

ON RESTARTING THE ARNOLDI METHOD FOR LARGE NONSYMMETRIC EIGENVALUE PROBLEMS

RONALD B. MORGAN

ABSTRACT. The Arnoldi method computes eigenvalues of large nonsymmetric matrices. Restarting is generally needed to reduce storage requirements and orthogonalization costs. However, restarting slows down the convergence and makes the choice of the new starting vector difficult if several eigenvalues are desired. We analyze several approaches to restarting and show why Sorensen's implicit QR approach is generally far superior to the others. Ritz vectors are combined in precisely the right way for an effective new starting vector. Also, a new method for restarting Arnoldi is presented. It is mathematically equivalent to the Sorensen approach but has additional uses.

1. INTRODUCTION

There are several methods for computing eigenvalues of a large nonsymmetric matrix A . The best known are the Arnoldi method [1,17,18,19,20,22,25], the nonsymmetric Lanczos algorithm [16,25], and subspace iteration [7,13,14]. These methods can be used with either the matrix A or with the shifted and inverted matrix $(A - \sigma I)^{-1}$ [5,15]. The second choice requires factorization of $A - \sigma I$, but it brings rapid convergence for the eigenvalues near σ . If factoring is not possible, an alternative approach is to use preconditioning methods for eigenvalue problems [3,9,10,12]. For this paper we will concentrate on just using the matrix A .

The Arnoldi and nonsymmetric Lanczos methods are both Krylov subspace methods. They generally give faster convergence than subspace iteration [13,18]. However, neither method is ideal. The nonsymmetric Lanczos algorithm involves a three-term recurrence but may encounter instabilities (see [16] for an improved version). The Arnoldi algorithm is stable but requires explicit orthogonalization against all previously computed basis vectors. So both the expense and the storage increase as the method proceeds. The Arnoldi method also requires calculation of the eigenvalues and eigenvectors of a Hessenberg matrix of order m at the cost of $O(m^3)$ operations, and this becomes prohibitive for large m . So restarting is generally necessary. However, restarting slows down the convergence. Here we analyze several approaches to restarting the Arnoldi algorithm and show why the Sorensen implicit QR approach is much better than the others. Then we give a new

Received by the editor July 3, 1991 and, in revised form, March 19, 1993, November 22, 1994, and February 13, 1995.

1991 *Mathematics Subject Classification.* Primary 65F15, 15A18.

Key words and phrases. Arnoldi, Krylov subspaces, eigenvalues, sparse matrices, nonsymmetric matrices.

This research was partially supported by the National Science Foundation under contract CCR-8910665.

©1996 American Mathematical Society

method that retains all of the desired approximate eigenvectors at the time of a restart and adds them to the subspace. This method is mathematically equivalent to the Sorensen method, but it has several additional uses, in particular when a priori eigenvector information is available. Section 2 discusses the problems with restarting Arnoldi. Section 3 looks at Sorensen's restarting. The new method is presented in §4, and §5 has some discussion on convergence rates.

2. RESTARTED ARNOLDI

Like many eigenvalue methods, the Arnoldi algorithm uses the Rayleigh-Ritz procedure [14, 19]. This procedure extracts approximate eigenvectors from a subspace of \mathbf{R}^n by reducing to a smaller eigenvalue problem.

THE RAYLEIGH-RITZ PROCEDURE

1. Let S be a j -dimensional subspace of \mathbf{R}^n .
2. Compute Q , an n by j orthonormal matrix whose columns span S .
3. Compute the j by j matrix $H = Q^T A Q$.
4. Find eigenvalues θ_i of H , and if desired, find eigenvectors g_i of unit length. The θ_i are approximate eigenvalues of A and are called Ritz values. The Ritz vectors, $y_i = Q g_i$, are approximate eigenvectors of A . The residual norm is $\|r_i\| = \|A y_i - \theta_i y_i\|$.

The particular subspace used by the Arnoldi method is a Krylov subspace $\text{Span}\{s, A s, A^2 s, \dots, A^{m-1} s\}$, where s is some starting vector [14, 19]. The reduced matrix H is upper-Hessenberg, and its elements are found during the orthogonalization phase.

THE RESTARTED ARNOLDI ALGORITHM

1. *Initialize:* Choose a starting vector s and let $q_1 = s/\|s\|$.
2. *Generation of Krylov subspace:* For $j = 1, 2, \dots, m$ do:
 - $h_{ij} = (q_i, A q_j), i = 1, 2, \dots, j,$
 - $\hat{q}_{j+1} = A q_j - \sum_{i=1}^j h_{ij} q_i,$
 - $h_{j+1,j} = \|\hat{q}_{j+1}\|,$ and
 - $q_{j+1} = \hat{q}_{j+1}/h_{j+1,j}.$
3. *Find approximate eigenvalues and eigenvectors:* For the m by m matrix H , compute the appropriate eigenvalues θ_i . If desired, compute Ritz vectors and residual norms, and check the residual norms for convergence.
4. *Restart:* Choose a new starting vector, possibly a Ritz vector or a combination of Ritz vectors. Normalize for q_1 , and go to 2 for the next run (we define a run to be each pass through the algorithm, in between restarts).

The Arnoldi recurrence formula [19, p. 173] is

$$(2.1) \quad A Q = Q H + h_{m+1,m} q_{m+1} e_m^*,$$

where e_m is the m th coordinate vector. A well-known formula for the residual norm

can be derived using (2.1):

$$\begin{aligned}
 r_i &\equiv Ay_i - \theta_i y_i \\
 &= AQg_i - \theta_i Qg_i \\
 &= QHg_i + h_{m+1,m}q_{m+1}e_m^*g_i - \theta_i Qg_i \\
 &= Q(Hg_i - \theta_i g_i) + h_{m+1,m}q_{m+1}e_m^*g_i \\
 &= h_{m+1,m}e_m^*g_i q_{m+1},
 \end{aligned}$$

since g_i is an eigenvector of H . Letting $\beta_{mi} = h_{m+1,m}e_m^*g_i$ gives

$$(2.2) \quad r_i = Ay_i - \theta_i y_i = \beta_{mi}q_{m+1},$$

and

$$(2.3) \quad \|r_i\| = |\beta_{mi}|.$$

The Arnoldi method is best at finding well-separated eigenvalues. Convergence properties for Krylov subspace methods can be analyzed using polynomials. Any vector from the Krylov subspace can be written as $y = p(A)s$, where p is a polynomial of degree $m - 1$ or less. Suppose λ_1 and z_1 are the desired eigenvalue and eigenvector. If A has a full set of eigenvectors z_1, z_2, \dots, z_n , then s can be expanded in terms of the eigenvectors:

$$s = \sum_{i=1}^n \alpha_i z_i.$$

Then

$$\begin{aligned}
 y &= \sum_{i=1}^n \alpha_i p(\lambda_i) z_i \\
 &= \alpha_1 p(\lambda_1) z_1 + \sum_{i=2}^n \alpha_i p(\lambda_i) z_i.
 \end{aligned}$$

For y to be an accurate approximation to z_1 , a polynomial is needed that is large at λ_1 and small at the other eigenvalues. This is difficult for a polynomial of limited degree, especially if there are eigenvalues close to λ_1 . See [14, 19, 21, 22] for more discussion.

Higher-degree polynomials give more accuracy. A problem with restarting is that it limits the size of the Krylov subspace and thus the degree of the polynomial. Another problem is that at each restart, a single vector must be chosen for the new starting vector. This choice is difficult if more than one eigenvalue is desired. And if a complex eigenvalue is needed, using a complex starting vector greatly increases costs.

Several ways of dealing with these problems are possible. For the rest of this section, we examine four restarting methods. The first approach is to use a weighted combination of Ritz vectors as the starting vector and to use only the real parts of complex Ritz vectors. Saad suggests a particular combination in [22]; the Ritz vectors are weighted by the corresponding residual norms. If the eigenvectors were known exactly and the starting vector was formed by combining the real parts, it

would only take a few steps of the Arnoldi method to compute all of the eigenvalues. However, in practice this method may not work well, especially for difficult problems [20]. Assume two eigenvalues, λ_1 and λ_2 , are desired. Let the corresponding eigenvectors be z_1 and z_2 , and let the approximate eigenvectors at the time of a restart in the Arnoldi method be y_1 and y_2 . Suppose the starting vector for the next run of Arnoldi is $s = y_1 + \alpha y_2$. After two steps of Arnoldi with starting vector $z_1 + \alpha z_2$, the subspace would contain z_1 . But after two steps with starting vector s , the error in the new approximation to z_1 can be much greater than the error in y_1 . We show this for the case of real Ritz values.

Theorem 1. *Assume θ_1 and θ_2 are real Ritz values with Ritz vectors y_1 and y_2 . Let $s = y_1 + \alpha y_2$ be the starting vector for the next run. Define $\delta \equiv \beta_{m1} + \alpha\beta_{m2}$ and let q_{m+1} be as defined in equation (2.1). Then every vector v in the subspace $\text{Span}\{s, As\}$ can be written in the form*

$$(2.4) \quad v = \sigma(y_1 + e),$$

with

$$(2.5) \quad \|e\| \geq \frac{|\alpha\delta|}{\sqrt{\delta^2 + \alpha^2(\theta_2 - \theta_1)^2}},$$

and with e in $\text{Span}\{y_2, q_{m+1}\}$ and σ a scalar.

Proof. We start by multiplying both sides of the equation $s = y_1 + \alpha y_2$ by A , then use the definition of the residual vector and equation (2.2):

$$(2.6) \quad \begin{aligned} As &= Ay_1 + \alpha Ay_2 \\ &= \theta_1 y_1 + r_1 + \alpha(\theta_2 y_2 + r_2) \\ &= \theta_1 y_1 + \beta_{m1} q_{m+1} + \alpha(\theta_2 y_2 + \beta_{m2} q_{m+1}) \\ &= \theta_1 y_1 + \alpha\theta_2 y_2 + (\beta_{m1} + \alpha\beta_{m2})q_{m+1}. \end{aligned}$$

Any combination of s and As is a multiple of

$$As - \gamma s = (\theta_1 - \gamma)y_1 + \alpha(\theta_2 - \gamma)y_2 + (\beta_{m1} + \alpha\beta_{m2})q_{m+1},$$

for some γ . So every vector v in $\text{Span}\{s, As\}$ can be expressed as follows, for some scalar σ :

$$\begin{aligned} v &= \frac{\sigma}{(\theta_1 - \gamma)} (As - \gamma s) \\ &= \sigma \left(y_1 + \alpha \frac{(\theta_2 - \gamma)}{(\theta_1 - \gamma)} y_2 + \frac{(\beta_{m1} + \alpha\beta_{m2})}{(\theta_1 - \gamma)} q_{m+1} \right) \\ &= \sigma \left(y_1 + \alpha \frac{(\theta_2 - \gamma)}{(\theta_1 - \gamma)} y_2 + \frac{\delta}{(\theta_1 - \gamma)} q_{m+1} \right). \end{aligned}$$

The vector v is in the form specified in equation (2.4). It would be desirable to have a value of γ that eliminates both the y_2 and q_{m+1} terms, but this is not generally possible. To eliminate y_2 , we need $\gamma = \theta_2$, but this causes the q_{m+1} term to be large if θ_1 is near θ_2 . Next, we find a γ that attempts to reduce both terms.

Since y_2 and q_{m+1} are orthogonal, we have

$$(2.7) \quad \|e\| = \sqrt{\alpha^2 \frac{(\theta_2 - \gamma)^2}{(\theta_1 - \gamma)^2} + \frac{\delta^2}{(\theta_1 - \gamma)^2}}.$$

The value of γ that minimizes this quantity can be found with calculus techniques. The result is

$$\gamma = \theta_2 + \frac{\delta^2}{\alpha^2(\theta_2 - \theta_1)}.$$

With this value of γ , there holds $(\theta_2 - \gamma)^2 = \frac{\delta^4}{\alpha^4(\theta_2 - \theta_1)^2}$ and $(\theta_1 - \gamma)^2 = \frac{(\alpha^2(\theta_2 - \theta_1)^2 + \delta^2)^2}{\alpha^4(\theta_2 - \theta_1)^2}$. After substitution into (2.7) and some algebra, the desired result (2.5) is obtained. \square

We now examine (2.5). If $\|e\|$ is significantly larger than the error already in y_1 , then accuracy is lost at the restart. First let $\alpha = 1$, and assume that the Ritz vectors are fairly accurate, so that δ is considerably smaller than $\theta_2 - \theta_1$. Then the lower bound on $\|e\|$ given by (2.5) is approximately $\frac{|\delta|}{\theta_2 - \theta_1}$. So the accuracy for the next run of Arnoldi depends largely on $\delta = \beta_{m1} + \beta_{m2}$ (the $\theta_2 - \theta_1$ term is partly a scaling factor to account for the fact that residual norms are used to bound an error norm). It is common for a Krylov subspace to give a better approximation to y_1 than to y_2 . Then β_{m2} is larger than β_{m1} . The accuracy for the new start is roughly the same as the accuracy of y_2 , instead of the desired accuracy of y_1 .

Next, note that the choice for α suggested in [22] is $\alpha = \frac{\|r_2\|}{\|r_1\|} = \left| \frac{\beta_{m2}}{\beta_{m1}} \right|$. With assumptions as before, the right-hand side of (2.5) is approximately $\frac{\beta_{m2}^2}{\beta_{m1}(\theta_2 - \theta_1)}$. This leads to even greater magnification of error.

If more than two Ritz vectors are combined, the accuracy can be set back to about the accuracy of the poorest approximation. Any setback can be disastrous, if there is not enough improvement during a run to make up for it.

Problems also occur with complex eigenvalues, even if only one is desired. While $\text{Re}(z_1)$ and $A(\text{Re}(z_1))$ can be combined to form z_1 , the vectors $\text{Re}(y_1)$ and $A(\text{Re}(y_1))$ generally cannot be combined to form either y_1 or any other approximation to z_1 of similar accuracy.

So it appears that the approach of using a combination of the real parts of Ritz vectors as the new starting vector can be ineffective. In [20], Saad suggests another approach. One Ritz vector is chosen as the new starting vector, and only one eigenvalue is computed at a time. After λ_1 has been found, it is deflated [14, p. 81] and the next eigenvalue is computed. One way of deflating is to replace A in the Arnoldi method with $A - \sigma z_1 z_1^*$, for some σ . The eigenvalues are left unchanged except for λ_1 , which shifts to $\lambda_1 - \sigma$ (see [20]). Thus the already computed eigenvalue can be shifted out of the way. But the disadvantage of this approach is that only one eigenvalue is computed at a time. Also there is still the problem of dealing with complex eigenvalues.

A third way of restarting is with a block method [6]. Several starting vectors are used, so it is possible to find several eigenvalues at the same time. With block Arnoldi, all of the approximate eigenvectors and even both their real and imaginary parts can be in the subspace. However, much lower-degree polynomials are generated. For example, suppose an m -dimensional subspace

is generated by the block method with blocks of size two. Then the subspace is $\text{Span}\{s_1, s_2, As_1, As_2, A^2s_1, A^2s_2, \dots, A^{\frac{m}{2}-1}s_1, A^{\frac{m}{2}-1}s_2\}$. So for the same size subspace, the powers of A reach only about half of what they are for the regular Arnoldi method. With lower-degree polynomials, there is much less accuracy.

The fourth approach to restarting is given by Sorensen in [24]. Like the method in [22] discussed earlier, he forms the new starting vector by combining Ritz vectors. His combination is different. An implicit QR approach is used; see [24] for the implementation details. This approach does not seem to be very promising, based on the results of Theorem 1. However, in the example that follows, it works quite well. This unexpected result is explained in the next section.

Example 1. For a test problem, let A be the tridiagonal matrix with $1, 2, \dots, 1000$ on the main diagonal, -0.1 in each superdiagonal position and 0.1 in each subdiagonal position. The three smallest eigenvalues are $1.0101, 1.9999, 3.0000$. The maximum size of subspace is 24. Four versions of Arnoldi are compared. The first uses a combination of three Ritz vectors weighted by residual norms [22]. The second restarts with just the one Ritz vector corresponding to the smallest Ritz value. The third approach is block Arnoldi with blocksize of three and again maximum size subspace of 24. The fourth method uses Sorensen's restarting [24] with a combination of three Ritz vectors. The initial starting vector is $(1, 1, 1, 0.1, 0.1, \dots, 0.1)^T$ and an additional two random vectors are used for block Arnoldi.

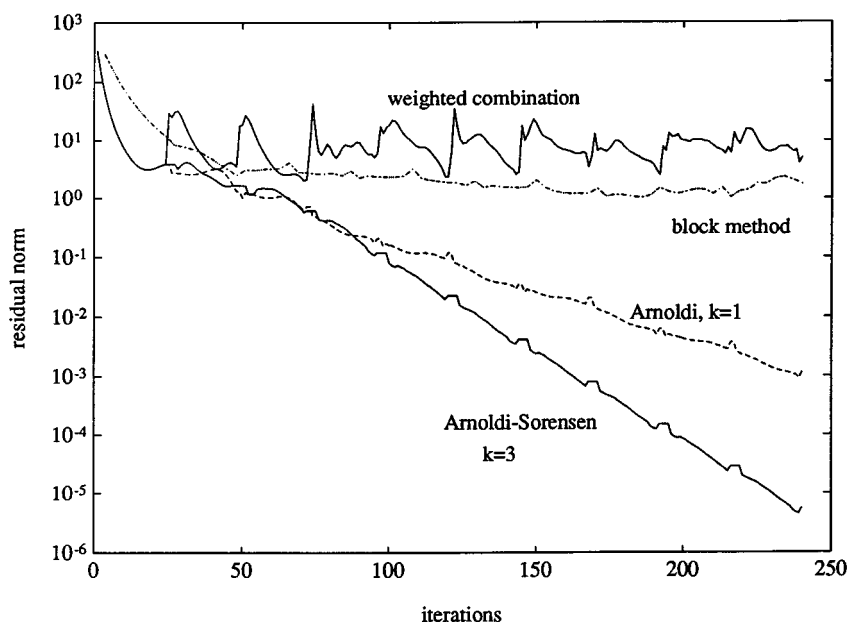


FIGURE 1. Comparison of restarts for Arnoldi

Figure 1 shows the residual norms for the smallest Ritz pair. The method using a combination of Ritz vectors weighted with residual norms is not very effective. We can see that the loss of accuracy around the time of each restart negates the improvement made during each run. Restarting with just one approximate eigenvector

is better. After 10 runs, the residual norm is 0.11e-2. But it will take quite a while to compute the first eigenvalue accurately, let alone find the next two eigenvalues. Block Arnoldi converges extremely slowly because with maximum size subspace of 24 and block size of 3, only degree-7 polynomials are generated in each run. This problem is difficult for low-degree polynomials.

Finally, the Sorensen approach to restarting is used with maximal dimension of subspace 24 and with a combination of three Ritz vectors at the restart (referring to [24], $p = 21$, $k = 3$ and “exact shifts” are used). This method converges much faster, especially over the last 6 runs. The residual norm for the smallest eigenvalue is 0.55e-5 after 10 runs. In addition, there are approximations to the next two eigenvalues with residual norms of 0.31e-3, and 0.12e-1. So not only does Sorensen’s method find the smallest eigenvalue faster, but it also gives approximations to other eigenvalues. In fact, in another test with $k = 6$ and subspaces of dimension 24, three eigenvalues are computed with residual norms of better than 10^{-6} after 15 runs. Restarting with one Ritz vector takes 19 runs to find just *one* eigenvalue with residual norm under 10^{-6} .

3. THE SORENSEN APPROACH

Now we reconsider Theorem 1, which seemed to indicate that restarting with a combination of Ritz vectors is ill-advised. Observe that there is a way of combining y_1 and y_2 that does not give loss of accuracy in Theorem 1. If we choose $\alpha = -\frac{\beta_{m1}}{\beta_{m2}}$, then $\delta = 0$. So the q_{m+1} term in equation (2.6) is eliminated, and As is just a combination of y_1 and y_2 . Thus $\text{Span}\{s, As\}$ is the same as $\text{Span}\{y_1, y_2\}$. So the two-dimensional subspace contains y_1 exactly. There is no accuracy lost at the restart. Only this combination of y_1 and y_2 , with $\alpha = -\frac{\beta_{m1}}{\beta_{m2}}$, eliminates the q_{m+1} term and gives no loss of accuracy. And this is exactly the choice that is made by the Sorensen method when exact shifts [24] are used.

Now consider combining more than two Ritz vectors, say y_1, y_2, \dots, y_k . There is a way to combine them as $s = \sum_{i=1}^k c_i y_i$ such that $\text{Span}\{s, As, \dots, A^{k-1}s\} = \text{Span}\{y_1, y_2, \dots, y_k\}$. As in deriving (2.6), the key is that $Ay_i - \theta_i y_i = \beta_{mi} q_{m+1}$. So for all i , $Ay_i - \theta_i y_i$ is a multiple of the same vector. For $k = 3$, the Ritz vectors can be combined with $c_1 = (\theta_3 - \theta_2)\beta_{m2}\beta_{m3}$, $c_2 = (\theta_1 - \theta_3)\beta_{m1}\beta_{m3}$, and $c_3 = (\theta_2 - \theta_1)\beta_{m1}\beta_{m2}$. The combination for a larger k can be found by solving a homogeneous underdetermined system of linear equations. The matrix is $k - 1$ by k with i th row $[\theta_1^{i-1}\beta_{m1}, \theta_2^{i-1}\beta_{m2}, \dots, \theta_k^{i-1}\beta_{mk}]$. The combination vector can be scaled so that it is a real vector. Sorensen’s implicit QR approach avoids explicitly computing the c_i ’s and is more efficient. The subspace of dimension k does not have to be generated with regular Arnoldi steps. It is formed from the previous subspace. This saves k matrix-vector products [24].

The next theorem says that the Sorensen procedure combines the Ritz vectors in the right way. It is proved in [24] that the new initial vector is a combination of Ritz vectors. It is stated that the particular linear combination of Ritz vectors is apparently different from others that had been used previously, but it is not stated that this combination gives different results.

Theorem 2. *Let the desired Ritz vectors be y_1, y_2, \dots, y_k . Suppose that Sorensen restarting is used with exact shifts. Then during the next Arnoldi run, the subspace of degree k is $\text{Span}\{y_1, y_2, \dots, y_k\}$.*

We omit the proof because it requires results from [24], and also it is mostly contained in the proof of Lemma 3.10 in [24]. (The only added difficulty is in showing that if $2 \leq i \leq k$, then q_i can be written as $\psi(H)w$, for some vector w . Then v_i^+ is, like v_1^+ , a combination of the desired Ritz vectors.) See [8, §4.4] for more on Lemma 3.10.

In the next theorem, p is the number of undesired Ritz values. During every run, Sorensen's method builds a Krylov subspace of dimension $m = p + k$. However, contained in that subspace are several interesting Krylov subspaces of smaller dimension.

Theorem 3. *The subspace generated during a run of the Sorensen method after restarting is*

$$(3.1) \quad \text{Span}\{y_1, y_2, \dots, y_k, Ay_i, A^2y_i, A^3y_i, \dots, A^py_i\},$$

for any i such that $1 \leq i \leq k$.

Proof. We give the proof of this theorem, although it also relies on a result from [24]. We first assert that the subspace generated during the new run in Sorensen's method is

$$(3.2) \quad \text{Span}\{y_1, y_2, \dots, y_k, q_{m+1}, Aq_{m+1}, \dots, A^{p-1}q_{m+1}\}.$$

This follows from Theorem 2 and from [24, p. 365, Eq. (7)]. Note that $\beta_k = 0$ in Eq. (7) of [24] when exact shifts are used, so the new q_{k+1} is the old q_{m+1} .

Next, we show that subspace (3.1) is contained in subspace (3.2). From equation (2.2),

$$Ay_1 = \beta_{m1}q_{m+1} + \theta_1y_1.$$

Then

$$\begin{aligned} A^2y_1 &= \beta_{m1}Aq_{m+1} + \theta_1Ay_1 \\ &= \beta_{m1}Aq_{m+1} + \theta_1\beta_{m1}q_{m+1} + \theta_1^2y_1. \end{aligned}$$

Since A^2y_1 is a combination of y_1 , q_{m+1} , and Aq_{m+1} , it is contained in subspace (3.2). Similarly for other j , A^jy_1 is in the subspace (3.2). For example, A^py_1 is a combination of y_1 , q_{m+1} , Aq_{m+1} , ..., $A^{p-1}q_{m+1}$. So the subspace in (3.1) is a subset of subspace (3.2). These subspaces have the same dimension, so they are the same. \square

According to Theorem 3, the subspace for Sorensen's method contains a Krylov subspace of dimension $p+1$ with each of the desired Ritz vectors as a starting vector. Thus the method works on approximations to all of the desired eigenpairs at the same time, without favoring one over the others. Nevertheless, owing to properties of Krylov subspaces, exterior and well-separated eigenvalues will converge more rapidly.

4. A NEW METHOD

In this section, we examine another approach to restarting. It is proposed that the new starting vector for Arnoldi be one approximate eigenvector, and that the other approximate eigenvectors be added to the subspace. If y_1 is chosen to be

the new starting vector and y_2, \dots, y_k are other approximate eigenvectors, then the vector basis $\{y_1, Ay_1, A^2y_1, \dots, A^p y_1, y_2, y_3, \dots, y_k\}$ is generated and the subspace spanning the vectors is used. So this is partly a Krylov subspace but with $k - 1$ extra basis vectors added. In the case of a complex approximate eigenvector, the real and imaginary parts are separated and both are added. The real part of y_1 can be used as the starting vector, while the imaginary part is added to the subspace.

Strickly speaking, this approach is no longer the Arnoldi algorithm, but it can be implemented as an addition at the end of the algorithm. We apply the Rayleigh-Ritz procedure to the subspace that has added approximate eigenvectors. This method was developed independently from Sorensen’s method and it is implemented quite differently, but it is mathematically equivalent.

In our implementation, the Krylov portion of the subspace is generated and orthogonalized first, then the old approximate eigenvectors are added to the subspace and orthogonalized. The reduced matrix H is not upper Hessenberg, but the next theorem shows that it still has a fairly nice form (see Figure 2).

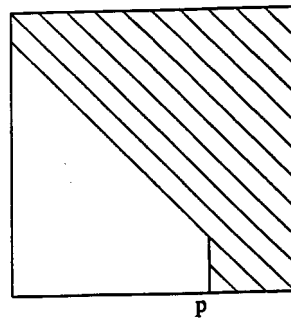


FIGURE 2. Form of H

Theorem 4. *Let $p + 1$ be the dimension of the Krylov portion of the subspace, and let the elements of H be h_{ij} . If $i > j + 1$ and $j \leq p$, then $h_{ij} = 0$.*

Proof. Let q_i be the i th column of Q . Note that $h_{ij} = q_i^T Aq_j$. Let the Krylov subspace of dimension j be $K^j = \text{Span}\{y_1, Ay_1, \dots, A^{j-1}y_1\}$. The first $p + 1$ columns of Q are from K^{p+1} . Therefore, since $j \leq p$, we have that $q_j \in K^j$ and

$$(4.1) \quad Aq_j \in K^{j+1}.$$

Since $i > j + 1$ and Q has orthonormal columns, q_i is orthogonal to q_1, q_2, \dots, q_{j+1} . Thus q_i is orthogonal to K^{j+1} . Using (4.1), we get $q_i^T Aq_j = 0$. \square

We now give the algorithm for this new method.

RESTARTED ARNOLDI WITH EIGENVECTOR APPROXIMATIONS

0. *First Run:* If initial approximate eigenvectors are known, proceed to 1. Otherwise, use the standard Arnoldi algorithm followed by steps 5 and 6 for the first run.
1. *Initial definitions and calculations:* The Krylov portion of the subspace has dimension $p + 1$, there are $k - 1$ extra vectors, and $m = p + k$. Let x_1, x_2, \dots, x_k

be the approximate eigenvectors or in the complex case, the real and imaginary parts of the approximate eigenvectors. Let $q_1 = x_1/||x_1||$.

2. *Generation of Krylov subspace:* For $j = 1, 2, \dots, p$ do:

$$h_{ij} = (q_i, Aq_j), i = 1, 2, \dots, j,$$

$$\hat{q}_{j+1} = Aq_j - \sum_{i=1}^j h_{ij}q_i,$$

$$h_{j+1,j} = ||\hat{q}_{j+1}||, \text{ and}$$

$$q_{j+1} = \hat{q}_{j+1}/h_{j+1,j}.$$
3. *Computations for the last vector in the Krylov subspace:*
compute and store Aq_{p+1} ,

$$h_{i,p+1} = (q_i, Aq_{p+1}), i = 1, 2, \dots, p + 1.$$
4. *Addition of approximate eigenvectors:* For $j = p + 2, \dots, m$ do:

$$c_{ij} = (q_i, x_{j-p}), i = 1, 2, \dots, j - 1,$$

$$\hat{q}_j = x_{j-p} - \sum_{i=1}^{j-1} c_{ij}q_i,$$

$$c_{jj} = ||\hat{q}_j||,$$

$$q_j = \hat{q}_j/c_{jj},$$

$$h_{ji} = (q_j, Aq_i), i = p + 1, \dots, j - 1,$$

$$h_{ij} = (q_i, Aq_j), i = 1, 2, \dots, j.$$
5. *Find approximate eigenvalues and eigenvectors:* Compute the appropriate eigenpairs, (θ_i, g_i) , of H . Compute Ritz vectors $y_i = Qg_i$. Norms of the residual vectors, $r_i = Ay_i - \theta_i y_i$, can be checked for convergence.
6. *Restart:* Let x_1 be the appropriate Ritz vector or in the complex case, the real part. Let $q_1 = x_1/||x_1||$ and let x_2 through x_k be the other Ritz vectors (including converged ones) or real or imaginary parts of Ritz vectors. Proceed to step 2.

Steps 2 and 4 should actually use modified Gram-Schmidt orthogonalization [6] and some reorthogonalization may be necessary [2]. Once an eigenvector has converged, some simplifications can be made in the implementation.

Computation of the residual norms is now considered. It is probably not necessary to know all of the residual norms at every restart. When needed, there are several ways to compute a residual norm. Another matrix-vector product can be used, or Ay_i can be formed from the vectors that have been saved (the additional cost is jn multiplications), but there is a cheaper formula that involves forming part of Ay_i and using the Arnoldi recurrence $AQ_p = Q_{p+1}H_{p+1,p}$ [19, 23]. We give the formula for the case of real eigenvalues: let θ and $y = Qg$ be the approximate eigenvalue and eigenvector of interest, where $||g|| = 1$. Divide g into two parts; let g_1 have the first p entries and g_2 have the last k . Let Q_2 have the last k columns of Q . Let G_1 be the leading $p + 1$ by p portion of $H - \theta I$. Let G_2 be the portion of $H - \theta I$ in the first $p + 1$ rows and the last k columns. Let H_3 be the intersection of the last k rows and the last k columns of H (not $H - \theta I$). Figure 3 shows the positions of these submatrices, though they are not all portions of the same matrix. Then

$$r^T r = g_1^T G_1^T G_1 g_1 + 2g_1^T G_1^T G_2 g_2 + ||AQ_2 g_2||^2 - 2\theta g_2^T H_3 g_2 + \theta^2 g_2^T g_2.$$

If it is desirable to avoid matrix-vector products, a different approach can be used in step 4. The entries of H can be computed without Aq_j . Form the Ay_i 's from vectors that have been saved and use the formula: for $j \geq p + 2$,

$$h_{ij} = (q_i^T Ay_{j-p} - \sum_{l=1}^{j-1} c_{lj} h_{il}) / c_{jj}.$$

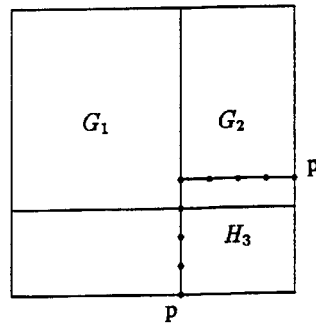


FIGURE 3. Positions of G_1 , G_2 , and H_3

There is another way to implement the method and avoid matrix-vector products. Orthogonalize Ay_i 's instead of y_i 's and use a generalized Rayleigh-Ritz procedure to develop a small generalized eigenvalue problem (see [11] for a version of this in conjunction with linear equations problems). These approaches appear effective in limited testing.

Next we show that this new method is mathematically equivalent to Sorensen's method. This assumes that no initial eigenvector approximations are used. Also, for the special case $\text{Re}(\beta_{m1}) = 0$, a change is needed in the algorithm. The result is a corollary of Theorem 3.

Theorem 5. *The Restarted Arnoldi with Eigenvector Approximations method generates the same subspace as the Arnoldi method with Sorensen's restarting.*

Proof. For the case where y_1 is a real Ritz vector, the subspace for the Restarted Arnoldi with Eigenvector Approximations method is constructed to be

$$(4.2) \quad \text{Span}\{y_1, Ay_1, A^2y_1, A^3y_1, \dots, A^py_1, y_2, \dots, y_k\}.$$

From Theorem 3, this is equivalent to the Sorensen subspace.

For the case of complex Ritz vectors, an inductive proof will be used. We assume that the previous subspace is the same as the subspace in Sorensen's method. Then the previous subspace is Krylov, so equation (2.2) can be used (the actual value of β_{m1} is not readily available because of the implementation). Let the complex Ritz pair be (θ_1, y_1) , with $\omega = \text{Re}(\theta_1)$, $\pi = \text{Im}(\theta_1)$, $x_1 = \text{Re}(y_1)$ and $x_2 = \text{Im}(y_1)$. The new subspace is constructed to be

$$(4.3) \quad \text{Span}\{x_1, Ax_1, A^2x_1, A^3x_1, \dots, A^px_1, x_2, \dots, x_k\}.$$

We need to show that this is the same as the subspace generated by the Sorensen method.

From (2.2), we have

$$Ay_1 = \theta_1 y_1 + \beta_{m1} q_{m+1}.$$

This becomes

$$(4.4) \quad Ax_1 + Ax_2i = \omega x_1 - \pi x_2 + (\pi x_1 + \omega x_2)i + \beta_{m1} q_{m+1}.$$

From (2.2) with $i = 2$, and using that y_2 is the complex conjugate of y_1 ,

$$(4.5) \quad Ax_1 - Ax_2i = \omega x_1 - \pi x_2 - (\pi x_1 + \omega x_2)i + \beta_{m2}q_{m+1}.$$

Combining (4.4) and (4.5) and using that β_{m1} and β_{m2} are conjugates gives

$$(4.6) \quad Ax_1 = \omega x_1 - \pi x_2 + \operatorname{Re}(\beta_{m1})q_{m+1}.$$

So Ax_1 is a combination of x_1 , x_2 and q_{m+1} . Similarly, $A^j x_1$ is a combination of x_1 , x_2 , q_{m+1} , Aq_{m+1} , \dots , $A^{j-1}q_{m+1}$, for any j . So subspace (4.3) is the same as

$$\operatorname{Span}\{x_1, q_{m+1}, Aq_{m+1}, \dots, A^{p-1}q_{m+1}, x_2, \dots, x_k\}.$$

Since the x_i 's are formed from the y_i 's, this subspace is

$$(4.7) \quad \operatorname{Span}\{y_1, y_2, \dots, y_k, q_{m+1}, Aq_{m+1}, \dots, A^{p-1}q_{m+1}\}.$$

For the case where $\operatorname{Re}(\beta_{m1}) = 0$, x_2 will already be in the subspace when the algorithm attempts to add it in step 4. In theory, this causes a breakdown. However, it can be shown for this case that $A^2 x_1$ is a combination of x_1 , x_2 and q_{m+1} . In order to generate subspace (4.7), the Krylov portion of the iteration can be allowed to go one more step. Then x_3, \dots, x_k , but not x_2 , can be added to the subspace. In practice, this case is probably not worth worrying about.

To conclude the proof, it was shown in the proof of Theorem 3 that subspace (4.7) is the same as the Sorensen subspace. So subspace (4.3) is also the same. \square

The consideration of the case where $\operatorname{Re}(\beta_{m1}) = 0$ in the proof of Theorem 5 leads to a suggestion for the method considered in §2 that restarts with just one Ritz vector [20]. If y_1 is complex, it can be rescaled so that $\operatorname{Re}(\beta_{m1}) = 0$. Then with $x_1 = \operatorname{Re}(y_1)$ as the new starting vector, the subspace of dimension two is $\operatorname{Span}\{x_1, Ax_1\} = \operatorname{Span}\{y_1, y_2\}$. This idea could also be used to combine several Ritz vectors as in the second paragraph of §3, but using only the real parts. So there are several ways of generating the Sorensen subspace. However, it seems that the implicit QR implementation is best. The rest of the section discusses the usefulness of the proposed new method.

The new method is not as efficient as the Sorensen approach. There is additional work for computing the y_i 's and Ay_i 's (they are not explicitly formed in the Sorensen method). Also, importantly, additional storage is needed for the Ay_i 's when matrix-vector products are avoided.

So it appears that for most eigenvalue problems, Sorensen's method should be used. The new method has simpler implementation, which might sometimes be of interest, but on the other hand, Sorensen has software for his method (the software is both remarkable and lengthy). However, there are a number of applications of the new method that make its study worthwhile. The method allows for adding any extra vectors to a subspace. They need not be Ritz vectors. Approximations to left eigenvectors [17, 25], singular vectors [6], or anything else can be included in the subspace. An important example is that in some applications, good approximations for eigenvectors are easily available. These approximations can be added to the subspace in the first run with the new method, but not with Sorensen's method. Another case where the new method is useful is with computing interior

eigenvalues using the modified Rayleigh-Ritz procedure in [9]. The vectors needed are not the standard Ritz vectors. The method can also be used to add approximate eigenvectors to the subspace for GMRES when solving linear equations [11]. The convergence of GMRES can be improved [11]. Next, for nonlinear eigenvalue problems the matrix changes. It is still easy to add approximate eigenvectors to a subspace for the new method, but it is not clear how a changing matrix could be handled by the Sorensen approach. Similarly, with a shift-and-invert Arnoldi method with variable shifts, the matrix would be changing.

5. ON CONVERGENCE OF SORENSEN'S METHOD

This section has some observations about convergence of the Sorensen method. The comments also apply to the new method introduced in §4. See [24] for a different discussion on convergence. Sorensen's method computes several eigenvalues at the same time. But Example 1 showed that computation of just one eigenvalue can be faster with the Sorensen method than with restarting with one Ritz vector. We look at how the convergence of an approximate eigenvector is improved by the presence of other approximate eigenvectors in the subspace.

Suppose for the moment that we are computing z_1 , but that we know the *exact* eigenvectors z_2, z_3, \dots, z_k . So assume that the subspace is

$$\text{Span}\{y_1, Ay_1, A^2y_1, \dots, A^p y_1, z_2, z_3, \dots, z_k\}.$$

Then the eigenvalues $\lambda_2, \dots, \lambda_k$ are effectively eliminated from the spectrum of A . The polynomial used to combine the vectors $y_1, Ay_1, A^2y_1, \dots, A^p y_1$ does not need to be small at these eigenvalues, because by properly combining z_2, \dots, z_k , the components in these directions can be eliminated. Now λ_1 is better separated from the remaining spectrum, and convergence is often much faster. It is perhaps surprising that adding *approximate* eigenvectors can be almost this beneficial, even before they are very accurate. This is a consequence of the following theorem, given for the case of one extra approximate eigenvector. Unlike in the rest of the paper, here we do not assume that y_1 and y_2 are normalized.

Theorem 6. *Suppose A has spectral decomposition $A \equiv Z\Lambda Z^{-1}$, with Λ diagonal. Assume the eigenvectors, the columns of Z , are normalized. Let $y_1 = z_1 + e_1$ and $y_2 = z_2 + e_2$, and further expand y_1 in terms of the eigenvectors as $y_1 = \sum_{i=1}^n \alpha_i z_i$, where $\alpha_1 = 1$. Then the subspace $\text{Span}\{y_1, Ay_1, \dots, A^{m-2}y_1, y_2\}$ contains the vector $z_1 + w$, where*

$$(5.1) \quad \|w\| \leq \|Z\| \|Z^{-1}\| \|e_1\| \left(\max_{i \neq 1, 2} |p(\lambda_i)| + \|e_2\| \right)$$

and where p is a polynomial of degree $m - 2$ or less such that $p(\lambda_1) = 1$ and $|p(\lambda_2)| \leq 1$.

Proof. Every vector v in the subspace can be written in the form

$$v = p(A)y_1 + \gamma y_2,$$

for p a polynomial of degree $m - 2$ or less and γ a scalar. A special choice of p and γ will be made to give a vector with the desired properties.

Using the expansions of y_1 and y_2 gives

$$\begin{aligned} v &= \sum_{i=1}^n \alpha_i p(\lambda_i) z_i + \gamma(z_2 + e_2) \\ &= p(\lambda_1) z_1 + (\alpha_2 p(\lambda_2) + \gamma) z_2 + \sum_{i=3}^n \alpha_i p(\lambda_i) z_i + \gamma e_2. \end{aligned}$$

Pick p to be 1 at λ_1 , less than 1 in magnitude at λ_2 and as small as possible at the other eigenvalues. Then pick $\gamma = -\alpha_2 p(\lambda_2)$. This removes the z_2 term. The resulting vector is

$$\begin{aligned} v &= z_1 + \sum_{i=3}^n \alpha_i p(\lambda_i) z_i - \alpha_2 p(\lambda_2) e_2 \\ &= z_1 + \hat{p}(A) e_1 - \alpha_2 p(\lambda_2) e_2 \\ &= z_1 + Z \hat{p}(\Lambda) Z^{-1} e_1 - \alpha_2 p(\lambda_2) e_2, \end{aligned}$$

where $\hat{p}(\lambda_i) = 0$ if $i = 1, 2$ and $\hat{p}(\lambda_i) = p(\lambda_i)$ for $i \neq 1, 2$. So we have the vector $v = z_1 + w$, where

$$(5.2) \quad \|w\| \leq \|Z\| \|Z^{-1}\| \max_{i \neq 1, 2} |p(\lambda_i)| \|e_1\| + |\alpha_2| \|e_2\|.$$

Note that w has a component in the direction of z_1 (from e_2), so v may be more accurate than (5.2) indicates.

With $e_1 = \sum_{i=2}^n \alpha_i z_i$, it can be shown that

$$|\alpha_2| \leq \|Z\| \|Z^{-1}\| \|e_1\|.$$

Along with (5.2), the desired result follows. \square

The $\|e_2\|$ term in equation (5.1) appears because y_2 is only an approximation. Fortunately, this term is not significant as long as $\|e_2\|$ is somewhat smaller than $\max_{i \neq 1, 2} |p(\lambda_i)|$. So if the accuracy of y_2 is greater than the improvement brought by p in one run of Arnoldi, then convergence is about the same as if the eigenvalue λ_2 had been removed from the spectrum.

We next consider some estimates for convergence. First assume we are computing λ_1 and do not have approximations to other eigenvectors. For a symmetric or a nearly symmetric matrix with real eigenvalues, an estimate of the convergence reduction factor is

$$(5.3) \quad \frac{1}{C_{m-1}(1+2\gamma)},$$

where m is the number of iterations in the Arnoldi method without restarting, $\gamma \equiv \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$ is the gap ratio, and C_{m-1} is a Chebyshev polynomial of degree $m-1$. (See [19, p. 200], and also [14], [21], and [22] for discussion of convergence for Krylov subspace methods.) So the error in the approximate eigenvector is reduced by a factor of approximately $\frac{1}{C_{m-1}(1+2\gamma)}$ during a run of Arnoldi with Krylov subspace of dimension m .

Now we also assume that the approximations to the eigenvectors z_2, z_3, \dots, z_k are fairly accurate, so that the convergence is as if the corresponding eigenvalues were deleted from the spectrum. So the effective spectrum for Sorensen's Arnoldi method is $\lambda_1, \lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_n$. Define the effective gap ratio, $\gamma_e \equiv \frac{\lambda_1 - \lambda_{k+1}}{\lambda_{k+1} - \lambda_n}$. An estimate for the convergence factor during a run of this method is

$$(5.4) \quad \frac{1}{C_p(1 + 2\gamma_e)}.$$

Consider as an example the case where A has equally spaced eigenvalues 1, 2, 3, ..., 1000. Then with $m = 24$, the convergence factor given by (5.3) is 0.44. Meanwhile, (5.4) gives 0.20 with $p = 21$ and $k = 3$. So convergence of the first eigenpair in the Sorensen method is expected to be about twice as fast as with restarting with just one Ritz vector. Looking back at Example 1, we see that the convergence is about twice as fast over the last 6 runs. However, it is possible for Sorensen's method to be slower at finding one eigenvalue. The next example is designed to demonstrate this.

Example 2. The matrix is the same as in Example 1, except that the diagonal elements are 1, 2, 2.05, 2.1, 3, 4, 5, ..., 998. The first five eigenvalues are 1.01, $2.05 + .13i$, $2.05 - .13i$, 2.05, and 3.00. Again we use $k = 3$. Unlike in Example 1, removing two eigenvalues close to λ_1 does not improve the effective gap ratio. Furthermore, it turns out that no eigenvalues are effectively removed. Good approximations do not develop for any of the eigenvalues near 2.05, because two Ritz vectors are used for a cluster of three eigenvalues. We give the results after 20 runs: the method restarting with one Ritz vector reaches residual norm 0.11e-6 while Sorensen's method reaches 0.32e-6. The Sorensen method is slower because it has a Krylov subspace of dimension 22 with starting vector y_1 as part of its space (see Theorem 3). The other Arnoldi method builds a Krylov subspace of dimension 24 with y_1 as starting vector. The higher dimension is an advantage. However, even in this difficult situation for Sorensen's method, it is not far behind.

The next example considers the question of how much the Arnoldi method is effected by the fact that it needs to be restarted. In this example, restarting does not slow the convergence very much. However, it is necessary to have the extra approximate eigenvectors.

Example 3. This example includes comparison with full, unrestarted Arnoldi. The matrix is the same as in Example 1. The Sorensen method has $k = 6$ and $p = 18$, so the subspaces are still of dimension 24. Figure 4 (next page) has plots of residual norm versus number of matrix-vector products for the smallest Ritz pair. With Sorensen restarting, matrix-vector products are not needed for the first six vectors in the subspace after a restart. In 10 runs with subspaces of size 24, the number of matrix-vector products needed is $24 + 9(18) = 186$. The large Krylov subspace used by full Arnoldi contains the Krylov subspaces used by the Sorensen method. For example, the full Arnoldi subspace of dimension 186 contains all of the Sorensen subspaces for 10 runs. Therefore, we expect full Arnoldi to be the better method. It is even reasonable to expect full Arnoldi to be much better, because this problem is fairly difficult (gap ratio of 0.001 for λ_1). Large subspaces are quite advantageous for difficult problems. However, we see from Figure 4, that convergence for the Sorensen method is very close to that of full Arnoldi. After 185

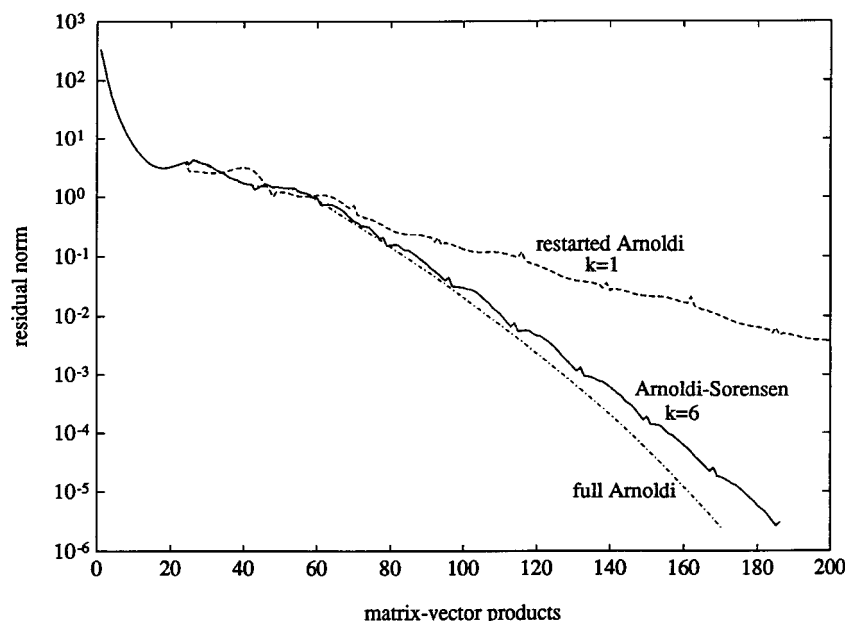


FIGURE 4. Comparison of restarted and full Arnoldi

matrix-vector products of the Sorensen method, the residual norm is the same as after 171 matrix-vector products of full Arnoldi. The results are surprising in view of the fact that full Arnoldi uses a polynomial of degree 170, instead of polynomials of degree 23. The next paragraph attempts to explain why the Sorensen method is so effective.

The Sorensen method develops approximations to the eigenvalues near 2, 3, 4, 5, and 6, in addition to the eigenvalue near 1. Once these approximations are accurate enough, the effective spectrum is 1, 7, 8, 9, ..., 999, 1000. This spectrum is not so difficult, because the eigenvalue near 1 is better separated. For easy problems, the degree of the polynomial is not as crucial. Having several low-degree polynomials is about as good as the corresponding high-degree polynomial. The approximate convergence factor for this effective spectrum (with $\gamma_e = .006$) and with a Krylov subspace of dimension 19 is $\frac{1}{C_{18}(1+2\gamma_e)} = 0.12$. So the error is expected to be reduced by almost an order of magnitude during each run (over each 18 matrix-vector products). From Figure 4, we see that the convergence is about this good over the last few runs.

If a restarted Arnoldi method competes well with full Arnoldi, then it will also contend well with the nonsymmetric Lanczos algorithm for the task of computing eigenvectors. Full Arnoldi uses the same subspace for finding eigenvectors as does the Lanczos algorithm. A study with further comparisons of full and restarted Arnoldi and also with comparisons to nonsymmetric Lanczos would be interesting.

6. CONCLUSION

Restarting the Arnoldi method with a combination of Ritz vectors is generally not effective. However, Sorensen's implicit QR method restarts with the right com-

mination. No accuracy is lost at the restart. Approximations to several eigenvalues can be found at the same time. In addition, convergence can often be improved even for finding one eigenvalue. Even though some information is discarded because of restarting, the most important information is retained. In an example, Sorensen's restarted Arnoldi method competes surprisingly closely with full, un-restarted Arnoldi.

A new method is introduced that is mathematically equivalent to the Sorensen method but has many possible additional uses. Of particular significance is that it can be used when initial approximate eigenvectors are available. It may also be of interest for nonlinear problems in both the symmetric and nonsymmetric cases.

ACKNOWLEDGEMENTS

The author would like to thank the referee for the many valuable suggestions. The paper has undergone great change because of these suggestions. The author also thanks Danny Sorensen for providing some information about performance of his method which helped in finding an implementation error.

REFERENCES

1. W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., **9** (1951), 17-29. MR **13**:163e
2. J. W. Daniel, W. B. Gragg, L. Kaufman and G. W. Stewart, *Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization*, Math. Comp. **30** (1976), 772-795. MR **55**:4638
3. E. R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices*, J. Comput. Phys. **17** (1975), 87-94. MR **52**:2168
4. I. S. Duff, R. G. Grimes, and J. G. Lewis, *Sparse matrix test problems*, ACM Trans. Math. Soft. **15** (1989), 1-14.
5. T. Ericsson and A. Ruhe, *The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems*, Math. Comp. **35** (1980), 1251-1268. MR **83b**:65034
6. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989. MR **90d**:65055
7. A. Jennings and W. J. Stewart, *A simultaneous iteration algorithm for real matrices*, ACM Trans. Math. Software **7** (1981), 184-198. MR **83d**:65118
8. R. B. Lehoucq, *Analysis and Implementation of an Implicitly Restarted Arnoldi Iteration*, Ph.D. thesis, Rice University, Houston, TX, 1995.
9. R. B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl. **154-156** (1991), 289-309. MR **92e**:65050
10. R. B. Morgan, *Generalizations of Davidson's method for computing eigenvalues of large non-symmetric matrices*, J. Comput. Phys. **101** (1992), 287-291. MR **93e**:65063
11. R. B. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl. **16** (1995), 1154-1171. CMP 96:01
12. R. B. Morgan and D. S. Scott, *Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices*, SIAM J. Sci. Statist. Comput. **7** (1986), 817-825. MR **87g**:65056
13. B. Nour-Omid, B. N. Parlett, R. L. Taylor, *Lanczos versus subspace iteration for solution of eigenvalue problems*, Inter. J. Num. Meth. Eng. **19** (1983), 859-871.
14. B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980. MR **81j**:65063
15. B. N. Parlett and Y. Saad, *Complex shift and invert strategies for real matrices*, Linear Algebra Appl. **88/89** (1987), 575-595. MR **88d**:65056

16. B. N. Parlett, D. R. Taylor and Z. A. Liu, *A look-ahead Lanczos algorithm for unsymmetric matrices*, Math. Comp., vol. 44, 1985, pp. 105-124. MR **86f**:65072
17. A. Ruhe, *Matrix pencils*, Lecture Notes in Mathematics, No. 973, Springer Verlag, New York, 1983, pp. 104-120. MR **84c**:65009
18. Y. Saad, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Math. Comp. **42** (1984), 567-588. MR **85i**:65049
19. Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Halsted Press, New York, NY, 1992. MR **93h**:65052
20. Y. Saad, *Numerical solution of large nonsymmetric eigenvalue problems*, Comput. Phys. Commun. **53** (1989), 71-90. MR **90f**:65064
21. Y. Saad, *On the rates of convergence of the Lanczos and the block Lanczos methods*, SIAM J. Numer. Anal. **17** (1980), 687-706. MR **82g**:65022
22. Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl. **34** (1980), 269-295. MR **81m**:65055
23. Y. Saad and M. H. Schultz, *GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput. **7** (1986), 856-869. MR **87g**:65064
24. D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl. **13** (1992), 357-385. MR **92i**:65076
25. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965. MR **32**:1894

DEPARTMENT OF MATHEMATICS, BAYLOR UNIVERSITY, WACO, TEXAS 76798-7328
E-mail address: morganr@baylor.edu