University of Maryland College Park

Institute for Advanced Computer Studies TR–2000–21
Department of Computer Science TR–4127

# An Arnoldi–Schur Algorithm for Large Eigenproblems*

G. W. Stewart[†]

April 2000

ABSTRACT

Sorensen's iteratively restarted Arnoldi algorithm is one of the most successful and flexible methods for finding a few eigenpairs of a large matrix. However, the need to preserve structure of the Arnoldi decomposition, on which the algorithm is based, restricts the range of transformations that can be performed on it. In consequence, it is difficult to deflate converged Ritz vectors from the decomposition. Moreover, the potential forward instability of the implicit QR algorithm can cause unwanted Ritz vectors to persist in the computation. In this paper we introduce a generalized Arnoldi decomposition that solves both problems in a natural and efficient manner.

*This report is available by anonymous ftp from thales.cs.umd.edu in the directory pub/reports or on the web at http://www.cs.umd.edu/~stewart/.

[†]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (stewart@cs.umd.edu). Work supported by the National Science Foundation under Grant No. 970909-8562

# An Arnoldi–Schur Algorithm for Large Eigenproblems

G. W. Stewart

ABSTRACT

Sorensen's iteratively restarted Arnoldi algorithm is one of the most successful and flexible methods for finding a few eigenpairs of a large matrix. However, the need to preserve structure of the Arnoldi decomposition, on which the algorithm is based, restricts the range of transformations that can be performed on it. In consequence, it is difficult to deflate converged Ritz vectors from the decomposition. Moreover, the potential forward instability of the implicit QR algorithm can cause unwanted Ritz vectors to persist in the computation. In this paper we introduce a generalized Arnoldi decomposition that solves both problems in a natural and efficient manner.

## 1. Introduction and background

In this paper we are going to describe a new implementation of the Arnoldi method that resolves some difficulties with the implicitly restarted Arnoldi method. To understand the difficulties and their solution requires a detailed knowledge of the Arnoldi process. We therefore begin with a survey, which will also serve to set the notation for this paper.

Let $A$ be a matrix of order $n$ and let $u_1$ be a vector of 2-norm one. Let $u_1, u_2, u_3 \ldots$ be the result of sequentially orthogonalizing the Krylov sequence $u_1, Au_1, A^2u_1, \ldots$. In 1950, Lanczos [5] showed that if $A$ is Hermitian then the vectors $u_i$ satisfy a three term recurrence of the form

$$\beta_k u_{k+1} = A_k u_k - \alpha_k u_k - \beta_{k-1} u_{k-1}, \tag{1.1}$$

a recursion that in principle allows the economical computation of the $u_j$.

There is an elegant representation of this recursion in matrix terms. Let

$$U_k = (u_1 \;\; u_2 \;\; \cdots \;\; u_k)$$

be the matrix formed from the Lanczos vectors $u_j$. Then there is a tridiagonal matrix $T$ formed from the $\alpha$'s and $\beta$'s in (1.1) such that

$$AU_k = U_k T_k + \beta_k u_{k+1} \mathbf{e}_k^{\mathrm{T}}, \tag{1.2}$$

where $\mathbf{e}_k$ is the vector whose last component is one and whose other components are zero. From the orthogonality of the $u_j$, it follows that $T_k$ is the Rayleigh quotient

$$T_k = U_k^{\mathrm{H}} A U_k.$$

We will call (1.2) a Lanczos decomposition.

Lanczos appreciated the fact that even for comparatively small $k$ the matrix $T_k$ could contain accurate approximations to the eigenvalues of $A$. When this happens, the column space $\mathcal{U}_k$ of $U_k$ will usually contain approximations to the corresponding eigenvectors. Such an approximation — call it $z$ — can be calculated by computing a suitable eigenpair $(\mu, w)$ of $T_k$ and setting $z = U_k w$. This process is called the Rayleigh–Ritz method; $\mu$ is called a Ritz value and $z$ a Ritz vector.

In 1951, Arnoldi [1], building on Lanczos's work, showed that if $A$ is non-Hermitian then the Lanczos decomposition becomes

$$AU_k = U_k H_k + \beta_k u_{k+1} \mathbf{e}_k^{\mathrm{T}}, \tag{1.3}$$

where $H_k$ is upper Hessenberg. We will call (1.3) an Arnoldi decomposition. Once again, $H_k$ may contain accurate approximations to the eigenvalues of $A$, especially those on the periphery of the spectrum of $A$. Moreover, approximations to the eigenvectors may be obtained by the natural generalization of the Rayleigh–Ritz process.

Since $H_k$ is not tridiagonal, the Arnoldi vectors do not satisfy a three term recurrence. To compute $u_{k+1}$ all the columns of $U_k$ must be readily available. If $n$ is large, these vectors will soon consume the available storage, and it the process must be restarted. The problem then becomes how choose a new $u_1$ that does not discard the information about the eigenvectors contained in $\mathcal{U}_k$. There have been several proposals, whose drawbacks have been nicely surveyed by Morgan [9].

In 1992, Sorensen [10] suggested an elegant way to use the QR algorithm to restart the Arnoldi process. Specifically, suppose we have an Arnoldi decomposition

$$AU_m = U_m H_m + \beta_m u_{m+1} \mathbf{e}_m^{\mathrm{T}} \tag{1.4}$$

of order $m$ that cannot be further expanded because of lack of storage. For some fixed $k$, choose $m-k$ shifts $\kappa_1, \ldots, \kappa_{m-k}$ and use them to perform $m-k$ steps of the implicitly shifted $QR$ algorithm on the Rayleigh quotient $H_m$. The effect is to generate an orthogonal matrix $Q$ such that $Q^{\mathrm{H}} H_m Q$ is upper Hessenberg. Then from (1.4)

$$A(U_m Q) = (U_m Q) Q^{\mathrm{H}} H_m Q + \beta_m u_{m+1} \mathbf{e}_m^{\mathrm{T}} Q.$$

or

$$A\tilde{U}_m = \tilde{U}_m \tilde{H}_m + u_{m+1} c^{\mathrm{H}}.$$

Sorensen then observed is that the structure of $Q$ is such that the first $k-1$ components of $c$ are zero. Consequently, if we let $\tilde{H}_k$ be the leading principal submatrix of $\tilde{H}_m$ of order $k$ and set

$$\beta_k \tilde{u}_{k+1} = \bar{\gamma}_k u_{m+1} + \tilde{h}_{k+1,k} u_{k+1}, \tag{1.5}$$

then

$$A\tilde{U}_k = \tilde{U}_k \tilde{H}_k + \tilde{u}_{k+1} e_k^{\mathrm{T}}$$

is an Arnoldi decomposition of order $k$. This process of truncating the decomposition is called implicit restarting.

A second key observation of Sorensen suggests a rationale for choosing the shifts. Specifically, if $p(t) = (t - \kappa_1 I) \cdots (t - \kappa_{m-k} I)$, then

$$\tilde{u}_1 = \frac{p(A)u_1}{\|p(A)u_1\|}.$$

It follows that if we choose the shifts to lie in the part of the spectrum that we are not interested in then the implicit restart process deemphasizes these very eigenvalues.

Each iteration of Sorensen's algorithm consists of two stages: an expansion stage, in which the decomposition is expanded until it is inconvenient to go further, and a contraction or purging stage, in which unwanted parts of the spectrum are suppressed. The contraction phase has two variants. In the exact variant, the shifts are taken to be unwanted eigenvalues of $H_m$. If, for example we were concerned with stability, we might choose to retain only the eigenvalues with largest real parts. In the general variant, the shifts are not necessarily eigenvalues of $H_m$. For example, they might be the zeros of a Chebyshev polynomial spanning an ellipse containing unwanted eigenvalues.

The implicitly restarted Arnoldi algorithm has been remarkably successful and has been implemented in the widely used ARPACK package [7]. However, the method has two important drawbacks.

First, for the exact restart procedure to be effective the unwanted Ritz values $\mu$ must be moved to the end of $H_m$, so that the Rayleigh quotient has the form illustrated below for $k = 3$ and $m = 6$:

$$\begin{pmatrix} h & h & h & h & h & h \\ h & h & h & h & h & h \\ 0 & h & h & h & h & h \\ 0 & 0 & 0 & \mu & h & h \\ 0 & 0 & 0 & 0 & \mu & h \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}. \tag{1.6}$$

If $H_m$ is unreduced — that is, if the elements of its first subdiagonal are nonzero — then mathematically $H_m$ must have the form (1.6). In the presence of rounding error, however, the process can fail (for a treatment of this phenomenon see [12]). This has lead Lehoucq and Sorensen to propose an elaborate method for permanently ridding the decomposition of persistent unwanted Ritz values [6].

The second problem is to move converged Ritz values $\mu$ to the beginning of $H_k$, so that it assumes the form illustrated below:

$$\begin{pmatrix} \mu & h & h & h & h & h \\ 0 & \mu & h & h & h & h \\ 0 & 0 & h & h & h & h \\ 0 & 0 & h & h & h & h \\ 0 & 0 & 0 & h & h & h \\ 0 & 0 & 0 & 0 & h & h \end{pmatrix}.$$

When the converged Ritz values are thus deflated (or locked), one does not have to update the corresponding eigenvectors $u_1$ and $u_2$ in the Arnoldi decomposition. Once again, Lehoucq and Sorensen have proposed a complicated deflation algorithm.[1]

Most of the complications in the purging and deflating algorithms come from the need to preserve the structure of the Arnoldi decomposition (1.3)—in particular, the Hessenberg form of the Rayleigh quotient and the zero structure of the vector $\mathbf{e}_k$. The purpose of this paper is to show that if we relax the definition of an Arnoldi decomposition, we can solve the purging and deflating problems in a natural and efficient way. Since the method is centered about the Schur decomposition of the Rayleigh quotient we will call the method the Arnoldi–Schur method.

We will be concerned with the exact-shift version of the algorithm. In the next section we introduce generalized Arnoldi decompositions and, in particular, the Arnoldi–Schur decomposition. Section 3 we will treat the expansion step, which is essentially the same for implicitly restarted Arnoldi and Arnoldi–Schur. In Section 4 we will treat the contraction step and in Section 5 treat the numerical stability of the combined steps. In Section 6 show how to deflate the Arnoldi–Schur decomposition. In Section 7 we will compare the work done by iteratively restarted Arnoldi and Arnoldi–Schur. We end with some general comments. Throughout this paper $\|\cdot\|$ will denote the vector and matrix 2-norm ([11, Section 1.4.1]).

## 2. Generalized Arnoldi decompositions

The structure of an Arnoldi decomposition restricts the operations we can perform on its Rayleigh quotient. The following definition introduces a less constraining decomposition.

**Definition 2.1.** *A generalized Arnoldi decomposition of order $k$ is a relation of the form*

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^{\mathrm{H}}, \tag{2.1}$$

[1] Lehoucq [personal communication] has written code that is related to the deflation method proposed here.

*where $B_k$ is of order $k$ and $(U_k \;\; u_{k+1})$ is orthonormal. If $B_k$ is upper triangular we say the decomposition is an Arnoldi–Schur decomposition.*

In the generalized Arnoldi decomposition the Rayleigh quotient $B_k = U_k^{\mathrm{H}} A U_k$ is no longer required to be Hessenberg, and the vector $\beta_k \mathbf{e}_k^{\mathrm{T}}$ is replaced by a full vector $b_{k+1}^{\mathrm{H}}$. This generality allows us to operate freely on $B_k$. Specifically, let $Q$ be unitary. Then we say that the generalized Schur decompositions

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^{\mathrm{H}} \quad \text{and} \quad A(U_k Q) = (U_k Q)(Q^{\mathrm{H}} B_k Q) + u_{k+1}(b_{k+1}^{\mathrm{H}} Q)$$

are (unitarily) similar.[2] Since generalized Arnoldi decompositions are closed under similarity transformations, we can reduce the Rayleigh quotient to any desirable form by unitary similarities. In particular, any generalized Arnoldi decomposition can be reduced to an Arnoldi–Schur decomposition by computing a Schur form of its Rayleigh quotient.

Before proceeding, we must dispose of the possibility that generalized Arnoldi decompositions are Arnoldi in name only. The following theorem shows that any generalized Arnoldi decomposition can be associated with an Arnoldi decomposition and hence a Krylov sequence. For convenience we drop the subscripts in $k$.

**Theorem 2.2.** *Let*

$$AU = UB + ub^{\mathrm{T}} \tag{2.2}$$

*be a generalized Arnoldi decomposition of order $k$. Then (2.2) is similar to an Arnoldi decomposition. If the Hessenberg part of the Arnoldi decomposition is unreduced, the transformation is essentially unique.*

**Proof.** The proof of the theorem is based on a variant of a standard theorem on the partial uniqueness of the reduction to Hessenberg form (e.g., see [3, Theorem 7.4.2]). Specifically, there is a unitary matrix $Q$ whose last column is $b/\|b\|_2$ such that $H = Q^{\mathrm{H}} B Q$ is upper Hessenberg. If $H$ is unreduced the transformation is unique up to the scaling of the columns of $Q$ by factors of modulus one. Since $b^{\mathrm{H}} Q = \|b\|_2 \mathbf{e}_k^{\mathrm{T}}$, (2.2) is similar to the Arnoldi decomposition

$$A(UQ) = (UQ)H + \|b\| u \mathbf{e}_k^{\mathrm{T}}. \quad \blacksquare$$

It should be stressed that the theorem is constructive, in the sense that the similarity transformation can be effected in a stable and efficient manner by Householder transformations.[3]

---

[2] We restrict ourselves to unitary similarities because they preserve the orthonormality of $U$. However, it is possible to define nonorthogonal generalized Arnoldi decompositions and manipulate them with nonorthogonal similarities.

[3] In brief, the reduction is started by choosing a Householder transformation $Q_1$ such that $b^{\mathrm{H}} Q_1 = \|b\| \mathbf{e}_k^{\mathrm{T}}$. The matrix $Q_1^{\mathrm{H}} B Q_1$ is then reduced to Hessenberg form by using Householder transformations

### 3. Expansion

The expansion phase of the Arnoldi–Schur method consists of the expansion proper, which destroys the Schur form, and a final reduction to Schur form. We will write the initial decomposition as

$$AU_k = U_k S_k + u_{k+1} b_{k+1}^{\mathrm{H}},$$

where the letter $S$ (for Schur) stresses the triangularity of the Rayleigh quotient. It will be more convenient to work with the equivalent factored form

$$AU_k = U_{k+1} \hat{S}_k,$$

where

$$\hat{S}_k = \begin{pmatrix} S_k \\ b_{k+1}^{\mathrm{H}} \end{pmatrix}.$$

The expansion proceeds as in the usual Arnoldi algorithm: the vector $Au_{k+1}$ is orthogonalized against $U_{k-1}$ and normalized to give $u_{k+2}$, after which $S_{k+1}$ is formed from $S_k$. The following algorithm implements this sketch. We assume that $U_{k+1}$ and $\hat{S}_k$ are contained in arrays $U$ and $S$.

$$\begin{array}{ll} 1. & v = A{*}U[:, k{+}1] \\ 2. & w = U^{\mathrm{H}}{*}v \\ 3. & v = v - U{*}w \\ 4. & \nu = \|v\|_2 \\ 5. & U = (U \;\; v/\nu) \\ 6. & \hat{S} = \begin{pmatrix} \hat{S} & w \\ 0 & \nu \end{pmatrix} \end{array} \tag{3.1}$$

Note that in a working implementation we would have to reorthogonalize to insure that the vector $v$ is orthogonal to the column space of $U$ to working accuracy (see [11, Algorithm 4.1.13]).

After this process the array $\hat{S}$ has the form illustrated below for $k = 3$:

$$\begin{pmatrix} s & s & s & h \\ 0 & s & s & h \\ 0 & 0 & s & h \\ b & b & b & h \\ 0 & 0 & 0 & h \end{pmatrix}.$$

---

to introduce zeros rowwise from the bottom up. These similarity transformations do not disturb the zeros in the vector $\|b\|\mathbf{e}_k^{\mathrm{T}}$.

Here the $s$'s stand for the elements of the original $S_k$ and the $b$'s for the elements of $b_{k+1}$. The process may be repeated. After $m-k$ steps, the array $S$ has the form illustrated below for $k=3$ and $m=6$:

$$\begin{pmatrix} s & s & s & h & h & h \\ 0 & s & s & h & h & h \\ 0 & 0 & s & h & h & h \\ b & b & b & h & h & h \\ 0 & 0 & 0 & h & h & h \\ 0 & 0 & 0 & 0 & h & h \\ 0 & 0 & 0 & 0 & h & h \\ 0 & 0 & 0 & 0 & 0 & h \end{pmatrix}. \tag{3.2}$$

At this point the Rayleigh quotient, which resides in $S(1{:}m,1{:}m)$, is reduced to Schur form to give the Arnoldi–Schur decomposition

$$AU_m = U_m S_m + u_{m+1} b_{m+1}^{\mathrm{H}}. \tag{3.3}$$

This reduction to Schur form begins with a reduction of the Rayleigh quotient to Hessenberg form, and some minor savings can be obtained at this stage by taking advantage the structure illustrated in (3.2). Although (3.3) suggests that we are computing computing the entire decomposition, including $U_m$, in fact it will be more efficient to defer the computation of the vectors $u_j$ until later. We will return to this point in Section 7.

## 4. Contraction

We now turn to the problem of purging the unwanted Ritz values from the Arnoldi–Schur decomposition (3.3). The key is the observation that an Arnoldi–Schur decomposition can be truncated at any point. Specifically, if we partition an Arnoldi–Schur decomposition in the form

$$A(U_1\ U_2) = (U_1\ U_2)\begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} + u(b_1^{\mathrm{H}}\ b_2^{\mathrm{H}}), \tag{4.1}$$

then

$$AU_{11} = U_1 S_{11} + u b_1^{\mathrm{H}}$$

is also an Arnoldi–Schur decomposition. Thus the purging problem can be solved by moving the unwanted Ritz values into the southeast corner of the Rayleigh quotient and truncating the decomposition.

The process of using unitary similarities to move eigenvalues around in a Schur form has been well studied (see [2] for references and the current front-running algorithm, which has been implemented by the LAPACK routine `xTREXC`). Consequently, our deflation algorithm consists of little more than moving the unwanted Ritz values, which are visible on the diagonals of $S_m$, to the southeast corner of the Rayleigh quotient and truncating the decomposition.

The following theorem shows just what a combined expansion and contraction step produces.

**Theorem 4.1.** *Let $\mathbb{P}$ be an unreduced Arnoldi–Schur decomposition and let $\mathbb{P}'$ be the results of applying the the expansion and contraction steps to $\mathbb{P}$ with exact shifts $\mu_1, \ldots, \mu_{m-k}$ that are distinct from the other Ritz values of $\mathbb{P}'$. Let $\mathbb{Q}$ be the Arnoldi decomposition corresponding to $\mathbb{P}$ and let $\mathbb{Q}'$ be the result of applying Sorensen's expansion and contraction algorithm to $\mathbb{Q}$ . If $\mathbb{Q}'$ is unreduced, then $\mathbb{P}'$ is a Schur form of $\mathbb{Q}'$.*

**Proof.** The proof is in the style of Morgan [9]. By Theorem 2.2, $\mathbb{Q}$ is uniquely determined, and (in an obvious nomenclature) has same U-space and u-vector as $\mathbb{P}$. Consequently, the expansion phase yields decompositions with the same U-spaces and u-vectors. Because exact shifts are used, the contraction phase for $\mathbb{Q}$ eliminates the Schur vectors corresponding to the unwanted eigenvalues and does not change the u-vector [$h_{k+1,k}$ in (1.5) is zero]. By the distinctness property of the $\mu_k$ the space spanned by the discarded Schur vectors is uniquely determined, and hence so is the U-space. Similarly, contraction phase on $\mathbb{P}$ gives the same U-space and u-vector. Since, $\mathbb{Q}'$ is unreduced, it must be similar to $\mathbb{P}'$. ∎

The import of this theorem is that no matter hwo you perform the expansion and contraction, mathematically you end up with a decomposition that has been filtered through the polynomial $(t - \mu_1) \ldots (t - \mu_{m-k})$. However, the procedure based on the Arnoldi–Schur form is numerically more reliable than the one based on implicit restarting.

## 5. Numerical stability

We now briefly consider the numerical stability of the algorithm. From standard techniques of rounding error analysis it can be shown that as the Arnoldi–Schur algorithm proceeds the computed generalized Arnoldi decompositions satisfy

$$AU = UB + ub^{\mathrm{H}} + R \tag{5.1}$$

where $\|R\|/\|A\|$ is of order of the rounding unit and grows slowly. If $U$ is computed with reorthogonalization in the expansion phase, $U^{\mathrm{H}}U = I + F$, where $\|F\|$ is the order of the rounding unit and also grows slowly. The following theorem shows that we can throw the error $R$ back on the matrix $A$.

**Theorem 5.1.** *Let* (5.1) *be satisfied and assume that* $U$ *is of full rank. Let* $E = RU^{\dagger}$, *where* $U^{\dagger} = (U^{\mathrm{H}}U)^{-1}U^{\mathrm{H}}$ *is the pseudo-inverse of* $U$. *Then*

$$(A + E)U = UB + ub^{\mathrm{H}}, \tag{5.2}$$

*and*

$$\frac{\|R\|}{\|U\|} \leq \|E\| \leq \|R\|\|U^{\dagger}\|.$$

*The lower bound holds for any matrix* $E$ *satisfying* (5.2).

**Proof.** The equation (5.2) is established by direct verification. The upper bound follows from taking norms in the definition of $E$. On the other hand, if $E$ is any matrix satisfying (5.2), then $EU = R$, and $\|R\| \leq \|E\|\|U\|$, which establishes the lower bound. ■

Since $U$ is nearly orthonormal, $\|U\|$ and $\|U^{\dagger}\|$ are near one. Hence the theorem shows that the computed generalized Arnoldi decomposition is an exact decomposition of a matrix near $A$. In this sense the Arnoldi–Schur algorithm (as well as the iteratively restarted Arnoldi algorithm) is backward stable.

## 6. Deflation

The conventional way of determining whether the Ritz pair $(\mu, z)$ has converged is to look at the residual norm

$$\|r\| \equiv \|Az - \mu z\|. \tag{6.1}$$

The justification for this is the fact that if $\|z\| = 1$ then there is a matrix $E = -rz^{\mathrm{H}}$ with $\|E\| = \|r\|$ such that $(A + E)z = \mu z$—i.e., $(\mu, z)$ is an exact eigenpair of $A + E$. If $E$ is small enough compared to $A$ and the Ritz pair $(\mu, z)$ is well conditioned, then it is accurate.

Generalized Arnoldi decompositions share with their ordinary counterparts the fact that residual norms like (6.1) are easy to compute. For let $AU = UB + ub^{\mathrm{H}}$ be a generalized Arnoldi decomposition and let $Bw = \mu w$ so that $(\mu, Uw)$ is a Ritz pair. Then

$$r = AUw - \mu Uw = UBw - \mu Uw + ub^{\mathrm{H}}w = ub^{\mathrm{H}}w.$$

Hence

$$\|r\| = |b^{\mathrm{H}}w|,$$

so that when the quantity $b^{\mathrm{H}}w$ is small, we can declare that the Ritz value has converged.

If we are working with an Arnoldi–Schur decomposition, $AU = US + ub^{\mathrm{H}}$, we can deflate a converged value from the problem as follows. Let $Q$ be a unitary matrix that moves the Ritz value $\mu$ to the $(1,1)$-element of $\tilde{S} = Q^{\mathrm{H}}SQ$, and let

$$A\tilde{U} = \tilde{U}\tilde{S} + u\tilde{b}^{\mathrm{H}} = UQ\tilde{S} + ub^{\mathrm{H}}Q \tag{6.2}$$

be the transformed Arnoldi–Schur decomposition. Since the eigenvector corresponding to $\mu$ in $\tilde{S}$ is $\mathbf{e}_1$,

$$\|r\| = |b^{\mathrm{H}}w| = |\tilde{b}^{\mathrm{H}}\mathbf{e}_1| = |\tilde{b}_1|.$$

Thus the modulus of the first component of $\tilde{b}$ is $\|r\|$. If $\|r\|$ is small enough — say $\|r\| \le \epsilon\|A\|$, where $\epsilon$ is a prescribed tolerance — then we may set the first component of $\tilde{b}$ to zero. The Arnoldi–Schur decomposition then assumes the partitioned form

$$A(\tilde{u}_1\ \ \tilde{U}_2) = (\tilde{u}_1\ \ \tilde{U}_2)\begin{pmatrix}\mu & \tilde{s}_{12}^{\mathrm{T}} \\ 0 & \tilde{S}_{22} \\ 0 & \tilde{b}_2\end{pmatrix}.$$

Thus the Ritz value $\mu$ has been decoupled from the decomposition. This allows us to save operations in the contraction phase.

This deflation technique amounts to replacing $\tilde{b}$ by $\tilde{b} + e$. The effect is to add quantities of size $\|e\|$ to the residual $R$ in (5.1). Theorem 5.1 says that $e$ contributes a like error to the backward error in $A$. Thus, if our criterion for deflation is sufficiently stringent, the deflation process will not affect the backward error unduly.

The problem becomes more difficult when more than one eigenvalue is involved. Suppose that we have moved $\ell$ eigenvalues to the beginning of the Rayleigh quotient, and partition it in the form

$$S = \begin{pmatrix}S_{11} & S_{12} \\ 0 & S_{22}\end{pmatrix}.$$

The matrix of eigenvectors corresponding to the first $\ell$ eigenvalues has the form

$$\begin{pmatrix}X_{11} \\ 0\end{pmatrix},$$

where $X_{11}$ is an upper triangular matrix of order $\ell$. If we partition $b^{\mathrm{H}} = (b_1^{\mathrm{H}}\ b_2)$ conformally, then the residual norms of the first $\ell$ Ritz vectors are the absolute values of the components of $g_1^{\mathrm{H}} = b_1^{\mathrm{H}}X_{11}$. Thus the components of $b_1$ can be as large as $\|X_{11}^{-\mathrm{H}}\|_\infty\|g_1\|_\infty$. It follows that if the system of deflated eigenvectors is ill conditioned, a small residual does not guarantee a small value of $b_1$.

Fortunately, we can monitor the components of $b$ as we attempt to deflate Ritz pairs. Theorem 5.1 shows that we should not deflate a pair with a large $b$ component, no matter how small its residual, since such a deflation corresponds to a large perturbation in $A$. This does not mean that the offending Ritz pairs cannot be kept around or that they cannot be returned to the user. It only means that they cannot provide us the benefits of deflation.[4]

Although we have focused on the deflation of Ritz pairs, the process can be applied to any pair $(\mu, Uz)$ which has a small residual, in particular to the refined Ritz pairs of Jia [4] and the harmonic Ritz pairs of Morgan [8]. Specifically, let the pair $(\mu, Uw)$ $(\|w\| = 1)$ have the residual,

$$r = AUw - \mu Uw.$$

Since the residual is minimized when

$$\mu = (Uw)^{\mathrm{H}} A (Uw) = w^{\mathrm{H}} B w \tag{6.3}$$

(see [13, p. 172]), we will assume that $\mu$ satisfies (6.3). Let $Q$ be a unitary matrix such that $Q^{\mathrm{H}} w = \mathbf{e}_1$ and transform the decomposition as in (6.2) to get the decomposition

$$A\tilde{U} = \tilde{U}\tilde{B} + u\tilde{b}^{\mathrm{H}},$$

in which the first column of $\tilde{U}$ is $Uw$ and the $(1,1)$-element of $\tilde{B}$ is $\mu$. Partition the column in the form

$$A(\tilde{u}_1\ \tilde{U}_2) = (\tilde{u}_1\ \tilde{U}_2)\begin{pmatrix} \mu & \tilde{b}_{12}^{\mathrm{H}} \\ \tilde{b}_{21} & \tilde{B}_{22} \end{pmatrix} + u(\tilde{\beta}_{k+1,1}\ \tilde{b}_{k+1,2}^{\mathrm{H}}).$$

From the first column of this partition it follows that

$$r = U_2\tilde{b}_{21} + \tilde{\beta}_{k+1,1}\tilde{u}.$$

and hence

$$\left\| \begin{pmatrix} \tilde{b}_{21} \\ \tilde{\beta}_{k+1,1} \end{pmatrix} \right\| = \|r\|.$$

Hence if $r$ is sufficiently small the decomposition deflates at its first column.

The caveats about the deflation of more than one vector apply here. An minor inconvenience with the method is that the matrix $\tilde{B}_{22}$ is no longer triangular. However, if we use plane rotations to reduce the vector $w$ to $\mathbf{e}_1$ from the bottom up, $\tilde{B}_{22}$ will be Hessenberg.

[4] One might be tempted to mark the deflated Ritz pairs as "good" in order to contrast them with "bad" pairs that would not deflate. But that is to miss the point. Any pair with a small residual is a good pair. It is only *sets* of pairs that cannot accurately determine their eigenspaces that are bad.

**7. Assessment**

In comparing the Arnoldi–Schur algorithm with implicitly restarted Arnoldi, we must distinguish the sources of work in the algorithms. The first is the multiplication of a vector by $A$. Since $A$ will usually be sparse, the cost of this product is unpredictable in general, but it is reasonable to assume that it forms a significant part — perhaps the dominant part — of the computation.

The second source of work is the expansion of the Arnoldi decomposition from one of order $k$ to one of order $m$. It is easily seen from (3.1) the the work is $2n(m^2 - k^2)$ floating-point adds and multiplies, assuming reorthogonalization is performed. This count is the same for both algorithms.

In the contraction step, both algorithms must transform the Rayleigh quotient and accumulate the transformations in $U$. For efficiency, we do not accumulate the transformations in $U$ as they are generated but instead accumulate them in an $m{\times}m$ matrix $Q$ and then compute the new $U_k$ in the form

$$U_m{*}Q[:,1{:}k]. \tag{7.1}$$

If $n \gg m$, the last step will dominate the transformations applied to the Rayleigh quotient and their accumulation in $Q$.

For the Arnoldi–Schur method we must compute the Schur decomposition of the Rayleigh quotient and transform the triangular factor. This means that $Q$ will be full, and the final accumulation step (7.1) will require $nkm$ floating-point additions and multiplications.

For the implicitly restarted Arnoldi we must also compute the Schur decomposition of the Rayleigh quotient $H_m$. But it is only used to determine the shifts, which are applied directly to $H_m$. The structure of the transformations is such that $Q[:,1{:}k]$ is zero below its $m-k$ subdiagonal. This means that the operation count for (7.1) is $nmk - \frac{1}{2}k^2$ additions and multiplication.

To put things together, if $m = 2k$ and reorthogonalization is performed during the expansion, the Arnoldi–Schur algorithm has an operation count of $7nk^2$ whereas implicitly restarted Arnoldi has a operation count of $6\frac{1}{2}nk^2$. Thus implicitly restarted Arnoldi is marginally superior to Arnoldi Schur when it comes to accumulation of transformations. Against this must be set the fact that Arnoldi–Schur deflates in an inexpensive and natural manner and does not require a special routine for purging.

**8. Concluding remarks**

The Arnoldi–Schur method admits variations. An important one is based on the observation that we can truncate an Arnoldi–Schur decomposition at any point where the Rayleigh quotient is block triangular [see (4.1)]. This means that when $A$ is real we can work with real Schur forms of the Rayleigh quotient and avoid the necessity of complex arithmetic. The algorithm for exchanging eigenvalues mentioned above will also move the $2\times 2$ blocks of the real Schur form so that the contraction phase proceeds as usual. In deflation, the block in question is moved to the position just after the previously deflated eigenvalues and blocks, and two components of $b$ are tested. An unusual feature of complex eigenvectors is that they may fail to deflate, not because they are dependent on other deflated vectors, but because the real and imaginary parts of their eigenvectors are not sufficiently independent.

When $A$ is Hermitian, the Arnoldi–Schur method becomes a restarted Lanczos algorithm. The Rayleigh quotient is diagonal, so that reordering of the eigenvalues reduces to simple permutations. Moreover, because the eigenvectors of the Rayleigh quotient are orthogonal, a Ritz pair with a small residual norm $\epsilon$ will deflate with backward error of order $\epsilon$.

Since the Arnoldi–Schur method works explicitly with the eigenvalues of the Rayleigh quotient, it is an exact-shift method. Nonetheless, it stands ready to help the general shift method to deflate Ritz pairs and to get rid of unwanted pairs. One simply computes an Arnoldi–Schur form of the current decomposition and performs the procedures described above. Theorem 2.2 assures us that we can then return to a pure Arnoldi decomposition.

In fact Theorem 2.2 is really the heart of the matter. It allows us to operate freely on the Rayleigh quotient with the knowledge that we are always attached to a Krylov sequence. It is hoped that this freedom will find other applications.

## Acknowledgement

I would like to thank Rich Lehoucq and Dan Sorensen for their comments on preliminary versions for this paper.

## References

[1] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9:17–29, 1951.

[2] Z. Bai and J. W. Demmel. On swapping diagonal blocks in real Schur form. *Linear Algebra and Its Applications*, 186:73–95, 1993.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, second edition, 1989.

[4] Z. Jia. Refined iterative algorithm based on Arnoldi's process for large unsymmetric eigenproblems. *Linear Algebra and Its Applications*, 259:1–23, 1997.

[5] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45:255–282, 1950.

[6] R. Lehoucq and D. C. Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17:789–821, 1996.

[7] R. B. Lehoucq, D. C. Sorensen, and C Yang. *ARPACK USERS' GUIDE: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods.* SIAM, Philadelphia, 1998.

[8] R. B. Morgan. Computing interior eigenvalues of large matrices. *Linear Algebra and Its Applications*, 154:289–309, 1991.

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

[10] D. C. Sorensen. Implicit application of polynomial filters in a $k$-step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications*, 13:357–385, 1992.

[11] G. W. Stewart. *Matrix Algorithms I: Basic Decompositions.* SIAM, Philadelphia, 1998.

[12] D. S. Watkins. Forward stability and transmission of shifts in the QR algorithm. *SIAM Journal on Matrix Analysis and Applications*, 16:469–487, 1995.

[13] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.