# MULTIGRID ARNOLDI FOR EIGENVALUES

RONALD B. MORGAN† AND ZHAO YANG‡

**Abstract.** A new approach is given for computing eigenvalues and eigenvectors of large matrices. Multigrid is combined with the Arnoldi method in order to solve difficult problems. First, a two-grid method computes eigenvalues on a coarse grid and improves them on the fine grid. On the fine grid, an Arnoldi-type method is used that, unlike standard Arnoldi methods, can accept initial approximate eigenvectors. This approach can vastly improve Krylov computations. It also succeeds for problems that regular multigrid cannot handle. Analysis is given to explain why fine grid convergence rates can sometimes unexpectedly match that of regular Arnoldi. Near-Krylov theory is developed for this. Finally, this new method is generalized for more grid levels to produce a Multiple-grid Arnoldi method.

**AMS subject classifications.** 65F15, 15A18

**1. Introduction.** We look at computing eigenvalues and eigenvectors of large, possibly nonsymmetric, matrices. Arnoldi methods [2, 22, 26, 15, 11, 19, 23] are Krylov subspace methods for nonsymmetric eigenvalue problems. Sorensen's implicitly restarted Arnoldi method [26] was a leap forward, because it can find several eigenvalues simultaneously. However, many eigenvalue problems are still challenging. For problems coming from discretization of differential equations, fine discretizations lead to large matrices, wide spectra and difficult computations. There is potential for another step forward with multigrid [3]. Using a coarse grid gives a smaller matrix and an easier spectrum, so the convergence can be much faster. However the information from the coarse grid must be translated to the fine grid and then improved to give acceptable fine grid eigenvalues and eigenvectors. Such an approach is proposed here.

In its simplest form, this new method uses two grids. Eigenvectors are computed on the coarse grid using a standard Arnoldi method, interpolated to the fine grid, and then improved on the fine grid with an Arnoldi method that can accept initial approximate eigenvectors. Examples are given showing this new aproach can dramatically improve eigenvalue calculations. For some difficult problems from partial differential equations, the improvement can be by a factor of ten or even a hundred. However, this method is best when the fine grid is fairly fine. Near-Krylov theory is developed to explain the effectiveness of this method. Finally, a more general method is given that uses multiple grids. For some problems, this is a significant improvement over just two grids.

Section 2 has background material including on needed Arnoldi methods. Section 3 has the new Two-grid Arnoldi, and it is analyzed in section 4. The multiple grids method is in section 5.

## 2. Background.

**2.1. Multigrid.** Partial differential equations need to be solved in many areas of science. Often the problem is discretized, turned from a continuous problem into

*The first author was supported by the National Science Foundation under grant DMS-1418677.

†Department of Mathematics, Baylor University, Waco, TX 76798-7328 (Ronald_Morgan@baylor.edu).

‡Department of Mathematics, Oklahoma State University, Stillwater, OK 74078 (zhao.yang@okstate.edu).

a discrete problem, by dividing the region with a grid. Finite difference and finite elements are two such methods. Frequently a large system of linear equations must then be solved. Multigrid methods solve the linear equations iteratively using several grids of varying fineness. There are many references for multigrid; see for example [7, 3, 5, 6, 31]. Sometimes multigrid is very fast. Much of the work is done on coarse grids which are cheaper to iterate on. However, multigrid methods do not work well for many problems. For example, standard multigrid fails when there is too much convection in a convection-diffusion equation. This also can happen for indefinite problems such as a Helmholtz equation with large enough wave number.

**2.2. Eigenvalue multigrid.** Multigrid methods have been proposed for eigenvalue problems; see for example [4, 13, 34, 12]. However, there do not seem to be any standard methods. Multigrid eigenvalue methods can suffer from two difficulties: 1) many methods compute one eigenvalue at a time and so do not have advantages that are possible if many eigenvalues are found simultaneously, and 2) generally there are the same limitations on the type of problem as just mentioned for linear equations multigrid. We note one multigrid eigenvalue method for later comparison. Shift-and-invert Arnoldi [23] with multigrid to solve the linear equations can find multiple eigenvalues simultaneously and so does not have difficulty 1) above.

**2.3. Krylov methods for eigenvalues.** Krylov subspaces such as $Span\{s, As, A^2s, A^3s, \ldots, A^{m-1}s\}$ are at the core of many iterative methods in linear algebra. Krylov subspace eigenvalue methods can be analyzed in terms of polynomials, and convergence is best for well-separated and exterior eigenvalues. For symmetric matrices, a Krylov subspace with Rayleigh-Ritz extraction becomes the Lanczos method [10, 21, 33, 1]. For nonsymmetric matrices, we get the Arnoldi algorithm [2, 22, 26, 15, 27, 11, 19, 23]. Except for easy problems, Arnoldi needs to be restarted to control orthogonalization expense and storage needs. Restarting with one vector slows convergence. Fortunately, Sorensen's implicitly restarted Arnoldi [26] allows for restarting that retains multiple approximate eigenvectors. At every cycle it uses the subspace

$$Span\{y_1, y_2, \ldots y_k, w, Aw, A^2w, A^3w, \ldots, A^{m-k-1}w\}. \tag{2.1}$$

where $\{y_1, y_2, \ldots, y_k\}$ are Ritz vectors computed at the end of the previous cycle and $w$ is the last Arnoldi vector, $v_{m+1}$, from the previous cycle and is a multiple of the residual vector for each of these Ritz vectors. This subspace is equivalent [15] to

$$Span\{y_1, y_2, \ldots y_k, Ay_j, A^2y_j, A^3y_j, \ldots, A^{m-k}y_j\}, \tag{2.2}$$

for each $y_j$. So the subspace contains a Krylov subspace with each Ritz vector as starting vector. Not only is convergence generally much better than when restarting with one vector, but several or even many eigenvalues can be computed simultaneously.

Another method using subspace (2.2) is given in [15]. It is called Arnoldi-E and has a different implementation, putting the $y_i$'s other than $y_j$ last. For example, if $y_1$ is chosen to be the starting vector for a particular cycle, then the vectors are orthogonalized in this order:

$$\{y_1, Ay_1, A^2y_1, A^3y_1, \ldots, A^{m-k}y_1, y_2, \ldots y_k, \}. \tag{2.3}$$

So the subspace includes a Krylov subspace with one of the current Ritz vectors as starting vector and then appends approximate eigenvectors at the end of the cycle. Arnoldi-E is normally equivalent to implicitly restarted Arnoldi at the end of each

cycle. However, initial approximate eigenvectors can be input at the beginning of the run. In this case, not only is the method no longer equivalent to implicitly restarted Arnoldi, but the choice of which Ritz vector is chosen as the starting vector for the Krylov portion now makes a difference. We give a sketch of the Arnoldi-E algorithm. For a more detailed implementation, see [15].

**Arnoldi-E(m,k)**

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 $nev$, the desired number of eigenpairs, and $rtol$, the convergence tolerance. Normally $nev < k$. Either choose an initial vector $v_1$ of unit length and go to step 2, or specify initial approximate eigenvectors $y_1, y_2, \ldots y_k$ and go to step 3.
2. **One cycle of regular Arnoldi:** Run a cycle of Arnoldi(m) with starting vector $v_1$. Compute desired Ritz vectors $y_1, y_2, \ldots y_k$.
3. **Arnoldi-E cycle:** Choose one of the approximate eigenvectors as starting vector, say $y_j$. Apply the Rayleigh-Ritz procedure to the vectors $\{y_j, Ay_j, A^2y_j, \ldots, A^{m-k}y_j, y_1, y_2, \ldots, y_{j-1}, y_{j+1}, \ldots, y_k\}$. Compute Ritz vectors $y_1, y_2, \ldots y_k$. If the $nev$ desired ones have converged to the tolerance, then stop. Otherwise, repeat this step.

The computation of the Rayleigh-Ritz reduced matrix $H$ is more expensive in Arnoldi-E than for regular restarted Arnoldi and there is some choice in how to implement it. As with implicitly restarted Arnoldi, only $m-k$ matrix-vector products are required for a cycle [15], though for quite sparse matrices it is simpler to use $m$ of them.

For our Two-grid Arnoldi method that is given in the next section, this Arnoldi-E method is used in the second phase with the fine grid. We next discuss the Arnoldi method that will be used for the first phase on the coarse grid. While it is equivalent to implicitly restarted Arnoldi at all iterations, the implementation is simpler.

Wu and Simon [33] (see also [1]) give the symmetric case of a method that uses the same subspace as Arnoldi-E, but puts the approximate eigenvectors at the beginning of the subspace instead of the end. This approach is in [19] for the nonsymmetric case. Stewart gives a more general method in [28]. The paper [19] also has a harmonic Rayleigh-Ritz version which is used in Example 4, but all other examples use the regular Rayleigh-Ritz version. The algorithm for this is given next. At the restart, an orthonormal basis is formed for $Span\{y_1, y_2, \ldots y_k, w\}$ in $V_{k+1}^{new}$ (recall $w$ is the vector $v_{m+1}$ from the previous cycle). Then this is built out with the Arnoldi iteration to an orthonormal basis for subspace (2.1).

**Restarted Arnoldi(m,k)**

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 $nev$, the desired number of eigenpairs, and $rtol$, the convergence tolerance. Normally $nev < k$. Choose an initial vector $v_1$ of unit length.
2. **Arnoldi iteration:** Apply the Arnoldi iteration from the current point to generate the Arnoldi-like recurrence $AV_m = V_{m+1}H_{m+1,m}$. The current point is either from $v_1$ if it is the first cycle or from $v_{k+1}$ on the other cycles. Here $H_{m+1,m}$ is upper-Hessenberg for the first cycle and for the others it is upper-Hessenberg except for a full leading $k+1$ by $k+1$ portion.

3. **Small eigenvalue problem:** Compute the $k$ desired eigenpairs $(\theta_i, g_i)$, with $g_i$ normalized, of $H_{m,m}$. The $\theta_i$ are the Ritz values.
4. **Check convergence:** Residual norms can be computed using $||r_i|| \equiv ||Ay_i - \theta_1 y_i|| = h_{m+1,m}|g_{m,i}|$. If all desired *numev* eigenpairs have acceptable residual norm, then stop, first computing eigenvectors, if desired, as $y_i = V_m g_i$. Otherwise continue. The next step begins the restart.
5. **Orthonormalize the first $k$ short vectors:** Orthonormalize $g_i$'s, for $1 \leq i \leq k$, 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. **Form P:** Extend the columns of $P_k$, called $p_1, \ldots p_k$, to length $m+1$ by appending a zero to each, then set $p_{m+1} = e_{m+1}$, the $(m+1)$st coordinate vector of length $m+1$. Let $P_{m+1,k+1}$ be the $m+1$ by $k+1$ matrix with $p_i$'s as columns.
7. **Form portions of new $H$ and $V$ using the old $H$ and $V$:** Let $H_{k+1,k}^{new} = P_{k+1}^T H_{m+1,m} P_k$ and $V_{k+1}^{new} = V_{m+1} P_{m+1,k+1}$. Then let $H_{k+1,k} = H_{k+1,k}^{new}$ and $V_{k+1} = V_{k+1}^{new}$.
8. **Reorthogonalize the long $k+1$ vector:** Orthogonalize $v_{k+1}$ against the earlier columns of the new $V_{k+1}$. Go to step 2.

Full reorthogonalization is generally used with Arnoldi eigenvalue methods. All experiments here do use this. For Step 3 of this implementation of Arnoldi, it is important to use the 'nobalance' option in older versions of Matlab and LAPACK to avoid errors.

**2.4. Restarted Arnoldi Example.** Even with implicitly restarted Arnoldi and related methods, some eigenvalue problems are still very difficult. We give an example that illustrates slow convergence of restarted Arnoldi. This shows the need for even better eigenvalue computation methods.

*Example 1.* We consider a matrix from finite difference discretization of the 2-D convection-diffusion equation $-u_{xx} - u_{yy} + 10u_x = \lambda u$ on the unit square with zero boundary conditions. The discretization size is $h = \frac{1}{700}$, leading to a matrix of dimension $n = 699^2 = 488,601$. The eigenvalues range from $9.13 * 10^{-5}$ to 8.0 with near-multiplicity adding to the difficulty. The Restarted Arnoldi(m,k) method given above is run with the goal of finding the ten smallest eigenvalues and the corresponding eigenvectors. We use subspaces of maximum size 30 and restart with 15 Ritz vectors so the method is denoted by Arnoldi(30,15). To find 10 eigenvalues with residual norms below $10^{-8}$ takes 2375 cycles or 35,640 matrix-vector products. The dash-dotted lines on Figure 3.1 show these residual norms. One reason for the somewhat erratic convergence is the presence of the nearly multiple eigenvalues.

Next we give a new approach for computing eigenvalues that combines Krylov subspaces with multigrid.

**3. Two-grid Arnoldi.** Multigrid methods use information from coarse grids to assist the desired computation on the fine grid. Here we wish to use eigenvector information from a coarse grid. These coarse grid eigenvectors can be extended to the fine grid but will only be approximate. They need to be improved on the fine grid. As mentioned earlier, the standard large matrix eigenvalue computation method is implicitly restarted Arnoldi. It uses steps of the shifted QR-iteration (part of the standard way of solving small eigenvalue problems). The implicit shifts for these QR

steps are chosen to be the unwanted Ritz values. This leads to a method which cannot be easily modified. Most notably, it is not possible to input approximate eigenvectors into the implicitly restarted Arnoldi method and have it improve them. However, we will use the Arnoldi-E method mentioned earlier that can accept initial inputed vectors.

The resulting Krylov multigrid method can compute many eigenvalues simultaneously and can be used on problems for which standard multigrid fails. It has potential to dramatically improve computation of eigenvalues and eigenvectors. We present the method first on two grids. It is given for finding the $nev$ eigenvalues smallest in magnitude, but other desired eigenvalues can be found instead.

**TWO-GRID ARNOLDI**

0. **Initial Setup:**
   Let the problem size be $nfg$, meaning the fine grid matrix is $nfg$ by $nfg$.
   Choose the coarse grid size $ncg$.
   Choose $m$ = the maximum subspace size,
   $k$ = the number of Ritz vectors retained at the restart,
   $nev$ = the number of desired eigenpairs,
   $rtol$ = the residual norm tolerance.
1. **Coarse Grid Computation:**
   Run restarted Arnoldi(m,k) on the coarse grid until the $nev$ smallest magnitude eigenvalues have converged to $rtol$.
2. **Move to Fine Grid:**
   Move the $k$ coarse grid Ritz vectors to the fine grid (we use spline interpolation).
3. **Fine Grid Computation:**
   Improve the approximate eigenvectors on the fine grid with the Arnoldi-E(m,k) method. For the starting vectors for the Krylov portion of each cycle, we alternate through the Ritz vectors $y_1$ through $y_{nev}$. However, converged Ritz vectors are skipped. Also, if there are complex vectors, they are split into real and imaginary parts. Stop when the $nev$ smallest Ritz pairs reach residual norms below $rtol$.

As noted in the algorithm, if Ritz vectors in the Arnoldi-E phase are complex, then they are split into real and imaginary parts. It seems this might degrade performance, because then the starting vector for the Krylov portion of the subspace is not a Ritz vector. However, not only does this prevent the bulk of computations needing complex arithmetic, but also the computational results are good. At the end of Example 3 there is a comparison with not splitting.

*Example 1 (cont.).* We apply the Two-grid Arnoldi approach to the eigenvalue problem in Example 1. For the coarse grid, we use discretization size of $h = \frac{1}{350}$, so the number of grid points and dimension of the matrix is $349^2 = 121,801$. This is about one-fourth of the dimension of the fine grid matrix. Only 665 cycles of Arn(30,15) are required to find the smallest 10 eigenvalues to accuracy of residual norm below $10^{-8}$. This is under a third of the cycles for the larger problem. However, the cost is actually much less than this, because with a smaller matrix and shorter vectors, the cost per cycle is about one-fourth as much. In this experiment, we run 665 cycles on the coarse grid, then move the coarse grid eigenvectors to the fine grid and improve them there. The Arnoldi-E method needs only 51 cycles on the fine grid for the eigenvectors to reach the desired level. The coarse grid is a good enough approximation to the fine

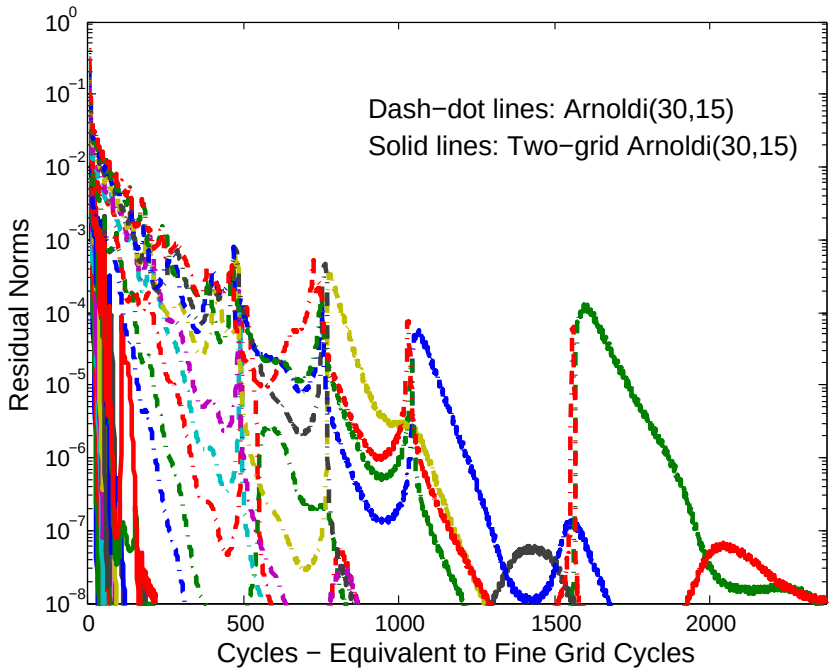


FIG. 3.1. *Standard Arnoldi compared to Two-grid Arnoldi. Fine grid matrix size is* $n = 488,601$ *and coarse grid matrix size is* $n = 121,801$.

grid that the approximations from the course grid are accurate to residual norms of $5 * 10^{-8}$ or better at the beginning of the fine grid work. To better compare the Two-grid approach, we multiply the number of coarse grid cycles by one-fourth and add this to the number of fine grid cycles. This gives 217 fine-grid-equivalent cycles compared to the 2375 for regular Arnoldi. Figure 3.1 shows with solid lines the Two-grid Arnoldi residual norms with coarse grid convergence scaled by one-fourth followed by fine grid results. As mentioned, they have all converged at 217. The regular Arnoldi convergence is there with dash-dot lines and takes ten times longer (note that some of the dash-dot lines look solid toward the end because of oscillations). One reason that Two-grid Arnoldi is better is that it deals with the nearly multiple eigenvalues on the coarse grid instead of having to do so on the fine grid. On the coarse grid, the residual norms jump up when an approximation appears to the second of a couple of nearly multiple eigenvalues, but there is no such problem on the fine grid because all approximations are already there.

We now consider the same fine grid eigenvalue problem, but compare different coarse grid sizes. We see that if fairly accurate eigenvectors are needed with residual norms below $10^{-8}$, then the coarser grids are not as competitive. However, for less accurate approximations, a coarser grid may be better. We note that all coarse grids give results significantly better than standard Arnoldi only on the fine grid. Figure 3.2 has the results with coarse grids using $h = \frac{1}{350}$, $h = \frac{1}{175}$ and $h = \frac{1}{88}$. So the three coarse grid matrices are dimension $ncg = 121,801$, $174^2 = 30,276$ and $87^2 = 7569$. The solid lines on the figure are for $ncg = 121,801$ and are the same as the solid lines in the previous figure. For this coarse grid, most of the work is done on the coarse grid and convergence is quick on the fine grid. The smaller coarse grid matrices need less effort on the coarse grid but use increasingly more cycles on the fine grid. It is

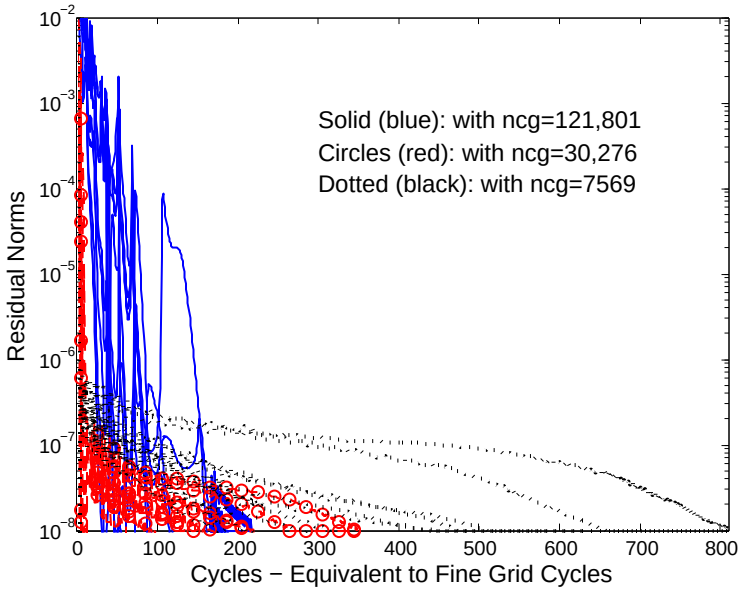


FIG. 3.2. *Two-grid Arnoldi with different coarse grids. Fine grid matrix size is* $n = 488,601$ *and coarse grid matrix sizes are* $n = 121,801$, $n = 30,276$ *and* $n = 7569$.

TABLE 3.1
*Comparison of several coarse grid sizes with regular Arnoldi. The number of fine-grid-equivalent cycles to reach different levels of accuracy is given.*

| rtol | reg. Arn. cycles | ncg = 121,801 fg-equiv. cycles | ncg = 30,276 fg-equiv. cycles | ncg = 7569 fg-equiv. cycles |
|---|---|---|---|---|
| $10^{-6}$ | 1865 | 167 | 12 | 2 |
| $10^{-7}$ | 1955 | 167 | 19 | 452 |
| $10^{-8}$ | 2375 | 217 | 349 | 811 |

interesting that the best method changes depending on the desired accuracy. Table 3.1 has the number of fine-grid-equivalent cycles for converging to different residual norm levels. Note that the iteration on the coarse grid is terminated at $10^{-8}$ for all of these tests, regardless of the desired level on the fine grid. For residual norms below $10^{-8}$ on the fine grid, the coarse grid of $ncg = 121,801$ is best. For $10^{-7}$, $ncg = 30,276$ is best. Then for residual norms below $10^{-6}$, $ncg = 7569$ should be used. It is about 1000 times faster than regular Arnoldi (about 2 fine-grid-equivalent cycles versus 1865 cycles).

For these lower accuracy levels, there is danger of missing some of the nearly multiple eigenvalues. For regular Arnoldi on the fine grid, if we request residual norms below $10^{-7}$, the iteration will stop at cycle 1297 and will miss one of the smallest 10 eigenvalues. For Two-grid Arnoldi, one eigenvalue is missed for $ncg = 30,276$ if the coarse grid tolerance is only $10^{-6}$. With tolerance of $10^{-7}$, all desired eigenvalues are found. With $ncg = 7569$ and $10^{-6}$ all eigenvalues are found, probably because this small problem is easier. These results point out a big advantage of the two-grid approach: since there is less expense on the coarse grid, more care can be given on that level to making sure we have all of the desired eigenvalues.

Two-grid Arnoldi may not be as effective relative to regular Arnoldi when the coarse grid does not give accurate approximations for the fine grid. This is particu-

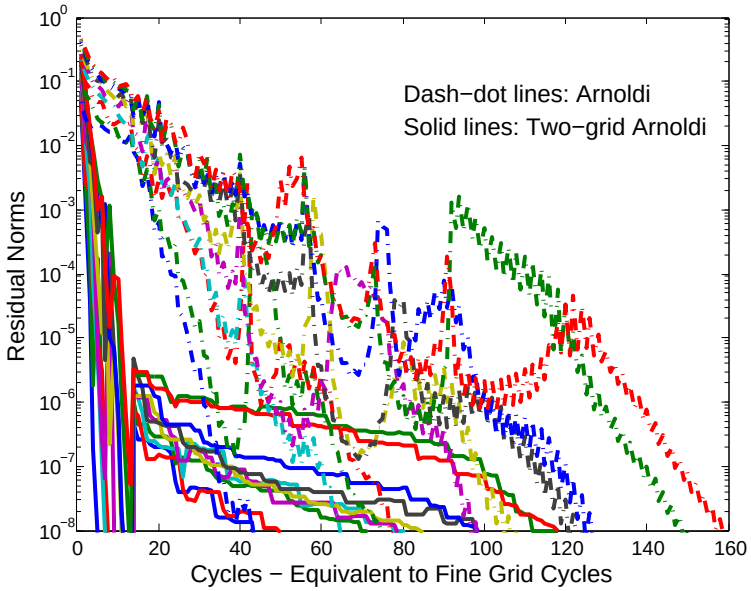


FIG. 3.3. *Standard Arnoldi compared to Two-grid Arnoldi with smaller matrices. Fine grid matrix size is* $n = 30,276$ *and coarse grid matrix size is* $n = 7569$.

lary likely if even the fine grid is fairly coarse. As an example, we let the fine grid problem that we want solved be the matrix of size $n = 30,276$ that was previously from a coarse grid. For the coarse grid, we use the $ncg = 7569$ matrix. The comparison of Arnoldi(30,15) with Two-grid Arnoldi(30,15) is given in Figure 3.3. For reaching accuracy of residual norm below $10^{-8}$, Two-grid Arnoldi is not as much of an improvement as for the previous larger matrix. The two-grid approach is significantly better if less accuracy is desired: after 25 fine-grid-equivalent cycles (53 cycles on the coarse grid and 12 cycles on the fine grid), all 10 approximations reach residual norm $10^{-6}$ or better. Regular Arnoldi uses 137 cycles. If, on the other hand, higher accuracy than $10^{-8}$ is needed, regular Arnoldi will likely beat Two-Grid. We conclude that while Two-grid Arnoldi may improve upon standard Arnoldi for smaller problems, there is more potential with large matrices coming from fine grids. The two reasons for this are first that when the problem is smaller, it is likely also easier, so there is not as much to improve upon. Second, when the problem is smaller, a coarse grid does not give as good of an approximation for it. So the approximate eigenvectors at the start of the fine grid phase are not as accurate and there is more work to be done on the fine grid.

We next do a comparison to an existing multigrid method for finding eigenvalues and eigenvectors. As mentioned in Subsection 2.2, this is Shift-and-invert Arnoldi with a multigrid linear equations solver.

*Example 2.* We begin with an example for which multigrid methods for linear equations work very well. The differential equation on the interval $[0, 1]$ is $-u'' = \lambda u$, so it is a one-dimensional Laplacian. The matrices from finite differences are symmetric. We let $h = \frac{1}{4096}$, so the matrix is dimension $n = 4095$. The desired accuracy is again residual norm below $10^{-8}$. For this problem, approximations from a coarse grid tend to be accurate on a finer grid. For example, with coarse grid matrix of size $ncg = 255$ for Two-grid Arnoldi, seven of the ten desired eigenpairs are already converged when the move is made from the coarse to the fine grid. It then takes

TABLE 3.2
*Two-grid Arnoldi vs. Inverted Arnoldi with Multigrid. Matrix is dimension $n = 4095$ from 1-D Laplacian.*

| Two-grid Arn. | coarse grid | cg cycles | fg cycles | mvp equiv's | time |
|---|---|---|---|---|---|
| | 127 | 11 | 8 | 158 | 0.47 |
| | 255 | 24 | 3 | 95 | 0.41 |
| | 511 | 64 | 0 | 132 | 0.61 |
| | 1023 | 203 | 0 | 775 | 1.90 |
| Inverted Arn. | rtol for lin. eq's | Arn. it.'s | mvp's per | mvp equiv's | time |
| | 1.e-8 | 27 | 47.3 | 1277 | 1.06 |
| | 1.e-4 | 27 | 26.6 | 719 | 0.71 |

three cycles on the fine grid to refine the other three. The number of fine grid cycle equivalents used is $\frac{24}{16}$ on the coarse grid and three on the fine grid for a total of $4\frac{1}{2}$. Standard Arnoldi(30,15) takes 2407 cycles or 36,120 matrix-vector products, so Two-grid Arnoldi is about 500 times faster.

We now compare the new approach against shift-and-invert Arnoldi with standard multigrid used to solve the linear equations. Since we are computing the eigenvalues nearest zero, the appropriate operator is $A^{-1}$. We compute 10 eigenvalues to residual norms below $10^{-8}$. The multigrid linear equations solver is implemented with V-cycles and one Jacobi relaxation weighted by $\frac{2}{3}$ on each grid. The linear equations are solved first to accuracy of residual norm below $10^{-8}$, which gives high accuracy eigenvalues, and then to $10^{-4}$, which gives eigenvalue residual norms just below $10^{-8}$. For both cases, the outer Arnoldi loop requires 27 iterations. We give the scaled number of matrix-vector products with coarser grids counting the appropriate fraction of a fine grid matrix-vector product. The results are in Table 3.2. For inverted Arnoldi and Arnoldi-E, they include one matrix-vector product to check the $10th$ residual norm at the end of each cycle and then nine more to check the others at the end of the process (this is needed for inverted Arnoldi but could possibly be avoided for Arnoldi-E with a shortcut residual formula [15]). The multigrid linear equations solves that are used to implement the inverted operator are very efficient for this matrix. If solving the linear equations to $10^{-8}$, only an average of 47.3 fine-grid equivalent matrix-vector products are needed per solve. A total of 1277 are needed for the entire inverted Arnoldi process. This is far less than the over 36 thousand for regular Arnoldi. However, as mentioned earlier, Two-grid Arnoldi with coarse grid of size $n = 255$ uses even less. The total of all coarse grid (scaled by one-sixteenth) and fine grid work is 95 fine-grid-equivalent matrix-vector products. So the new approach is much better in terms of matrix-vector products. The time required is also reduced, 0.41 seconds compared to 1.06. The improvement on time for Two-grid Arnoldi is less because the matrix is very sparse and the greater orthogonalization expense of Two-grid Arnoldi is significant. Stopping the multigrid linear equations solution at $10^{-4}$ reduces the time for shift-and-invert Arnoldi to 0.71 (but limits the accuracy of the eigenvalues).

We have shown that the new method can compete with a standard multigrid approach even for an ideal multigrid matrix. Next we modify the matrix so that it is nonsymmetric and see that standard multigrid is not robust. We use the one-dimensional convection-diffusion equation $-u'' + \beta u' = \lambda u$ and again discretize to get a matrix of size $n = 4095$. We increase the $\beta$ and observe how this affects the performance of the multigrid linear equations solver used for inverted Arnoldi. Around

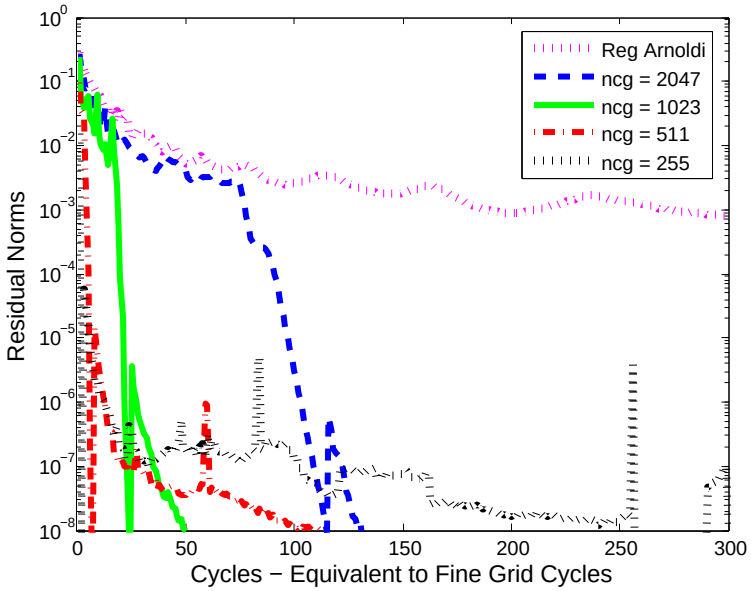


FIG. 3.4. *Two-grid Arnoldi with varying course grid sizes for 1-D Convection-diffusion Problem with Convection Coefficient of 204.8. Fine grid matrix size is* $n = 4095$ *and coarse grid matrices range from* $ncg = 255$ *up to* $ncg = 2047$. *The residual norm for the tenth eigenvalue is shown.*

$\beta = 20$, the multigrid solution of the linear equations begins having problems. By $\beta = 25$, the multigrid iteration diverges and inverted Arnoldi fails. Meanwhile, Two-grid Arnoldi at $\beta = 25$ works about the same as for $\beta = 0$. For the next example, we try a much larger $\beta$ in order to show the robustness of the method.

*Example 3.* We use the convection-diffusion equation as in the previous example, but increase $\beta$ to 204.8. We also switch to Arnoldi(30,16), because then if the $16th$ and $17th$ Ritz values are a complex conjugate pair, they are excluded and the $k$ is temporarily reduced to 15. Figure 3.4 shows the residual norm convergence with several choices of $ncg$ for the $10th$ eigenvalue (the $9th$ and $10th$ are a complex pair and are the last to converge of the first 10). Regular Arnoldi requires 741 cycles to reach residual norm below $10^{-8}$. Two-grid Arnoldi with $ncg = 1023$ uses just 51 fine-grid-equivalent cycles (95 coarse grid cycles and 27 on the fine grid). So even with convection about ten times larger than standard multigrid can handle, Two-grid Arnoldi is effective. For less accurate eigenpairs, the comparison is even more extreme. With both $ncg = 255$ and 511, only 12 fine-grid-equivalent cycles are needed for residual norm below $10^{-6}$. This compares to 650 cycles of regular Arnoldi. For this example, there is sometimes erratic convergence due to the high non-normality. For instance, if $ncg = 1023$, then Two-grid Arnoldi with (30,15) instead of (30,16) converges in 105 fine-grid-equivalent cycles instead of 49.

As mentioned in the Two-grid Arnoldi algorithm and the paragraph after, splitting complex Ritz vectors into real and imaginary parts in the Arnoldi-E portion avoids complex arithmetic. We continue the example by quickly comparing with not splitting. With $ncg = 1023$, the fine-grid-equivalent cycles actually go up from 49 to 53 if they are not split. In other tests, splitting is not always better, but is competitive. Further study of these results is needed.

We next consider a problem with an indefinite matrix. Standard multigrid methods do not work for this matrix, because it is far too indefinite. However, Two-grid

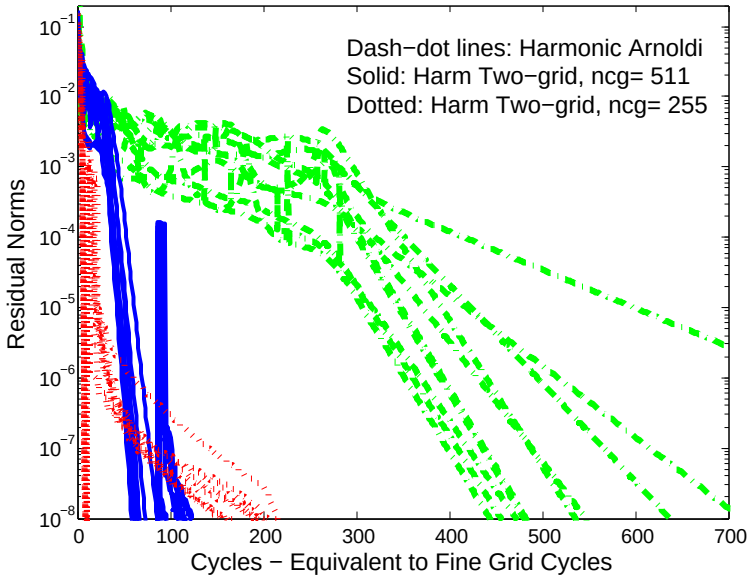


FIG. 3.5. *Standard Arnoldi compared to Two-grid Arnoldi for a 1-D simple Helholtz matrix. Fine grid matrix size is* $n = 1023$ *and coarse grid matrix sizes are* $n = 511$ *and* $n = 255$.

Arnoldi does work.

*Example 4.* We consider a one-dimensional Helmholtz problem $-u'' - 40,000u = \lambda u$. For simplicity, we use zero boundary conditions. The fine grid matrix is size $n = 1023$ and it has 63 negative eigenvalues. Our goal is to compute the 10 eigenvalues closest to the origin, so this is an interior eigenvalue problem. Therefore we switch to harmonic restarted Arnoldi [19] in the first phase of Two-grid Arnoldi. For the second phase, we use harmonic Arnoldi-E [18]. These methods use harmonic Rayleigh-Ritz [14, 20] which makes convergence more reliable for interior eigenvalues. Figure 3.5 has harmonic Arnoldi compared to two tests of harmonic Two-grid Arnoldi. Figure 3.6 has a close-up of Two-grid Arnoldi with $ncg = 511$. Harmonic Arnoldi uses 1148 cycles for 10 eigenvalues to converge to residual norms below $10^{-8}$. However, it misses one of the 10 smallest eigenvalues in magnitude (this is better than non-Harmonic which takes 3058 cycles and misses two of the 10 smallest). Harmonic Two-grid Arnoldi needs 124 fine-grid-equivalent cycles with $ncg = 511$ and 217 for $ncg = 255$. Both find all of the 10 smallest eigenvalues. As mentioned earlier, Two-grid Arnoldi can do much of its work on the coarse grid where the problem is easier. This makes it more reliable.

We also tried a larger fine grid matrix with $nfg = 2047$, and the harmonic two-grid approach with $ncg = 511$ improves by a factor of almost 100 (59 fine-grid-equivalent cycles compared to 5636 for harmonic Arnoldi).

## 4. Fine Grid Convergence Theory.

**4.1. Special properties of vectors from coarse grid.** The Arnoldi-E method does not always work well at improving approximate eigenvectors. The next example demonstrates that the approximate eigenvectors that come from the coarse grid have special properties.

*Example 5.* We give an example with a symmetric matrix of size $n = 1023$ from the 1-D Laplacian. We compare Arnoldi-E for improving approximate eigenvectors

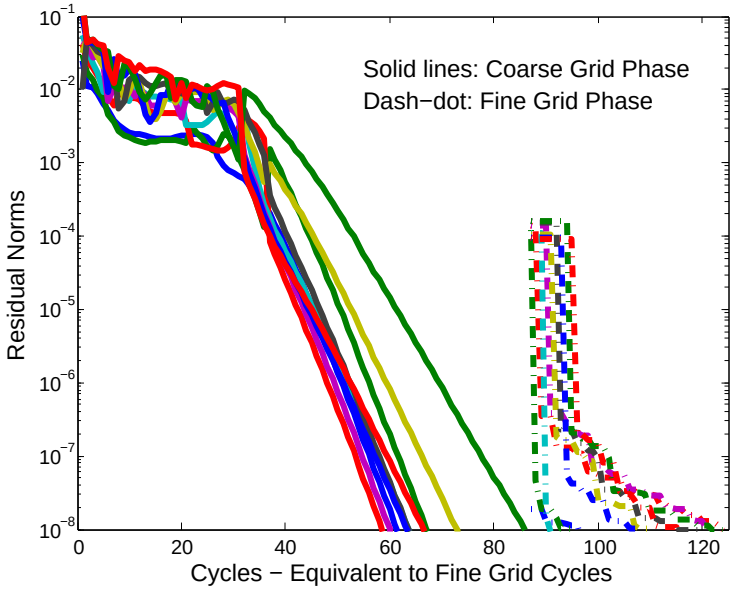


FIG. 3.6. *Two-grid Arnoldi for a 1-D simple Helholtz matrix. Fine grid matrix size is* $n = 1023$ *and coarse grid matrix size is* $n = 511$.

that come from a coarse grid with $ncg = 255$ versus from random perturbation of the true fine grid eigenvectors. The approximate eigenvectors from the coarse grid come from a run of Arnoldi(30,15) with $nev = 10$ and $rtol$ of only $10^{-3}$ and then as usual they are interpolated to the fine grid. The $rtol$ for the second phase is $10^{-10}$. The approximate eigenvectors from perturbation of true vectors use random Normal(0,1) numbers to fill perturbation vectors that are scaled to norm $2 * 10^{-4}$ and then added to the true eigenvectors. Figure 4.1 shows the comparison of the convergence. The eventual convergence rate is almost five times faster for improving the coarse grid vectors. If one looks closeup at the graph, each perturbed eigenvector improves only every tenth cycle, when it is the starting vector for the cycle. Meanwhile the approximate eigenvectors from the coarse grid initially improve at every cycle and later improve at most cycles. We also tried perturbing the fine grid eigenvectors with continuous functions (combinations of exponentials and sine's) and the results were similar to those from the random perturbations.

Next we show an even more dramatic demonstration by restricting the starting vector for each Arnoldi-E cycle to be the first Ritz vector $y_1$. Figure 4.2 has the same comparison of perturbed eigenvectors versus approximate eigenvectors from the coarse grid. This time only $y_1$ converges for the perturbed eigenvectors. Meanwhile, all vectors from the coarse grid converge initially. By the end, their converge curves flatten out except for $y_1$.

We conclude that approximate eigenvectors from the coarse grid have properties that make them work better in Arnoldi-E. We will characterize this property as being near-Krylov. Previous work on inexact Krylov [25, 32] focused on accuracy of matrix-vector products with different analysis.

We first look at the relation between near-Krylov and having nearly parallel residuals. Then it is shown that under idealized conditions, near-Krylov properties do not degrade. Next, we give theorems about convergence for near-Krylov subspaces. Finally, some examples are given of how near-Krylov can help.

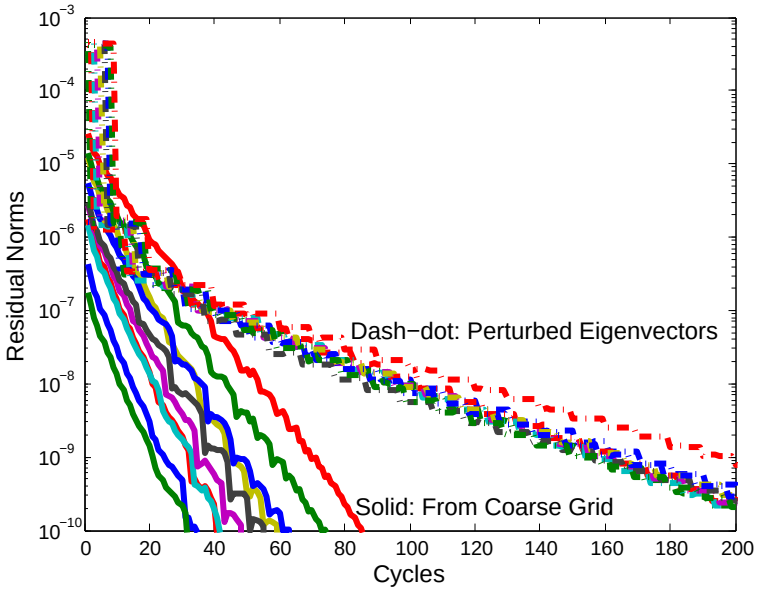


FIG. 4.1. *Comparison of Arnoldi-E convergence with approximate eigenvectors from coarse grid vs. from perturbation of true eigenvectors. 1-D Laplacian with matrix of size* $n = 1047$ *and coarse grid matrix of size* $ncg = 255$.

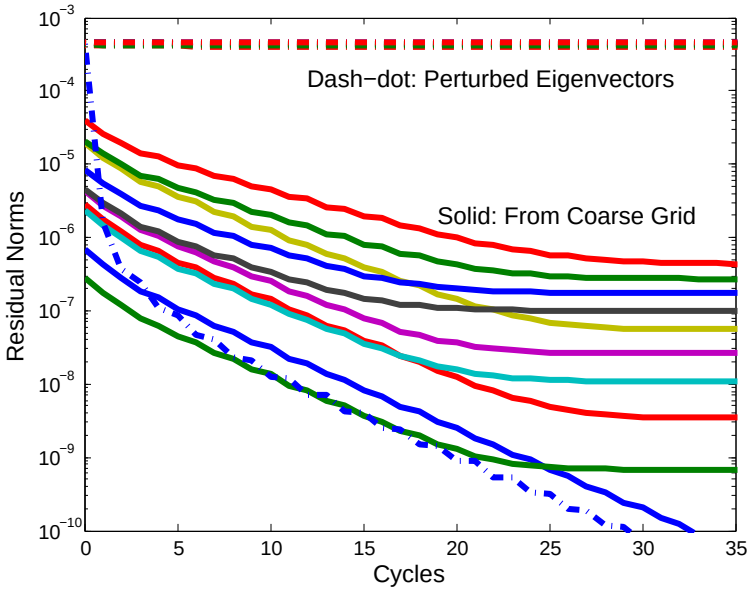


FIG. 4.2. *Comparison of Arnoldi-E convergence with approximate eigenvectors from coarse grid vs. from perturbation of true eigenvectors. Now the first Ritz vector is the starting vector for every cycle.*

**4.2. Near-Krylov and nearly parallel residuals.** One way to characterize a Krylov subspace is that the residual vectors associated with Ritz pairs are parallel [16, 28, 17]. In fact, for the case of regular Arnoldi, $r_i \equiv Ay_i - \theta_i y_i = \gamma_i v_{m+1}$, where $v_{m+1}$ is the last Arnoldi vector. Going the other way, if a subspace has Ritz pairs $(\theta_i, y_i)$, for $i = 1, 2, \ldots, m$ and the associated residual vectors satisfy $r_i \equiv Ay_i - \theta_i y_i = \gamma_i w$,

then $Span\{y_1, \dots y_m\}$ and $Span\{y_1, \dots y_m, w\}$ are Krylov subspaces [17] .

For the Two-grid Arnoldi method, the Krylov property is lost when the Ritz vectors are moved from the coarse grid to the fine grid. However, the span of these vectors on the fine grid has the property of being near-Krylov. They also will have nearly parallel residual vectors. These concepts are defined next. Note that Stewart defines Krylov decomposition in [28] and near-Krylov in [29].

DEFINITION 4.1. *For a subspace of dimension $m$, let the Ritz values be $\theta_i$ for $i = 1, \dots, m$, the corresponding Ritz vectors of unit length be $y_i$'s, and the residuals be $r_i = Ay_i - \theta_i y_i$. Suppose there is a vector $w$ such that for all $y_i$,*

$$Ay_i = \theta_i y_i + \gamma_i w + f_i, \tag{4.1}$$

*with $f_i$ small, Then we say the residuals of the $y_i$ are nearly parallel.*

*Let $Y = [y_1, y_2, \cdots, y_m]$, and $\Theta$ be a diagonal matrix with corresponding $\theta_i$ on the diagonal. Then the near parallel property in the matrix form is*

$$AY = Y\Theta + wa^T + F. \tag{4.2}$$

Next we define near-Krylov.

DEFINITION 4.2. *Let*

$$AU_m = U_m B_m + u_{m+1} b_{m+1}^T + R, \tag{4.3}$$

*where the columns of $(U_m, u_{m+1})$ are linear independent. If $R$ is small, then we say this is a near-Krylov decomposition. $R$ is called the Krylov residual.*

Sometimes we want also to have the columns of $U_m$ be orthonormal, and possibly to have them orthogonal to $u_{m+1}$ and $R$. Note that given a subspace spanned by the columns of $U_m$, Stewart [29] shows how to find the near-Krylov decomposition with the lowest $||R||$.

For some of the analysis that follows, we focus on nearly parallel residuals, while sometimes we use the near-Krylov property. However, the two are related. It is obvious that (4.2) is a special case of (4.3), which means that nearly parallel residuals of the $k$ Ritz vectors implies a near-Krylov decomposition of dimension $k$. Theorem 4.3 shows that a near-Krylov decomposition implies nearly parallel residuals.

THEOREM 4.3. *Let $AU_m = U_m B + u_{m+1} b^T + R$, where $U_m$ and $R$ are $n$ by $m$, $B$ is $m$ by $m$, and $u_{m+1}$ and $b$ are vectors of length $n$ and $m$ respectively. Suppose $(\theta_i, g_i), i = 1, \cdots, k$ are eigenpairs of $B$, denoted by $BG = G\Theta$. Let $y_i = U_m g_i$ and $Y = [y_1, \cdots, y_k]$. Then $AY_k = Y_k\Theta + u_{m+1}a^T + F$ with $\|F\| \leq \|R\|\|G\|$. If $A$ is symmetric and $G$ has orthonormal columns, then $\|F\| \leq \|R\|$.*

*Proof.*

$$\begin{aligned} AU_m &= U_m B + u_{m+1} b^T + R, \\ AU_m G &= U_m BG + u_{m+1} b^T G + RG, \\ AY &= U_m G\Theta + u_{m+1} b^T G + RG. \end{aligned}$$

Let $a^T = b^T G$ and $F = RG$, then

$$\begin{aligned} AY &= Y\Theta + u_{m+1}a^T + F, \\ \text{and } \|F\| &= \|RG\| \leq \|R\|\|G\|. \end{aligned}$$

If $A$ is symmetric, then normally $G$ has orthonormal columns. In this case, $\|G\| = 1$ and $\|F\| \leq \|R\|$. ∎

**4.3. Maintaining near-Krylov.** Since the Arnoldi-E phase starts with a near-Krylov subspace, it is natural to ask if this property is mantained or lost during the iteration. This is now explored.

The first cycle of Arnoldi-E starts with the vectors that were moved from the coarse grid, and the other cycles start with Ritz vectors generated by the previous cycle. Then Arnoldi-E generates a Krylov subspace with a Ritz vector $y_j$ as starting vector. This is joined with the other Ritz vectors to form the overall subspace

$$Span\{y_1, \cdots, y_k, Ay_j, A^2y_j, \cdots A^{m-k}y_j\} \tag{4.4}$$

Let $u = Ay_j$ (the $u$ vector is analogous to using the Ritz residual vector in the regular Arnoldi case for which residuals are all parallel and subspace (4.4) is Krylov). We can consider subspace (4.4) as having the portion $Span\{y_1, \cdots, y_k\}$ and the Krylov portion $Span\{u, Au, A^2u, \cdots A^{m-k-1}u\}$. The first portion has a near-Krylov decomposition and this is (4.5) in the next theorem.

In this theorem, we show that when the subspace $Span\{y_1, \cdots, y_k, u\}$ is expanded out from dimension $k$ to $m$ during Arnoldi-E, the Krylov residual does not increase in norm. So the near-Krylov property is maintained. The theorem after that shows that the contraction back to a subspace of size $k$ also does not increase the Krylov residual norm. However, there can be an increase in Krylov residual when the next vector is chosen, i.e. when we go from $Span\{y_1, \cdots, y_k\}$ to $Span\{y_1, \cdots, y_k, u\}$, because $u$ may not be the optimal choice for the next vector (as mentioned, we choose $u$ to be $A * y_j$ for some Ritz vector $y_j$).

THEOREM 4.4. *Suppose there is a near-Krylov decomposition of a dimension $k$ subspace:*

$$AU_{n\times k} = U_{n\times k}B + ub^T + R_{n\times k}, \tag{4.5}$$

*where the columns of $(U_{n\times k}, u)$ are independent, and*

$$\left[U_{n\times k}, u\right]^T R_{n\times k} = 0.$$

*Suppose there is an Arnoldi decomposition of a dimension $p$ subspace:*

$$AV_{n\times p} = V_{n\times p}H + \eta v_{p+1}e_p^T, \tag{4.6}$$

*and $u$ can be written as*

$$u = V_{n\times p}d. \tag{4.7}$$

*Let $m = p + k$. Assume columns of $(U_{n\times k}, V_{n\times p})$ are linear independent, then there is a near-Krylov decomposition of dimension $m$*

$$A\hat{U}_{n\times m} = \hat{U}_{n\times m}\hat{B} + \hat{u}\hat{b}^T + \hat{R}_{n\times m},$$

*where*

$$\hat{U}_{n\times m} = [V_{n\times p}, U_{n\times k}],$$

*whose columns are linearly independent. Furthermore,*

$$\hat{U}_{n\times m}^T\hat{u} = 0, \quad \hat{U}_{n\times m}^T\hat{R}_{n\times m} = 0, \quad \textit{and } \|\hat{R}\| \le \|R\|.$$

*Proof.* For ease of presentation, we let $V = V_{n\times p}$, $U = U_{n\times k}$ and $R = R_{n\times k}$. Combining (4.5) and (4.6),

$$A\begin{bmatrix}V,U\end{bmatrix} = \begin{bmatrix}V,U\end{bmatrix}\begin{bmatrix}H & \\ & B\end{bmatrix} + \begin{bmatrix}\eta v_{p+1}e_p^T & 0\end{bmatrix} + \begin{bmatrix}0 & ub^T\end{bmatrix} + \begin{bmatrix}0 & R\end{bmatrix}.$$

Since $u = Vd$ from (4.7),

$$\begin{aligned} A\begin{bmatrix}V,U\end{bmatrix} &= \begin{bmatrix}V,U\end{bmatrix}\begin{bmatrix}H & \\ & B\end{bmatrix} + \begin{bmatrix}\eta v_{p+1}e_p^T & 0\end{bmatrix} + \begin{bmatrix}0 & Vdb^T\end{bmatrix} + \begin{bmatrix}0 & R\end{bmatrix} \\ &= \begin{bmatrix}V,U\end{bmatrix}\begin{bmatrix}H & db^T \\ & B\end{bmatrix} + \begin{bmatrix}\eta v_{p+1}e_p^T & 0\end{bmatrix} + \begin{bmatrix}0 & R\end{bmatrix}. \end{aligned}$$

With orthogonal decomposition of $v_{p+1}$ and $R$, we get

$$\begin{aligned} v_{p+1} &= v_0 + \begin{bmatrix}V & U\end{bmatrix}c, \\ \text{and } R &= R_0 + \begin{bmatrix}V & U\end{bmatrix}K, \end{aligned}$$

such that $v_0$ and columns of $R_0$ are orthogonal to the columns of $V$ and $U$, so

$$U^T v_0 = 0, V^T v_0 = 0, \tag{4.8}$$

$$U^T R_0 = 0, V^T R_0 = 0. \tag{4.9}$$

Then

$$A\begin{bmatrix}V,U\end{bmatrix} = \begin{bmatrix}V,U\end{bmatrix}\begin{bmatrix}H & db^T \\ & B\end{bmatrix} + \begin{bmatrix}\eta(v_0 + \begin{bmatrix}V & U\end{bmatrix}c)e_p^T & 0\end{bmatrix} + \begin{bmatrix}0 & (R_0 + \begin{bmatrix}V & U\end{bmatrix}K)\end{bmatrix},$$

$$A\begin{bmatrix}V,U\end{bmatrix} = \begin{bmatrix}V,U\end{bmatrix}(\begin{bmatrix}H & db^T \\ & B\end{bmatrix} + \eta ce_p^T + K) + \begin{bmatrix}\eta v_0 e_p^T & 0\end{bmatrix} + \begin{bmatrix}0 & R_0\end{bmatrix}.$$

Let $\hat{U}_{n\times m} = \begin{bmatrix}V & U\end{bmatrix}$, $\hat{B}_{n\times m} = \begin{bmatrix}H & db^T \\ & B\end{bmatrix} + \eta ce^T + K$, $\hat{u} = \eta v_0$, $\hat{b} = e_p$, $\hat{R} = \begin{bmatrix}0 & R_0\end{bmatrix}$, then we have

$$A\hat{U}_{n\times m} = \hat{U}_{n\times m}\hat{B} + \hat{u}\hat{b}^T + \hat{R}.$$

From the construction of $\hat{U}$, $\hat{u}$ and $\hat{R}$ and using (4.8) and (4.9), we have

$$\begin{aligned} \hat{U}^T\hat{u} &= \begin{bmatrix}V & U\end{bmatrix}^T \eta v_0 = 0, \\ \hat{U}^T\hat{R} &= \begin{bmatrix}V & U\end{bmatrix}^T \begin{bmatrix}0 & R_0\end{bmatrix} = 0. \end{aligned}$$

And

$$\|\hat{R}\| = \|\begin{bmatrix}0 & R_0\end{bmatrix}\| = \|R_0\| \leq \|R\|. \tag{4.10}$$

□

Equation (4.10) tells us that for $\|\hat{R}\|$ to be small, we need $\|R_0\|$ to be small. Since $R = R_0 + \begin{bmatrix}V & U\end{bmatrix}K$, this means if the Krylov residual $R$ of the near-Krylov subspace portion $\mathcal{S}$ can be expanded in terms of the vectors of the Krylov subspace portion $\mathcal{K}$ as $\begin{bmatrix}V & U\end{bmatrix}K$, then the Krylov residual of the overall subspace $\mathcal{W}$ can potentially be reduced.

The next theorem shows that the Krylov residual will not increase during one cycle from $Span\{y_1, \cdots, y_k, u\}$ out to a subspace of dimension $m$ and then back to a subspace $Span\{y_1^{new}, \cdots, y_k^{new}, u^{new}\}$.

THEOREM 4.5. *Assume there is a near-Krylov decomposition*

$$AU_{n\times k} = U_{n\times k}B + ub^T + R_{n\times k}$$

*corresponding to the basis* $\{y_1, y_2, \cdots, y_k, u\}$*, and with* $\left[U_{n\times k}, u\right]^T R_{n\times k} = 0$*. Suppose the subspace we generate for Arnoldi-E procedure is*

$$span\{y_1, y_2, \cdots, y_k, u, Au, A^2u, \cdots, A^{m-k-1}u\},$$

*from which the new k Ritz vectors are* $\{y_1^{new}, y_2^{new}, \cdots, y_k^{new}\}$*. Then there is a near-Krylov decomposition*

$$AU_{n\times k}^{new} = U_{n\times k}^{new}B^{new} + u^{new}(b^{new})^T + R_{n\times k}^{new},$$

*where the columns of* $U_{n\times k}^{new}$ *span the same subspace as* $\{y_1^{new}, y_2^{new}, \cdots, y_k^{new}\}$*, and*

$$\|R_{n\times k}^{new}\| \leq \|R_{n\times k}\|.$$

*Proof.* For the subspace $\{y_1, y_2, \cdots, y_k\}$, there is a near-Krylov decomposition from the assumption

$$AU_{n\times k} = U_{n\times k}B + ub^T + R_{n\times k},$$

For the subspace $span\{u, Au, A^2u, \cdots, A^{m-k-1}u\}$, with as before $p = m - k$, there is an Arnoldi decomposition

$$AV_{n\times p} = V_{n\times p}H + ve_{n\times p}^T.$$

And $u = Ve_1$ since it is the starting vector of the Krylov subspace portion. According to Theorem 4.4, there is a near-Krylov decomposition

$$A\hat{U}_{n\times m} = \hat{U}_{n\times m}\hat{B} + \hat{u}\hat{b}^T + \hat{R}_{n\times m}, \tag{4.11}$$
$$\text{where } \hat{U}_{n\times m} = \begin{bmatrix} V_{n\times p} & U_{n\times k} \end{bmatrix},$$

and we also have

$$\|\hat{R}_{n\times m}\| \leq \|R\|, \tag{4.12}$$
$$\hat{U}_{n\times m}^T\hat{u} = 0,\ \hat{U}_{n\times m}^T\hat{R} = 0.$$

It can be shown (see Lemma 5.4 in [35]), that $\hat{B}$ is similar to the matrix $Q^TAQ$ where columns of $Q$ are orthonormal basis of

$$span\{y_1, y_2, \cdots, y_k, u, Au, A^2u, \cdots, A^{m-k-1}u\}.$$

Hence eigenvalues of $B$ are Ritz values corresponding to the subspace.

Here we assume the $k$ Ritz values of $\hat{B}$ that we want are separated from the other $m - k$ unwanted Ritz values, meaning $\{\theta_1, \cdots, \theta_k\} \cap \{\theta_{k+1}, \cdots, \theta_m\} = \emptyset$. We write the Schur decomposition of $\hat{B}$ as

$$\hat{B}[G_1, G_2] = [G_1, G_2]\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

where the eigenvalues of $T_{11}$ are the new $k$ Ritz values we want. And hence

$$\hat{B}G_1 = G_1 T_{11}.$$

Multiplying both sides of (4.11) by $G_1$, we get

$$\begin{aligned} A\hat{U}_{n\times m}G_1 &= \hat{U}_{n\times m}\hat{B}G_1 + \hat{u}\hat{b}^T G_1 + \hat{R}_{n\times m}G_1, \\ &= \hat{U}_{n\times m}G_1 T_{11} + \hat{u}\hat{b}^T G_1 + \hat{R}_{n\times m}G_1. \end{aligned}$$

Let $U^{new} = \hat{U}_{n\times m}G_1$, $B^{new} = T_{11}$, $u^{new} = \hat{u}$, $(b^{new})^T = \hat{b}^T G_1$ and $R^{new} = \hat{R}_{n\times m}G_1$. Then

$$AU^{new} = U^{new}B^{new} + u^{new}(b^{new})^T + R^{new}.$$

The subspace spanned by the columns of $U^{new}$ is $span\{y_1^{new}, y_2^{new}, \cdots, y_k^{new}\}$. Use (4.12) and $\|G_1\| = 1$,

$$\|R^{new}\| = \|\hat{R}_{n\times(k+m)}G_1\| \leq \|\hat{R}_{n\times(k+m)}\|\|G_1\| \leq \|R_{n\times k}\|.$$

□

The next example shows that in spite of this theorem, the Krylov residual can go up. This is because instead of using the $u^{new}$ vector, we use $u = Ay_j$, where $y_j$ is a Ritz vector. Nevertheless, the norm of the Krylov residual generally does not increase by much during the Arnoldi-E run.

*Example 6.* Here we first use the same matrix as in Example 5, the symmetric matrix of size $n = 1023$ from the 1-D Laplacian. Again $ncg = 255$, but $rtol$ for the first phase is $10^{-8}$. The starting vector for each Arnoldi-E cycle is the first Ritz vector $y_1$. The top part of Figure 4.3 has the Krylov residuals for the $k$-dimensional subspace with $u = A * y_1$ as the $k+1$ vector and for the $m$-dimensional subspace with Stewart's optimal choice [29] of $m+1$ vector. Both of these residuals slightly decrease in norm as the iteration proceeds. The situation is different when a convection term with coefficient of $\beta = 102.4$ is added. The bottom of the figure shows that the Krylov residual norms often jump up some for the $k$-dimensional subspace, with $Ay_j$ as $k+1$ vector. Then the norm goes down as the subspace is built out. Also noticeable is that the initial Krylov residual is larger as there is a much greater departure from Krylov during the move from coarse to fine grid for this matrix than for the symmetric one.

**4.4. Convergence.** It is shown in [17] that if a set of vectors $y_1, \ldots, y_k$ have parallel residuals, then subspace (4.4) is a Krylov subspace. Also it contains the Krylov subspaces with each $y_i$ as starting vector, $span\{y_j, Ay_j, \cdots, A^{m-k}y_j\}$, for $j$ between 1 and $k$. So in the regular Arnoldi method, all eigenvalues are improved at the same time. In this subsection, the plan is to explain the convergence of Arnoldi-E with similar ideas. We show that the subspace still contains Krylov subspaces with each Ritz vector as the starting vector, however with a perturbed matrix. The proof here is from the nearly parallel perspective. See the second part of Thm. 5.10 in [35] for a similar theorem from the near-Krylov perspective.

We focus on only two vectors and how being nearly parallel can help them converge together. With $y_1$ as starting vector, its residual vector is $r_1 = \gamma_1 w$, where $w$ is norm 1. For $y_2$, the residual vector is broken into a multiple of $w$ and a vector $f$ that indicates the deviation from having parallel residuals: $r_2 = \gamma_2 w + f$.

THEOREM 4.6. *Suppose*

$$Ay_1 = \theta_1 y_1 + \gamma_1 w \quad \textit{and} \quad Ay_2 = \theta_2 y_2 + \gamma_2 w + f,$$

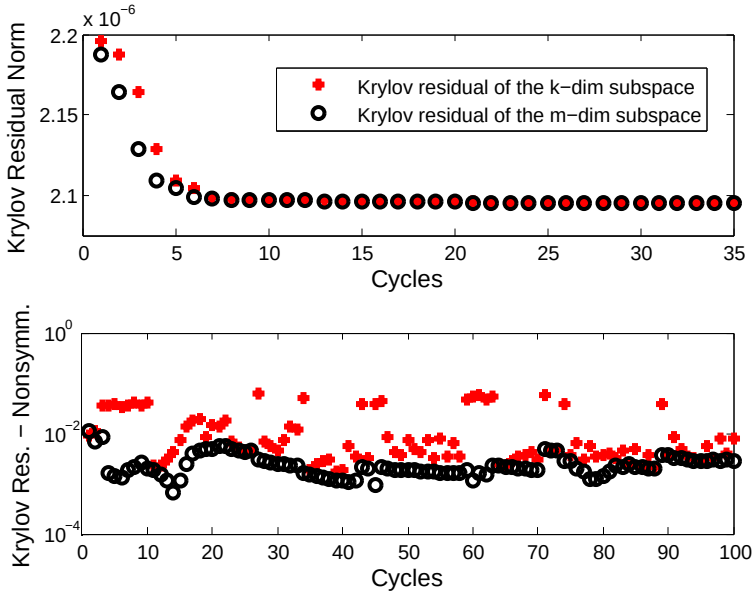


FIG. 4.3. *Plot of Krylov residual norms showing how near the subspaces are to being Krylov. Top matrix is symmetric from 1-D Laplacian with* $ncg = 1023$ *and* $nfg = 255$*. Bottom plot has nonsymmetric matrices due to convection term with* $\beta = 102.4$*.*

*with* $\|y_1\| = \|y_2\| = \|w\| = 1$*. There is a matrix* $E$ *such that*

$$span\{y_2, (A+E)y_2, \cdots, (A+E)^p y_2\} \subset span\{y_1, Ay_1, \cdots, A^p y_1, y_2\}.$$

*Let*

$$y_2 = \alpha y_2^K + \beta y_2^{\perp K}, \tag{4.13}$$

*where* $\|y_2^K\| = \|y_2^{\perp K}\| = 1$*,* $y_2^K \in \mathcal{K} = span\{y_1, Ay_1, \cdots, A^p y_1\}$ *and* $y_2^{\perp K} \perp \mathcal{K}$*, then one choice of* $E$ *is:*

$$E = -\frac{1}{\beta} f (y_2^{\perp K})^T \text{ and } \|E\| \le \frac{\|f\|}{\beta}.$$

*Proof.* Let $E = -\frac{1}{\beta} f (y_2^{\perp K})^T$. Since $y_2^{\perp K} \perp y_1$,

$$(A+E)y_1 = Ay_1 - \frac{1}{\beta} f (y_2^{\perp K})^T y_1 = Ay_1 = \theta_1 y_1 + \gamma_1 w,$$

$$(A+E)y_2 = Ay_2 - \frac{1}{\beta} f (y_2^{\perp K})^T y_2 = Ay_2 - \frac{1}{\beta} f (y_2^{\perp K})^T y_2 = Ay_2 - f = \theta_2 y_2 + \gamma_2 w,$$

So $y_1$ and $y_2$ have parallel residuals under the multiplication of $A+E$. As mentioned at the beginning of this subsection, parallel residuals gives us that a Krylov subspace with $y_2$ is contained in the subspace starting with $y_1$ but augmented with $y_2$ [17]. So

$$span\{y_2, (A+E)y_2, \cdots, (A+E)^p y_2\} \subset span\{y_1, (A+E)y_1, \cdots, (A+E)^p y_1, y_2\}.$$

Next we want to show that
$span\{y_1, (A+E)y_1, \cdots, (A+E)^p y_1, y_2\} = span\{y_1, Ay_1, \cdots, A^p y_1, y_2\}$.

We have $(A+E)y_1 = Ay_1$ from earlier in the proof. Suppose

$$\begin{aligned}(A+E)^j y_1 &= A^j y_1,\\ \text{then } (A+E)^{j+1} y_1 &= (A+E)A^j y_1\\ &= A^{j+1}y_1 + EA^j y_1\\ &= A^{j+1}y_1 - \frac{1}{\beta} f (y_2^{\perp K})^T A^j y_1\\ &= A^{j+1}y_1, \text{ since } y_2^{\perp K} \perp A^j y_1 \text{ for } j = 1, \cdots, p-1.\end{aligned}$$

$$\begin{aligned}\text{So } span\{y_2, (A+E)y_2, \cdots, (A+E)^p y_2\} &\subset span\{y_1, (A+E)y_1, \cdots, (A+E)^p y_1, y_2\}\\ &= span\{y_1, Ay_1, \cdots, A^p y_1, y_2\}.\end{aligned}$$

And

$$\|E\| = \| - \frac{1}{\beta} f (y_2^{\perp K})^T \| \leq \frac{\|f\| \|(y_2^{\perp K})^T\|}{\|\beta\|} = \frac{\|f\|}{\|\beta\|}.$$

□

The theorem indicates that when $y_1$ and $y_2$ are nearly parallel, $y_2$ will converge along with $y_1$ even though $y_1$ is the starting vector in Arnoldi-E. The Krylov subspace for $y_2$ uses $A+E$, but $\|E\|$ may be small if the residuals are nearly parallel and thus $\|f\|$ is small. When $A$ is symmetric or nearly symmetric, the projection of $y_2$ on $\{y_1, Ay_1, \cdots, A^p y_1\}$ will tend to be small. Then $\alpha$ in (4.13) is small and $\beta$ close to 1. In this case, $\|E\|$ is mainly determined by $\|f\|$.

We wish to understand how much the perturbed matrix in Theorem 4.6 can affect the convergence. This is a difficult question, but one way to study this is to use the Cauchy integral to express the polynomials of a matrix. The ideas can be found in [9, 8, 24]. We give a theorem that bounds the difference between the approximate eigenvector from the ideal subspace with $A$ and the subspace actually used with $A+E$. However, we leave out the proof and discussion because they are similar to Theorem 2.1 in [24], adjusted for eigenvectors instead of linear equations.

THEOREM 4.7. *Suppose there are two Krylov subspaces* $K_1 = span\{y_2, Ay_2, \cdots, A^{m-k}y_2\}$ *and* $K_2 = span\{y_2, (A+E)y_2, \cdots, (A+E)^{m-k}y_2\}$ *with the perturbation matrix* $\|E\| = \epsilon$ *and* $\|y_2\| = 1$. *Let* $\delta = \epsilon$. *Let the curve* $\Gamma$ *be the boundary of a* $\delta$*-psedospectrum of* $A$. *If the best approximation of an eigenvector* $z$ *is* $\hat{y} = p(A)y_2$ *from* $K_1$, *where* $p$ *is a polynomial, then* $\tilde{y} = p(A+E)y_2$ *is an approximation of* $z$ *in* $K_2$ *with* $\|\hat{y} - \tilde{y}\| \leq (\frac{\epsilon}{\delta - \epsilon})(\frac{L_\delta}{2\pi\delta}) \max_{z \in \Gamma} |p(z)|$, *where* $L_\delta$ *is the arclength of* $\Gamma$.

**4.5. Examples and further analysis.** Here we look at the Arnoldi-E residual vectors and how their being nearly parallel affects convergence. In the first experiment, we fix the first Ritz vector $y_1$ as the starting vector for each cycle. The residual vector for $y_1$ is $r_1 = Ay_1 - \theta_1 y_1$. We let $r_1 = \gamma_1 w$, where $w$ is norm one. Then we look at the orthogonal projection of $r_2$ onto $w$: $r_2 = \gamma_2 w + f_2$. So we have, as in Theorem 4.6,

$$\begin{aligned}Ay_1 &= \theta_1 y_1 + \gamma_1 w,\\ Ay_2 &= \theta_2 y_2 + \gamma_2 w + f_2, \quad \text{where } f_2 \perp w.\end{aligned}$$

*Example 7.* We consider the test in Example 5 that had $y_1$ the starting vector for every cycle. The curve for $\|r_2\|$ in Figure 4.2 is the lowest solid line until it levels out

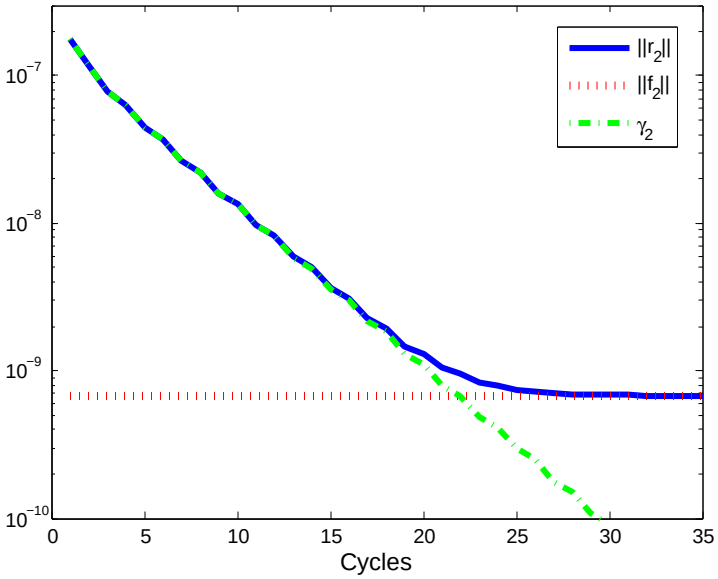


FIG. 4.4. *Breakdown of the residual vector for $y_2$ during the Arnoldi-E phase with $y_1$ as the starting vector for every cycle. Convergence stops when deviation from parallel becomes significantly larger than the rest of the residual. Matrix is symmetric from 1-D Laplacian of size $nfg = 1023$ and $ncg = 255$.*

and is passed by $\|r_1\|$. Figure 4.4 has this curve again along with curves for $\gamma_2$ and $||f_2||$. The $\|r_1\|$ curve keeps converging since $y_1$ is the starting vector for the Krylov portion of the Arnoldi-E subspace. But fortunately $y_2$ also converges for a while. According to Theorem 4.6, for polynomial $p$ of degree $m-k$, we have $p(A+E)y_2$ in the subspace, where $||E||$ is almost the same as $||f_2||$ in the symmetric case. So $y_2$ converges until its residual norm reaches the level of $||f_2||$.

We next give a simpler way than Theorem 4.6 to analyze the convergence. The simplest Arnoldi-E subspace with $y_1$ as starting vector is $S = Span\{y_1, Ay_1, y_2\}$. In order for this to be fully effective at improving $y_2$, it is needed that $Ay_2$ is also in this subspace. This is the case when the residual vectors are parallel, so when $f_2 = \vec{0}$. The $\gamma_2 w$ vector is contained in the subspace $S$ and can be thought of as a correction to the $\theta_2 y_2$ term since it will often be smaller. When $||f_2||$ is larger that $\gamma_2$, then it can wash out this correction. When $||f_2||$ is significantly smaller than $\gamma_2$, it should not have much effect, and the subspace will have an accurate approximation to $Ay_2$. In Figure 4.4, the $||r_2||$ curve starts to level out when $\gamma_2$ gets down near the level of $||f_2||$.

We continue this simple analysis with again the symmetric matrix of Example 5 and with rotating through the desired 10 Ritz vectors as starting vectors, as is done for the solid lines in Figure 4.1. However, this time we continue to rotate through all ten even after some have converged. The residual curves are shown in Figure 4.5, and because of this change, they are concave up while the ones in Figure 4.1 converge fairly consistently. We wish to look at why the convergence slows down as the iteration proceeds. For a particular cycle, let $y_j$ be the starting vector for the Krylov portion of the Arnoldi-E subspace where $j$ rotates from 1 to 10. Let the corresponding residual be $r_j = \gamma_j w$, with $||w|| = 1$. Let the orthogonal decomposition of $r_2$ from a projection

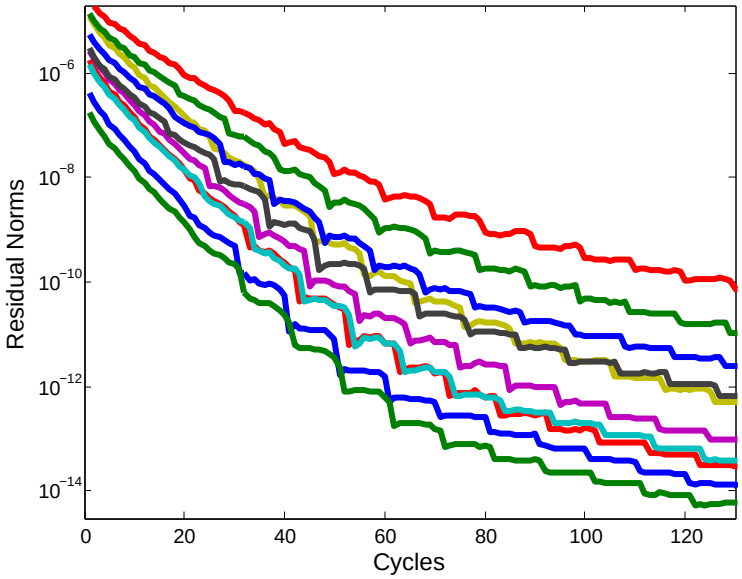


FIG. 4.5. *Convergence for the second Arnoldi-E phase of Two-grid Arnoldi. Starting vectors for Arnoldi-E are cycled through the 10 smallest Ritz vectors including converged ones. Convergence slows down as the iteration proceeds. Matrix is from 1-D Laplacian with* $nfg = 1023$. *Also* $ncg = 255$

onto $w$ be $r_2 = \gamma_2 w + f_2$. So we have

$$Ay_j = \theta_j y_j + \gamma_j w,$$
$$Ay_2 = \theta_2 y_2 + \gamma_2 w + f_2, \quad \text{where } f_2 \perp w.$$

The top part of Figure 4.6 has the residual norm curve for $y_2$ along with $\gamma_2$ and $||f_2||$. These quantities are computed at the end of each cycle, so a $||f_2||$ smaller than $\gamma_2$ helps on the next cycle. Every tenth cycle, $y_2$ is the starting vector in Arnoldi-E and so $||f_2||$ is zero. Initially with $\gamma_2$ well less than $||f_2||$, $y_2$ converges rapidly. Here $||r_2||$ does not stall out as it does in Figure 4.4, because every tenth cycle $y_2$ is the starting vector. Then the size of $f_2$ also is reduced, since it is one component of $r_2$. So $\gamma_2$ is able to stay around $||f_2||$ for a while. However, it slowly gets further below $||f_2||$ (except every tenth cycle) and so convergence slows. The bottom part of Figure 4.5 has a closeup from the top part for cycles 30 through 50. When $\gamma_2$ comes up near $||f_2||$, the residual improves for the next cycle.

Next we change the *rtol* in the intial phase of regular Arnoldi from $10^{-3}$ to $rtol = 10^{-8}$ and look at how this changes the convergence in the Arnoldi-E phase. The convergence for 20 cycles of Arnoldi-E is shown in the top of Figure 4.7. The vectors at the beginning of this phase are more accurate, because the first phase was run longer. This accuracy is now limited mainly by the transfer from coarse to fine grid, and so the deviation from parallel is generally larger than the rest of the residual. The bottom part of Figure 4.7 shows this for $y_8$. At the beginning, $\|f_8\|$ is much larger than $\gamma_8$ and there is no improvement in $y_8$ until it becomes the starting vector in cycle 8. During that cycle, the residual norm is reduced enough that $\gamma_8$ is near to $\|f_8\|$ and there is some slight improvement in $y_8$ during cycles 9 through 17.

Next we change the matrix to be nonsymmetric by adding a convection term $\beta = 102.8$. This changes the eigenvectors in that they are skewed somewhat in the

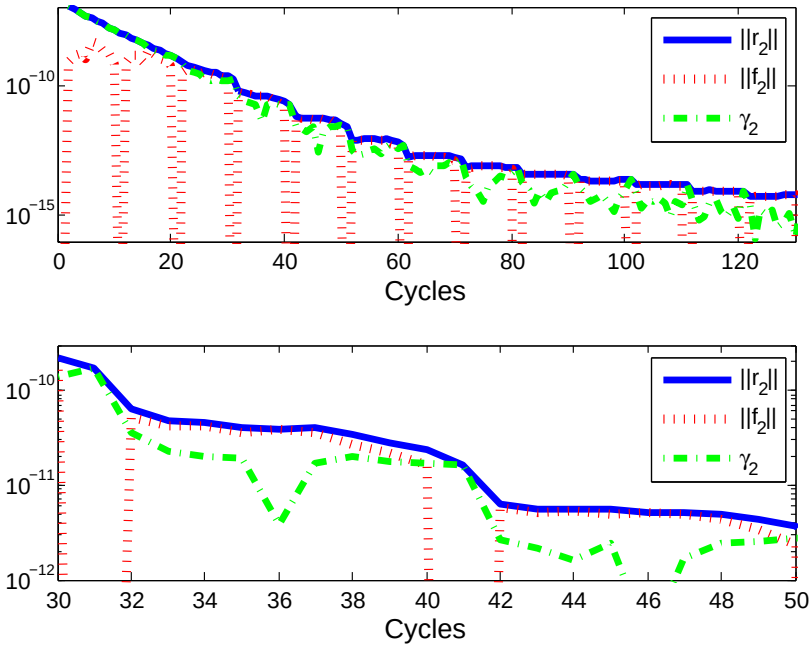


FIG. 4.6. *Breakdown of the residual vector for $y_2$ with starting vectors for Arnoldi-E cycled through 10 Ritz vectors. Convergence slows as the deviation from parallel generally becomes larger relative to the rest of the residual. There is less convergence when vectors other than $y_2$ are the starting vector. Matrix is symmetric from 1-D Laplacian. Lower graph is a closeup of a portion of the upper graph.*

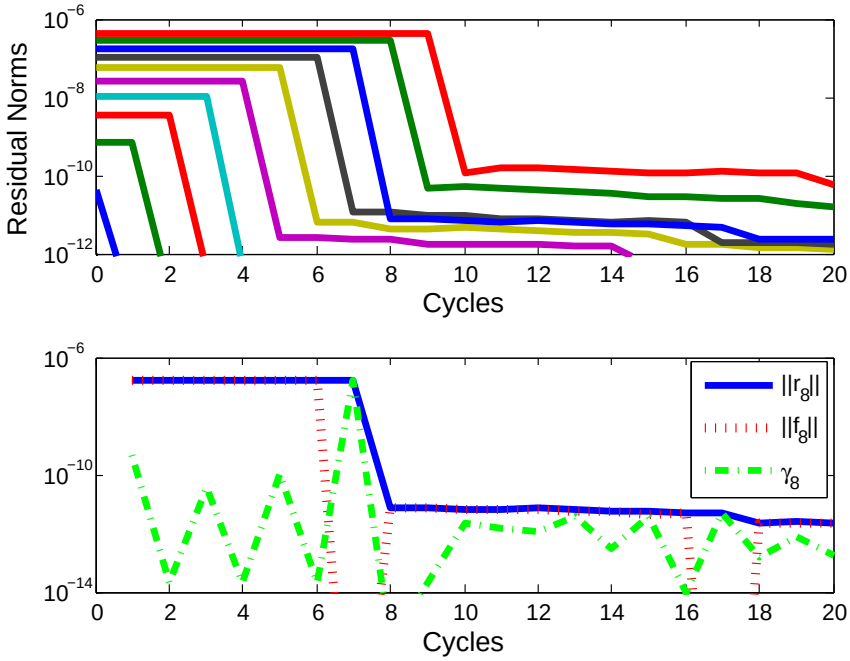


FIG. 4.7. *Top has convergence of Arnoldi-E with starting vectors cycled through 10 Ritz vectors and after a first phase with rtol $= 10^{-8}$. The matrix is from the 1-D Laplacian. The bottom graph has a breakdown of the residual vector for $y_8$. Initially there is no convergence because the deviation from parallel dominates the residual.*

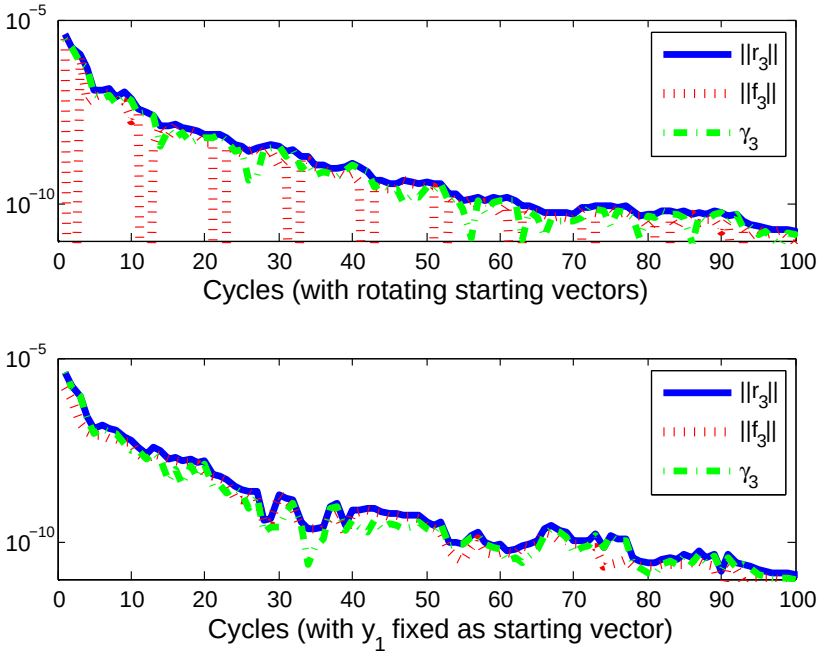


FIG. 4.8. *Breakdown of the residual vector for $y_3$. Matrix is nonsymmetric from the 1-D convection-diffusion equation with $\beta = 102.4$. The top figure has rotating through all 10 Ritz vectors and the bottom keeps $y_1$ as the starting vector. It is notable that even with $y_1$ as starting vector, the deviation from parallel keeps going down, and the residual for $y_3$ keeps improving.*

same direction. We still have $ncg = 255$ and $nfg = 1023$ and $10^{-8}$ for the first phase $rtol$, and we cycle through 10 $y_j$'s as starting vectors, even if converged. The convergence of the third residual vector is given in Figure 4.8 with rotating through starting vectors on top and fixing $y_1$ as starting vector on the bottom. The behavior is very different than in the symmetric case. The $\|f_3\|$ term showing deviation from parallel residuals stays around the same size as the $\gamma_3$ term. Convergence of $y_3$ is irregular but improves not just when it is the starting vector. Even when $y_1$ is always the starting vector, $y_3$ keeps improving. Unlike in Figure 4.4, $\|f_3\|$ keeps being reduced. Perhaps this happens because the eigenvectors are related to each other instead of being orthogonal (see Figure 12.3 of [30]). Pushing toward $y_1$ is also partially going toward $y_3$.

**5. Multiple Grids for Arnoldi.** We now investigate whether using more than two grids may be an improvement. More grid levels makes it possible to get from grid to grid with smaller changes in grid size.

As before, the problem size is $nfg$. We let $nl$ be the number of grid levels used, with grid level 1 being the coarsest and grid level $nl$ being the finest one corresponding to matris size $nfg$. We now give the algorithm. Basically it is the Two-grid Arnoldi method, except we repeat Steps 2 and 3 for each of grid levels 2 through $nl$, from second coarsest up to finest grid.

**MULTIPLE-GRID ARNOLDI**

0. **Initial Setup:**
   Let the problem size be $nfg$.
   Choose the grid levels. Let $nl$ be the number of grids ordered from coarsest

TABLE 5.1
*Two-grid Arnoldi vs. Multiple-grid Arnoldi. Matrix is dimension $n = 4095$ from 1-D Conv-diff with $\beta = 51.2$.*

| Coarsest grid matrix size | 2047 | 1023 | 511 | 255 | 127 | 63 | 31 |
|---|---|---|---|---|---|---|---|
| Two-grid Arnoldi cycle equiv's | 227 | 50.8 | 56.4 | 55.7 | 108 | 728 | 514 |
| Multiple-grid Arn. cycle equiv's | 227 | 41.8 | 15.6 | 9.56 | 11.9 | 9.86 | 10.1 |

to finest.
Choose $m$ = the maximum subspace size,
$k$ = the number of Ritz vectors retained at the restart,
$nev$ = the number of desired eigenpairs,
$rtol$ = the residual norm tolerance.

1. **Coarsest Grid Computation:**
   Run restarted Arnoldi(m,k) on the coarsest grid until the $nev$ smallest magnitude eigenvalues have converged to $rtol$.
2. **For** $grid\ level = 2 \dots nl$ :
   A. **Move to next finer grid:**
      Move the $k$ current Ritz vectors to the next finer grid (we use spline interpolation).
   B. **Finer grid computation:**
      Improve the approximate eigenvectors on the finer grid with the Arnoldi-E(m,k) method. For the starting vectors for the Krylov portion of each cycle, we alternate through the Ritz vectors $y_1$ through $y_{nev}$. However, converged Ritz vectors are skipped. Also, if there are complex vectors, they are split into real and imaginary parts. Stop when the $nev$ smallest Ritz pairs reach residual norms below $rtol$. Then done if this is the finest grid, otherwise go back to A.

In the examples that follow, we are in 1-D, and we let the decreasing sizes of the matrices be $nfg$, $\frac{nfg+1}{2} - 1$, $\frac{nfg+1}{4} - 1$, ..., $\frac{nfg+1}{2^{nl}} - 1$. Other choices are possible, such as skipping some levels.

*Example 8.* We return to a matrix from the 1-D convection-diffusion equation, but now with convection of $\beta = 51.2$. The size is $nfg = 4095$. Standard Arnoldi(30,15) takes 1574 cycles for 10 Ritz pairs to converge to residual norm below $10^{-8}$. Table 5.1 has the results with different choices of coarsest grid. As mentioned above, the number of subintervals in the grid is increased by a factor of 2 at each new grid level. The Multiple-grid Arnoldi result with coarsest grid of 2047 uses only two grids, while with coarsest of 31, there are eight grid levels. The best Two-grid Arnoldi(30,15) result is 50.75 fine-grid-equivalent cycles with $ncg = 1023$. With Multiple-grid Arnoldi, we can get below 10 fine-grid-equivalents. So while Two-grid improves by a factor of 30 compared to regular Arnoldi, Multiple-grid can be over 150 times better than regular Arnoldi. We are getting the significant speedup that is characteristic of multigrid methods for linear equations on problems with low convection, but here the convection is higher. Multiple-grid Arnoldi also is very consistent for the choice of smallest matrix from size 31 up to 255. Two-grid is consistent for coarse grid matrix of size 255 up to 1023, but struggles with smaller ones.

We next give an example for which Multiple-grid Arnoldi does not work as well.

*Example 9.* As in the previous example, we have a matrix from the 1-D convection-diffusion equation, but the convection is increased to $\beta = 102.4$ and the size of the

TABLE 5.2
*Two-grid Arnoldi vs. Multiple-grid Arnoldi. Matrix is dimension* $n = 1023$ *from 1-D Conv-diff with* $\beta = 102.4$.

| Coarsest grid matrix size | 511 | 255 | 127 | 63 | 31 |
|---|---|---|---|---|---|
| Two-grid Arnoldi cycle equiv's | 47 | 34.8 | 71.9 | 73.6 | 513 |
| Multiple-grid Arnoldi cycle equiv's | 47 | 39.8 | 55.9 | 72.1 | 264 |

matrix is reduced to $nfg = 1023$. We use Arnoldi(30,16) since the matrix is more non-normal. Standard Arnoldi(30,16) takes 109 cycles for 10 Ritz pairs to converge to residual norm below $10^{-8}$. Table 5.2 has the results with different choices of coarsest grid and increasing the number of subintervals in the grid by a factor of 2 at each new phase. Multiple-grid Arnoldi beats Two-grid on some of the choices, but not by as much as in the previous example. The important thing to note is that using too small of a coarsest grid can make things worse. For coarsest grid of size 31, the multiple-grid method takes over twice as long as regular Arnoldi. The method is not as effective as in the previous example, because approximations from a coarse grid to the next are not as accurate with the increased convection. Also, because the matrix is smaller, there is not the same opportunity versus regular Arnoldi. The finer grids are missing which are difficult for regular Arnoldi and for which approximations from the next coarser grid are particulary accurate.

*Example 10.* For this example we go back to the larger fine grid matrix of size $n = 4095$. The convection is varied to see when Multiple-grid Arnoldi is better. Figure 5.1 has convection coefficients from $\beta = 0$ up to $\beta = 204.8$. Multiple-grid Arnoldi is implemented with coarsest grid of size 255 and number of subintervals increasing by a factor of two for each finer grid. Compared with this method is regular Arnoldi and two types of Two-grid Arnoldi, first with coarse grid matrix of size $ncg = 1023$ and then with $ncg = 255$. For the case of a symmetric matrix, $\beta = 0$, both Multiple-grid Arnoldi and Two-grid Arnoldi with $ncg = 255$ are very efficient with only about four fine-grid equivalent cycles. For moderate values of convection, there is a sweet spot for Multiple-grid Arnoldi as unlike Two-grid, it is still very efficient. Then for $\beta = 204.8$, Multiple-grid is not optimal.

**6. Conclusion.** Eigenvalue problems from differential equations are easier for smaller grids, because smaller matrices mean less work per iteration, but also the spectrum is easier. Here we gave Krylov methods that can use the power of coarse grids for eigenvalue problems. On the coarsest grid, standard Arnoldi is applied, then on finer grids, the Arnoldi-E method allows inputed approximate eigenvectors. This can significantly improve upon regular Arnoldi, especially for problems with very fine grids. Compared to traditional multigrid, the use of Krylov subspaces makes this approach more robust. It was shown that effectiveness can be explained with near-Krylov properties of the approximate eigenvectors that are passed from coarse to fine grids.

This approach can be implemented with just two grids or also using a sequence of increasingly finer grids. For moderate convection in a simple convection-diffusion equation, multiple grids were better than just two. This should be studied more, but the best method is likely problem dependent.

Future work should include investigating the choice of starting vector for the cycles of Arnoldi-E and why splitting complex Ritz vectors is effective. Also, an algebraic multigrid version of this work should be developed. We plan to work on

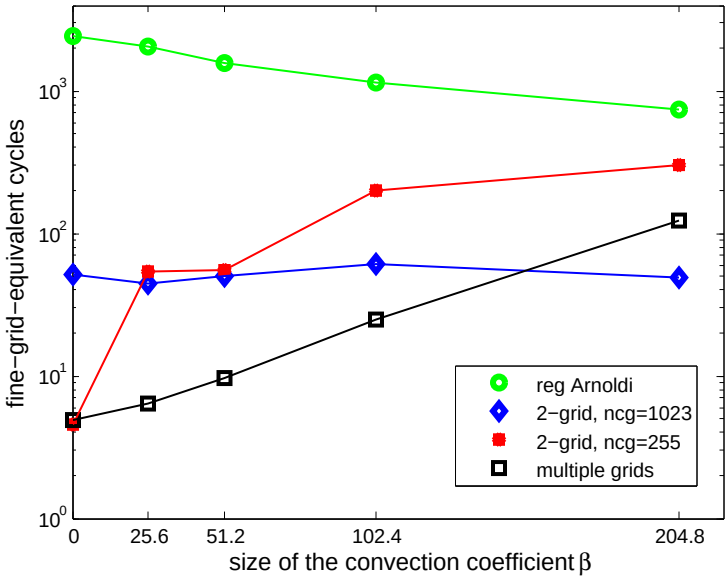


FIG. 5.1. *Multiple-grid Arnoldi with coarsest grid of size 255 compared to Two-grid Arnoldi and regular Arnoldi. The convection coeffient varies from 0 to 204.8. Fine grid matrix size is* $n = 4095$.

multigrid deflation of eigenvalues for linear equations, with eigenvectors computed on coarse grids used to improve convergence during the linear equations solution.

## REFERENCES


[1] A. M. Abdel-Rehim, R. B. Morgan, D. A. Nicely, and W. Wilcox. Deflated and restarted symmetric Lanczos methods for eigenvalues and linear equations with multiple right-hand sides. *SIAM J. Sci. Comput.*, 32:129–149, 2010.

[2] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.

[3] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Math. Comp.*, 31:333–390, 1977.

[4] A. Brandt, S. McCormick, and J. Ruge. Multigrid methods for differential eigenproblems. *SIAM J. Sci. Statist. Comput.*, 4:244–260, 1983.

[5] S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods.* Springer-Verlag, New York, NY, 1994.

[6] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial, 2nd Edition.* SIAM, Philadelphia, PA, USA, 2000.

[7] R. P. Fedorenko. The speed of convergence of one iterative process. *USSR Comput. Math. Math. Phys.*, 4:227–235, 1964.

[8] N. J. Higham. *Functions of Matrices: Theory and Computation.* SIAM, Philadelphia, 2008.

[9] T. Kato. *Perturbation Theory for Linear Operators, 2nd ed.* Springer-Verlag, Berlin, 1980.

[10] C. Lanczos. An iterative method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards*, 45:255–282, 1950.

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

[12] Q. Lin and H. Xie. A multi-level correction scheme for eigenvalue problems. *Math. Comp.*, 84:71–88, 2015.

[13] J. Mandel and S. McCormick. A multilevel variational method for Au = $\lambda$ Bu on composite grids. *J. Comput. Phys.*, 80:442–452, 1989.

[14] R. B. Morgan. Computing interior eigenvalues of large matrices. *Linear Algebra Appl.*, 154-156:289–309, 1991.

[15] R. B. Morgan. On restarting the Arnoldi method for large nonsymmetric eigenvalue problems.

*Math. Comp.*, 65:1213–1230, 1996.
[16] R. B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM J. Matrix Anal. Appl.*, 21:1112–1135, 2000.
[17] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 24:20–37, 2002.
[18] R. B. Morgan and M. Zeng. Harmonic projection methods for large non-symmetric eigenvalue problems. *Numer. Linear Algebra Appl.*, 5:33–55, 1998.
[19] R. B. Morgan and M. Zeng. A harmonic restarted Arnoldi algorithm for calculating eigenvalues and determining multiplicity. *Linear Algebra Appl.*, 415:96–113, 2006.
[20] C. C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Num. Lin. Alg. with Appl.*, 2:115–133, 1995.
[21] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, N.J., 1980.
[22] Y. Saad. Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices. *Linear Algebra Appl.*, 34:269–295, 1980.
[23] Y. Saad. *Numerical Methods for Large Eigenvalue Problems, 2nd Edition*. SIAM, Philadelphia, PA, 2011.
[24] J. A. Sifuentes, M. Embree, and R. B. Morgan. GMRES convergence for perturbed coefficient matrices, with application to approximate deflation preconditioning. *SIAM J. Matrix Anal. Appl.*, 34:1066–1088, 2013.
[25] V. Simoncini and D. B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM J. Sci. Comput.*, 25:454–477, 2003.
[26] D. C. Sorensen. Implicit application of polynomial filters in a $k$-step Arnoldi method. *SIAM J. Matrix Anal. Appl.*, 13:357–385, 1992.
[27] A. Stathopoulos, Y. Saad, and K. Wu. Dynamic thick restarting of the Davidson, and the implicitly restarted Arnoldi methods. *SIAM J. Sci. Comput.*, 19:227–245, 1998.
[28] G. W. Stewart. A Krylov–Schur algorithm for large eigenproblems. *SIAM J. Matrix Anal. Appl.*, 23:601 – 614, 2001.
[29] G. W. Stewart. Backward error bounds for approximate Krylov subspaces. *Linear Algebra Appl.*, 340:81–86, 2002.
[30] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton Univ. Press, Princeton, NJ, 2005.
[31] U. Trottenberg, C. W. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, London, 2001.
[32] J. van den Eshof and G. L.G. Sleijpen. Inexact Krylov subspace methods for linear systems. *SIAM J. Matrix Anal. Appl.*, 26:125–153, 2004.
[33] K. Wu and H. Simon. Thick-restart Lanczos method for symmetric eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 22:602 – 616, 2000.
[34] J. Xu and A. Zhou. A two-grid discretization scheme for eigenvalue problems. *Math. Comp.*, 70:1725, 2001.
[35] Z. Yang. *A Multigrid Krylov Method for Eigenvalue Problems*. PhD thesis, Baylor University, Waco, Texas, 2015.