A harmonic restarted Arnoldi algorithm for calculating eigenvalues and determining multiplicity

Ronald B. Morgan [a],[*],[1], Min Zeng [b]

[a]*Department of Mathematics, Baylor University, Waco, TX 76798-7328, United States*
[b]*Consumnes River College, 8401 Center Parkway, Sacramento, CA 95823, United States*

Abstract

A restarted Arnoldi algorithm is given that computes eigenvalues and eigenvectors. It is related to implicitly restarted Arnoldi, but has a simpler restarting approach. Harmonic and regular Rayleigh–Ritz versions are possible.

For multiple eigenvalues, an approach is proposed that first computes eigenvalues with the new harmonic restarted Arnoldi algorithm, then uses random restarts to determine multiplicity. This avoids the need for a block method or for relying on roundoff error to produce the multiple copies.
© 2005 Elsevier Inc. All rights reserved.

---

* Corresponding author.
 E-mail address: Ronald_Morgan@baylor.edu (R.B. Morgan).

1. Introduction

We are interested in computing a few eigenvalues and eigenvectors of a large, non-symmetric matrix. We propose a new harmonic restarted Arnoldi method and also give a regular Rayleigh–Ritz version. These methods are closely related to other restarted eigenvalue methods such as implicit restarting and thick-restart Lanczos as will be discussed.

Then an approach is given that checks the multiplicity of the eigenvalues. The first phase computes some eigenvalues with the new harmonic restarted Arnoldi algorithm. Then a second phase looks for double eigenvalues. A random starting vector is used for a second harmonic restarted Arnoldi iteration. The eigenvectors from the first phase are combined with approximate eigenvectors from the second iteration to give the multiple eigenvalues and eigenvectors. Convergence of the second phase appears to act as if the previously computed single eigenvalues have been removed from the problem. More phases can be added in order to search for higher multiplicities.

Section 2 gives background information that will be needed. Section 3 presents the new restarted harmonic Arnoldi algorithm. Then multiple eigenvalues are dealt with in Section 4.

2. Background

We review ways of restarting the Arnoldi algorithm and then mention the harmonic approach to extracting approximate eigenpairs from a subspace.

2.1. Restarted Arnoldi methods

The Arnoldi method for eigenvalues [1,16,17] finds approximate eigenvalues using a Krylov subspace. The Arnoldi recurrence is

$$
\begin{aligned}
AV_m &= V_m H_m + h_{m+1,m} v_{m+1} e_m^{\mathrm{T}} \\
&= V_{m+1} \overline{H}_m,
\end{aligned}
\tag{2.1}
$$

where V_m is the orthonormal matrix whose columns span the dimension m Krylov subspace and H_m is an m by m upper-Hessenberg matrix. $\overline{H}_m$ is $m+1$ by m. The mth coordinate vector is e_m. The Rayleigh–Ritz procedure [15,17] is used to compute approximate eigenpairs from the Krylov subspace generated by the Arnoldi recurrence. If we let the eigenpairs of H_m be (θ_i, g_i), then the approximate eigenvalues, called Ritz values, are the θ_i's. The approximate eigenvectors are $y_i \equiv V g_i$ and are called Ritz vectors.

Storage and expense for Arnoldi grows as the size of the Krylov subspace increases. This makes restarting [17] necessary. However, restarting the Arnoldi recurrence is difficult. It is natural to start with only one vector, but several approximate

eigenvectors need to be retained. That way more than one eigenvalue can be computed at a time. Also, convergence can be better to a specified eigenvector if nearby approximate eigenvectors are in the subspace and cause the corresponding eigenvalues to deflate from the problem. Block methods [17] are one possible remedy, but they can have disadvantages in storage and expense [9].

The implicitly restarted Arnoldi method (IRA) given by Sorensen [19] solves this problem by generating a subspace that contains several Ritz vectors. Specifically, if "exact shifts" are chosen (the unwanted Ritz values are shifts), it can be shown [9,21] that the IRA subspace is

$$\text{Span}\{y_1, y_2, \ldots y_k, v_{m+1}, Av_{m+1}, A^2 v_{m+1}, A^3 v_{m+1}, \ldots, A^{m-k-1} v_{m+1}\},$$
(2.2)

where $y_1, \ldots, y_k$ are Ritz vectors from the previous Arnoldi cycle and v_{m+1} is the $m + 1$st Arnoldi vector, also from the previous cycle. Subspace (2.2) has a Ritz vector portion and a Krylov portion, however, surprisingly, the entire subspace is a Krylov subspace. See [11,22] for other proofs that (2.2) is a Krylov subspace, independent of implicit restarting. It is shown in [9] that subspace (2.2) is equivalent to

$$\text{Span}\{y_1, y_2, \ldots y_k, Ay_i, A^2 y_i, A^3 y_i, \ldots, A^{m-k} y_i\}$$
(2.3)

for $1 \leqslant i \leqslant k$. This helps explain the effectiveness of IRA, since for each Ritz vector, the IRA subspace contains a smaller Krylov subspace with the Ritz vector as starting vector. See also [19] and [6] for convergence of IRA.

Another approach to restarting with several vectors is given in [9]. The subspace

$$\text{Span}\{y_1, Ay_1, A^2 y_1, A^3 y_1, \ldots, A^{m-k} y_1, y_2, \ldots y_k\}$$

is generated in a method called Arnoldi with eigenvectors (Arnoldi-E). It is equivalent to IRA at the end of each cycle, but is not as efficient. However, the vectors $y_2, \ldots, y_k$ can be replaced with any vectors, so initial estimates for the eigenvectors can be used if they are available. The whole subspace is then no longer a Krylov subspace, and there can be consequences; see Section 4.3 in [11].

Wu and Simon [24] propose another restarted version of Arnoldi that is mathematically equivalent to IRA. Stewart [22] analyzes a related approach that is presented in a more general setting and thus can apply to various new methods. We discuss the Wu and Simon approach. They consider only the symmetric case and call their method thick-restart Lanczos. Instead of using implicit restarting, they orthonormalize the vectors $y_1, y_2, \ldots y_k, v_{m+1}, Av_{m+1}, A^2 v_{m+1}, A^3 v_{m+1}, \ldots, A^{m-k-1} v_{m+1}$ in order to generate an orthonormal matrix V_m with columns spanning subspace (2.2). This approach is like the Arnoldi-E method, except that the Ritz vectors are all put first. It ends up the same as IRA, except for the form of the first k basis vectors (and the QR approach is not needed to generate them). The good properties that were discussed for IRA still exist; namely the entire subspace is a Krylov subspace, and it includes smaller Krylov subspaces with Ritz vectors as starting vectors.

2.2. Harmonic Arnoldi

Harmonic Ritz approximations [8,4,14,18,13] are an alternative to the regular Rayleigh–Ritz procedure. Given a subspace S and an orthonormal matrix V whose columns span S, regular Rayleigh–Ritz projects over the subspace using operator A. The small eigenvalue problem $V^T A V g = \theta g$ is solved. Meanwhile, harmonic Rayleigh–Ritz projects over subspace $(A - \sigma I)S$ using the operator $(A - \sigma I)^{-1}$, where σ is a shift (possibly complex) in the region where eigenvalues are desired. The small eigenvalue problem becomes

$$V^T(A - \sigma I)^H(A - \sigma I)V\tilde{g} = (\tilde{\theta} - \sigma)V^T(A - \sigma I)^H V\tilde{g}. \tag{2.4}$$

Harmonic Ritz pairs are $(\tilde{\theta}_i, \tilde{y}_i)$, where $\tilde{y}_i = V\tilde{g}_i$.

Stewart [23] shows from (2.4) that if $\tilde{y}_i$ has unit norm, then

$$\|(A - \sigma I)\tilde{y}_i\| \leqslant |\tilde{\theta}_i - \sigma|. \tag{2.5}$$

So it is guaranteed that if $\tilde{\theta}_i$ is near σ, then the corresponding harmonic Ritz pair has a small residual. This means that harmonic Ritz pairs near σ are meaningful, even when σ is in the interior of the spectrum. This is not always the case for regular Ritz pairs.

The Rayleigh quotients of the harmonic Ritz vectors can be computed. We call these the harmonic Rayleigh quotients or ρ values. They often are more accurate than the harmonic Ritz values, particularly at early stages. However, the harmonic Rayleigh quotients do not have a property like the harmonic Ritz values do in (2.5), so they may not be as reliable. See [8,13] for more on these ρ values.

For harmonic Rayleigh–Ritz applied to the Arnoldi method, the small eigenvalue problem can be rewritten as

$$\left(H_m + h_{m+1,m}^2 f e_m^T\right)\tilde{g} = \tilde{\theta}\tilde{g}, \tag{2.6}$$

where $f = (H_m - \sigma I)^{-H}e_m$. So the difference computationally between regular and harmonic Arnoldi is that we replace finding eigenpairs of H_m with finding eigenpairs from (2.6).

3. New restarted Arnoldi algorithms

3.1. Harmonic restarted Arnoldi

We present a new restarted Arnoldi method that adapts Wu and Simon's approach to non-symmetric matrices and also uses harmonic Ritz approximations. We call it harmonic restarted Arnoldi (HRA). It is mathematically equivalent at the end of each cycle to the less efficient method called interior Arnoldi with eigenvectors in [13]. The subspace is

$$\text{Span}\{\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_k, r, Ar, A^2 r, \ldots, A^{m-k-1}r\}, \tag{3.1}$$

where $\tilde{y}_1, \ldots, \tilde{y}_k$ are harmonic Ritz vectors from the previous cycle and r is a multiple of the harmonic residual vectors. It is shown in [10] that the harmonic residuals are all multiples of each other, so we can say

$$A\tilde{y}_i - \tilde{\theta}_i \tilde{y}_i = \gamma_i r \tag{3.2}$$

for some scalars γ_i. At first glance, (3.1) has a Krylov portion and a separate augmenting part made up of approximate eigenvectors. However, it is proved in [10] that the entire subspace is a Krylov subspace. The proof uses implicit restarting. See [11] for a shorter proof in which the Krylov subspace is constructed and [3] for yet another approach.

Subspace (3.1) is equivalent to

$$\text{Span}\{\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_k, A\tilde{y}_i, A^2\tilde{y}_i, A^3\tilde{y}_i, \ldots, A^{m-k}\tilde{y}_i\} \tag{3.3}$$

for $1 \leqslant i \leqslant k$ [10], so it contains Krylov subspaces with each of the desired harmonic Ritz vectors as starting vectors.

We next give the algorithm. Note that because the first $k+1$ vectors of the new V_m are formed from the previous subspace, the orthonormalization can be done with short vectors of length m or $m+1$. However, it has been noticed that for numerical reasons, v_{k+1} needs to be reorthogonalized. As with other Arnoldi algorithms, it may be desirable to reorthogonalize all basis vectors. Other notes about the algorithm follow the listing.

Harmonic restarted Arnoldi (HRA)

1. *Start*: Choose m, the maximum size of the subspace, and k, the number of approximate eigenvectors that are retained from one cycle to the next. Also pick *numev*, the desired number of eigenpairs. Specify σ, the target around which eigenvalues are desired. Choose an initial vector v_1 of unit length.
2. *Arnoldi iteration*: Apply the Arnoldi iteration from the current point to form the rest of V_{m+1} and $\overline{H}_m$. The current point is either from v_1 if it is the first cycle or from v_{k+1} on the other cycles.
3. *Small eigenvalue problem*: Compute eigenpairs $(\tilde{\theta}_i, \tilde{g}_i)$, with $\tilde{g}_i$ normalized, of $(H_m + h_{m+1,m} f e_m^T)\tilde{g} = \tilde{\theta}\tilde{g}$, where $f = (H_m - \sigma I)^{-H} e_m$. Order the eigenpairs so that the first k are the desired ones. They normally would be the ones with $\tilde{\theta}_i$'s nearest σ. If desired, the harmonic Rayleigh quotients can be computed: $\rho_i = \tilde{g}_i^H H_m \tilde{g}_i$.
4. *Check convergence*: Residual norms can be computed (see (3.6) below) and convergence can be checked. If all desired eigenvalues have acceptable residual norm, then stop, first computing eigenvectors, if desired, as $\tilde{y}_i = V_m \tilde{g}_i$. Otherwise continue. The next step begins the restart.
5. *Orthonormalization of first k short vectors*: Orthonormalize $\tilde{g}_i$'s, for $1 \leqslant i \leqslant k$, first separating into real and imaginary parts if complex, in order to form a real m by k matrix P_k. Both parts of complex vectors need to be included, so temporarily reduce k by 1 if necessary (or k can be increased by 1).

6. *Orthonormalization of the $k + 1$ short vector*: Extend $p_1, \ldots p_k$ to length $m + 1$ by appending a zero to each, then orthonormalize $s = (-h_{m+1,m} f^{\mathrm{T}}, 1)^{\mathrm{T}}$ against $p_1, \ldots, p_k$ to form p_{k+1}. P_{k+1} is $m + 1$ by $k + 1$.
7. *Form portions of new H and V using the old H and V*: Let $\overline{H}_k^{\mathrm{new}} = P_{k+1}^{\mathrm{T}} \overline{H}_m P_k$ and $V_{k+1}^{\mathrm{new}} = V_{m+1} P_{k+1}$. Then let $\overline{H}_k = \overline{H}_k^{\mathrm{new}}$ and $V_{k+1} = V_{k+1}^{\mathrm{new}}$. Converged eigenvectors can be locked in by zeroing out part of $\overline{H}_k$; see the note below.
8. *Reorthogonalization of long $k + 1$ vector*: Orthogonalize v_{k+1} against the earlier columns of the new V_{k+1}. Go to step 2.

We give the main expense of one cycle this algorithm, considering only length n vector operations. The Arnoldi iteration requires about $m^2 - k^2$ vector operations for orthogonalization (we neglect smaller terms). There are also $m - k$ matrix–vector multiplications per cycle at a cost of $(m - k)matvec$, where *matvec* is the cost of a matrix–vector multiplication with A. Forming the basis for approximate eigenvectors in V_{k+1}^{new} takes about km vector operations. So the total is roughly $m^2 + km - k^2 + (m - k)matvec$.

For the rest of the section, we discuss some other aspects of the HRA algorithm. In step 1, *numev* can be the same as k, but it is generally better to have k larger so that more approximate eigenvectors are retained than the number actually being computed. This helps convergence, particularly of the last desired eigenvectors.

Step 3 of the algorithm uses (2.6), which appears to be dangerous when $(H - \sigma I)$ is nearly singular. While we have not observed any trouble from this, it should be noted that there are other possible formulas from (2.4) that could be used instead [13].

As noted in the algorithm, converged eigenvectors can be locked in. This reduces expense a little. It also prevents converged eigenvectors from losing accuracy as the iteration continues. This has been observed to happen sometimes after an eigenvector converges to very high accuracy. So eigenvectors can be locked once they have reached the prescribed accuracy, or they can be allowed to grow more accurate as long as they do not go to extremely high accuracy. More study of this is needed. The locking is accomplished by putting the converged eigenvectors first in the ordering and then zeroing out part of $\overline{H}_k$ (for the jth converged eigenvector, in step 7, zero out the $j + 1$ through $k + 1$ entries in the jth column of $\overline{H}_k$ if the eigenvector is real, and for the j and $j + 1$ converged eigenvectors in a complex pair, zero out the $j + 2$ through $k + 1$ entries in the jth and $j + 1$st columns). This causes these eigenvectors to be fixed.

At each cycle after the first, a recurrence somewhat similar to the Arnoldi recurrence (2.1) is generated by the HRA algorithm:

$$\begin{aligned} A V_m &= V_{m+1} \overline{H}_m, \\ &= V_m H_m + h_{m+1,m} v_{m+1} e_m^{\mathrm{T}}, \end{aligned} \qquad (3.4)$$

where H_m and $\overline{H}_m$ are upper-Hessenberg, *except* for a full leading $k + 1$ by $k + 1$ portion (as before H_m is square and $\overline{H}_m$ is $m + 1$ by m). If a Schur form is used for the eigenvectors in step 3, then the recurrence can have a different form:

$$AV_m = V_{m+1}\overline{H}_m,$$

where $\overline{H}_m$ is upper-Hessenberg, except for the $k + 1$ row. See Stewart [22] for more on using Schur forms.

We now explain the choice of the $k + 1$ vector s in step 6 of the algorithm. To do this, a formula is developed for the harmonic residual vectors. Using Eq. (3.4), then the fact that $(\tilde{\theta}_i, \tilde{g}_i)$ satisfy (2.6),

$$
\begin{aligned}
A\tilde{y}_i - \tilde{\theta}_i\tilde{y}_i &= AV_m\tilde{g}_i - \tilde{\theta}_i V_m\tilde{g}_i \\
&= V_m(H_m - \tilde{\theta}_i I)\tilde{g}_i + h_{m+1,m}v_{m+1}e_m^T\tilde{g}_i \\
&= -h_{m+1,m}^2 V_m f e_m^T\tilde{g}_i + h_{m+1,m}v_{m+1}e_m^T\tilde{g}_i \\
&= h_{m+1,m}e_m^T\tilde{g}_i(v_{m+1} - h_{m+1,m}V_m f) \\
&= h_{m+1,m}e_m^T\tilde{g}_i(V_{m+1}s), \quad (3.5)
\end{aligned}
$$

where s is the length $m + 1$ vector with $-h_{m+1,m}f$ in the first m positions and 1 in the last entry; recall in (2.6) that $f = (H_m - \sigma I)^{-H}e_m$. So we see that the vector r in (3.1) and (3.2) can be chosen to be $V_{m+1}s$. An important thing about this choice is that this vector is real as long as σ is real, even when some harmonic residual vectors are complex. Thus s is used in forming the $k + 1$ column of the new H. We also get from (3.5) an easy formula for the norm of the harmonic residual vector:

$$\|A\tilde{y}_i - \tilde{\theta}_i\tilde{y}_i\| = h_{m+1,m}|e_m^T\tilde{g}_i|\sqrt{h_{m+1,m}^2\|f\|^2 + 1}.$$

As mentioned in step 3, the harmonic Rayleigh quotients can be computed as $\rho_i = \tilde{g}_i^H H_m\tilde{g}_i$ ($\tilde{g}_i$ is unit length). A formula from [13] for the norm of the residual associated with the approximate eigenpair $(\rho_i, \tilde{y}_i)$ is

$$\|A\tilde{y}_i - \rho_i\tilde{y}_i\| = \sqrt{\|(H_m - \rho_i I)\tilde{g}_i\|^2 + h_{m+1,m}^2(e_m^T\tilde{g}_i)^2}. \quad (3.6)$$

This formula is derived with (3.4).

3.2. A restarted Arnoldi iteration with regular Rayleigh–Ritz

We now give the non-harmonic version of this algorithm. It is a generalization of Wu and Simon's thick-restart Lanczos [24]. It is also mathematically equivalent to IRA.

Restarted Arnoldi algorithm
   Same as HRA, except:

3. *Small eigenvalue problem*: Compute eigenpairs (θ_i, g_i) of H nearest σ.
6. *Orthonormalization of the $k + 1$ short vector*: Let $p_{k+1} = e_{k+1}$. This vector is already orthonormal.

Residual norms can be computed with the standard Arnoldi formula $\|A\tilde{y}_i - \rho_i\tilde{y}_i\| = h_{m+1,m}|e_m^T g_i|$.

The expense and storage for this restarted Arnoldi method is the same as given earlier for HRA. It is also essentially the same as for the IRA method, although this depends to some degree on the implementation of IRA. The new algorithms are simpler than IRA in two ways. First the restarting is less complicated, since it does not use the QR algorithm. Second there is no need for the purging routine that IRA uses to control some of the effects of roundoff error [7,20]. Locking converged eigenvectors is fairly simple in HRA.

3.3. Experiments

Example 1. The first test matrix is a 2500 by 2500 full matrix with random entries distributed normally with mean 0 and standard deviation 1, except the main diagonal has entries $1, 2, 3, \ldots, 2449, 2450, -21, -22, \ldots, -70$. The 11 eigenvalues nearest zero are $1.00 \pm .73i$, 2.85, 3.76, 5.96, $6.55 \pm 1.00i$, $8.09 \pm 3.87i$, and $10.1 \pm 2.38i$. Note these eigenvalues are in the interior of the spectrum but are near a gap. Without such a gap, computing interior eigenvalues is very difficult with a Krylov method. The harmonic and regular Rayleigh–Ritz restarted Arnoldi algorithms are compared for computing the three eigenvalues closest to zero. We use $m = 30$, $k = 10$, and 200 cycles. The residual norms for the five approximations nearest zero are plotted in Fig. 1. There are only four lines for each method, because complex pairs have the same residual norm. Note that in this case, the harmonic approach works better. However, in some examples with interior eigenvalues, harmonic is not significantly better. For exterior eigenvalues, it is not clear which approach will be best. The next example has exterior eigenvalues that are close together.

Example 2. We consider the matrix CK656 from the Test Matrix Collection for Non-Hermitian Eigenvalue Problems [2]. It is of size $n = 656$. The smallest eigenvalues are double eigenvalues at $9.0268e-5$, $9.303e-5$, $9.314e-5$, $9.597e-5$, $8.29e-4$, and $8.37e-4$, and the largest is a double eigenvalue at 1.6. Finding the small eigenvalues is a tough problem since they are not well separated. This time we have $m = 30$, $k = 10$, and use 25 cycles. Fig. 2 shows only the five smallest residual norms. The harmonic approach gives a little better results at most of the cycles. But the Ritz values at cycle 25 for regular Rayleigh–Ritz are more accurate than the harmonic Rayleigh quotients in the HRA method. For example the smallest Ritz value of $9.0263e-5$ is closer to the true value of $9.0268e-5$ than is the harmonic value of $9.0296e-5$. So it is not clear which approach is better. Neither method works extremely well for this problem. After the 25 cycles, only three of the smallest four eigenvalues have been computed and the fact that they are double eigenvalues has not been determined. The fourth small eigenvalue appears (accurate to two significant digits) at cycle 30 for the regular method and at cycle 35 for HRA.
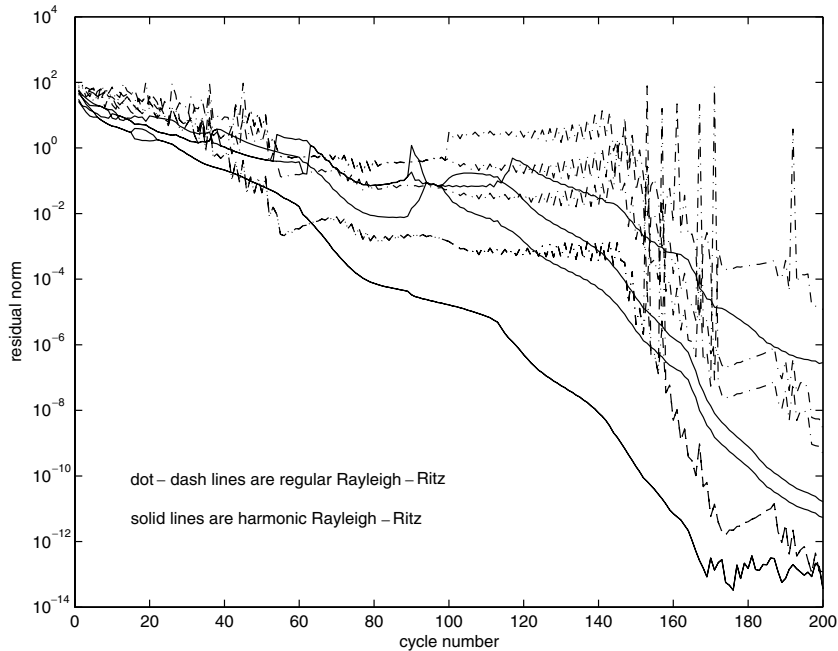
Fig. 1. Comparison of harmonic to regular Rayleigh–Ritz.

4. Finding multiple eigenvalues

A Krylov subspace will in theory produce only one eigenvector corresponding to a multiple eigenvalue. So determining multiplicity is difficult. One approach is to continue a restarted algorithm until roundoff error causes extra copies to appear for multiple eigenvalues. Using a block method [17] is also possible. We consider a third approach of restarting with a random vector. Note that others have also considered using random vectors. For example, Sorensen considers restarting with a random vector in an IRA algorithm [20]. Jia [5] has a random restart along with a singular value decomposition of a set of eigenvectors to determine multiplicity. We will instead use properties of harmonic Ritz values to monitor multiplicity.

We will mention here several possible approaches for using the random restart along with the HRA iteration. We will call it a "new phase" after each restart with a random vector. The eigenvalues that have already been computed in the previous phase of the algorithm must in some way be included in the new phase. One possibility is that these eigenvectors can be added to the new subspaces that are generated, using the Arnoldi-E method from [9]. The HRA recurrence cannot be used when vectors from outside the current Krylov subspaces are introduced. Some efficiency is lost when Arnoldi-E is used instead of HRA. We did not find this approach to be competitive.
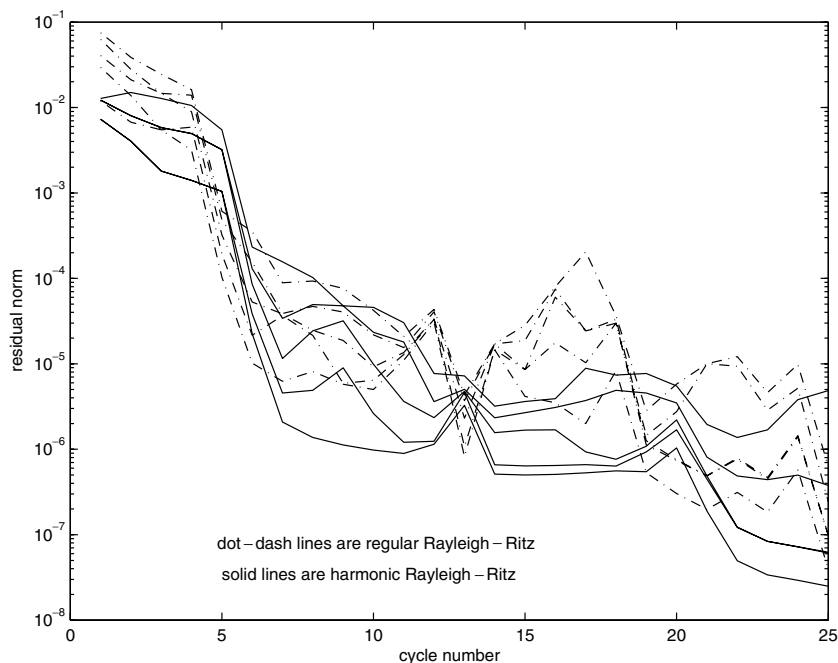
Fig. 2. Comparison for the matrix CK656.

A second possibility is to deflate in a fairly traditional way. The eigenvectors that have been computed in the previous phase can be used to modify the matrix and turn it into a deflated operator for HRA. For example, we will consider the operator $(I + \sigma V V^{\mathrm{T}})A$, where the columns of V are an orthonormal basis for the approximate eigenvectors, and σ is chosen to shift the computed eigenvalues out of the way. This approach will be tested at the end of Example 3.

Finally, we suggest another approach on which we will concentrate. We run a second HRA iteration with the new random starting vector. Separate from this, multiplicity is monitored by occasionally, at the end of cycles, combining the new HRA approximate eigenvectors with the eigenvectors from the earlier phase. A small harmonic Rayleigh–Ritz procedure is used for this [13], and it is fairly efficient thanks to having both the previously determined eigenvectors and the new approximate eigenvectors represented in the form of the compact Arnoldi recurrence (3.4). We do not modify the approximate eigenvectors for the next cycle of HRA, so the HRA recurrence remains in place. The HRA portion by itself may not appear to be generating accurate eigenvalues for a while. However, good approximations can appear after the combining with previous vectors. Convergence can be quick to multiple eigenvalues, because in the combination step, eigenvalues that were previously computed are

essentially deflated. And significantly, there is no requirement that the eigenvectors from the earlier HRA phase be very accurately determined.

In the sketch of the multiple eigenvalue algorithm that follows, it is assumed that *numev* < *k*. We take the new HRA far enough to help give multiple eigenvalues if they exist, then look at the results of the combination step to see if they actually do exist. We rely on the fact that there will be no spurious harmonic Ritz values near the target σ. So we can monitor whether new or multiple eigenvalues are being found by looking at the number of harmonic Ritz values that are near σ.

Approach for multiple eigenvalues

1. *Initial*: Begin from where the HRA algorithm stops. Save the *k* harmonic Ritz values closest to the target σ from the last cycle of HRA, and call them $\tilde{\theta}_i^1$, for $1 \leqslant i \leqslant k$. Also, save V_{k+1} and $\overline{H}_k$ and call them V_{k+1}^1 and $\overline{H}_k^1$. So we have the short recurrence $AV_k^1 = V_{k+1}^1 \overline{H}_k^1$, where V_k^1 has columns spanning a set of approximate eigenvectors. Note that $\overline{H}_k^1$ is a full matrix. Let *mev* be the maximum multiplicity that you want to look for. Set *mevlevel* = 2.

2. *HRA combined with a projection over previous eigenvectors*: Generate a new random starting vector and apply HRA. We will call this the second HRA iteration. At the end of HRA cycles (this can be done at the end of each cycle or done only occasionally), check for multiplicity as follows. Use V_{numev} developed by HRA. V_{numev} should have columns spanning the first *numev* approximate eigenvectors (the ones corresponding to the *numev* harmonic Ritz values nearest σ). Apply harmonic Rayleigh–Ritz [13] to the subspace spanned by the columns of V_k^1 and V_{numev}. This will be called the combination step (the implementation is discussed following this algorithm). The resulting approximate eigenvalues can be examined for multiplicity and to see if this step can be stopped (as discussed in step 3). If this step is not stopped, discard the results of the small harmonic Rayleigh–Ritz procedure and continue the second HRA iteration with no modifications.

3. *Stopping test for step 2*: There are various ways of stopping step 2. We give a fairly conservative way here that should generally find multiple eigenvalues if they exist. Note that in this step we will be considering two different groups of harmonic Ritz values that are generated in step 2, those from the second HRA iteration and then the ones from the combination step. We consider stopping the process in step 2 once the second HRA generates *numev* harmonic Ritz values that are almost as close to σ as $\tilde{\theta}_{numev}^1$. For example, we can wait until *numev* harmonic Ritz values are within $0.5|\tilde{\theta}_{numev}^1 - \sigma| + 0.5|\tilde{\theta}_{numev+1}^1 - \sigma|$ of σ. This is the midpoint of the distance from σ to $\tilde{\theta}_{numev}^1$ and the distance from σ to $\tilde{\theta}_{numev+1}^1$. At that point, examine the number of harmonic Ritz values from the combination step. Determine if more than *numev* of these harmonic Ritz values are clearly closer to σ than $\tilde{\theta}_{numev+1}^1$: for example, we could stop if there are only *numev* harmonic Ritz

values nearer to σ than $0.5|\tilde{\theta}^1_{numev} - \sigma| + 0.5|\tilde{\theta}^1_{numev+1} - \sigma|$. Otherwise there are some additional eigenvalues being found, so continue step 2 until the accuracy is improved as needed. Then go to the next step.

4. *Higher multiplicity*: Once steps 2 and 3 are concluded and additional eigenvalues are found (these will usually be additional copies of multiple eigenvalues, but may sometimes be eigenvalues that were missed originally), stop unless it is desired to determine a higher degree of multiplicity. So if *mevlevel* = *mev*, stop. Otherwise add 1 to *mevlevel*, and save the current $V_{numev+1}$ and $\overline{H}_{numev}$ and call them $V^2_{numev+1}$ and $\overline{H}^2_{numev}$. Now go to step 2, but for the combination step, project over the subspace spanned by the columns of V^1_k, V^2_{numev} and the current V_{numev} from the third HRA iteration. For determining higher multiplicity than three, continue this idea.

The regular Rayleigh–Ritz Arnoldi approach can be substituted for the HRA iterations in this algorithm. However the stopping test for the multiple eigenvalue phase may not be as reliable if spurious eigenvalues can occur.

We now look at the implementation of the combination step in step 2 of the algorithm. There are many ways to do this. One possibility is to form the matrix $[V^1_k, V_{numev}]$ and then use the Arnoldi-like recurrences (see (3.4)) to form $[(A - \sigma I)V^1_k, (A - \sigma I)V_{numev}]$. The matrix $[V^1_k, V_{numev}]$ can then be orthonormalized (only the last *numev* vectors need to be modified) to form a matrix we call V_{combo}. The matrix $(A - \sigma I)V_{combo}$ is then available by performing the same operations to $[(A - \sigma I)V^1_k, (A - \sigma I)V_{numev}]$ that were used to orthonormalize. Then the generalized eigenvalue problem for harmonic Rayleigh–Ritz is

$$((A - \sigma I)V_{combo})^{\mathrm{T}}(A - \sigma I)V_{combo}g = (\theta - \sigma)((A - \sigma I)V_{combo})^{\mathrm{T}}V_{combo}g.$$

The upper k by k portions of the matrices in this eigenvalue problem can be found using the Arnoldi-like recurrence for V^1_k. A possibly more efficient (but complicated) approach for the combination step would involve combining the Arnoldi-like recurrences for V^1_k and V_{numev} into one big Arnoldi-like recurrence.

The expense for a cycle of the multiple eigenvalue algorithm is about $m^2 + km - k^2 + (m - k)matvec$, as given earlier for HRA, plus the cost for the combination step. For second multiplicity, this can be done for a one-time expense of about mk to form $(A - \sigma I)V^1_k$, and then for each combination step, about $m(numev) + 6k(numev) + 3.5(numev)^2$ vector operations plus the cost of the generalized eigenvalue problem. So if a number of eigenvalues are being computed, then the combination step should not be performed more than necessary.

Example 3. For a test matrix we use the tridiagonal matrix of dimension 1000 with main diagonal entries $3, 3, 1, 2, 3, 4, \ldots, 998$. The superdiagonal elements are all 1's, and the subdiagonal has 1 for the first entry and then all zeros. This matrix has eigenvalues from 1 to 998 including double eigenvalues at 2 and 4. We first discuss

the new multiple eigenvalue approach just outlined. Then comparisons are given with other approaches: waiting for roundoff error, a block method and a deflated operator.

For the new multiple eigenvalue approach, the HRA iteration uses $m = 25$, $k = 8$, *numev* $= 5$, and the desired residual norm tolerance is 10^{-5}. It takes 15 cycles to compute the five desired eigenvalues, then only nine cycles to get both second copies of the multiple eigenvalues. The combination step is needed only for the last three of the nine cycles (using the test suggested in step 2 of the algorithm).

We now observe that the multiple eigenvalues appear long before the stopping test suggested in the algorithm is activated. There are six values from the combination step closer to $\sigma = 0$ than $0.5|\tilde{\theta}^1_{numev} - \sigma| + 0.5|\tilde{\theta}^1_{numev+1} - \sigma| = 5.5$ after two cycles. It takes seven cycles for five harmonic Ritz values from the second HRA iteration to be less than 5.5. So the fact there are multiple eigenvalues is determined long before the stopping test would be triggered. This suggests a better, less conservative stopping test may be possible. We next look at a graph that shows further that the combination step produces much better approximations to the second copies of the double eigenvalues than does the second HRA iteration. In Fig. 3, the second and fourth smallest harmonic Ritz values are given for the second HRA iteration. They eventually converge to give second copies of the double eigenvalues. Also plotted are the third and sixth harmonic Ritz values for the combination step. They also converge to the second copies of the double eigenvalues. We see that the approximations are much better in the combination step. This will be discussed further after Example 4.

It can be difficult to compute multiple eigenvalues to the same accuracy that the first eigenvalue has been determined. For this example that we are considering, the first eigenvalue at 4 has residual norm of 5.5e−7 when the first HRA iteration is stopped. The second eigenvalue at 4 reaches the desired accuracy in nine cycles of the second HRA and has residual norm of 7.7e−6. But then convergence slows, reaching 7.5e−7 at cycle 13. However, if the first eigenvalue was determined to greater accuracy of 9.4e−9 (in 18 cycles), then the second one converges faster toward the end, reaching residual norm of 5.6e−7 in 11 cycles. So first eigenvalues should be determined to greater accuracy that *rtol* (if all eigenvectors corresponding to a multiple eigenvalue are needed to full accuracy). Eigenvectors for the first eigenvalues should not be locked in when they have just reached *rtol*. Locking in eigenvectors when the residual norm has reached $0.1 * rtol$ has worked in a number of examples including a case with triple eigenvalues, but this needs to be studied more.

It was mentioned earlier that the eigenvectors from the first HRA phase do not need to be accurately determined. To show this, we use a tolerance of residual norm less than 10^{-2} for this first phase. The five eigenvalues are computed to this accuracy in 12 cycles. Then the multiple eigenvalue phase determines the multiple eigenvalues accurate to residual norm 10^{-2} in only five cycles.

We now continue this example with comparisons to other multiple eigenvalues methods. With the approach of waiting for roundoff error, the five eigenvalues again converge after 15 cycles with $rtol = 10^{-5}$. Then the first HRA iteration is simply
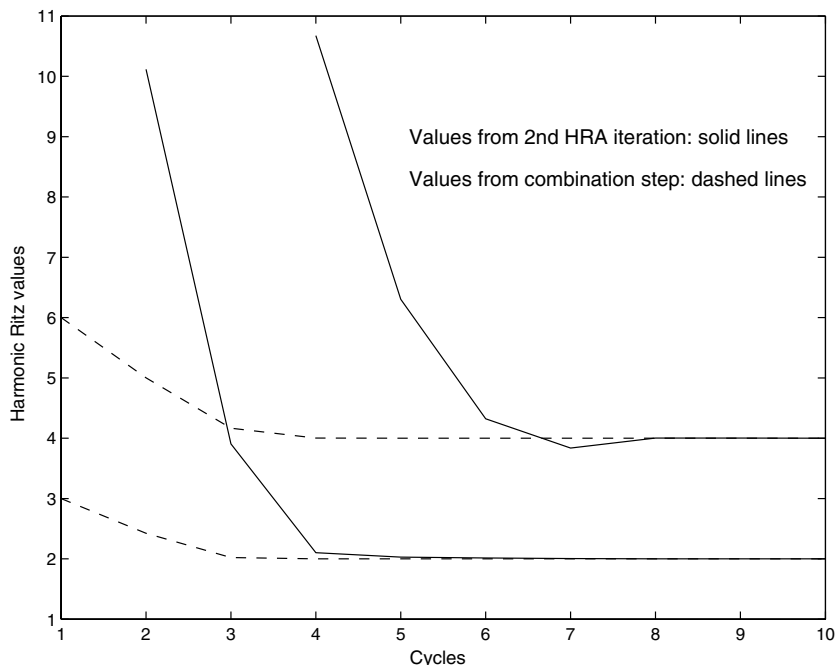
Fig. 3. Comparison of harmonic Ritz values from 2nd HRA and combination step.

continued until double eigenvalues appear. This takes a while, but both approxima-
tions are accurate with residual norms under 10^{-5} after 23 additional cycles. This
compares to seven additional cycles of the new random restart method.

To compare with a block approach, we use a block version of HRA that builds
a subspace that includes approximate eigenvectors along with the multiple Krylov
subspaces (see [12] for a linear equations version). The block size is 2, so there is an
assumption that we are searching only for double eigenvalues. First we use subspaces
of dimension $m = 26$ including $k = 8$ approximate eigenvectors. It takes 39 cycles
to compute both double eigenvalues to the desired residual tolerance. This compares
to a total of 24 cycles for the new HRA multiple eigenvalue approach. The number
of matrix–vector products is 710 for the block method versus 424. The block method
has slower convergence, because it uses smaller Krylov subspaces as part of the
subspaces it builds: two of dimension 9 in each cycle instead of one of dimension
17 for HRA(25,8). We also try block Arnoldi with $m = 42$ and $k = 8$, which gives
Krylov subspaces of dimension 17. The block method then converges in 15 cycles.
Interestingly, this is the same number of cycles as HRA takes to compute only first
copies of eigenvalues. The costs of the block method are greater. It requires 518
matrix–vector products versus 424 for the new approach. Also, the orthogonalization
expense due to the larger subspaces of dimension 42 may be significant.

Next we compare against using the deflated operator $(I + \sigma V V^{\mathrm{T}})A$ for the second HRA iteration. Here we let $\sigma = 100$ and $V = V_k^1$. The convergence toward the second copy of the eigenvalue at 4 is slightly better for this approach than for the combination step approach given in the algorithm above. It reaches residual norm of 1.1e−5 after eight HRA cycles versus 2.5e−5 with the combination step approach. However, it does not improve any from there, while the combination step approach does improve further (as discussed above). This deflated operator approach requires an extra $2mk$ length n vector operations per cycle to implement the deflated operator and some extra expense to compute residual norms (since a modified operator is used). On the other hand, it does not require the combination step.

Example 4. Next we test the matrix CK656 from Example 2. HRA uses $m = 30$, $k = 10$, *numev* $= 5$ and the desired residual norm tolerance is 10^{-8}. The first copies of the smallest five double eigenvalues converge in 46 cycles. To compute the second copies to eight decimal places (four significant digits) using the roundoff error approach requires an additional 140 cycles. Using the restart with a random vector approach is much better, as it requires only an additional 48 cycles.

The rest of the section discusses convergence of the multiple eigenvalue phase. The combination step combines fairly accurate eigenvectors from the first HRA phase with not as accurate approximations from the second HRA iteration. However, these second approximations do not need to be as accurate, because they are combined with the good vectors. What is needed is that these approximations be rich in components in the directions of eigenvectors corresponding to the already computed eigenvalues.

In the theorem that follows, we show that if an approximation from the second HRA iteration has components only in those directions, then the second eigenvector for a double eigenvalue is computed exactly in the combination step.

Theorem 4.1. *Assume A has a full set of eigenvectors. Let $\lambda_1, \lambda_1, \lambda_3, \lambda_4, \ldots, \lambda_n$ be the eigenvalues, ordered in distance from the target σ. Let $z_1, z_2, \ldots, z_n$ be normalized eigenvectors, with $\mathrm{Span}\{z_1, z_2\}$ the eigenspace corresponding to the double eigenvalue at λ_1. If harmonic Rayleigh–Ritz is applied to $\mathrm{Span} \left\{ z_1, z_3, z_4, \ldots, z_{k+1}, \sum_{i=1}^{k+1} \alpha_i z_i \right\}$, where the α_i's are any scalars, then the harmonic Ritz values are $\lambda_1, \lambda_1, \lambda_3, \lambda_4, \ldots, \lambda_{k+1}$ with associated harmonic Ritz vectors $\mathrm{Span}\{z_1, z_2\}, z_3, z_4, \ldots, z_{k+1}$.*

Proof. The harmonic reduced problem is

$$V^{\mathrm{T}}(A - \sigma I)^H (A - \sigma I) V \tilde{g} = (\tilde{\theta} - \sigma) V^{\mathrm{T}} (A - \sigma I)^H V \tilde{g}, \qquad (4.1)$$

where $\tilde{\theta}$ is a harmonic Ritz value and $V \tilde{g}$ is a harmonic Ritz vector. Eq. (4.1) becomes

$$V^{\mathrm{T}}(A - \sigma I)^H (A - \tilde{\theta}) V \tilde{g} = 0.$$

This is equivalent to

$$V^T(A - \sigma I)^H (A - \tilde{\theta}) \tilde{y} = 0 \tag{4.2}$$

for $\tilde{y}$ in the span of the columns of V. Every pair (λ_1, z) for $z \in \text{Span}\{z_1, z_2\}$ is a solution of Equation (4.2). Also each pair (λ_i, z_i) for $i = 3, 4, \ldots, k + 1$ is a solution. $\square$

Results can also be given for more realistic situations. For example, standard theorems on polynomials from the Arnoldi iteration can bound the size of the polynomial at the different eigenvalues of A (see, for example, Theorem 6.5 in [17]). We can put the corresponding vector into the combination step to remove components corresponding to already computed eigenvectors. So what matters is the size of the polynomial on the deflated spectrum. What cannot be guaranteed is that the vector actually produced by the second HRA iteration is as good as what the subspace contains.

In the theorem, we assumed that the second HRA iteration has produced an approximation $\left(\text{the } \sum_{i=1}^{k+1} \alpha_i z_i \text{ vector}\right)$ that is a combination of a few eigenvectors. Even before there is an approximation that accurate, the convergence of the combined step tends to be as if the previously computed single eigenvalues have been removed from the spectrum. The next example shows this type of convergence.

Example 5. We want to look at how similar the convergence toward the second copies of multiple eigenvalues is to that of a matrix with all the computed eigenvalues deflated out. To test this, we compare convergence toward the multiple eigenvalues of the matrix from Example 3 with convergence for a matrix that has eigenvalues at 2, 4, 9, 9.1, 9.2, …, 9.9, 10, 11, …, 996, 997. This matrix has the single eigenvalues computed during the first phase removed (one each of double eigenvalues remains). The cluster of eigenvalues from 9 to 10 is put there so that the eigenvalues from 9 and up will not naturally deflate from the problem as the iteration proceeds (the eigenvalues 9, 10 and up do not quickly deflate while determining multiple eigenvalues in Example 3 with the new algorithm). After 10 cycles of HRA applied to this matrix (with $m = 25$, $k = 8$), the residual norms for the eigenvalues 2 and 4 are 2.5e−8 and 6.5e−7. After 10 cycles of the second HRA phase for the matrix from Example 3, the residual norms of the second approximate eigenvectors near 2 and 4 are similar at 1.0e−8 and 1.4e−6. So the multiple eigenvalue algorithm converges as if eigenvalues were simply removed from the matrix.

5. Conclusion

We have proposed a new algorithm that computes eigenvalues and can determine multiplicity. It uses a harmonic restarted Arnoldi approach that is simpler than implicitly restarted Arnoldi. Also, the harmonic approach can be better for interior

eigenvalues. We also mentioned a restarted Arnoldi method with regular Rayleigh–Ritz instead of harmonic. It is equivalent to IRA, but does not use implicit restarting.

The multiple eigenvalue portion of the new method uses a random restart for HRA and thus generates new approximate eigenvectors. Once these are combined with the previously determined eigenvectors, multiple eigenvalues can appear. There is extra expense for the combination step, but it does not have to be performed after every cycle.

This new approach is an alternative to relying on roundoff error to generate multiple eigenvalues and to using block methods. The new approach seems to give multiple eigenvalues quicker than waiting for roundoff, and it also has the advantage that it is easier to know when to terminate. Multiple eigenvalues will generally begin to appear in less cycles of the second HRA iteration than were used for the first HRA iteration (and usually far less). Compared to a block Arnoldi method, convergence of the new approach can be faster if the second HRA iteration and the block method uses the same size overall subspaces. Also, a block method needs to know the maximum multiplicity ahead of time.

Acknowledgments

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.

[2] A. Bai, D. Day, J. Demmel, J. Dongarra, A test matrix collection for non-Hermitian eigenvalue problems, Technical Report CS-97-355, University of Tennessee, Knoxville, TN, 1997.

[3] M. Eiermann, O.G. Ernst, O. Schneider, Analysis of acceleration strategies for restarted minimal residual methods, J. Comput. Appl. Math. 123 (2000) 261–292.

[4] R.W. Freund, Quasi-kernel polynomials and their use in non-Hermitian matrix iterations, J. Comput. Appl. Math. 43 (1992) 135–158.

[5] Z. Jia, Arnoldi type algorithms for large unsymmetric multiple eigenvalue problems, J. Comput. Math. 17 (1999) 257–274.

[6] R.B. Lehoucq, Implicitly restarted Arnoldi methods and subspace iteration, SIAM J. Matrix Anal. Appl. 23 (2001) 551–562.

[7] R.B. Lehoucq, D.C. Sorensen, Deflation techniques for an implicitly restarted Arnoldi iteration, SIAM J. Matrix Anal. Appl. 17 (1996) 789–821.

[8] R.B. Morgan, Computing interior eigenvalues of large matrices, Linear Algebra Appl. 154–156 (1991) 289–309.

[9] R.B. Morgan, On restarting the Arnoldi method for large nonsymmetric eigenvalue problems, Math. Comp. 65 (1996) 1213–1230.

[10] R.B. Morgan, Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations, SIAM J. Matrix Anal. Appl. 21 (2000) 1112–1135.
[11] R.B. Morgan, GMRES with deflated restarting, SIAM J. Sci. Comput. 24 (2002) 20–37.
[12] R.B. Morgan, Restarted block-GMRES with deflation of eigenvalues, Appl. Numer. Math. 54 (2005) 222–236.
[13] R.B. Morgan, M. Zeng, Harmonic projection methods for large non-symmetric eigenvalue problems, Numer. Linear Algebra Appl. 5 (1998) 33–55.
[14] C.C. Paige, B.N. Parlett, H.A. van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces, Numer. Linear Algebra Appl. 2 (1995) 115–133.
[15] B.N. Parlett, The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[16] Y. Saad, Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices, Linear Algebra Appl. 34 (1980) 269–295.
[17] Y. Saad, Numerical Methods for Large Eigenvalue Problems, Halsted Press, New York, NY, 1992.
[18] G.L.G. Sleijpen, H.A. van der Vorst, A Jacobi–Davidson iteration method for linear eigenvalue problems, SIAM J. Matrix Anal. Appl. 17 (1996) 401–425.
[19] D.C. Sorensen, Implicit application of polynomial filters in a *k*-step Arnoldi method, SIAM J. Matrix Anal. Appl. 13 (1992) 357–385.
[20] D.C. Sorensen, Deflation for implicitly restarted Arnoldi methods, Technical Report TR98-12, Rice University, Houston, TX, 1998.
[21] A. Stathopoulos, Y. Saad, K. Wu, Dynamic thick restarting of the Davidson and the implicitly restarted Arnoldi methods, SIAM J. Sci. Comput. 19 (1998) 227–245.
[22] G.W. Stewart, A Krylov–Schur algorithm for large eigenproblems, SIAM J. Matrix Anal. Appl. 23 (2001) 601–614.
[23] G.W. Stewart, Matrix Algorithms II: Eigensystems, SIAM, Philadelphia, 2001.
[24] K. Wu, H. Simon, Thick-restart Lanczos method for symmetric eigenvalue problems, SIAM J. Matrix Anal. Appl. 22 (2000) 602–616.