrix} 0 & .57735 & .57735 & .57735 \end{bmatrix},$$
$$\mathbf{s}_2^T = -\mathbf{s}_1^T + 1.8\cdot 10^{-18}\mathbf{e}_1^T,$$

corresponding to the two eigenvalues equal to one. (Note that in exact arithmetic $\mathbf{s}_1 = \mathbf{s}_2$.) Choosing $\mathbf{y}_4$ to be an unit vector in the linear span of

$$\left[\begin{array}{cccc} \|\mathbf{A}\mathbf{V}_4\mathbf{s}_1 - \mathbf{V}_4\mathbf{s}_1\| & \|\mathbf{A}\mathbf{V}_4\mathbf{s}_2 - \mathbf{V}_4\mathbf{s}_2\| & 0 & 0 \end{array} \right]^T$$

gives that $\mathbf{V}_4 = \pm\mathbf{e}_1$.

The explanation is simple. Although $\mathbf{e}_1$ is orthogonal to the eigenspace associated with the eigenvalue one of $\mathbf{A}$, it is not orthogonal to the invariant subspace associated with the unit eigenvalue. Hence, why implicit restarting works where explicitly doing so does not.

The major drawback of using a linear combination of the eigenvectors of $\mathbf{H}_m$ is that they may form a poor choice for the starting vector. If $\mathbf{H}_m$ is defective, there might not be enough eigenvectors associated with the wanted eigenvalues. A pair of approximate eigenvectors is produced that are aligned to working precision. On the other hand, using an expansion in terms of the Schur vectors of $\mathbf{H}_m$ gives a "richer" starting vector.

We remark that both restarting techniques differ only in the polynomial filter applied. For a detailed computational study comparing software based on these two different restarting mechanisms, we refer the reader to [23]. For an unblocked Arnoldi reduction, Morgan [27] shows that an implicit restarting mechanism is a better behaved numerical process than an explicit one.

10. Convergence of a Truncated QR Algorithm. Orthonormal iteration with a projection step converges at a linear rate. Unlike orthonormal iteration without the projection step, the jth column ($1 \leq j \leq b$) of $\mathbf{V}_1\mathbf{W}_1$ converges to the j Schur vector (ordered in decreasing order of magnitude) at a rate of $|\lambda_{b+1}/\lambda_j|$. See [52] for details on more general shifting strategies.

For the collection of subspaces underlying a block Arnoldi reduction, the situation is considerably more complicated. Saad [35] considered the distance between a Ritz vector drawn from a block Lanczos reduction to an eigenvector of $\mathbf{A}$ as a function of the length m of the reduction. He extended his result to a $b = 1$ Arnoldi reduction in [36] with the assumption that $\mathbf{A}$ is diagonalizable. Jia [18] 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 invariant subspace.

Suppose that $k = b\ell$, where k is at least as large as the actual number of Ritz values needed. Since $\mathbf{A}\mathbf{V}_\ell = \mathbf{V}_\ell\mathbf{H}_\ell + \mathbf{F}_\ell\mathbf{E}_\ell^T$, a truncated QR algorithm is attempting to drive $\|\mathbf{G}_{\ell+1,\ell}\|$ to zero so that $\mathbf{V}_\ell$ approaches an invariant subspace. Suppose we complete this block Arnoldi reduction to the full one:

$$\mathbf{A} \left[\begin{array}{cc} \mathbf{V}_\ell & \mathbf{W} \end{array} \right] = \left[\begin{array}{cc} \mathbf{V}_\ell & \mathbf{W} \end{array} \right] \left[\begin{array}{cc} \mathbf{H}_\ell & \mathbf{M}_\ell \\ \mathbf{G}_{\ell+1,\ell}\mathbf{E}_\ell^T & \mathbf{C}_\ell \end{array} \right]. \tag{10.1}$$

Watkins and Elsner [52] discuss the rate of convergence of $\|\mathbf{G}_{\ell+1,\ell}\|$ as a function of the shifts given mild conditions on the initial block $\mathbf{U}_1$ within the context of a QR algorithm.

When $\|\mathbf{G}_{\ell+1,\ell}\|$ is small, the $\mathcal{R}(\mathbf{V}_\ell)$ is an exact invariant subspace for $\mathbf{A} - \mathbf{F}_\ell\mathbf{U}_\ell^H$ where $\|\mathbf{F}_\ell\mathbf{U}_\ell^H\| = \|\mathbf{G}_{\ell+1,\ell}\|$. However, whether the distance from $\mathbf{V}_\ell$ to an invariant subspace is small depends upon the sensitivity of $\mathbf{A}$'s invariant subspaces.

Stewart [46] considers how close $\mathbf{V}_\ell$ is to an invariant subspace of $\mathbf{A}$ for small $\|\mathbf{G}_{\ell+1,\ell}\|$. He considers whether an orthonormal matrix $\mathbf{Y}$ deviating little from $\mathbf{I}_n$ can be found so that $\mathbf{V}_\ell\mathbf{Y}$ is an invariant subspace for $\mathbf{A}$. Stewart chooses

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

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

THEOREM 10.1. *Suppose that* $\mathbf{A}\mathbf{V}_\ell = \mathbf{V}_\ell\mathbf{H}_\ell + \mathbf{F}_\ell\mathbf{E}_\ell^T$ *is a length* ℓ *block Arnoldi reduction. Suppose the reduction is completed to a band Hessenberg decomposition of* $\mathbf{A}$ *given by Equation (10.1), where* $\|\mathbf{G}_{\ell+1,\ell}\| = \|\mathbf{F}_\ell\|$. *Let*

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

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

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

$$\|\mathbf{P}\| \leq 2\frac{\beta_{\ell+1}}{\delta_\ell}$$

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

Proof. The conclusion now follows directly from Theorem 4.1 of Stewart [46]. □

The size of γ_ℓ measures the amount of coupling between the $\mathcal{R}(\mathbf{V}_\ell)$ and $\mathcal{R}(\mathbf{W})$. The reciprocal of δ_ℓ measures the sensitivity of the $\mathcal{R}(\mathbf{Z}_\ell)$ as an invariant subspace. Varah [49] shows that if the matrices involved are highly nonnormal, the smallest difference between the spectrums of $\mathbf{H}_\ell$ and $\mathbf{C}_\ell$ may be an overestimate of the true separation.

Theorem 10.1 shows the dependence of $\beta_{\ell+1}$ upon γ_ℓ and δ_ℓ in determining the quality of the $\mathcal{R}(\mathbf{V}_\ell)$ as an eigenspace of $\mathbf{A}$. Since $\mathbf{V}_\ell^H\mathbf{Z}_\ell = (\mathbf{I}+\mathbf{P}^H\mathbf{P})^{-1/2}$, Stewart [46] shows that the singular values of $\mathbf{P}$ are the tangents of the canonical, or principal, angles [8, 13, 46] between the two spaces spanned by the columns of $\mathbf{V}_\ell$ and $\mathbf{Z}_\ell$, respectively.

Golub and Wilkinson [16] also examine the many practical difficulties involved when computing invariant subspaces. They conclude that working with a basis of Schur vectors is a better-behaved numerical process. Within the context of subspace iteration, Stewart [48] also arrives at the same conclusion.

In conclusion, that a Ritz vector drawn from a block Arnoldi reduction never "settles down" implies that we must enlarge our view. What we should try to approximate is a well-conditioned invariant subspace. A truncated QR algorithm allows us to do this by varying b, ℓ, and r.

11. Practical Considerations. We briefly discuss two important issues required for a practical truncated QR algorithm. They are a shift selection strategy and deflation scheme.

11.1. Shift Selection. The shift strategy that leads to cubic and quadratic convergence rates of convergence for the QR algorithm on Hermitian and non-Hermitian eigenvalue problems cannot be adopted for a truncated QR algorithm. The shift strategy requires information available only when a full band Hessenberg reduction is at hand.

An exact shift strategy was already mentioned in § 9.2. By using the unwanted eigenvalues as shifts, the interesting Ritz values are places in $\mathbf{H}_r^+$. This is the strategy used by `ARPACK`. However, it is not clear which of the $bm-k$ eigenvalues should be used as shifts when a block Arnoldi reduction is used. In this situation, $m-1$ shifts can be applied, resulting in a length one Arnoldi reduction. Equations (8.7)–(8.8) explain how to apply the mth shift. Thus, even if k eigenvalues are of interest, the value r may be varied in the algorithm of Figure 9.2 from k for an adaptive strategy. This is a variation on a strategy proposed by Baglama, Calvetti, and Reichel [3] for symmetric $\mathbf{A}$. They employ Leja shifts and demonstrate how this strategy can outperform an exact shift strategy for small m. They have also extended their results to a block formulation [4].

The question of a near-optimal shift strategy is still the work of current research. However, it is clear from the results in [52] that an approximation $\psi_p(\lambda)$ to the minimization problem (9.2) is required. This should help guide the selection of m and r relative to k. The recent report [44] discusses an adaptive strategy for symmetric eigenvalue problems.

11.2. Deflation. Saad [38, pp. 234–235] explains how a deflation scheme leads to a far more reliable and efficient algorithm. When a Ritz pair $(\mathbf{z},\theta)$ has a small residual, it is *locked* into the leading portion of an Arnoldi reduction. Subsequent Arnoldi reductions are computed so that $\mathbf{V}_m$ is orthonormal to this Ritz vector. This is equivalent to working with the deflated Krylov subspace $\mathcal{K}_m((\mathbf{I}-\mathbf{z}\mathbf{z}^H)\mathbf{A},\mathbf{U}_1)$. As Ritz pairs of this deflated matrix are computed, a partial Schur decomposition is incrementally computed. This is an outline of the procedure discussed in [38, pp. 179–182]. Scott [40] employs this deflation scheme within a block Arnoldi reduction.

However, these explicit deflation schemes require that the Arnoldi reduction be restarted from scratch in order to deflate the Ritz pair. The recent paper [24] instead explains how to deflate the Ritz pair in an implicit fashion, thus avoiding the need to build a new reduction. Although this implicit deflation technique assumed an unblocked Arnoldi reduction, the scheme is easily extended to a block reduction. The report [22] considers a practical implementation of a truncated QR algorithm in detail.

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] 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*, technical report, Department of Mathematics and Computer science, Kent State University, 1997.

[5] Z. Bai and J. W. Demmel, *On a block implementation of Hessenberg multishift QR iteration*, International Journal of High Speed Computation, 1 (1989), pp. 97–112.

[6] Z. Bai and G. W. Stewart, *SRRIT — A FORTRAN subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix*, Technical Report 2908, Department of Com-

puter Science, University of Maryland, 1992. Submitted to *ACM Transactions on Mathematical Software*.

[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] J. Cullum and W. E. Donath, *A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace for large, sparse symmetric matrices*, in Proceedings of the 1974 IEEE Conference on Decision and Control, New York, 1974, pp. 505–509.

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

[11] J. G. F. Francis, *The QR transformation—part 1*, The Computer Journal, 4 (1961), pp. 265–271.

[12] ———, *The QR transformation—part 2*, The Computer Journal, 4 (1962), pp. 332–345.

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

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

[15] G. H. Golub and R. Underwood, *The block Lanczos method for computing eigenvalues*, in Mathematical Software III, J. R. Rice, ed., New York, 1977, Academic Press, pp. 361–377.

[16] 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.

[17] 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.

[18] Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*, SIAM J. Matrix Analysis and Applications, 15 (1994).

[19] W. Karush, *An iterative method for finding characteristics vectors of a symmetric matrix*, Pacific J. Mathematics, 1 (1951), pp. 233–248.

[20] V. N. Kublanovskaya, *On some algorithms for the solution of the complete eigenvalue problem*, USSR Comput. Math. Phys., 3 (1961), pp. 637–657.

[21] 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.

[22] R. B. Lehoucq and K. J. Maschhoff, *A practical implementation of an implicitly re-started block arnoldi method*, Preprint MCS-P649-0297, Argonne National Laboratory, Argonne, Ill, 1997. In preparation.

[23] R. B. Lehoucq and J. A. Scott, *An evaluation of software for computing eigenvalues of sparse nonsymmetric matrices*, Preprint MCS-P547-1195, Argonne National Laboratory, Argonne, Ill, 1996.

[24] R. B. Lehoucq and D. C. Sorensen, *Deflation techniques for an implicitly restarted Arnoldi iteration*, SIAM J. Matrix Analysis and Applications, 17 (1996), pp. 789–821.

[25] R. B. Lehoucq, D. C. Sorensen, P. Vu, and C. Yang, *ARPACK:* An implementation of an implicitly re-started Arnoldi method for computing some of the eigenvalues and eigenvectors of a large sparse matrix, 1996. See the URL http://www.caam.rice.edu/ARPACK/.

[26] T. A. Manteuffel, *Adaptive procedure for estimating parameters for the nonsymmetric Tchebychev iteration*, Numerische Mathematik, 31 (1978), pp. 183–208.

[27] R. B. Morgan, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Mathematics of Computation, 65 (1996), pp. 1213–1230.

[28] C. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, PhD thesis, University of London, London, England, 1971.

[29] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.

[30] A. Ruhe, *Rational Krylov sequence methods for eigenvalue computations*, Linear Algebra and Its Applications, 58 (1984), pp. 391–405.

[31] H. Rutihauser, *Solution of eigenvalue problems with the LR-transformation*, National Bureau of Standards Applied Mathematics Service, 49 (1958), pp. 47–81.

[32] ———, *Computational aspects of F.L. Bauer's simultaneous iteration method for symmetric matrices*, Numerische Mathematik, 13 (1969), pp. 4–13.

[33] ———, *Simultaneous iteration method for symmetric matrices*, Numerische Mathematik, 16 (1970), pp. 205–223.

[34] Y. Saad, *On the rates of convergence of the Lanczos and the block Lanczos methods*, SIAM Journal of Numerical Analysis, 17 (1980), pp. 687–706.

[35] ———, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra and Its Applications, 34 (1980), pp. 269–295.
[36] ———, *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, Berlin, 1982, Springer–Verlag, pp. 121–144.
[37] ———, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Mathematics of Computation, 42 (1984), pp. 567–588.
[38] ———, *Numerical Methods for Large Eigenvalue Problems*, Halsted Press, 1992.
[39] M. Sadkane, *A block Arnoldi–Chebyshev method for computing the leading eigenpairs of large sparse unsymmetric matrices*, Numerische Mathematik, 64 (1993), pp. 181–193.
[40] J. A. Scott, *An Arnoldi code for computing selected eigenvalues of sparse real unsymmetric matrices*, ACM Trans. Mathematical Software, 21 (1995), pp. 432–475.
[41] V. Simoncini, *Ritz and pseudo-ritz values using matrix polynomials*, Linear Algebra and Its Applications, 241–243 (1996), pp. 787–802.
[42] 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*.
[43] 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.
[44] A. Stathopoulos, Y. Saad, and K. Wu, *Dynamic thick restarting of the Davidson, and the implicitly restarted Arnoldi methods*, Tecnical Report 96/123, University of Minnesota Supercomputing Institute, June 1996.
[45] G. W. Stewart, *Accelerating the orthogonal iteration for the eigenvectors of a Hermitian matrix*, Numerische Mathematik, 13 (1969), pp. 362–376.
[46] ———, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review, 15 (1973), pp. 727–764.
[47] ———, *Introduction to Matrix Computations*, Academic Press, San Diego, California, 1973.
[48] ———, *Simultaneous iteration for computing invariant subspaces of non-Hermitian matrices*, Numerische Mathematik, 25 (1976), pp. 123–136.
[49] J. M. Varah, *On the separation of two matrices*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 216–222.
[50] D. S. Watkins, *Understanding the QR algorithm*, SIAM Review, 24 (1982), pp. 427–439.
[51] ———, *Fundamentals of Matrix Computations*, Wiley, 1991.
[52] 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.
[53] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, U.K., 1965.