Truncated QZ Methods for Large Scale Generalized Eigenvalue Problems

D.C. Sorensen

CRPC-TR98774
January 1998

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted September 1998; Available as Rice CAAM TR98-01; To appear in *Electronic Transactions on Numerical Analysis*

TRUNCATED QZ METHODS FOR LARGE SCALE GENERALIZED EIGENVALUE PROBLEMS[*]

D. C. SORENSEN[†]

Abstract. This paper presents three methods for the large scale generalized eigenvalue problem

$$\mathbf{Ax} = \mathbf{Bx}\lambda.$$

These methods are developed within a subspace projection framework as a truncation and modification of the QZ-algorithm for dense problems that is suitable for computing partial generalized Schur decompositions of the pair $(\mathbf{A}, \mathbf{B})$. A generalized partial reduction to condensed form is developed by analogy with the Arnoldi process. Then truncated forward and backward QZ iterations are introduced to derive generalizations of the Implicitly Restarted Arnoldi Method and the Truncated RQ method for the large scale generalized problem. These two methods require accurate solutions of linear systems at each step of the iteration. Relaxing these accuracy requirements forces us to introduce non-Krylov projection spaces that lead most naturally to block variants of the QZ iterations. A two-block method is developed that incorporates k approximate Newton corrections at each iteration. An important feature is the potential to utilize k matrix vector products for each access of the matrix pair $(\mathbf{A}, \mathbf{B})$. Preliminary computational experience is presented to compare the three new methods.

Key words. Generalized eigenvalue problem, Krylov projection methods, Arnoldi method, Lanczos method, QZ method, block methods, preconditioning, implicit restarting

AMS subject classifications. Primary 65F15, Secondary 65G05

1. Introduction.

This paper presents three methods for the large scale generalized eigenvalue problem

$$\mathbf{Ax} = \mathbf{Bx}\lambda. \tag{1.1}$$

The methods are developed within a Krylov subspace projection framework as truncations of the QZ-algorithm [13] for dense problems. These techniques provide natural extensions of the Implicitly Restarted Arnoldi Method [20] and the Truncated RQ Method [21] to the generalized problem. Relaxing the accuracy level for the solutions of required linear systems leads naturally to a non-Krylov block projection method. This block method does not require accurate solution of shift-invert equations and it makes efficient use of each matrix access by performing k matrix-vector products instead of one.

The first two methods require accurate solutions of linear systems at each step of the iteration. However, these methods are developed within a projection a framework that can accommodate inexact solves of the shift invert equations if the standard Krylov relations are relaxed. Introducing inexact solves forces us to introduce non-Krylov projection spaces. Once the Krylov property has been given up, it is natural to consider block variants of the QZ iterations. Therefore, we have developed a two-block method that incorporates k approximate Newton corrections at each iteration. An important feature is the potential to utilize k matrix vector products for each access of the matrix pair $(\mathbf{A}, \mathbf{B})$.

[*] This work was supported in part by NSF cooperative agreement CCR-9120008, NSF contract ASC-9408795 and by ARPA contract number DAAL03-91-C-0047 (administered by the U.S. Army Research Office).

[†] Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892, (sorensen@caam.rice.edu).

For some time, there has been considerable interest in improving eigenvalue methods either by making better use of spectral transformation through multi-shift Rational Krylov methods [16] or by utilizing some sort of pre-conditioned iterative solution of these shift-invert equations at a relaxed accuracy level [11, 10, 14, 2, 19]. The ultimate goal is to achieve the enhanced convergence properties of the spectral transformation without the cost of an accurate direct or iterative solution of the shift-invert equations. Generalization of the Davidson method [4] to a wider class of problems has received a lot of attention and the Jacobi-Davidson method of Sleijpen and Van der Vorst [18] has emerged as an effective variant. The methods developed here (particularly the backward form of truncated QZ) have a great deal in common with Ruhe's *RKS* method. This truncated backward form also appears to be closely related to the work of De Samblanx, Meerbergen and Bultheel on implicit applications of a rational filter in *RKS* [17].

The paper begins in Sections 2 and 3 with the development of simultaneous projections of the matrices $\mathbf{A}$ and $\mathbf{B}$ onto two subspaces to achieve a partial reduction to condensed form

$$\begin{aligned} \mathbf{AV}_k &= \mathbf{W}_k\mathbf{H}_k + \mathbf{F}_k \ , \quad \text{with} \quad \mathbf{W}_k^T\mathbf{F}_k = \mathbf{0}, \\ \mathbf{BV}_k &= \mathbf{W}_k\mathbf{R}_k, \end{aligned}$$

through a generalized Arnoldi process. Here, $\mathbf{V}_k$ and $\mathbf{W}_k$ are both $n \times k$ orthogonal matrices, $\mathbf{H}_k$ is a $k \times k$ upper Hessenberg matrix and $\mathbf{R}_k$ is upper triangular. With this reduction, approximate generalized eigenvalues of the pair $(\mathbf{A}, \mathbf{B})$ are obtained from the projected pair $(\mathbf{H}_k, \mathbf{R}_k)$.

As with the standard Arnoldi process, storage and arithmetic costs are prohibitive for large k. Thus, restarting schemes are essential and two possibilities are developed. In Section 4, forward and backward variants of the implicitly shifted QZ iteration are developed for dense generalized problems. These are analogous to the QR and RQ iterations for the standard problem. Truncated forms of these forward and backward QZ -iterations are developed in Section 5. The forward form is analogous to implicit restarting [20] while the backward form generalizes the truncated RQ iteration [21]. These developments result in methods that are effective in computing a few (k) selected eigenvalues and corresponding eigenvectors within a fixed pre-determined storage requirement proportional to $n \cdot k$ and work proportional to $n \cdot k^2 + \mathcal{O}(k^3)$.

The generalized Arnoldi process requires the solution of a linear system at each step regardless of how it is organized. Depending on certain choices, this amounts to applying a mathematically equivalent standard Arnoldi process to one of the following matrix operators:

$$\mathbf{B}^{-1}\mathbf{A}, \quad \mathbf{AB}^{-1}, \quad \text{or} \ \ (\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{B}.$$

The backward variant of the truncated QZ iteration makes the most economical use of storage but tends to require more LU-factorizations than the forward variant. Very limited computational experience with all three methods shall be presented in Section 7. No reliable conclusions on comparative performance can be drawn from these limited tests.

Throughout this paper, capital and lower case Latin letters denote matrices and vectors respectively, while lower case Greek letters denote scalars. The j-th canonical basis vector is denoted by $\mathbf{e}_j$. The Euclidean norm is used exclusively and is denoted by $\| \cdot \|$. The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^T$ and conjugate transpose by $\mathbf{A}^H$. Upper Hessenberg matrices will appear frequently and are usually denoted by the letter $\mathbf{H}$. The notation $\mathbf{M}(:,1:k)$ and $\mathbf{M}(1:k,1:k)$ denote the leading k columns and the leading $k \times k$ principal submatrix of $\mathbf{M}$.

2. Subspace Projection.

Certainly, projection methods are prominent for the iterative solution of linear systems and for computing a few eigenvalues of a large matrix or matrix pencil. In the case of the standard problem $\mathbf{Ax} = \mathbf{x}\lambda$, Krylov subspace projection results in the Lanczos/Arnoldi class of methods. These may be viewed as systematic ways to extract additional eigen-information from the sequence of vectors produced by a power iteration. If one hopes to obtain additional information through various linear combinations of the power sequence, it is natural to formally consider the *Krylov* subspace

$$\mathcal{K}_k(\mathbf{A},\mathbf{v}_1) = \text{Span}\,\{\mathbf{v}_1, \mathbf{Av}_1, \mathbf{A}^2\mathbf{v}_1, \ldots, \mathbf{A}^{k-1}\mathbf{v}_1\}$$

and to attempt to formulate the best possible approximations to eigenvectors from this subspace.

Approximate eigenpairs are constructed by imposing a Galerkin condition: A vector $\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{v}_1)$ is called a *Ritz vector* with corresponding *Ritz value* θ if the Galerkin condition

$$\langle \mathbf{w}, \mathbf{Ax} - \mathbf{x}\theta \rangle = 0 \ , \quad \text{for all} \ \ \mathbf{w} \in \mathcal{K}_k(\mathbf{A},\mathbf{v}_1)$$

is satisfied. It is well known that the Lanczos/Arnoldi iteration computes an orthonormal basis $\mathbf{V}_k$ for this Krylov subspace along with a small projected matrix $\mathbf{H}_k = \mathbf{V}_k^H\mathbf{AV}_k$ of order k from which Ritz values and vectors may be obtained: $(\mathbf{x},\theta)$ is a Ritz pair if and only if $\mathbf{H}_k\mathbf{y} = \mathbf{y}\theta$ and $\mathbf{x} = \mathbf{V}_k\mathbf{y}$.

Several schemes have been developed to extend the Krylov subspace idea to the generalized problem (1.1). These extensions are generally based upon a conversion of the generalized problem to a standard one. Perhaps the most successful variant [5] is to use the *spectral transformation*

$$(\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{Bx} = \mathbf{x}\nu.$$

An eigenvector $\mathbf{x}$ of this transformed problem is also an eigenvector of the original problem (1.1) with the corresponding eigenvalue given by $\lambda = \sigma + \frac{1}{\nu}$. In applications, $\mathbf{B}$ is often symmetric and positive (semi-)definite and then it is helpful to work with the $\mathbf{B}$ (semi-)inner product in the Lanczos/Arnoldi process [5, 8, 12]. With this transformation, convergence of the Lanczos/Arnoldi iteration is very rapid to eigenvalues near the shift σ because they are transformed to extremal well-separated eigenvalues and also since eigenvalues far from σ are damped (mapped near zero).

To utilize this transformation in a Lanczos/Arnoldi process, the repeated operation $\mathbf{w} \leftarrow \mathbf{Av}$ is replaced by repeated solutions of a shift invert equation $(\mathbf{A} - \sigma\mathbf{B})\mathbf{w} = \mathbf{Bv}$ at each step of the iteration. If a sparse-direct factorization of the shifted matrix $(\mathbf{A} - \sigma\mathbf{B})$ is possible then this single factorization may be re-used at each step of the iteration. This approach is certainly the method of choice but may not be practical or even possible in many important applications.

In some cases, it may be effective to use a pre-conditioned iterative method to solve the shift-invert equations but there are a number of pitfalls to this approach. Typically, the shifted matrix is very ill-conditioned because σ will be chosen to be near an interesting eigenvalue. Moreover, this shifted matrix will usually be indefinite (or have indefinite symmetric part). These are the conditions that are most difficult for iterative solution of linear systems. Finally, these difficulties are exacerbated by the fact that each linear system must be solved to a considerably greater accuracy than the desired accuracy of the eigenvalue calculation. Otherwise, each step of the Lanczos/Arnoldi process will essentially involve a different matrix operator.

The underlying Krylov subspace projections associated with the Lanczos/Arnoldi process provide a number of important approximation properties related to convergence and accuracy. Unfortunately, if it is not possible to solve the shift-invert equations accurately then these desirable properties are generally lost. However, it is possible to retain the projection idea in a way that generalizes the Arnoldi process when the shift invert equations can be solved accurately and yet can accommodate inaccurate solution of the shift-invert equations. To do this, we must consider more general subspaces and relax the Krylov requirement. The development of this projection framework is the primary topic of this paper. It is influenced by the following well known result.

LEMMA 2.1. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are complex matrices of order* n *then there are unitary matrices* $\mathbf{V}$,$\mathbf{W}$ *an upper Hessenberg matrix* $\mathbf{H}$ *and an upper triangular matrix* $\mathbf{R}$ *all of order* n *such that*

$$\begin{aligned} \mathbf{AV} &= \mathbf{WH}, \\ \mathbf{BV} &= \mathbf{WR}. \end{aligned} \tag{2.1}$$

This factorization can be computed in a finite number ($\mathcal{O}(n^3)$) *of rational arithmetic and square root opertions.*

Proof. See [7]. □

For the standard problem ($\mathbf{B} = \mathbf{I}$) this lemma reduces to the statment that $\mathbf{A}$ may be put in condensed form by unitary similarity transformations. The Arnoldi process produces a partial reduction of $\mathbf{A}$ to condensed (Hessenberg) form

$$\mathbf{AV}_k = \mathbf{V}_k\mathbf{H}_k + \mathbf{F}_k,$$

with $\mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_k$ and $\mathbf{V}_k^T\mathbf{F}_k = \mathbf{0}$. This may be interpreted simply as a truncation of the full reduction. It turns out that $\mathbf{F}_k = \mathbf{f}_k\mathbf{e}_k^T$ is a rank one matrix and this property is intrinsically tied to the fact that $\{\mathbf{V}_j : j = 1, 2, \dots, k\}$ is a sequence of orthonormal bases for the nested sequence of Krylov subspaces $\mathcal{K}_j(\mathbf{A}, \mathbf{v}_1)$. The Hessenberg matrix $\mathbf{H}_k = \mathbf{V}_k^T\mathbf{AV}_k$ is the orthogonal projection of $\mathbf{A}$ onto the subspace $\mathcal{K}_k(\mathbf{A}, \mathbf{v}_1)$ as represented in the basis $\mathbf{V}_k$ and

$$\mathbf{F}_k = (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^T)\mathbf{AV}_k.$$

If $k = n$ then $\mathbf{F}_k = \mathbf{0}$ and this provides a complete reduction of $\mathbf{A}$ to condensed (Hessenberg) form.

The generalization suggested by Lemma (2.1) is

$$\begin{aligned} \mathbf{F}_k &= (\mathbf{I} - \mathbf{W}_k\mathbf{W}_k^T)\mathbf{AV}_k, \\ \mathbf{BV}_k &= \mathbf{W}_k\mathbf{R}_k. \end{aligned} \tag{2.2}$$

where $\mathbf{W}_k^T\mathbf{W}_k = \mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_k$. This projection makes the residual $\mathbf{F}_k$ orthogonal to *Range*($\mathbf{BV}_k$), since the columns of $\mathbf{W}_k$ form an orthonormal basis for that space. The Arnoldi process for the standard problem systematically produces the columns of $\mathbf{V}_k, k = 1, 2, ..., n$ at the cost of a matrix vector product $\mathbf{y} \leftarrow \mathbf{Av}$ and an orthogonal decompostion of this vector into a component in the existing Krylov space and one that is orthogonal to it.

The precise extension of this process to the generalized problem still requires the solution of a linear system at each step. Nevertheless, it is interesting to develop this analogous generalized Arnoldi process along with two restarting variants. These will be developed in Sections 3,4,5. These algorithms are significant in themselves, but they may also be viewed as laying the groundwork developing schemes that can relax the accuracy requirement on the shift-invert equations and yet retain the projection propertes in the framework of a truncated reduction to condensed form.

3. Generalizing the Arnoldi reduction. The projection in equations (2.2) are well defined for any specification of the matrix $\mathbf{V}_k$, but it is not clear which specifications will provide good approximations to eigenvalues. The success of the implicitly restarted Lanczos/Arnoldi processes viewed as truncated *QR* iterations provides considerable motivation to develop a truncation of the *QZ* iteration in this projection framework.

The factorization expressed in (2.1) provides an initial reduction of the pair $(\mathbf{A}, \mathbf{B})$ to an equivalent pair $(\mathbf{H}, \mathbf{R})$ in condensed form. This reduction precedes the *QZ* iteration just as reduction to Hessenberg form precedes the *QR* iteration. In fact, the two reductions are identical when $\mathbf{B} = \mathbf{I}$. The Arnoldi process may be derived (for $\mathbf{B} = \mathbf{I}$) simply by equating the leading k columns on both sides of (2.1). Therefore, this Arnoldi idea is easily generalized by doing the same thing when $\mathbf{B}$ is not the identity matrix. This is fairly straightforward, but a little manipulation must be done to place this truncation within the projection framework of the previous section.

Truncating the relations (2.1) after k-steps provides

$$\begin{aligned} \mathbf{A}\mathbf{V}_k &= \mathbf{W}_k\mathbf{H}_k + \mathbf{f}_k\mathbf{e}_k^T \\ \mathbf{B}\mathbf{V}_k &= \mathbf{W}_k\mathbf{R}_k, \end{aligned} \tag{3.1}$$

with $\mathbf{V}_k, \mathbf{W}_k$ representing the leading k columns of $\mathbf{V}, \mathbf{W}$, $\mathbf{H}_k, \mathbf{R}_k$ representing the leading $k \times k$ principal submatrices of $\mathbf{H}, \mathbf{R}$ and $\mathbf{f}_k = \mathbf{w}_{k+1}\gamma_{k+1,k}$ where $\mathbf{w}_{k+1}$ is the $k+1$-st column of $\mathbf{W}$ and $\gamma_{k+1,k}$ is the k-th subdiagonal element of $\mathbf{H}$.

To advance this k-step factorization one step, the relations

$$\begin{aligned} \mathbf{A}[\mathbf{V}_k, \mathbf{v}] &= [\mathbf{W}_k, \mathbf{w}] \begin{bmatrix} \mathbf{H}_k & \mathbf{h} \\ \gamma\mathbf{e}_k^T & \alpha \end{bmatrix} + \mathbf{f}_{k+1}\mathbf{e}_{k+1}^T \\ \mathbf{B}[\mathbf{V}_k, \mathbf{v}] &= [\mathbf{W}_k, \mathbf{w}] \begin{bmatrix} \mathbf{R}_k & \mathbf{r} \\ 0 & \rho \end{bmatrix}, \end{aligned} \tag{3.2}$$

must be obtained to give the new columns $\mathbf{v}_{k+1} = \mathbf{v}, \mathbf{w}_{k+1} = \mathbf{w}$ and to update the matrices $\mathbf{H}_{k+1}$ and $\mathbf{R}_{k+1}$.

Equating the leading k columns on both sides implies $\gamma = \|\mathbf{f}_k\|$ and $\mathbf{w} = \mathbf{f}_k/\gamma$. The direction $\mathbf{v}$ must satisfy

$$\mathbf{B}\mathbf{v} = \mathbf{W}_k\mathbf{r} + \mathbf{w}\rho \quad \text{and} \quad \mathbf{V}_k^T\mathbf{v} = 0. \tag{3.3}$$

This implies that

$$0 = [\mathbf{V}_k^T\mathbf{B}^{-1}\mathbf{W}_k, \mathbf{V}_k^T\mathbf{B}^{-1}\mathbf{w}] \begin{bmatrix} \mathbf{r} \\ \rho \end{bmatrix}.$$

Now, $\mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_k$ and $\mathbf{B}\mathbf{V}_k = \mathbf{W}_k\mathbf{R}_k$ gives $\mathbf{V}_k^T\mathbf{B}^{-1}\mathbf{W}_k = \mathbf{R}_k^{-1}$ and thus

$$\mathbf{R}_k^{-1}\mathbf{r} = -\mathbf{V}_k^T\mathbf{B}^{-1}\mathbf{w}\rho. \tag{3.4}$$

```
GENARN: Generalized Arnoldi Reduction

Input: $[\mathbf{A}, \mathbf{B}, \mathbf{v}, k]$ with $\|\mathbf{v}\| = 1$.
Output: $[\mathbf{V}_k, \mathbf{W}_k, \mathbf{H}_k, \mathbf{R}_k, \mathbf{f}_k]$ such that $\mathbf{A}\mathbf{V}_k = \mathbf{W}_k\mathbf{H}_k + \mathbf{f}_k\mathbf{e}_k^T$, $\mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_k$,
        $\mathbf{B}\mathbf{V}_k = \mathbf{W}_k\mathbf{R}_k$, $\mathbf{W}_k^T\mathbf{W}_k = \mathbf{I}_k$, $\mathbf{W}_k^T\mathbf{f}_k = 0$,
        $\mathbf{H}_k$ upper Hessenberg and $\mathbf{R}_k$ upper triangular.

        1. $\mathbf{V}_1 \leftarrow [\mathbf{v}]$; $\mathbf{w} = \mathbf{B}\mathbf{v}$; $\rho = \|\mathbf{w}\|$; $\mathbf{R}_1 = [\rho]$; $\mathbf{W}_1 = [\mathbf{w}/\rho]$;
        2. $\mathbf{z} \leftarrow \mathbf{A}\mathbf{v}$;  $\mathbf{H}_1 \leftarrow [\mathbf{W}_1^T\mathbf{z}]$;  $\mathbf{f}_1 \leftarrow \mathbf{z} - \mathbf{W}_1\mathbf{H}_1$;
        3. for $j = 1, 2, 3, ..., k$
          3.1. $\gamma \leftarrow \|\mathbf{f}_j\|$; $\mathbf{w} \leftarrow \mathbf{f}_j/\gamma$;
          3.2. $\mathbf{W}_{j+1} \leftarrow [\mathbf{W}_j, \mathbf{w}]$;  $\mathbf{H}_j \leftarrow \begin{bmatrix} \mathbf{H}_j \\ \gamma\mathbf{e}_j^T \end{bmatrix}$;
          3.3. Solve $\mathbf{B}\hat{\mathbf{v}} = \mathbf{w}$;
          3.4. $\mathbf{z} \leftarrow \mathbf{V}_j^T\hat{\mathbf{v}}$;  $\hat{\mathbf{v}} \leftarrow \hat{\mathbf{v}} - \mathbf{V}_j\mathbf{z}$;  $\rho \leftarrow 1/\|\hat{\mathbf{v}}\|$;
          3.5. $\mathbf{v} \leftarrow \hat{\mathbf{v}}\rho$; $\mathbf{r} \leftarrow -\mathbf{R}_j\mathbf{z}\rho$;
          3.6. $\mathbf{V}_{j+1} \leftarrow [\mathbf{V}_j, \mathbf{v}]$;  $\mathbf{R}_{j+1} \leftarrow \begin{bmatrix} \mathbf{R}_j & \mathbf{r} \\ 0 & \rho \end{bmatrix}$;
          3.7. $\mathbf{z} \leftarrow \mathbf{A}\mathbf{v}$;  $\mathbf{h} \leftarrow \mathbf{W}_{j+1}^T\mathbf{z}$;
          3.8. $\mathbf{f}_{j+1} \leftarrow \mathbf{z} - \mathbf{W}_{j+1}\mathbf{h}$;  $\mathbf{H}_{j+1} \leftarrow [\mathbf{H}_j, \mathbf{h}]$;
        4. end
```

FIG. 3.1. *Generalized Arnoldi Reduction*

Combining (3.2), (3.3) and (3.4) gives

$$\begin{aligned} \mathbf{v} &= \mathbf{B}^{-1}\mathbf{W}_k\mathbf{r} + \mathbf{B}^{-1}\mathbf{w}\rho \\ &= -\mathbf{V}_k\mathbf{R}_k^{-1}\mathbf{r} + \mathbf{B}^{-1}\mathbf{w}\rho \\ &= -\mathbf{V}_k\mathbf{V}_k^T\mathbf{B}^{-1}\mathbf{w}\rho + \mathbf{B}^{-1}\mathbf{w}\rho \\ &= (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^T)\mathbf{B}^{-1}\mathbf{w}\rho, \end{aligned} \tag{3.5}$$

with $\rho \equiv 1/(\|(\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^T)\mathbf{B}^{-1}\mathbf{w}\|)$ so that $\mathbf{V}_k^T\mathbf{v} = 0$ and $\|\mathbf{v}\| = 1$. Now that the new $\mathbf{v}$ has been determined, it follows that

$$\begin{bmatrix} \mathbf{h} \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{W}_k^T\mathbf{A}\mathbf{v} \\ \mathbf{w}^T\mathbf{A}\mathbf{v} \end{bmatrix}$$

and

$$\mathbf{f}_{k+1} = \mathbf{A}\mathbf{v} - (\mathbf{W}_k\mathbf{h} + \mathbf{w}\alpha).$$

This completes the update and leads to the the generalized Arnoldi process *GENARN* shown in Fig. 3.1.

Remark 1: The substitution $\mathbf{V}_k = \mathbf{B}^{-1}\mathbf{W}_k\mathbf{R}_k$ gives

$$(\mathbf{A}\mathbf{B}^{-1})\mathbf{W}_k = \mathbf{W}_k\hat{\mathbf{H}}_k + \hat{\mathbf{f}}_k\mathbf{e}_k^T$$

and

$$(\mathbf{B}^{-1}\mathbf{A})\mathbf{V}_k = \mathbf{V}_k\tilde{\mathbf{H}}_k + \tilde{\mathbf{f}}_k\mathbf{e}_k^T$$

where $\hat{\mathbf{H}}_k = \mathbf{H}_k\mathbf{R}_k^{-1}$, $\tilde{\mathbf{H}}_k = \mathbf{R}_k^{-1}\mathbf{H}_k$ and $\hat{\mathbf{f}}_k = \mathbf{f}_k/\rho_{kk}$, $\tilde{\mathbf{f}}_k = \mathbf{B}^{-1}\mathbf{f}_k$, are both Arnoldi processes that are mathematically equivalent to Algorithm 3.1.

Remark 2: Replacing $\mathbf{A}$ with $\mathbf{B}$ and replacing $\mathbf{B}$ with $\mathbf{A} - \sigma\mathbf{B}$ in this algorithm is mathematically equivalent to shift-invert Arnoldi method applied to $(\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{B}$. With this substitution, the second relation in the previous remark would be

$$(\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{B}\mathbf{V}_k = \mathbf{V}_k\tilde{\mathbf{H}}_k + \tilde{\mathbf{f}}_k\mathbf{e}_k^T.$$

This generalized Arnoldi iteration does nothing more than produce a partial reduction of the pair $(\mathbf{A}, \mathbf{B})$ to condensed form $(\mathbf{H}_k, \mathbf{R}_k)$. Just as with the standard Arnoldi process, there is no active mechanism to search for desired eigenvalues. However, methods that are analogous to implicit restarting [20] and truncated *RQ* [21] are possible and these shall be developed in the following section.

4. Implicitly Shifted *QZ*-Iterations.

Forward and backward versions of implicitly shifted *QZ* iterations are developed here as simple extensions of the of the *QR* and *RQ* iterations. This leads naturally to truncated *QZ* iterations that generalize the truncated *QR* and *RQ* iterations developed in ([20, 21]).

In the following discussion, assume that there is a complete reduction of $(\mathbf{A}, \mathbf{B})$ to condensed form

$$\mathbf{A}\mathbf{V} = \mathbf{W}\mathbf{H},$$
$$\mathbf{B}\mathbf{V} = \mathbf{W}\mathbf{R}.$$

Forward *QZ* Iteration:

A forward *QZ* iteration may be developed from the following observations:
For a given shift μ, factor

$$\mathbf{H} - \mu\mathbf{R} = \mathbf{Z}\mathbf{T} \tag{4.1}$$

where $\mathbf{Z}$ is unitary and $\mathbf{T}$ is upper triangular. Now, factor

$$\mathbf{Z}^H\mathbf{R} = \mathbf{R}^+\mathbf{Q}, \tag{4.2}$$

where $\mathbf{R}^+$ is upper triangular and $\mathbf{Q}$ is unitary. As with the *QR* iteration, it is straightforward to show that $\mathbf{Z}$ is upper Hessenberg in (4.1). Since both $\mathbf{R}$ and $\mathbf{R}^+$ are upper triangular, the relation (4.2) implies that $\mathbf{Q}^H$ is also upper Hessenberg. It follows that

$$\begin{aligned}(\mathbf{A} - \mu\mathbf{B})\mathbf{V} &= \mathbf{W}\mathbf{Z}\mathbf{T},\\ \mathbf{B}\mathbf{V} &= \mathbf{W}\mathbf{Z}(\mathbf{Z}^H\mathbf{R}) = \mathbf{W}\mathbf{Z}\mathbf{R}^+\mathbf{Q}.\end{aligned} \tag{4.3}$$

Multiplying both sides of (4.3) on the right by $\mathbf{Q}^H$ and rearranging terms gives

$$\mathbf{A}\mathbf{V}^+ = \mathbf{W}^+\mathbf{H}^+,$$
$$\mathbf{B}\mathbf{V}^+ = \mathbf{W}^+\mathbf{R}^+,$$

where $\mathbf{V}^+ = \mathbf{V}\mathbf{Q}^H$, $\mathbf{W}^+ = \mathbf{W}\mathbf{Z}$ and $\mathbf{H}^+ = \mathbf{Z}^H\mathbf{H}\mathbf{Q}^H = \mathbf{T}\mathbf{Q}^H + \mu\mathbf{R}^+$ is upper Hessenberg. This sequence of operations comprises a forward *QZ* step. It may be accomplished implicitly when $\mathbf{Q}$ and $\mathbf{Z}$ are represented as products of Givens' transformations.

```
FQZ: Forward Implicitly Shifted QZ-iteration

Input: [V, W, H, R] with AV = WH, BV = WR,
          H upper Hessenberg, R upper triangular,
          V^H V = W^H W = I.
Output: [V, W, H, R] such that AV = WH, BV = WR,
          H and R both upper triangular,
          V^H V = W^H W = I.

          1. for $j = 1, 2, 3, \ldots$ until convergence,
             1.1. Select a shift $\mu \leftarrow \mu_j$;
             1.2. Factor $[\mathbf{Z}, \mathbf{T}] = qr(\mathbf{H} - \mu\mathbf{R})$;
             1.3. Factor $[\mathbf{R}^+, \mathbf{Q}] = rq(\mathbf{Z}^H\mathbf{R})$;
             1.4. $\mathbf{H} \leftarrow \mathbf{Z}^H\mathbf{H}\mathbf{Q}^H$ ; $\mathbf{R} \leftarrow \mathbf{R}^+$;
             1.5. $\mathbf{V} \leftarrow \mathbf{V}\mathbf{Q}^H$ ; $\mathbf{W} \leftarrow \mathbf{W}\mathbf{Z}$;
          2. end;
```

FIG. 4.1. *Forward Implicitly Shifted* QZ-*iteration.*

From (4.3) it follows that

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v}_1 = \mathbf{W}\mathbf{Z}\mathbf{T}\mathbf{e}_1 = \mathbf{B}\mathbf{V}^+(\mathbf{R}^+)^{-1}\mathbf{T}\mathbf{e}_1 = \mathbf{B}\mathbf{v}_1^+\tau$$

so that

$$(\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I})\mathbf{v}_1 = \mathbf{v}_1^+\tau$$

where τ is the (1,1) element of the upper triangular matrix $(\mathbf{R}^+)^{-1}\mathbf{T}$. Thus, the new starting vector $\mathbf{v}_1^+$ is the result of the application of a linear polynomial factor $(\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I})$ to the old starting vector $\mathbf{v}_1$.

Backward *QZ* Iteration:

A similar development leads to a backward *QZ* iteration:
For a given shift μ, factor

$$\mathbf{H} - \mu\mathbf{R} = \mathbf{T}\mathbf{Z}$$

where $\mathbf{Z}$ is unitary and $\mathbf{T}$ is upper triangular. Now, factor

$$\mathbf{R}\mathbf{Z}^H = \mathbf{Q}\mathbf{R}^+,$$

where $\mathbf{R}^+$ is upper triangular and $\mathbf{Q}$ is unitary. As before, $\mathbf{Z}$ and $\mathbf{Q}^H$ are upper Hessenberg. It follows that

$$\begin{aligned}(\mathbf{A} - \mu\mathbf{B})\mathbf{V}\mathbf{Z}^H &= \mathbf{W}\mathbf{T},\\ \mathbf{B}\mathbf{V}\mathbf{Z}^H &= \mathbf{W}\mathbf{Q}\mathbf{R}^+.\end{aligned}$$

Thus

$$\begin{aligned}\mathbf{A}\mathbf{V}^+ &= \mathbf{W}^+\mathbf{H}^+,\\ \mathbf{B}\mathbf{V}^+ &= \mathbf{W}^+\mathbf{R}^+,\end{aligned}$$

```
BQZ: Backward Implicitly Shifted QZ-iteration

Input: $[\mathbf{V}, \mathbf{W}, \mathbf{H}, \mathbf{R}]$ with $\mathbf{AV} = \mathbf{WH}$, $\mathbf{BV} = \mathbf{WR}$,
        $\mathbf{H}$ upper Hessenberg, $\mathbf{R}$ upper triangular,
        $\mathbf{V}^H\mathbf{V} = \mathbf{W}^H\mathbf{W} = \mathbf{I}$.
Output: $[\mathbf{V}, \mathbf{W}, \mathbf{H}, \mathbf{R}]$ such that $\mathbf{AV} = \mathbf{WH}$, $\mathbf{BV} = \mathbf{WR}$,
        $\mathbf{H}$ and $\mathbf{R}$ both upper triangular,
        $\mathbf{V}^H\mathbf{V} = \mathbf{W}^H\mathbf{W} = \mathbf{I}$.

    1. for $j = 1, 2, 3, \ldots$ until convergence,
      1.1. Select a shift $\mu \leftarrow \mu_j$;
      1.2. Factor $[\mathbf{T}, \mathbf{Z}] = rq(\mathbf{H} - \mu\mathbf{R})$;
      1.3. Factor $[\mathbf{Q}, \mathbf{R}^+] = qr(\mathbf{RZ}^H)$;
      1.4. $\mathbf{H} \leftarrow \mathbf{Q}^H\mathbf{HZ}^H$; $\mathbf{R} \leftarrow \mathbf{R}^+$;
      1.5. $\mathbf{V} \leftarrow \mathbf{VZ}^H$ ; $\mathbf{W} \leftarrow \mathbf{WQ}$;
    2. end;
```

FIG. 4.2. *Backward Implicitly Shifted* QZ*-iteration.*

where $\mathbf{V}^+ = \mathbf{VZ}^H$, $\mathbf{W}^+ = \mathbf{WQ}$ are unitary and $\mathbf{H}^+ \equiv \mathbf{Q}^H\mathbf{HZ}^H = \mathbf{TZ}^H + \mu\mathbf{R}^+$ is upper Hessenberg to complete the backwards *QZ* step.

This time, observe that

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v}_1^+ = \mathbf{WTe}_1 = \mathbf{BVR}^{-1}\mathbf{Te}_1 = \mathbf{Bv}_1\tau$$

so that

$$\begin{aligned} \mathbf{v}_1^+ &= (\mathbf{A} - \mu\mathbf{B})^{-1}\mathbf{Bv}_1\tau \\ &= (\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{v}_1\tau \end{aligned}$$

where τ is the (1,1) element of $\mathbf{R}^{-1}\mathbf{T}$. Hence, the leading column of two successive $\mathbf{V}$ matrices are in an inverse iteration relationship.

5. Truncated Forward and Backward *QZ*-Iterations. With these versions of the *QZ* iteration, one can develop generalizations of truncated *QR* and *RQ* iterations for the generalized Arnoldi process. The truncated forward iteration will correspond to implicit restarting (truncated *QR*) developed in [20] while the truncated backward iteration will correspond to the truncated *RQ* iteration developed in [21]. These are recovered from the methods developed here when $\mathbf{B} = \mathbf{I}$.

Assume now that there is a partial k-step reduction to condensed form

$$\begin{aligned} \mathbf{AV}_k &= \mathbf{W}_k\mathbf{H}_k + \mathbf{f}_k\mathbf{e}_k^T, \\ \mathbf{BV}_k &= \mathbf{W}_k\mathbf{R}_k, \end{aligned} \tag{5.1}$$

as in (3.1).

Truncated FQZ:

Select a shift μ and apply one forward *QZ* step to the projected pair $(\mathbf{H}_k, \mathbf{R}_k)$ to obtain $k \times k$ unitary upper Hessenberg matrices $\mathbf{Q}_k^H$ and $\mathbf{Z}_k$ and an upper triangular $\mathbf{T}_k$ such that $\mathbf{H}_k - \mu\mathbf{R}_k = \mathbf{Z}_k\mathbf{T}_k$. Completion of the *FQZ* step will give

$$\begin{aligned}\mathbf{H}_k\mathbf{Q}_k^H &= \mathbf{Z}_k\mathbf{H}_k^+\\ \mathbf{R}_k\mathbf{Q}_k^H &= \mathbf{Z}_k\mathbf{R}_k^+,\end{aligned}$$

where $\mathbf{H}_k^+$ and $\mathbf{R}_k^+$ are order k upper Hessenberg and triangular matrices respectively. Then

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{V}_k = \mathbf{W}_k\mathbf{Z}_k\mathbf{T}_k + \mathbf{f}_k\mathbf{e}_k^T$$

and just as in the full iteration, equating the first column of both sides implies that

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v}_1 = \mathbf{W}_k\mathbf{Z}_k\mathbf{T}_k\mathbf{e}_1 = \mathbf{B}\mathbf{V}_k^+(\mathbf{R}_k^+)^{-1}\mathbf{T}_k\mathbf{e}_1 = \mathbf{B}\mathbf{v}_1^+\tau.$$

Thus,

$$(\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I})\mathbf{v}_1 = \mathbf{v}_1^+\tau,$$

where τ is the (1,1) element of $(\mathbf{R}_k^+)^{-1}\mathbf{T}_k$. Now,

$$\begin{aligned}\mathbf{A}\mathbf{V}_k^+ &= \mathbf{W}_k^+\mathbf{H}_k^+ + \mathbf{f}_k\mathbf{e}_k^T\mathbf{Q}_k^H\\ \mathbf{B}\mathbf{V}_k^+ &= \mathbf{W}_k^+\mathbf{R}_k^+\end{aligned} \tag{5.2}$$

and since $\mathbf{Q}_k^H$ is upper Hessenberg, it follows that the last row of $\mathbf{Q}_k^H$ has the form $\mathbf{e}_k^T\mathbf{Q}_k^H = [\sigma\mathbf{e}_{k-1}^T, \gamma]$. Hence, the leading $k-1$ columns on both sides of (5.3) remain in a generalized Arnoldi relation

$$\begin{aligned}\mathbf{A}\mathbf{V}_{k-1}^+ &= \mathbf{W}_{k-1}^+\mathbf{H}_{k-1}^+ + \hat{\mathbf{f}}_{k-1}\mathbf{e}_{k-1}^T\\ \mathbf{B}\mathbf{V}_{k-1}^+ &= \mathbf{W}_{k-1}^+\mathbf{R}_{k-1}^+\end{aligned}$$

where $\hat{\mathbf{f}}_{k-1} = \mathbf{W}_k^+\mathbf{e}_k\beta + \mathbf{f}_k\sigma$. Now, one additional generalized Arnoldi step may be performed to return this to an implicitly restarted k-step reduction.

Just as with the IRA iteration, this idea may be cast in the form of repeating the following steps: (1) Extend to a $k+p$ step factorization, (2) Apply p shifts with *FQZ* sweeps, (3) Truncate the last p-colums to return to a k step factorization. This will define a generalized implicitly restarted Arnoldi method.

Truncated BQZ:

To truncate the backwards *QZ* iteration, it will be necessary to derive relationships existing in column $k+1$ on both sides of (2.2). The required theory for the standard problem has been derived in [21] and this will generalize in a straightforward way to obtain a corresponding truncated backwards *QZ* equation. However, the details for completing a backward *QZ* sweep once this equation has been solved are a bit more intricate than in the *TRQ* iteration.

Following the develpment of the *TRQ* iteration, given a shift μ and the partial k-step reduction, the truncated *BQZ* is initiated by constructing vectors $\mathbf{v}$ and $\mathbf{w}$ of unit length that are orthogonal to the columns of $\mathbf{V}_k$ and $\mathbf{W}_k$ respectively, with $(\mathbf{A} - \mu\mathbf{B})\mathbf{v} \in \mathit{Range}([\mathbf{W}_k, \mathbf{w}])$. Then, a relation of the form

$$(\mathbf{A} - \mu\mathbf{B})[\mathbf{V}_k, \mathbf{v}] = [\mathbf{W}_k, \mathbf{w}]\begin{bmatrix}\mathbf{H}_k - \mu\mathbf{R}_k & \mathbf{h}\\ \beta\mathbf{e}_k^T & \alpha\end{bmatrix}. \tag{5.3}$$

is obtained to intitiate a truncated BQZ step. To develop this further, assume for purposes of the following discussion that $\mathbf{v}$, $\mathbf{w}$, $\mathbf{h}$ and α have been constructed to satisfy these relations. Let us postpone the construction of these quantities and first show how to complete the truncated BQZ step assuming that $(\mathbf{A}-\mu\mathbf{B})\mathbf{v} = \mathbf{W}_k\mathbf{h}+\mathbf{w}\alpha$. At this point, it is important to realize that the bordered Hessenberg matrix in (5.3) is precisely the leading principal submatrix that would appear if the full matrix $\mathbf{H}-\mu\mathbf{R}$ were partially factored into an RQ factorization from right to left using Givens' transformations up to the $k{+}1$ st column. The subsequent computations amount to arranging the remaining relations in the $\mathbf{W}$ and $\mathbf{R}$ matrices that would be in place had the first $n-k$ steps of a BQZ sweep been done. The idea is to anticipate this configuration and then complete the sweep in the leading k columns without ever computing the remaining $n-k$ columns of the BQZ relations.

At this point, the relationships for $\mathbf{B}$ must be brought up to date. Equations must be derived that will keep $\mathbf{B}$ in a triangular relation with the two basis sets. We first construct a vector $\mathbf{w}^+$ such that

$$\mathbf{B}\mathbf{v} = \mathbf{W}_k\mathbf{r}+\mathbf{w}^+\rho \quad \text{with} \quad \mathbf{W}_k^H\mathbf{w}^+ = 0$$

using classical Gram-Schmidt with the orthogonality correction scheme proposed in [3] Once this is done, we have

$$\mathbf{B}[\mathbf{V}_k,\mathbf{v}] = [\mathbf{W}_k,\mathbf{w}^+]\begin{bmatrix}\mathbf{R}_k & \mathbf{r}\\ 0 & \rho\end{bmatrix}. \tag{5.4}$$

From Equations (5.3) and (5.4) we may derive

$$\begin{aligned}\mathbf{A}[\mathbf{V}_k,\mathbf{v}] &= [\mathbf{W}_k,\mathbf{w}]\begin{bmatrix}\mathbf{H}_k-\mu\mathbf{R}_k & \mathbf{h}\\ \beta\mathbf{e}_k^T & \alpha\end{bmatrix} + [\mathbf{W}_k,\mathbf{w}^+]\begin{bmatrix}\mu\mathbf{R}_k & \mathbf{r}\mu\\ 0 & \rho\mu\end{bmatrix}\\ &= [\mathbf{W}_k,\mathbf{w}^+]\begin{bmatrix}\mathbf{H}_k & \mathbf{h}+\mathbf{r}\mu\\ \beta\theta\mathbf{e}_k^T & \alpha\theta+\rho\mu\end{bmatrix} + \mathbf{z}[\beta\mathbf{e}_k^T,\alpha],\end{aligned}$$

where $\mathbf{w}$ has been written as $\mathbf{w}=\mathbf{w}^+\theta+\mathbf{z}$ with $\mathbf{z}^H\mathbf{w}^+ = 0$.

At this point, in the full factorization, the leading principal $(k+1)\times(k+1)$ submatrices of the $\mathbf{H}-\mu\mathbf{R}$ and $\mathbf{R}$ matrices are of the form

$$\hat{\mathbf{H}}_{k+1}-\mu\hat{\mathbf{R}}_{k+1} = \begin{bmatrix}\mathbf{H}_k-\mu\mathbf{R}_k & \mathbf{h}\\ \beta\theta\mathbf{e}_k^T & \alpha\theta\end{bmatrix} \tag{5.5}$$

and

$$\hat{\mathbf{R}}_{k+1} = \begin{bmatrix}\mathbf{R}_k & \mathbf{r}\\ 0 & \rho\end{bmatrix}.$$

To complete the BQZ step, factor

$$\hat{\mathbf{H}}_{k+1}-\mu\hat{\mathbf{R}}_{k+1} = \mathbf{T}_{k+1}\mathbf{Z}_{k+1} \tag{5.6}$$

where $\mathbf{T}_{k+1}$ is upper triangular and $\mathbf{Z}_{k+1}$ is unitary. Now, factor

$$\mathbf{Q}_{k+1}\mathbf{R}_{k+1}^+ = \hat{\mathbf{R}}_{k+1}\mathbf{Z}_{k+1}^H,$$

where $\mathbf{Q}_{k+1}$ is unitary and $\mathbf{R}_{k+1}^+$ is upper triangular . As before, $\mathbf{Z}_{k+1}$ and $\mathbf{Q}_{k+1}^H$ are both upper Hessenberg.

From Equations (5.5) and (5.6), we observe that $(\beta \mathbf{e}_k^T, \alpha)\mathbf{Z}_{k+1}^H = (0, \tilde{\alpha})$ where $\tilde{\alpha}\theta$ is the $(k+1, k+1)$ element of $\mathbf{T}_{k+1}$.

It follows that

$$\begin{aligned} (\mathbf{A} - \mu\mathbf{B})[\mathbf{V}_k, \mathbf{v}]\mathbf{Z}_{k+1}^H &= [\mathbf{W}_k, \mathbf{w}^+]\mathbf{T}_{k+1} + \mathbf{z}(0, \tilde{\alpha}) \\ \mathbf{B}[\mathbf{V}_k, \mathbf{v}]\mathbf{Z}_{k+1}^H &= [\mathbf{W}_k, \mathbf{w}^+]\mathbf{Q}_{k+1}\mathbf{R}_{k+1}^+, \end{aligned}$$

and then

$$\begin{aligned} (\mathbf{A} - \mu\mathbf{B})[\mathbf{V}_k, \mathbf{v}]\mathbf{Z}_{k+1}^H &= [\mathbf{W}_k, \mathbf{w}^+]\mathbf{Q}_{k+1}\mathbf{Q}_{k+1}^H\mathbf{T}_{k+1} + \mathbf{z}(0, \tilde{\alpha}) \\ \mathbf{B}[\mathbf{V}_k, \mathbf{v}]\mathbf{Z}_{k+1}^H &= [\mathbf{W}_k, \mathbf{w}^+]\mathbf{Q}_{k+1}\mathbf{R}_{k+1}^+. \end{aligned} \tag{5.7}$$

As in the full case, the relations

$$\mathbf{Q}_{k+1}^H\mathbf{T}_{k+1} + \mu\mathbf{R}_{k+1}^+ = \mathbf{Q}_{k+1}^H\hat{\mathbf{H}}_{k+1}\mathbf{Z}_{k+1}^H$$

hold and imply that $\mathbf{H}_{k+1}^+ \equiv \mathbf{Q}_{k+1}^H\hat{\mathbf{H}}_{k+1}\mathbf{Z}_{k+1}^H$ is upper Hessenberg. Therefore, deleting the $k+1$-st column on both sides of (5.7) will give

$$\begin{aligned} \mathbf{A}\mathbf{V}_k^+ &= \mathbf{W}_k^+\mathbf{H}_k^+ + \mathbf{f}_k^+\mathbf{e}_k^T, \\ \mathbf{B}\mathbf{V}_k^+ &= \mathbf{W}_k^+\mathbf{R}_k^+, \end{aligned}$$

where $\mathbf{V}_k^+$ is the matrix consisting of the leading k columns of $[\mathbf{V}_k, \mathbf{v}]\mathbf{Z}_{k+1}^H$ and $\mathbf{W}_k^+$ is the matrix consisting of the leading k columns of $[\mathbf{W}_k, \mathbf{w}^+]\mathbf{Q}_{k+1}$. The matrices $\mathbf{R}_k^+$ and $\mathbf{H}_k^+$ are the leading principal order k submatrices of $\mathbf{R}_{k+1}^+$ and $\mathbf{H}_{k+1}^+$, and $\mathbf{f}_k^+$ is the last column of $[\mathbf{W}_k, \mathbf{w}^+]\mathbf{Q}_{k+1}$ scaled by the $(k+1, k)$ element of $\mathbf{H}_{k+1}^+$.

This time, observe that Equation (5.7) implies that

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v}_1^+ = \mathbf{W}_k\mathbf{T}_k\mathbf{e}_1 = \mathbf{B}\mathbf{V}_k\mathbf{R}_k^{-1}\mathbf{T}_k\mathbf{e}_1$$

so that

$$\begin{aligned} \mathbf{v}_1^+ &= (\mathbf{A} - \mu\mathbf{B})^{-1}\mathbf{B}\mathbf{v}_1\tau \\ &= (\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{v}_1\tau \end{aligned}$$

where τ is the (1,1) element of $\mathbf{R}_k^{-1}\mathbf{T}_k$. Hence, just as in the full case, the leading columns of two succesive $\mathbf{V}$ matrices are in an inverse iteration relationship.

Now that the truncated *BQZ* step is understood, it is time to develop the truncated *BQZ* equation needed to construct $\mathbf{v}, \mathbf{h}$ and α in equation (5.3) so that

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v} = \mathbf{W}_k\mathbf{h} + \mathbf{w}\alpha$$

with $\mathbf{w} = \mathbf{f}_k/\|\mathbf{f}_k\|$, $\mathbf{v}^H\mathbf{V}_k = 0$ and $\|\mathbf{v}\| = 1$. Existence and uniqueness for the case $\mathbf{B} = \mathbf{I}$ was developed in [21] and easily generalizes to this setting. Of the various possibilites developed there, the following seems most appropriated in this setting: First, compute a solution $\hat{\mathbf{v}}$ to the equation

$$(\mathbf{A} - \mu\mathbf{B})\hat{\mathbf{v}} = \mathbf{W}_k\mathbf{t} + \mathbf{f}_k\eta \tag{5.8}$$

where $(\mathbf{t}^H, \eta)^H$ is an essentially arbitrary $k+1$ vector. Then set

$$\mathbf{v} = (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}\tau \tag{5.9}$$

where $\tau = 1/\|(\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}\|$. Now put

$$\mathbf{h} = \mathbf{W}_k^H(\mathbf{A} - \mu\mathbf{B})\mathbf{v} \quad \text{and} \quad \alpha = \mathbf{w}^H(\mathbf{A} - \mu\mathbf{B})\mathbf{v}. \tag{5.10}$$

The following lemma indicates why this will work.

LEMMA 5.1. *Assume $\mathbf{A} - \mu\mathbf{B}$ is nonsingular and that there is a partial reduction of $(\mathbf{A}, \mathbf{B})$ to condensed form as in (5.1). If $\mathbf{H}_k - \mu\mathbf{R}_k$ is nonsingular, put*

$$\mathbf{t} = (\mathbf{H}_k - \mu\mathbf{R}_k)\mathbf{s} \quad \textit{and} \quad \textit{choose}\ \eta \neq \mathbf{e}_k^T\mathbf{s},$$

where $\mathbf{s}$ is any k-vector. Otherwise, let $\mathbf{t} \neq 0$ be a left null vector so that

$$0 = \mathbf{t}^H(\mathbf{H}_k - \mu\mathbf{R}_k) \quad \textit{and} \quad \textit{choose}\ \eta \ \textit{to} \ \textit{be} \ \textit{arbitrary}.$$

Let $\hat{\mathbf{v}}$ be the unique solution to (5.8). Then $0 \neq (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}$, so the vector $\mathbf{v}$ can be constructed by projection and normalized as in (5.9). Moreover,

$$(\mathbf{A} - \mu\mathbf{B})\mathbf{v} = \mathbf{W}_k\mathbf{h} + \mathbf{w}\alpha,$$

i.e. $(\mathbf{A} - \mu\mathbf{B})\mathbf{v} \in \textit{Range}([\mathbf{W}_k, \mathbf{w}])$.

Proof. Suppose $\mathbf{t}$, η, and $\hat{\mathbf{v}}$ are constructed as prescribed in the hypothesis. If $0 = (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}$, then $\hat{\mathbf{v}} = \mathbf{V}_k\mathbf{y}$ must hold for some nonzero k-vector $\mathbf{y}$. Now, this would imply

$$\begin{aligned}(\mathbf{A} - \mu\mathbf{B})\hat{\mathbf{v}} &= (\mathbf{A} - \mu\mathbf{B})\mathbf{V}_k\mathbf{y} \\ &= \mathbf{W}_k(\mathbf{H}_k - \mu\mathbf{R}_k)\mathbf{y} + \mathbf{f}_k\mathbf{e}_k^T\mathbf{y}.\end{aligned}$$

Substituting this on the left side of (5.8) and using orthogonality gives

$$(\mathbf{H}_k - \mu\mathbf{R}_k)\mathbf{y} = \mathbf{t} \quad \text{and} \quad \mathbf{e}_k^T\mathbf{y} = \eta. \tag{5.11}$$

If $\mathbf{H}_k - \mu\mathbf{R}_k$ is nonsingular, then $\mathbf{y} = \mathbf{s}$ and (5.11) would contradict the choice of η. Otherwise, the choice of $\mathbf{t}$ as a null vector would lead to the following contradiction:

$$0 = \mathbf{t}^H(\mathbf{H}_k - \mu\mathbf{R}_k)\mathbf{y} = \mathbf{t}^H\mathbf{t} \neq 0.$$

This shows $0 \neq (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}$, so that $\mathbf{v}$ can be constructed by projection and normalized as in (5.9). It remains to show $(\mathbf{A} - \mu\mathbf{B})\mathbf{v} \in \textit{Range}([\mathbf{W}_k, \mathbf{w}])$. However, this follows easily from the relations

$$\begin{aligned}(\mathbf{A} - \mu\mathbf{B})\mathbf{v} &= (\mathbf{A} - \mu\mathbf{B})\hat{\mathbf{v}} - (\mathbf{A} - \mu\mathbf{B})\mathbf{V}_k\mathbf{V}_k^H\hat{\mathbf{v}} \\ &= \mathbf{W}_k\mathbf{t} + \mathbf{f}_k\eta - [\mathbf{W}_k(\mathbf{H}_k - \mu\mathbf{R}_k) + \mathbf{f}_k\mathbf{e}_k^T]\mathbf{V}_k^H\hat{\mathbf{v}}.\end{aligned} \tag{5.12}$$

This completes the proof. □

Since $(\mathbf{A} - \mu\mathbf{B})$ is nonsingular and $[\mathbf{W}_k, \mathbf{w}]$ is unitary, $\mathbf{v}$,$\mathbf{h}$ and α are uniquely determined once $\mathbf{t}$ and η have been specified. This justifies using (5.8) and (5.9) to compute them. However, it is remarkable that $\mathbf{v}$,$\mathbf{h}$ and α are unique, regardless of the choice for $\mathbf{t}$ and η so long as $0 \neq (\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^H)\hat{\mathbf{v}}$. This result is a fairly straightforward modification of the results in Section 2 of [21].

Typically, $\mathbf{t} = \mathbf{e}_k$ is chosen because this corresponds to the standard Arnoldi process for $\mathbf{B} = \mathbf{I}$, but many other interesting choices are possible.

Remark: We may choose to cast (5.8) in the form

$$(\mathbf{I} - \mathbf{X}\mathbf{X}^H)(\mathbf{A} - \mu\mathbf{B})(\mathbf{I} - \mathbf{Z}\mathbf{Z}^H)\hat{\mathbf{v}} = \mathbf{W}_k\mathbf{t} + \mathbf{f}_k\eta, \tag{5.13}$$

where $\mathbf{X} \equiv \mathbf{W}_k\mathbf{Y}$ and $\mathbf{Z} \equiv \mathbf{V}_k\mathbf{S}$ with $\mathbf{Y}^H\mathbf{Y} = \mathbf{S}^H\mathbf{S} = \mathbf{I}_j$. Here $\mathbf{Y}$ and $\mathbf{S}$ may be of dimension $k \times j$ for any $j = 1, 2, \cdots, k$. Once $\hat{\mathbf{v}}$ is determined, (5.13) may be rearranged to obtain a relation of the form

$$(\mathbf{A} - \mu\mathbf{B})\hat{\mathbf{v}} = \mathbf{W}_k\hat{\mathbf{t}} + \mathbf{f}_k\hat{\eta},$$

since

$$(\mathbf{X}\mathbf{X}^H)(\mathbf{A}-\mu\mathbf{B})(\mathbf{I}-\mathbf{Z}\mathbf{Z}^H)\hat{\mathbf{v}} \in \mathit{Range}(\mathbf{W}_k) \text{ and } (\mathbf{A}-\mu\mathbf{B})(\mathbf{Z}\mathbf{Z}^H)\hat{\mathbf{v}} \in \mathit{Range}([\mathbf{W}_k, \mathbf{w}]).$$

Observe that there is no need to actually compute $\hat{\mathbf{t}}$ and $\hat{\eta}$. One may simply project and normalize as in (5.9) to get $\mathbf{v}$ and then obtain $\mathbf{h}$ and α as in (5.10).

This remark may have computational significance in case we choose to compute $\hat{\mathbf{v}}$ with an iterative method. In particular, if μ is a nearly converged Ritz value, then it may be a good idea to take $\mathbf{X} = \mathbf{W}_k\mathbf{y}$ where $\mathbf{y}^H(\mathbf{H}_k - \mu\mathbf{R}_k) = 0$ and $\mathbf{Z} = \mathbf{V}_k\mathbf{s}$ where $(\mathbf{H}_k - \mu\mathbf{R}_k)\mathbf{s} = 0$. This choice would tend to project out the near singularity of $(\mathbf{A} - \mu\mathbf{B})$ as suggested in [18] along the directions of the converging eigenvectors. Another possibility is to take $\mathbf{X} = \mathbf{W}_k$ and $\mathbf{Z} = \mathbf{V}_k$ as suggested in [21] to project out all of the current subspace. The latter choice is computationally more expensive (per iteration in the linear solve) but may have other advantages in the presence of clustered eigenvalues.

6. Inexact Arnoldi Processes.

In the previous two sections, algorithms have been developed to generalize the Arnoldi process and to derive truncated forms of the forward and backward *QZ* iterations. Unfortunately, these algorithms require the accurate solution of linear systems. However, the accuracy requirement for computing the direction $\mathbf{v}$ through Steps (3.3)-(3.4) may be relaxed. A projection algorithm is still obtained but the Krylov property will be lost.

To relax the exact solution requirement indicated at Step(3.3), simply replace the computation of $\mathbf{y}$ from $\mathbf{B}\mathbf{y} = \mathbf{w}$ with $\mathbf{y} = \mathit{itsol}(\mathbf{B}, \mathbf{M}, \mathbf{w})$ where $\mathbf{M}$ represents a preconditioner for $\mathbf{B}$ and *itsol* represents a few steps of a pre-conditioned iterative method for the solution of the linear system $\mathbf{B}\mathbf{y} = \mathbf{w}$. Formally, there is no accuracy requirement here and as little as one step of the iterative method may be specified. However, the rank-one nature of the residual $\mathbf{F}_k$ will be lost along with the Hessenberg form for $\mathbf{H}_k$ when this accuracy is relaxed.

Of course, there are algorithmic consequences of relaxing the accuracy requirements. The relations (3.5) are no longer valid. Therefore, the relationship $\mathbf{B}\mathbf{v} = \mathbf{W}_k\mathbf{r} + \mathbf{w}\rho$ must be forced explicitly once the direction $\mathbf{v}$ has been determined. The resulting algorithm *INXARN* is described in Fig. 6.1.

Generating Directions and the Newton Step:

Once the decision has been made to relax the Krylov property, a more general point of view may be taken. The sequence of vectors $\{\mathbf{v}_j\}$ may just as well be generated by some arbitrary process unrelated to the projections. Certainly, some relation to the shift-invert equations is desirable and the remainder of this discussion will focus on properties of the generated sequence $\{\mathbf{v}_j\}$ required for rapid convergence. With this end in mind, let us consider an arbitrary sequence of generated vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j, \ldots\}$ and assume that these vectors are orthonormal in some convenient norm.

```
TBQZ: Truncated Backward QZ-iteration

Input: $[\mathbf{A}, \mathbf{B}, \mathbf{v}, k]$ with $\mathbf{A}, \mathbf{B}$ matrices of order $n$
       $\mathbf{v}$ an $n$-vector with $\|\mathbf{v}\| = 1$
       $k << n$ the desired number of eigenvalues.

Output: $[\mathbf{V}, \mathbf{W}, \mathbf{H}, \mathbf{R}]$ such that $\mathbf{AV} = \mathbf{WH}$, $\mathbf{BV} = \mathbf{WR}$,
        $\mathbf{H}$ and $\mathbf{R}$ both $k \times k$ upper triangular,
        $\mathbf{V}^H\mathbf{V} = \mathbf{W}^H\mathbf{W} = \mathbf{I}_k$.

    1. $[\mathbf{V}, \mathbf{W}, \mathbf{H}, \mathbf{R}, \mathbf{f}] = \text{genarn}(\ \mathbf{A}, \mathbf{B}, \mathbf{v}, k)$;
    2. $\beta = \|\mathbf{f}\|$;  $\mathbf{w} = \mathbf{f}/\beta$;
    3. for $j = 1, 2, 3, \ldots$ until convergence,
      3.1. $\mu = \text{select\_shift}(\mathbf{H}, \mathbf{R})$; $\mathbf{t} = \text{select\_vector}(\mathbf{H}, \mathbf{R})$;
      3.2. Solve $(\mathbf{A} - \mu\mathbf{B})\hat{\mathbf{v}} = \mathbf{Wt}$ for $\hat{\mathbf{v}}$;
      3.3. $\mathbf{h} = \mathbf{V}^H\hat{\mathbf{v}}$;  $\hat{\mathbf{v}} \leftarrow \hat{\mathbf{v}} - \mathbf{Vh}$;  $\mathbf{v} = \hat{\mathbf{v}}/\|\hat{\mathbf{v}}\|$;
      3.4. $\mathbf{f} = \mathbf{Av}$;  $\mathbf{g} = \mathbf{Bv}$;  $\mathbf{f} \leftarrow \mathbf{f} - \mathbf{g}\mu$;
      3.5. $\mathbf{h} = \mathbf{W}^H\mathbf{f}$;  $\alpha = \mathbf{w}^H\mathbf{f}$;  $\theta = \mathbf{w}^H\mathbf{g}$;
      3.6. $\mathbf{r} = \mathbf{W}^H\mathbf{g}$;  $\mathbf{w} = \mathbf{g} - \mathbf{Wr}$;
      3.7. $\rho = \|\mathbf{w}\|$;  $\theta \leftarrow \theta/\rho$;  $\mathbf{w} = \mathbf{w}/\rho$;
      3.8. $\mathbf{H} \leftarrow \begin{bmatrix} \mathbf{H} & \mathbf{h} + \mathbf{r}\mu \\ \theta\beta\mathbf{e}_k^T & \theta\alpha + \rho\mu \end{bmatrix}$;  $\mathbf{R} \leftarrow \begin{bmatrix} \mathbf{R} & \mathbf{r} \\ 0 & \rho \end{bmatrix}$;
      3.9. $[\mathbf{T}, \mathbf{Z}] = rq(\mathbf{H} - \mu\mathbf{R})$;
      3.10. $[\mathbf{Q}, \mathbf{R}^+] = qr(\mathbf{RZ}^H)$;
      3.11. $\hat{\mathbf{H}} \leftarrow \mathbf{Q}^H\mathbf{HZ}^H$;

      3.12. $\mathbf{V} \leftarrow [\mathbf{V}, \mathbf{v}]\mathbf{Q}(:, 1:k)$ ; $[\mathbf{W}, \mathbf{w}] \leftarrow [\mathbf{W}, \mathbf{w}]\mathbf{Z}$;
      3.13. $\beta = \mathbf{H}(k+1, k)$;  $\mathbf{H} \leftarrow \mathbf{H}(1:k, 1:k)$;  $\mathbf{R} \leftarrow \mathbf{R}(1:k, 1:k)$;

    4. end;
```

FIG. 5.1. *Truncated Backward* QZ-*iteration*

Given this sequence, it is straightforward to obtain a derived sequence of orthogonal vectors $\{\mathbf{w}_j\}$ along with a sequence of projections that provide a partial reduction of the pair $(\mathbf{A}, \mathbf{B})$ to condensed form at each step:

$$\begin{aligned} \mathbf{V}_j &\leftarrow [\mathbf{V}_{j-1}, \mathbf{v}_j]; \\ \mathbf{BV}_j &= \mathbf{W}_j\mathbf{R}_j; \\ \mathbf{AV}_j &= \mathbf{W}_j\mathbf{H}_j + \mathbf{F}_j; \end{aligned}$$

with $\mathbf{W}_j^T\mathbf{W}_j = \mathbf{V}_j^T\mathbf{V}_j = \mathbf{I}_j$, $\mathbf{W}_j^T\mathbf{F}_j = 0$ as before through classical Gram Schmidt othogonalization.

How should the sequence $\{\mathbf{v}_j\}$ be generated to achieve or to accelerate convergence of the Ritz values (eigenvalues of $(\mathbf{H}_j, \mathbf{R}_j)$) to selected eigenvalues of the pair $(\mathbf{A}, \mathbf{B})$?

```
INXARN: Inexact Arnoldi Process

Input: $(\mathbf{A}, \mathbf{B}, \mathbf{v}, k)$ such that $\|v\| = 1$.
Output: $(\mathbf{V}_k, \mathbf{H}_k, \mathbf{R}_k, \mathbf{F}_k\ )$ such that $\mathbf{AV}_k = \mathbf{W}_k\mathbf{H}_k + \mathbf{F}_k\ ,\ \mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_k$,
        $\mathbf{BV}_k = \mathbf{W}_k\mathbf{R}_k,\ \mathbf{W}_k^T\mathbf{W}_k = \mathbf{I}_k,\ \mathbf{W}_k^T\mathbf{F}_k = 0$
        with $\mathbf{H}_k$ upper Hessenberg and $\mathbf{R}_k$ upper triangular.

        1. $\mathbf{V}_1 \leftarrow (\mathbf{v})$; $\mathbf{w} = \mathbf{Bv}$; $\rho = \|\mathbf{w}\|$; $\mathbf{R}_1 = (\rho)$; $\mathbf{W}_1 = (\mathbf{w}/\rho)$;
        2. $\mathbf{y} \leftarrow \mathbf{Av}$; $\mathbf{H}_1 \leftarrow (\mathbf{W}_1^T\mathbf{w})$; $\mathbf{f}_1 \leftarrow \mathbf{y} - \mathbf{W}_1\mathbf{H}_1$; $\mathbf{F}_1 = (\mathbf{f}_1)$;
        3. for $j = 1, 2, 3, ..., k-1$
           3.1. $\gamma \leftarrow \|\mathbf{f}_j\|$; $\mathbf{w} \leftarrow \mathbf{f}_j/\gamma$;
           3.2. $\hat{\mathbf{v}} = itsol(\mathbf{B}, \mathbf{M}, \mathbf{w})$;
           3.3. $\mathbf{z} \leftarrow \mathbf{V}_j^T\hat{\mathbf{v}}$;  $\hat{\mathbf{v}} \leftarrow \hat{\mathbf{v}} - \mathbf{V}_j\mathbf{z}$;
           3.4. $\mathbf{v} \leftarrow \hat{\mathbf{v}}/\|\hat{\mathbf{v}}\|$;
           3.5. $\hat{\mathbf{w}} \leftarrow \mathbf{Bv}$; $\mathbf{r} \leftarrow \mathbf{W}_j^T\hat{\mathbf{w}}$;  $\hat{\mathbf{w}} \leftarrow \hat{\mathbf{w}} - \mathbf{W}_j\mathbf{r}$;  $\rho \leftarrow \|\hat{\mathbf{w}}\|$;
           3.6. $\mathbf{w} \leftarrow \hat{\mathbf{w}}/\rho$;  $\mathbf{W}_{j+1} \leftarrow (\mathbf{W}_j, \mathbf{w})$;  $\hat{\mathbf{R}}_{j+1} \leftarrow \begin{pmatrix} \hat{\mathbf{R}}_j & \mathbf{r} \\ 0 & \rho \end{pmatrix}$;
           3.7. $\mathbf{y} \leftarrow \mathbf{Av}$;   $\mathbf{h} \leftarrow \mathbf{W}_{j+1}^T\mathbf{y}$;   $\mathbf{c}^T = \mathbf{w}^T\mathbf{F}_j$;
           3.8. $\mathbf{H}_j \leftarrow \begin{pmatrix} \mathbf{H}_j \\ \mathbf{c}^T \end{pmatrix}$; $\mathbf{V}_{j+1} \leftarrow (\mathbf{V}_j, \mathbf{v})$;
           3.9. $\mathbf{f}_{j+1} \leftarrow \mathbf{y} - \mathbf{W}_{j+1}\mathbf{h}$; $\mathbf{H}_{j+1} \leftarrow (\mathbf{H}_j, \mathbf{h})$; $\mathbf{F}_{j+1} \leftarrow [\mathbf{F}_j - \mathbf{wc}^T, \mathbf{f}_{j+1}]$;
        4. end
```

FIG. 6.1. *An Inexact Arnoldi Process.*

Certainly, it would be helpful to develop a connection with Newton's method and then perhaps modify those choices to reduce computional cost while retaining reasonable convergence properties. To this end, suppose $\mathbf{Hy} = \mathbf{Ry}\theta$ and $\mathbf{x} = \mathbf{Vy}$ with $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. Let $\lambda \in \sigma(\mathbf{A}, \mathbf{B})$ be the closest eigenvalue to θ and let $\mathbf{q}$ be the corresponding eigenvector normalized so that $\mathbf{x}^H\mathbf{q} = 1$ (hence $\|\mathbf{q}\| \geq 1$).

With these assumptions, let us represent

$$\mathbf{q} = \mathbf{x} + \mathbf{z}, \quad \lambda = \theta + \delta,$$

with $\mathbf{x}^H\mathbf{z} = 0$ and derive the standard second order approximation from the relation $\mathbf{Aq} = \mathbf{Bq}\lambda$. Substituting, combining and rearranging terms gives

$$(\mathbf{A} - \theta\mathbf{B})\mathbf{z} = -(\mathbf{A} - \theta\mathbf{B})\mathbf{x} + \mathbf{Bx}\delta + \mathbf{Bz}\delta \tag{6.1}$$

At this point, several alternatives are available to approximate the correction vector $\mathbf{z}$. Two possibilities shall be examined here. The first of these gives the correction developed in [18, 6]. Since $\mathbf{x} = \mathbf{Vy}$, it follows that

$$\begin{aligned} -(\mathbf{A} - \theta\mathbf{B})\mathbf{x} + \mathbf{Bx}\delta &= -\mathbf{W}(\mathbf{H} - \theta\mathbf{R})\mathbf{y} - \mathbf{Fy} + \mathbf{WRy}\delta \\ &= -\mathbf{Fy} + \mathbf{WRy}\delta. \end{aligned}$$

Now, if both sides of equation (6.1) are multiplied on the left by $\mathbf{I} - \mathbf{WW}^H$ the resulting equation is

$$(\mathbf{I} - \mathbf{WW}^H)(\mathbf{A} - \theta\mathbf{B})(\mathbf{I} - \mathbf{xx}^H)\mathbf{z} = -\mathbf{Fy} + (\mathbf{I} - \mathbf{WW}^H)\mathbf{Bz}\delta, \tag{6.2}$$

since $0 = \mathbf{W}^H\mathbf{F}$ and $0 = \mathbf{x}^H\mathbf{z}$. From this, it also follows that equation (6.2) is consistent and there is a unique minimum norm solution $\mathbf{z}$. Hence the direction $\mathbf{v}$ obtained by finding the minimum norm solution to

$$(\mathbf{I} - \mathbf{W}\mathbf{W}^H)(\mathbf{A} - \theta\mathbf{B})(\mathbf{I} - \mathbf{x}\mathbf{x}^H)\mathbf{v} = -\mathbf{F}\mathbf{y}$$

will assure that the second order correction is a member of the updated spaces $\mathcal{S}_\mathbf{V} \equiv Range(\mathbf{V})$ and $\mathcal{S}_\mathbf{W} \equiv Range(\mathbf{W})$ when $\mathbf{v}$ is adjoined and the corresponding $\mathbf{w}$ is obtained.

An alternative to the solution just developed is to treat equation (6.1) in a straightforward way assuming that the matrix $\mathbf{A} - \theta\mathbf{B}$ is nonsingular. Then

$$\mathbf{z} = -\mathbf{x} + (\mathbf{A} - \theta\mathbf{B})^{-1}\mathbf{B}\mathbf{x}\delta + (\mathbf{A} - \theta\mathbf{B})^{-1}\mathbf{B}\mathbf{z}\delta. \tag{6.3}$$

Now, using the facts $0 = \mathbf{x}^H\mathbf{z}$ and $0 = (\mathbf{I} - \mathbf{x}\mathbf{x}^H)\mathbf{x}$ gives

$$\mathbf{z} = (\mathbf{I} - \mathbf{x}\mathbf{x}^H)(\mathbf{A} - \theta\mathbf{B})^{-1}\mathbf{B}\mathbf{x}\delta + (\mathbf{I} - \mathbf{x}\mathbf{x}^H)(\mathbf{A} - \theta\mathbf{B})^{-1}\mathbf{B}\mathbf{z}\delta,$$

when both sides of equation (6.3) are multiplied on the left by the projection $(\mathbf{I} - \mathbf{x}\mathbf{x}^H)$. Now, the second order correction will be included in the updated spaces if the new direction $\mathbf{v}$ obtained by finding the solution $\hat{\mathbf{z}}$ to

$$(\mathbf{A} - \theta\mathbf{B})\hat{\mathbf{z}} = \mathbf{B}\mathbf{x},$$

and then projecting and normalizing to get

$$\mathbf{z} = (\mathbf{I} - \mathbf{x}\mathbf{x}^H)\hat{\mathbf{z}} \quad \text{and} \quad \mathbf{v} = \mathbf{z}/\|\mathbf{z}\|.$$

Note the advantage here of adjoining the direction $\mathbf{z}$ to the existing space. We do not need to explicitly compute δ in (6.3) as would be needed in an explicit Newton method. This projection process assures that the Newton correction is in the updated subspace so that the new Ritz vector and Ritz value will be at least as good as those obtained through an explicit Newton step.

The methods of Davidson [4], Olsen et. al, [15], Sleijpen and Van der Vorst [18] and those introduced and discussed by Knyazev [10] can all be placed within this Newton-like framework.

Blocked Formulation: Futher consideration of the previous development would suggest that a block formulation is more appropriate than a single vector approach when the Krylov property is no longer enforced. To develop this, we assume a partial decomposition of the form

$$\begin{aligned} \mathbf{A}\mathbf{V}_1 &= \mathbf{W}_1\mathbf{H}_{11} + \mathbf{F}_1, \quad \text{with} \quad \mathbf{W}_1^H\mathbf{F}_1 = 0, \\ \mathbf{B}\mathbf{V}_1 &= \mathbf{W}_1\mathbf{R}_{11}, \end{aligned} \tag{6.4}$$

where $\mathbf{V}_1, \mathbf{W}_1, \mathbf{F}_1$ are $n \times k$ matrices and $\mathbf{H}_{11}, \mathbf{R}_{11}$ are $k \times k$ matrices. We then construct the $n \times k$ matrix $\mathbf{V}_2$ as follows:

$$\begin{aligned} \mathbf{V} &= (\mathbf{I} - \mathbf{V}_1\mathbf{V}_1^H)p(\mathbf{A}, \mathbf{B})\mathbf{F}_1, \\ [\mathbf{V}_2, \mathbf{T}] &= qr(\mathbf{V}), \end{aligned}$$

(i.e.,$\mathbf{V}_2\mathbf{T} = \mathbf{V}$ with $\mathbf{V}_2$ orthogonal and $\mathbf{T}$ upper triangular). Obtain additional basis vectors $\mathbf{W}_2$ via

$$\mathbf{B}\mathbf{V}_2 = \mathbf{W}_1\mathbf{R}_{12} + \mathbf{W}_2\mathbf{R}_{22} \quad \text{with} \quad \mathbf{W}_1^H\mathbf{W}_2 = 0, \quad \mathbf{W}_2^H\mathbf{W}_2 = \mathbf{I}_k.$$

Then compute $\mathbf{H}_{12}$, $\mathbf{H}_{21}$, $\mathbf{H}_{22}$, $\mathbf{F}_1^+$ and $\mathbf{F}_2$ such that

$$\begin{aligned} \mathbf{A}\mathbf{V}_1 &= \mathbf{W}_1\mathbf{H}_{11} + \mathbf{W}_2\mathbf{H}_{21} + \mathbf{F}_1^+, \\ \mathbf{A}\mathbf{V}_2 &= \mathbf{W}_1\mathbf{H}_{12} + \mathbf{W}_2\mathbf{H}_{22} + \mathbf{F}_2. \end{aligned}$$

Finally, apply the *QZ* method (say) to the pair $(\mathbf{H}, \mathbf{R})$ to obtain unitary matrices $\mathbf{Q}$, $\mathbf{Z}$, an upper-triangular $\mathbf{H}^+$ and an upper triangular $\mathbf{R}^+$ such that

$$\begin{aligned} \mathbf{H}\mathbf{Q} &= \mathbf{Z}\mathbf{H}^+, \\ \mathbf{R}\mathbf{Q} &= \mathbf{Z}\mathbf{R}^+, \end{aligned}$$

where $\mathbf{H} = (\mathbf{H}_{ij})$ and $\mathbf{R} = (\mathbf{R}_{ij})$, $i = 1, 2; j = 1, 2$, with the best approximations to the desired eigenvalues appearing as eigenvalues of $(1, 1)$ block of the pair $(\mathbf{H}^+, \mathbf{R}^+)$. Now, update

$$\begin{aligned} \mathbf{V}_1 &\leftarrow [\mathbf{V}_1, \mathbf{V}_2]\mathbf{Q}(:, 1:k), \quad \mathbf{W}_1 \leftarrow [\mathbf{W}_1, \mathbf{W}_2]\mathbf{Z}(:, 1:k), \\ \mathbf{H}_{11} &\leftarrow \mathbf{H}^+(1:k, 1:k), \quad \mathbf{R}_{11} \leftarrow \mathbf{R}^+(1:k, 1:k), \\ \mathbf{F}_1 &\leftarrow [\mathbf{F}_1^+, \mathbf{F}_2]\mathbf{Q}(:, 1:k). \end{aligned}$$

In this development, $p(\mathbf{A}, \mathbf{B})$ represents a matrix polynomial in $\mathbf{A}$ and $\mathbf{B}$ generated by a (preconditioned) iterative method designed to solve

$$(\mathbf{A} - \theta\mathbf{B})\hat{\mathbf{V}} = \mathbf{F}_1.$$

In fact, $\mathbf{G} \equiv p(\mathbf{A}, \mathbf{B})\mathbf{F}_1$ could easily represent a much more general object with each column of $\mathbf{G}$ representing a separate iterative solution of the form

$$\mathbf{g}_j \approx (\mathbf{A} - \theta_j\mathbf{B})^{-1}\mathbf{F}_1\mathbf{y}_j, \quad j = 1, 2, \cdots, k.$$

This could be made very efficient in terms of data movement per matrix-vector product. Each separate column would need two operations of the form $\mathbf{A}\mathbf{g}_j$ and $\mathbf{B}\mathbf{g}_j$. For example, a Richardson's iteration could take the form

```
$\mathbf{G} = \mathbf{F}_1\mathbf{Y}$;
for $j = 1, 2, \cdots$
    $\mathbf{G} \leftarrow \mathbf{G}\mathbf{\Gamma} - \mathbf{A}\mathbf{G} - \mathbf{B}\mathbf{G}\mathbf{\Theta}$;
end
```

where $\mathbf{\Gamma} \equiv diag(\gamma_1, \gamma_2, ..., \gamma_k)$ with reciprocal Richardson parameters γ_j and $\mathbf{\Theta} \equiv diag(\theta_1, \theta_2, ..., \theta_k)$ and $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k]$ the current Ritz approximations to desired eigenvalues and vectors, i.e. $\mathbf{H}\mathbf{Y} = \mathbf{R}\mathbf{Y}\mathbf{\Theta}$.

We may express the above discussion formally as the algorithm *BLKQZ* shown in Figure 6.2.

```
**BLKQZ:** Block Inexact $QZ$ Process

Input: $(\mathbf{A},\mathbf{B},\mathbf{V}_1,k)$ such that $\mathbf{V}_1^H\mathbf{V}_1 = \mathbf{I}_k$,
Output: $(\mathbf{V}_1,\mathbf{H}_{11},\mathbf{R}_{11})$ such that $\mathbf{A}\mathbf{V}_1 = \mathbf{W}_1\mathbf{H}_{11}$ , $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}_k$.
        $\mathbf{B}\mathbf{V}_1 = \mathbf{W}_1\mathbf{R}_{11}$, $\mathbf{W}_1^H\mathbf{W}_1 = \mathbf{I}_k$,
        with $\mathbf{H}_{11}$ upper upper triangular and $\mathbf{R}_{11}$ upper triangular.

        1. $\hat{\mathbf{W}}_1 = \mathbf{B}\mathbf{V}_1$; $[\mathbf{W}_1,\mathbf{R}_{11}] = qr(\hat{\mathbf{W}}_1)$;
        2. $\mathbf{F}_1 \leftarrow \mathbf{A}\mathbf{V}_1$; $\mathbf{H}_{11} \leftarrow (\mathbf{W}_1^T\mathbf{F}_1)$; $\mathbf{F}_1 \leftarrow \mathbf{F}_1 - \mathbf{W}_1\mathbf{H}_1$;
        3. for $j = 1,2,3,...,k-1$
          3.1. $\hat{\mathbf{V}}_2 = itsol(\mathbf{A},\mathbf{B},\mathbf{M},\mathbf{F}_1,\mathbf{Y}_1)$;
          3.2. $\mathbf{S} \leftarrow \mathbf{V}_1^H\hat{\mathbf{V}}_2$; $\hat{\mathbf{V}}_2 \leftarrow \hat{\mathbf{V}}_2 - \mathbf{V}_1\mathbf{S}$;
          3.3. $[\mathbf{V}_2,\mathbf{S}] = qr(\mathbf{V}_2)$;
          3.4. $\hat{\mathbf{W}}_2 \leftarrow \mathbf{B}\mathbf{V}_2$; $\mathbf{R}_{12} \leftarrow \mathbf{W}_1^T\mathbf{W}_2$;
          3.5. $\hat{\mathbf{W}}_2 \leftarrow \hat{\mathbf{W}}_2 - \mathbf{W}_1\mathbf{R}_{12}$; $[\mathbf{W}_2,\mathbf{R}_{22}] = qr(\hat{\mathbf{W}}_2)$;
          3.6. $\mathbf{H}_{21} \leftarrow \mathbf{W}_2^H\mathbf{F}_1$; $\mathbf{F}_1 \leftarrow \mathbf{F}_1 - \mathbf{W}_2\mathbf{H}_{21}$;
          3.7. $\mathbf{F}_2 \leftarrow \mathbf{A}\mathbf{V}_2$; $\mathbf{H}_{12} \leftarrow \mathbf{W}_1^H\mathbf{F}_2$;
          3.8. $\mathbf{F}_2 \leftarrow \mathbf{F}_2 - \mathbf{W}_1\mathbf{H}_{12}$;  $\mathbf{H}_{22} = \mathbf{W}_2^H\mathbf{F}_2$;  $\mathbf{F}_2 \leftarrow \mathbf{F}_2 - \mathbf{W}_2\mathbf{H}_{22}$;
          3.9. $\mathbf{H} \leftarrow \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}$;  $\mathbf{R} \leftarrow \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ 0 & \mathbf{R}_{22} \end{pmatrix}$;
          3.10. $[\mathbf{Q},\mathbf{Z},\mathbf{H},\mathbf{R}] = qziter(\mathbf{H},\mathbf{R},'sort')$;
          3.11. $\mathbf{V}_1 \leftarrow [\mathbf{V}_1,\mathbf{V}_2]\mathbf{Q}[:,1:k]$;  $\mathbf{W}_1 \leftarrow [\mathbf{W}_1,\mathbf{W}_2]\mathbf{Z}[:,1:k]$;
          3.12.$\mathbf{F}_1 \leftarrow [\mathbf{F}_1,\mathbf{F}_2]\mathbf{Q}[:,1:k]$;
          3.13. $\mathbf{H}_{11} \leftarrow \mathbf{H}(1:k,1:k)$;  $\mathbf{R}_{11} \leftarrow \mathbf{R}(1:k,1:k)$;
        4. end
```

FIG. 6.2. *A Block Inexact* QZ *Process*

7. Computational Results and Conclusions. We shall present some very preliminary computational results to give some indication of the relative performance of three methods: *TFQZ*, *TBQZ*, *BLKQZ*. The purpose of these results is mainly to indicate that the methods have been programmed and will solve a difficult problem. There are many implementation details to consider and a number of parameter choices to be made. A thorough computational study including comparison with other methods is certainly called for.

Our results will consist of a comparison of the three methods on a single problem. The problem we consider is a symmetric generalized problem from the Harwell–Boeing collection. The matrices are stiffness and mass matrices were obtained through the Matrix Market from

`http://math.nist.gov/MatrixMarket/data/Harwell-Boeing/bcsstruc1/`

to form a generalized eigenvalue problem $\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{x}\lambda$. The is matrix $\mathbf{A}$ is BCSSTK12 and the matrix $\mathbf{B}$ is BCSSTM12 from the BCSSTRUC1 set. BCSSTK12 and BCSSTM12 represent the consistent mass formulation for an ore car model. The consistent mass formulation leads to a non-diagonal mass matrix. All computations

TABLE 7.1
Eigenvalues calculated by TBQZ.

Eigenvalues	Error/λ_{min}	Error/λ_{max}
3.469305448324274e+03	2.8e-11	4.3e-16
3.670875661790737e+03	2.27e-11	3.4e-16
5.538220406841684e+03	3.7e-10	5.5e-15
6.410197672779293e+03	1.0e-09	1.5e-14

were done in Matlab Version 5.1.0.421 on a Sun SparcStation 20 Model 61 with 64 megabytes of RAM.

For these matrices,$n = 1473$ and $\mathbf{A}$ has 17857 nonzero entries. The smallest four generalized eigenvalues are

```
3.469305448042201e+03
3.670875662014555e+03
5.538220410502827e+03
6.410197662646212e+03
```

and the largest generalized eigenvalue is on the order of 6.55e+08.

Here, we list estimates of the computational and storage costs of the three routines and indicate the performance of each of them on this test problem. The term “matvec” stands for a matrix-vector product and the term “LU-solve” stands for solving the two successive triangular linear systems first with $\mathbf{L}$ and then with $\mathbf{U}$ as coefficient matrices.

TBQZ:

For a k-step factorization, the work and storage required for **TBQZ** is

- Storage: $2n(k+1)$ plus storage for $\mathbf{A}, \mathbf{B}, \mathbf{L}, \mathbf{U}$
- Initial work:
  1 sparse LU-factorization,
  $k+1$ LU-solves,
  $4n(k+1)^2$ flops for orthogonalization.
- Work per iteration:
  1 LU-solve,
  1 matvec with (A, B),
  $4n(k+1)^2$ flops for orthogonalization,
  sparse LU-factorization if there is a shift change.

For our run, $k = 9$ and the iteration was halted after four Ritz values had converged. The code took 14 iterations and 7 matrix factorizations. The eigenvalues computed by *TBQZ* are shown in Table (7.1).

TFQZ:

For an m step factorization that retains a k step factorization after each implicit restart, the work and storage required is

- Storage: $2nm$ plus storage for $\mathbf{A}, \mathbf{B}, \mathbf{L}, \mathbf{U}$
- Initial work:

TABLE 7.2
Eigenvalues calculated by TFQZ.

Eigenvalues	Error/λ_{min}	Error/λ_{max}
3.469305447658971e+03	3.8e-11	5.8e-16
3.670875661610020e+03	4.1e-11	6.1e-16
5.538220410338460e+03	1.6e-11	2.5e-16
6.410197662356884e+03	2.9e-11	4.4e-16

TABLE 7.3
Eigenvalues calculated by BLKQZ.

Eigenvalues	Error/λ_{min}	Error/λ_{max}
3.469305447907588e+03	1.4e-11	2.0e-16
3.670875661903084e+03	1.1e-11	1.7e-16
5.538220410459639e+03	4.4e-12	6.6e-17
6.410197662585929e+03	6.1e-12	9.2e-17

1 sparse LU-factorization,
m LU-solves,
$4nm^2$ flops for orthogonalization.

- Work per iteration:
  $m-k$ LU-solves,
  $m-k$ matvecs with (A,B),
  $4nm^2$ flops for orthogonalization,

For our run, $k=4$ and $m=12$ with $tol = 1.0e-09$. The code took two iterations and 20 LU-solves. The eigenvalues computed by *TFQZ* are shown in Table (7.2).

BLKQZ:

The work and storage required with blocksize k is

- Storage: $4n(2k)$ plus storage for $\mathbf{A},\mathbf{B},\mathbf{L},\mathbf{U}$
- Initial work:
  1 incomplete sparse LU-factorization,
  1 block ILU-solve,
  $4n(2k)^2$ flops.
- Work per iteration:
  1 block ILU-solve,
  1 block matvec with (A,B),
  $30n(2k)^2$ flops,

For our run, $k=4$. The code took 43 matrix accesses, 43 block matvecs (A,B) and 443 individual matrix-vector products. The eigenvalues computed by *BLKQZ* are shown in Table (7.3).

In each routine, we used a reference shift of $\sigma = 3.4e{+}3$ and in the call to `tfqz` we passed $\mathbf{A}-\sigma\mathbf{B}$ in place of $\mathbf{B}$ and $\mathbf{B}$ in place of $\mathbf{A}$ in the calling sequence. This is mathematically equivalent to using implicit restarting with the shift-invert operator $(\mathbf{A}-\mathbf{B})^{-1}\mathbf{B}$ and the convergence results confirm that. For the *BLKQZ* method we used a block variant of BICGSTAB that we constructed from the single vector code in the templates collection [1] and with an incomplete LU preconditioner from Matlab. We were able to arrange the code so that each column of the right hand side represented a residual of the form

$$\mathbf{r}_j = (\mathbf{A} - \mu_j \mathbf{B})\mathbf{x}_j$$

but used the same pre-conditioner for the whole block. Typically, not all of the column equations converged and our cut off was 10 iterations. As the results show, this was sufficient for convergence.

With these results, it is difficult to choose between the methods. Here, *TFQZ* seems to be the winner but that is in absence of any architecture considerations and without specific comparison between ILU and complete LU costs. We did not report flop counts or timings because the implementations are fairly crude at this point in time. These results only indicate that the three methods are indeed implementable and that they work on a challenging problem.

The real value of the *TBQZ* may lie in its applicability to rational interpolation with respect to constructing reduced order models of state space control systems as explored in [9]. More investigation and testing needs to be done with respect to shift selection and selecting the right hand side of the BQZ equations. The pre-conditioned *BLKQZ* is very promising with respect to parallel performance but is far from robust at this time.

Acknowledgement. The author would like to thank R. B. Lehoucq and C. Yang for several discussions and comments. He would also like to thank the anonymous referee for several useful comments and for catching some mistakes in the section on inexact iterations.

REFERENCES

[1] Richard Barrett, Michael Berry, Tony F. Chan, James Demmel, June Donato, Jack Dongarra, Victor Eijkhout, Roldan Pozo, Charles Romine, and Henk van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods.* SIAM, Philadelphia, USA, 1993.

[2] M. Crouzeix, B. Philippe, and M. Sadkane. The Davidson method. *SIAM J. Scientific Computing*, 15:62–76, 1994.

[3] J. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart. Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization. *Mathematics of Computation*, 30:772–795, 1976.

[4] Ernest R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17:87, 1975.

[5] T. Ericsson and A. Ruhe. The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems. *Mathematics of Computation*, 35:1251–1268, October 1980.

[6] D.R. Fokkema, G.L.G. Sleijpen, and H.A. van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the partial reduction of matrix pencils. Technical Report Preprint 941, Department of Mathematics, Utrecht University, Utrecht, The Netherlands, January 1996.

[7] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, third edition, 1996.

[8] 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(1):228–272, January 1994.

[9] E.J. Grimme. *Krylov Projection Methods for Model Reduction.* PhD thesis, University of Illinois, Urbana-Champaign, 1997.

[10] A.V. Knyazev. Convergence rate estimates for iterative methods for a mesh symmetric eigenvalue problem. *Sov. J. Numer. Anal. Math. Modeling*, 2(5):371–396, 1987.

[11] R. B. Lehoucq and Karl Meerbergen. Using generalized cayley transformations within an inexact rational krylov sequence method. *SIAM J. Matrix Analysis and Applications*, To appear, 1998. Revised version of the Argonne National Laboratory Preprint MCS-P612-1096.

[12] K. Meerbergen and A. Spence. Implicitly restarted Arnoldi with purification for the shift–invert transformation. *Mathematics of Computation*, 218:667–689, 1997.

[13] C.B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM J. Num. Anal.*, 10:241–256, 1973.

[14] R. B. Morgan and D. S. Scott. Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices. *SIAM J. Scientific and Statistical Computing*, 7:817–825, 1986.

[15] J. Olsen, P. Jorgensen, and J. Simons. Passing the one-billion limit in full configuration interaction (fci) calculations. *Chemical Physics Letters*, 169(6):463–472, 1990.

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

[17] G. De Samblanx, K. Meerbergen, and A. Bultheel. The implicit application of a rational filter in the rks method. *BIT*, 37:925–947, 1997.

[18] G. L. G. Sleijpen and H.A. van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Analysis and Applications*, 17(2):401–425, 1996.

[19] G.L.G. Sleijpen, J.G.L. Booten, D.R. Fokkema, and H.A. van der Vorst. Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT*, 36(3):595–633, 1996.

[20] D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Analysis and Applications*, 13(1):357–385, January 1992.

[21] D.C. Sorensen and C. Yang. A truncated RQ-iteration for large scale eigenvalue calculations. *SIAM J. Matrix Analysis and Applications*, to appear. Available as Rice U. Rept. CAAM-TR96-06.