SIAM J. SCI. COMPUT.
Vol. 40, No. 5, pp. A3470–A3494

TWO-GRID AND MULTIPLE-GRID ARNOLDI FOR EIGENVALUES*

RONALD B. MORGAN† AND ZHAO YANG†

Abstract. A new approach is given for computing eigenvalues and eigenvectors of large matrices. Multiple grids are combined with the Arnoldi method in order to solve difficult problems. First, a two-grid method computes eigenvectors 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 two-grid 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 on the fine grid in spite of having many initial guess vectors. Near-Krylov theory is developed for this. Finally, this new method is generalized for more grid levels to produce a multiple-grid Arnoldi method.

Key words. eigenvalues, Arnoldi, multigrid, Krylov

AMS subject classifications. 65F15, 15A18

DOI. 10.1137/16M1062260

1. Introduction.

We look at computing eigenvalues and eigenvectors of large, possibly nonsymmetric, matrices. Arnoldi methods [2, 21, 25, 14, 10, 18, 22] are Krylov subspace methods for nonsymmetric eigenvalue problems. Sorensen's implicitly restarted Arnoldi method [25] 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 multiple grid levels. 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 eigenpairs. Such an approach is proposed here.

In its simplest form, this new method uses two grids. Eigenvectors are computed on the coarse grid using standard Arnoldi, then interpolated to the fine grid and improved there with an Arnoldi method that can accept initial approximate eigenvectors. Examples show this approach can dramatically improve eigenvalue calculations. However, this method is best when eigenvectors from a coarse grid are fairly accurate on the fine grid. The method may not be useful if the fine grid is not very fine or if highly accurate eigenvectors must be computed. Near-Krylov theory is developed to explain the effectiveness of this method. Finally, a more general method is given that uses multiple 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.

*Submitted to the journal's Methods and Algorithms for Scientific Computing section February 19, 2016; accepted for publication (in revised form) July 26, 2018; published electronically October 16, 2018.

http://www.siam.org/journals/sisc/40-5/M106226.html

Funding: The first author's work was supported by the National Science Foundation under grant DMS-1418677.

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

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 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, 30]. Sometimes multigrid is very effective and solves linear equations much faster than other methods. Much of the work is done on coarse grids where the problem is easier and the iteration is cheaper. However, multigrid methods do not work well for some 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 a large enough wave number.

2.2. Eigenvalue multigrid. Multigrid methods have been proposed for eigenvalue problems; see, for example, [4, 12, 33, 8, 11]. Multigrid eigenvalue methods generally have the same limitations as just mentioned for linear equations multigrid. Another disadvantage for some methods is that they compute one eigenvalue at a time. However, other methods do compute several eigenvalues simultaneously. For example, in [4] the simultaneous iteration block approach is used. The multigrid preconditioned LOBPCG method in [8] also has blocks. Shift-and-invert Arnoldi [22] with multigrid to solve the linear equations finds multiple eigenvalues with just one Krylov subspace. However, it can find eigenvalues only near a chosen shift.

2.3. Krylov methods for eigenvalues. For symmetric matrices, a Krylov subspace with Rayleigh–Ritz extraction becomes the Lanczos method [9, 20, 32, 1]. For nonsymmetric matrices, we get the Arnoldi algorithm [2, 21, 25, 14, 26, 10, 18, 22]. Arnoldi generally needs to be restarted to control orthogonalization expense and storage. Sorensen's implicitly restarted Arnoldi [25] has 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.

THEOREM 2.1 (see [14]). *For each y_j, subspace* (2.1) *is equivalent to*

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

So the subspace contains a Krylov subspace with each Ritz vector as starting vector. Another result needed later is that a subspace being Krylov is equivalent to having parallel residual vectors [15, 27, 16]. It is well known that a Krylov subspace gives parallel residuals. For Arnoldi, the residual vector for the Ritz pair (θ_i, y_i), $r_i \equiv Ay_i - \theta_i y_i$, has $r_i = \gamma_i v_{m+1}$. This goes in the opposite direction as well.

THEOREM 2.2 (see [16]). *With pairs (θ_i, y_i) and $r_i = Ay_i - \theta_i y_i$, if $r_i = \gamma_i w$, for some vector w and for each $i = 1, \ldots, k$, then the subspace $Span\{y_1, y_2, \ldots, y_k, w\}$ is Krylov.*

Another method using subspace (2.2) is given in [14]. 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. Then the method is not equivalent to implicitly restarted Arnoldi, and the choice of 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 [14].

Arnoldi-E(m,k).

1. **Start:** Choose m, the max size of the subspace, and k, the number of approximate eigenvectors that are retained from one cycle to the next. Pick nev, the desired number of eigenpairs, and $rtol$, the convergence tolerance. Normally $nev < k$. Choose 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 Rayleigh–Ritz reduced matrix H is more expensive to compute in Arnoldi-E than for regular Arnoldi and there is some choice in implementing it. As with implicitly restarted Arnoldi, only $m-k$ matrix-vector products are required for a cycle [14], though for quite sparse matrices it is simpler to use m of them. Arnoldi-E will be used in our new method, as will the Arnoldi method that will be mentioned next.

Wu and Simon [32] (see also [1]) give the symmetric case of a method called thick-restart. It uses the same subspace as Arnoldi-E but puts the approximate eigenvectors at the beginning of the subspace instead of the end. Stewart gives a more general method, Krylov–Schur, in [27]. A nonsymmetric version of Wu and Simon's method is given in [18]. This is the Arnoldi algorithm used in experiments.

Also, full reorthogonalization in the Arnoldi iteration is used in all experiments here. We count cycles of Arnoldi-E and restarted Arnoldi as if they are the same cost. This is approximately true for not very sparse matrices, but for our tests with very sparse matrices (and with $m = 30$, $k = 15$), a cycle of Arnoldi-E costs roughly 70% more.

3. Two-grid Arnoldi. 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. We will use the Arnoldi-E method, since it can accept initial inputed vectors. 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 n_f, meaning the fine grid matrix is n_f by n_f.
   Choose the coarse grid size n_c.

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 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, complex Arnoldi-E Ritz vectors are split into real and imaginary parts to keep the subspace real. It seems this might degrade performance, because the starting vector for the Krylov portion of the subspace is not a Ritz vector. However, the results seem good. There is a comparison at the end of Example 3.

Example 1. We consider a matrix from finite difference discretization of the two-dimensional (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. First, the restarted Arnoldi(m,k) method is run with the goal of finding the 10 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 dotted lines in the top part of Figure 3.1 show these residual norms. One reason for the somewhat erratic convergence is the presence of the nearly multiple eigenvalues.

Next, we apply the two-grid Arnoldi approach. For the coarse grid, we use a discretization size of $h = \frac{1}{350}$, so the number of grid points and dimension of the matrix are $349^2 = 121{,}801$. 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. Next, the Arnoldi-E method on the fine grid needs only 51 cycles for the eigenvectors to reach the desired level. The coarse grid is a good enough approximation to the fine 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 fine grid cycles. This gives 217 fine-grid-equivalent cycles compared to the 2375 for regular Arnoldi. The solid lines in the top part of Figure 3.1 show the two-grid Arnoldi residual norms with coarse grid convergence scaled by one-fourth followed by fine grid results. The bottom part of the figure has a close-up. Note the coarse grid Arnoldi converges at cycle 166 and the fine grid convergence is from cycle

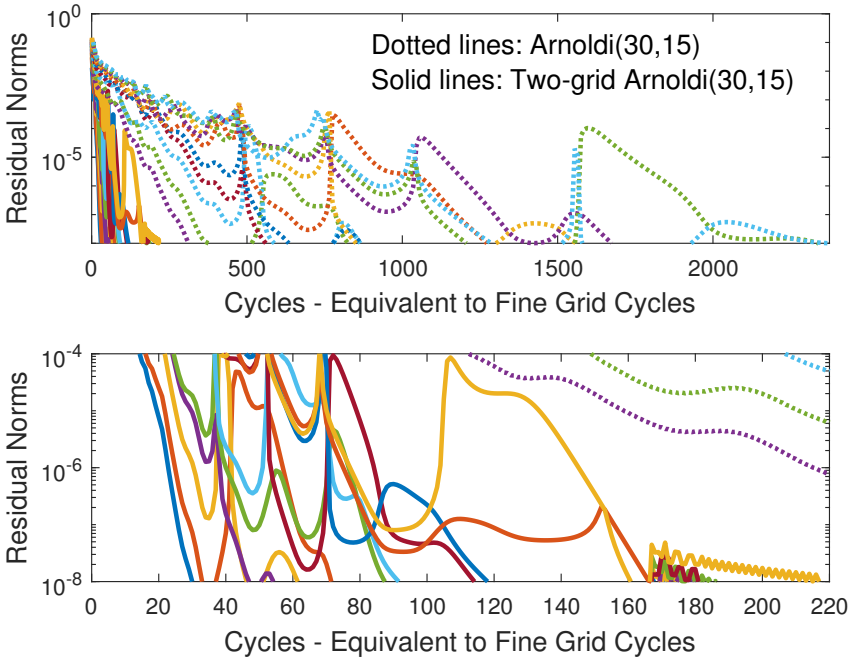

FIG. 3.1. *The top plot has standard Arnoldi compared to two-grid Arnoldi. The fine grid matrix size is* $n = 488{,}601$ *and the coarse grid matrix size is* $n = 121{,}801$. *The bottom plot has a close-up view of a portion of the top.*

167 to 217. Two-grid Arnoldi has the advantage of dealing with the nearly multiple eigenvalues on the coarse grid. If higher accuracy is needed, the two-grid approach does not work as well. For example, if the *rtol* is set to 10^{-10}, then regular Arnoldi uses 2877 cycles and two-grid takes 1547 fine-grid-equivalent cycles (747 on the coarse grid and 1360 on the fine grid).

We now consider the same fine grid eigenvalue problem but compare different coarse grid sizes. 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 $n_c = 121{,}801$, $174^2 = 30{,}276$, and $87^2 = 7569$. The solid lines on the figure are for $n_c = 121{,}801$ and are the same as the solid lines in the previous figure. 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. The best choice of coarse grid changes depending on the desired accuracy. Table 3.1 has on its second line the accuracy of the approximations on the fine grid that come from the coarse grid (the max of the 10 residual norms of the desired eigenvectors). Then the table 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; there is danger of missing some of the nearly multiple eigenvalues if the coarse grid is not solved accurately. For residual norms below 10^{-8} on the fine grid, the coarse grid $n_c = 121{,}801$ is best. For 10^{-7}, $n_c = 30{,}276$ is best. Then for residual norms below 10^{-6}, $n_c = 7569$ should be used. It needs only two fine-grid-equivalent cycles versus 1865 cycles for regular Arnoldi.

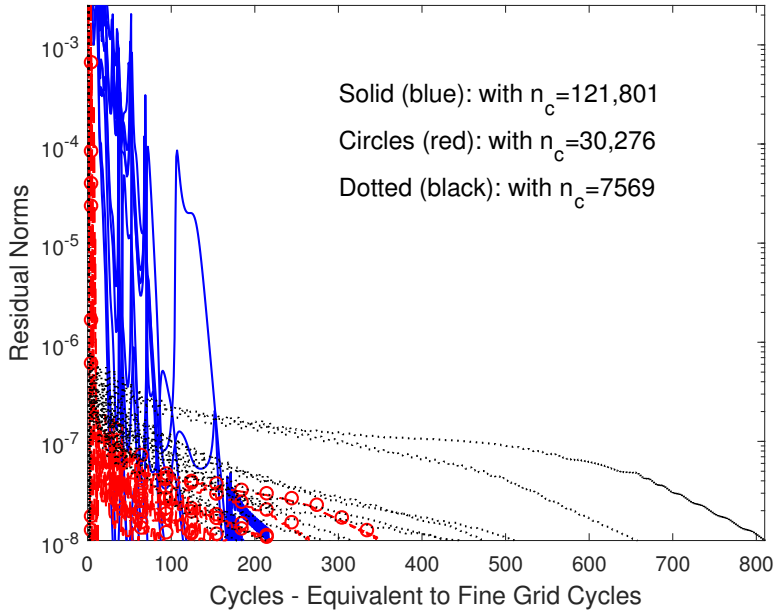

FIG. 3.2. *Two-grid Arnoldi with different coarse grids. Fine grid matrix size is* $n = 488{,}601$ *and coarse grid matrix sizes are* $n = 121{,}801$*,* $n = 30{,}276$*, and* $n = 7569$*.*

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.

		$n_c = 121{,}801$	$n_c = 30{,}276$	$n_c = 7569$
	Accur. of approx's	$4.4 * 10^{-8}$	$2.1 * 10^{-7}$	$7.8 * 10^{-7}$
rtol	reg. Arn. cyc's	fg-equiv. cycles	fg-equiv. cycles	fg-equiv. cycles
10^{-6}	1865	167	12	2
10^{-7}	1955	167	19	452
10^{-8}	2375	217	349	811

This last discussion points out that using two-grid Arnoldi involves an optimization problem of determining the best coarse grid based on two things: first, the accuracy of the approximations that come from the coarse grid to the fine grid, and second, the desired *rtol* on the fine grid. For higher accuracy on the fine grid, a less coarse first grid is likely to be needed.

We next compare two-grid Arnoldi with two existing multigrid methods.

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 1-D 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 $n_c = 255$ for two-grid Arnoldi, 7 of the 10 desired eigenpairs are already converged when the move is made from the coarse to the fine grid. It then takes three cycles on the fine grid to refine the other three. The number of fine-grid-equivalent cycles used is $\frac{24}{16}$ on the

TABLE 3.2
Two-grid Arnoldi versus inverted Arnoldi and LOBPCG with multigrid. Matrix is dimension $n = 4095$ *from* 1*-D Laplacian.*

Two-grid Arnoldi	Coarse grid	cg cycles	fg cycles	mvp equiv's	Time
	127	11	8	158	0.45
	255	24	3	95	0.45
	511	64	0	132	0.58
	1023	203	0	775	1.53
Inverted Arnoldi	rtol for lin. eq's	Arnoldi iter's	mvp equiv's per solve (average)	mvp equiv's total	Time
	1.e-6 & decr.	27	29.8	840	0.57
LOBPCG	rtol for lin. eq's	Precond calls	mvp equiv's per solve (average)	mvp equiv's total	Time
	1.e-1	85	4.4	390	0.58

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 against inverted Arnoldi (using operator A^{-1}) with standard multigrid used to solve the linear equations. This is implemented with V-cycles and one Jacobi relaxation weighted by $\frac{2}{3}$ on each grid. Interpolation and restriction between grids is linear. The first 15 linear equations are solved to residual norm below 10^{-6}. Then for the other systems, the variable tolerance [24] $max\{10^{-6}/rnmax, 10^{-1}\}$ is used, where $rnmax$ is the maximum residual norm of the 10 desired eigenpairs. The outer Arnoldi loop requires only 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 (for Arnoldi-E, we could instead use a shortcut residual formula [14]). A total of 840 fine-grid-equivalent matrix-vector products are needed for the entire inverted Arnoldi process. This is far less than regular Arnoldi but more than two-grid Arnoldi. With coarse grid of size $n = 255$, two-grid has 95 fine-grid-equivalent matrix-vector products. The cpu times are similar to inverted Arnoldi (timings are the average of five runs) for the smaller coarse grid sizes in the table. The time is not improved like the matrix-vector products, because with this sparse matrix, the greater orthogonalization expense of two-grid Arnoldi is significant.

Now we compare to the LOBPCG method with multigrid preconditioning [8] which is for symmetric problems. Accurate multigrid linear equation solves are not needed for the preconditioning and the residual tolerance is set to 10^{-1}. More linear equations are solved than for inverted Arnoldi, but there are fewer total sweeps. The cpu time is about the same. For comparison with a large matrix, see Example 11.

Next we make the matrix nonsymmetric and see that standard multigrid is not robust. We use the convection-diffusion equation $-u'' + \beta u' = \lambda u$ and again discretize for $n = 4095$. As β is increased, around $\beta = 20$ 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 $\beta = 25$ is similar to $\beta = 0$.

We examine why multigrid has problems. The 3 by 3 unweighted Jacobi iteration matrix from the grid with $h = \frac{1}{4}$ has real eigenvalues for $\beta = 0$. The eigenvalues move together as β increases and then after $\beta = 8$ move out on the imaginary axis. By

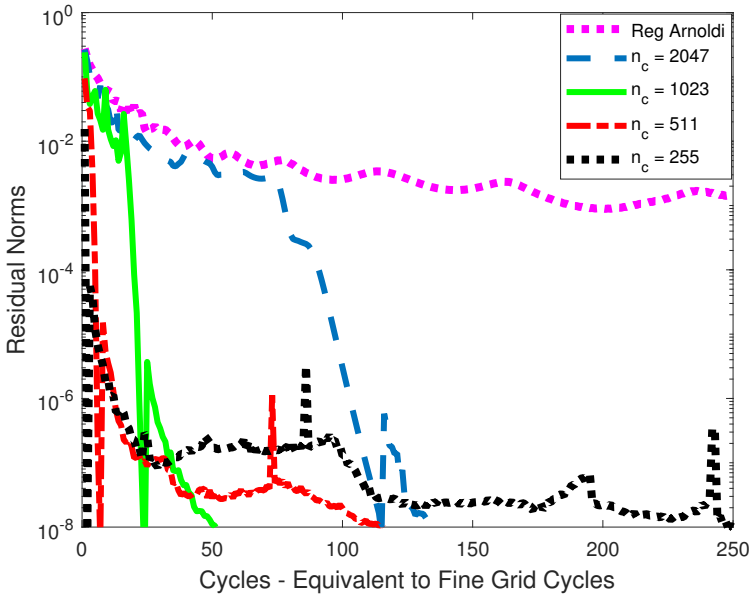

FIG. 3.3. *Two-grid Arnoldi with varying coarse grid sizes for* 1*-D convection-diffusion problem with convection coefficient of* 204.8*. Fine grid matrix size is* $n = 4095$ *and coarse grid matrices range from* $n_c = 255$ *up to* $n_c = 2047$*. The residual norm for the* 10*th eigenvalue is shown.*

$\beta = 14$, two eigenvalues have passed 1 in magnitude. For the weighted Jacobi used above, the eigenvalues are not purely imaginary, but they pass magnitude 1 at about $\beta = 18$. For finer grids, this problem does not occur until β is larger. For instance, the largest eigenvalues of the 7 by 7 weighted Jacobi iteration matrix pass magnitude 1 about $\beta = 30$. This suggests that this multigrid iteration can be more reliable if some of the coarsest grids are left out of the iteration. However, in Example 4, multigrid cannot be fixed by restricting to only the finer grids.

Example 3. Now we try a much larger β in order to show the robustness of the two-grid method. We use the convection-diffusion equation as in the previous example, but increase β 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 3.3 shows the residual norm convergence with several choices of n_c for the complex 10th eigenvalue. Regular Arnoldi requires 741 cycles and two-grid Arnoldi with $n_c = 1023$ uses just 51 fine-grid-equivalent cycles (95 coarse grid cycles and 27 on the fine grid). For this example, there is sometimes erratic convergence due to the high nonnormality.

As mentioned in the two-grid Arnoldi algorithm, splitting complex Ritz vectors into real and imaginary parts in the Arnoldi-E portion avoids complex arithmetic. We quickly compare with not splitting. With $n_c = 1023$, 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.

Example 4. We use the 1-D Helmholtz problem $-u'' - 40{,}000u = \lambda u$ with zero boundary conditions. The fine grid matrix of size $n = 1023$ has 63 negative eigen-

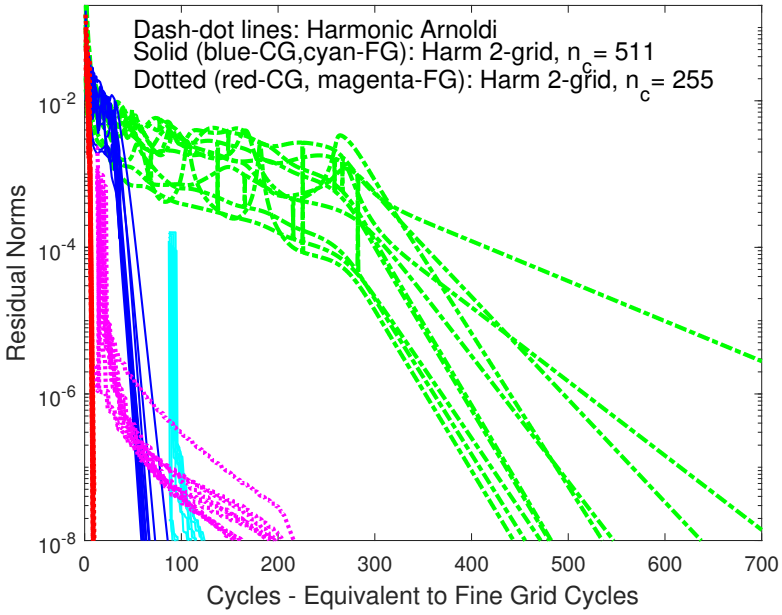

FIG. 3.4. *Standard Arnoldi compared to two-grid Arnoldi for a* 1*-D simple Helmholtz matrix. Fine grid matrix size is* $n = 1023$ *and coarse grid matrix sizes are* $n = 511$ *and* $n = 255$.

values. We compute the 10 eigenvalues closest to the origin, so this is an interior eigenvalue problem. Thus harmonic Rayleigh–Ritz [13, 19] is more reliable. We switch to harmonic restarted Arnoldi [18] followed by harmonic Arnoldi-E [17]. Figure 3.4 has harmonic Arnoldi, which takes 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 nonharmonic, which takes 3058 cycles and misses two of the 10 smallest). In two tests of harmonic two-grid Arnoldi, both find all of the 10 smallest eigenvalues. With $n_c = 511$, 124 fine-grid-equivalent cycles are used, and 217 for $n_c = 255$.

Care must be taken because for a larger shift of the Laplacian (a higher wave number) and too coarse a grid, the eigenvectors for the smallest magnitude eigenvalues may not correspond to the eigenvectors on the fine grid. In fact, this happens even for the current wave number if a coarse grid of size $n_c = 127$ is used. The vectors moved from the coarse grid are not good approximations to the desired fine grid eigenvectors, and the convergence is then slow on the fine grid (1147 cycles) and misses 4 of the smallest 10 eigenvalues.

4. Fine grid convergence theory. Arnoldi-E with inputed initial vectors in two-grid Arnoldi is not a Krylov subspace. In this section we look at why sometimes the fine grid convergence is nevertheless similar to that of standard Arnoldi.

4.1. Special properties of vectors from the coarse grid. The Arnoldi-E method does not always do as well at improving approximate eigenvectors as it does for the two-grid method. The next example shows this.

Example 5. We use a symmetric matrix of size $n = 1023$ from the 1-D Laplacian. We compare Arnoldi-E for improving approximate eigenvectors that come from a

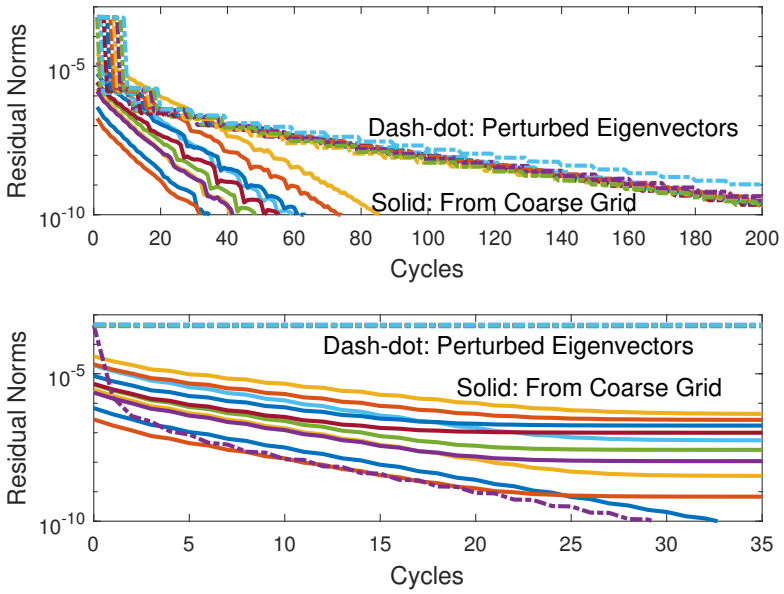

FIG. 4.1. *Top: Comparison of Arnoldi-E convergence with approximate eigenvectors from coarse grid versus from perturbation of true eigenvectors. 1-D Laplacian with matrix of size* $n = 1047$ *and coarse grid matrix of size* $n_c = 255$. *Bottom: Now the first Ritz vector is the starting vector for every cycle.*

coarse grid versus from random perturbation of the true fine grid eigenvectors. The coarse grid calculation has *rtol* of only 10^{-3} and after interpolation their initial fine-grid residual norms range from $6.8{*}10^{-7}$ to $3.9{*}10^{-5}$. The *rtol* for the second phase is 10^{-10}. Second, the eigenvectors are perturbed with random vectors normed to $2{*}10^{-4}$. The top portion of Figure 4.1 shows that the eventual convergence rate is almost five times faster for improving the vectors derived from the coarse grid. If one looks closely at the graph, each perturbed eigenvector improves only every 10th 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. Perturbing the fine grid eigenvectors with continuous functions (combinations of exponentials and sines) gives similar results to those from the random perturbations.

Next we fix the starting vector for each Arnoldi-E cycle to be the first Ritz vector y_1. The bottom portion of Figure 4.1 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 convergence curves flatten out except for y_1.

We conclude that approximate eigenvectors from the coarse grid have a property that makes them work better in Arnoldi-E. We will characterize this property as being near-Krylov. Previous work on inexact Krylov [24, 31] 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 a theorem about convergence for near-Krylov subspaces. Finally, some examples are given of how near-Krylov can help.

4.2. Near-Krylov and nearly parallel residuals. 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 [27] and near-Krylov in [28].

DEFINITION 4.1. *For a subspace of dimension m, let the Ritz values be θ_i for $i = 1, \ldots, m$, 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 $Y = [y_1, y_2, \ldots, y_m]$, Θ be the diagonal matrix with θ_i's on the diagonal, and g be the vector with γ_i's as entries. Then the near parallel property in matrix form is

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

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 have them orthogonal to u_{m+1} and R. Given a subspace spanned by the columns of U_m, Stewart [28] 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 imply 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 $(\theta_i, g_i), i = 1, \ldots, k$, are eigenpairs of B, denoted by $BG = G\Theta$. Let $y_i = U_m g_i$ and $Y = [y_1, \ldots, y_k]$. Then $AY_k = Y_k\Theta + u_{m+1} g^T + F$ 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 $g^T = b^T G$ and $F = RG$, then

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

With A symmetric, G normally is orthogonal, so $\|G\| = 1$ and $\|F\| \leq \|R\|$. ∎

4.3. Maintaining near-Krylov. If the Arnoldi-E phase starts with a near-Krylov subspace, one may wonder whether this property is maintained 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 the starting vector. This is joined with the other Ritz vectors to form the overall subspace

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

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 $Span\{y_1, \ldots, y_k\}$ and the Krylov portion $Span\{u, Au, A^2u, \ldots, A^{m-k-1}u\}$. 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 $Span\{y_1, \ldots, y_k, u\}$ 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 $Span\{y_1, \ldots, y_k\}$ to $Span\{y_1, \ldots, y_k, u\}$, 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,

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

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),

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

Putting in $u = Vd$, from (4.7), and combining,

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

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

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

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

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

Let $\hat{U}_{n\times m} = \begin{bmatrix}V & U\end{bmatrix}$, $\hat{B} = \left[\begin{smallmatrix} H & db^T \\ & B\end{smallmatrix}\right] + \eta c e^T + K$, $\hat{u} = \eta v_0$, $\hat{b} = e_p$, $\hat{R} = \begin{bmatrix}0 & R_0\end{bmatrix}$; 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

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

And

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

For $\|\hat{R}\|$ to be small, (4.10) says that $\|R_0\|$ needs 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 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 $Span\{y_1, \ldots, y_k, u\}$ out to a subspace of dimension m and then back to a subspace $Span\{y_1^{new}, \ldots, y_k^{new}, u^{new}\}$.

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 $\{y_1, y_2, \ldots, y_k, u\}$, *and with* $\begin{bmatrix}U_{n\times k}, u\end{bmatrix}^T R_{n\times k} = 0$. *Suppose the subspace we generate for Arnoldi-E is*

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

from which the new k Ritz vectors are $\{y_1^{new}, y_2^{new}, \ldots, y_k^{new}\}$. *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* $\{y_1^{new}, y_2^{new}, \ldots, y_k^{new}\}$, *and*

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

Proof. For the subspace $\{y_1, y_2, \ldots, y_k\}$, 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 $span\{u, Au, A^2u, \ldots, A^{m-k-1}u\}$, 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

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

and we also have

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

It can be shown (see Lemma 5.4 in [34] that $\hat{B}$ is similar to the matrix Q^TAQ where columns of Q are orthonormal basis of

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

Hence eigenvalues of $\hat{B}$ 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 $\{\theta_1, \ldots, \theta_k\} \cap \{\theta_{k+1}, \ldots, \theta_m\} = \emptyset$. 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

$$\hat{B}G_1 = G_1T_{11}.$$

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

$$\begin{aligned} A\hat{U}_{n\times m}G_1 &= \hat{U}_{n\times m}\hat{B}G_1 + \hat{u}\hat{b}^TG_1 + \hat{R}_{n\times m}G_1, \\ &= \hat{U}_{n\times m}G_1T_{11} + \hat{u}\hat{b}^TG_1 + \hat{R}_{n\times m}G_1. \end{aligned}$$

Let $U^{new} = \hat{U}_{n\times m}G_1$, $B^{new} = T_{11}$, $u^{new} = \hat{u}$, $(b^{new})^T = \hat{b}^TG_1$, and $R_{n\times k}^{new} = \hat{R}_{n\times m}G_1$. Then

$$AU^{new} = U^{new}B^{new} + u^{new}(b^{new})^T + R_{n\times k}^{new}.$$

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

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

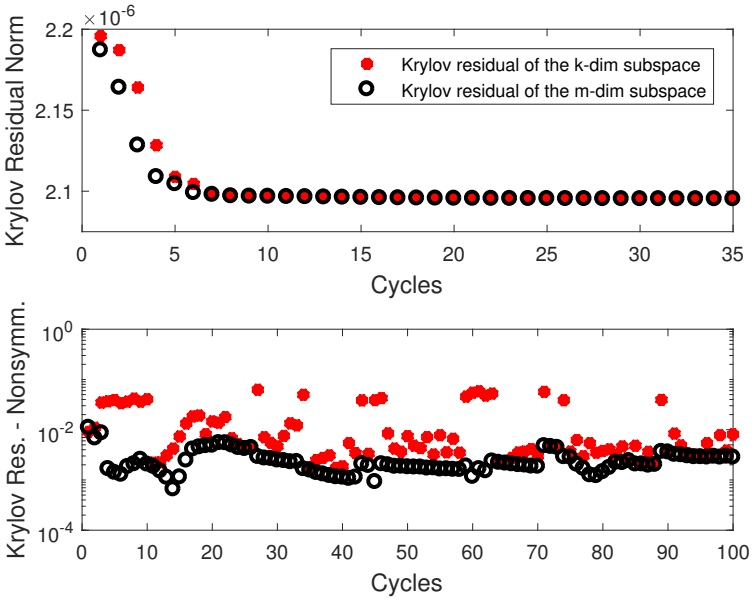

FIG. 4.2. *Plot of Krylov residual norms showing how near the subspaces are to being Krylov. Top matrix is symmetric from* 1*-D Laplacian with* $n_c = 1023$ *and* $n_f = 255$*. Bottom plot has nonsymmetric matrices due to convection term with* $\beta = 102.4$*.*

As mentioned before Theorem 4.4, the Krylov residual can go up with u coming from Ay_j. But the norm of the Krylov residual generally does not increase by much.

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 $n_c = 255$, 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 4.2 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 [28] 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 the $k+1$ 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. Using Theorem 2.2, if a set of vectors $\{y_1, \ldots, y_k\}$ has parallel residuals, then subspace (2.1) is a Krylov subspace, and from Theorem 2.1, so is subspace (2.2). These contain Krylov subspaces with each y_j as starting vector, $span\{y_j, Ay_j, \ldots, A^{m-k}y_j\}$, for j from 1 to 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. 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 Theorem 5.10 in [34] for a similar theorem from the near-Krylov perspective.

We focus on only two vectors and show how being nearly parallel can help them converge together. Let y_1 be starting vector, and let its residual vector be $r_1 = \gamma_1 w$,

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*

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

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

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

Let

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

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

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

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

$$(A+E)y_1 = Ay_1 - \frac{1}{\beta} f (y_2^{\perp K})^T y_1 = Ay_1 = \theta_1 y_1 + \gamma_1 w,$$
$$(A+E)y_2 = Ay_2 - \frac{1}{\beta} f (y_2^{\perp K})^T y_2 = Ay_2 - f = \theta_2 y_2 + \gamma_2 w.$$

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 [16]. So

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

Next we want to show that

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

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

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

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

And

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

The theorem indicates that if y_1 and y_2 are nearly parallel, y_2 will converge along with y_1 even with y_1 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 $\{y_1, Ay_1, \ldots, A^p y_1\}$ tends to be small. For example, with the experiment in Example 5, except with $m = 10$ and $k = 2$ (alternating between y_1 and y_2 as starting vectors), α in (4.13) is $3.9 * 10^{-4}$ after 100 cycles. If a convection term of size 25.6 is added, then α is 0.99 at that point. If α is small and β close to 1, $\|E\|$ is mainly determined by $\|f\|$.

We wish to understand how much the perturbed matrix in Theorem 4.6 can affect the convergence. We leave this for future work, but one possible way to study this is to use the Cauchy integral to express the polynomials of a matrix (a theorem related to Theorem 2.1 in [23] can be developed).

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,

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

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 the bottom of Figure 4.1 is the lowest solid line until it levels out and is passed by $\|r_1\|$. Figure 4.3 has this curve again along with curves for γ_2 and $||f_2||$. The $\|r_1\|$ curve keeps converging since y_1 is the starting vector

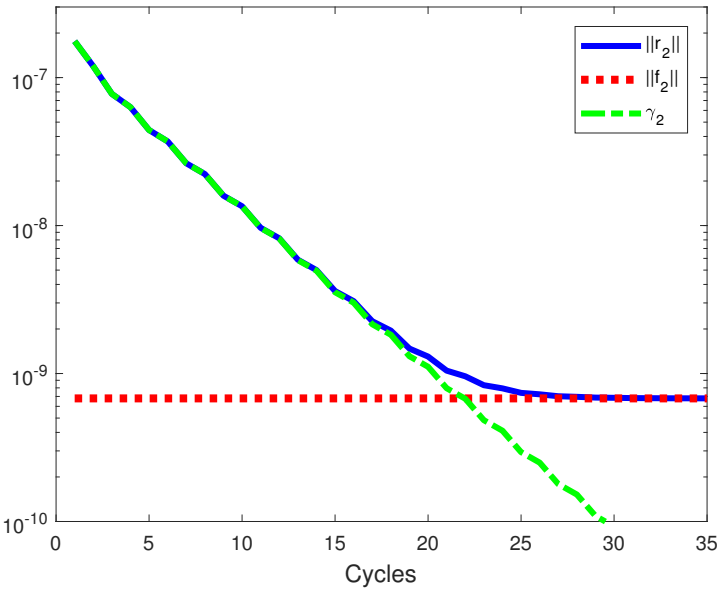

FIG. 4.3. *Decomposition of the residual vector for y_2 during the Arnoldi-E phase with y_1 as the starting vector for every cycle. Convergence stops when deviation from parallel becomes significantly larger than the rest of the residual. Matrix is symmetric from the 1-D Laplacian of size $n_f = 1023$ and $n_c = 255$.*

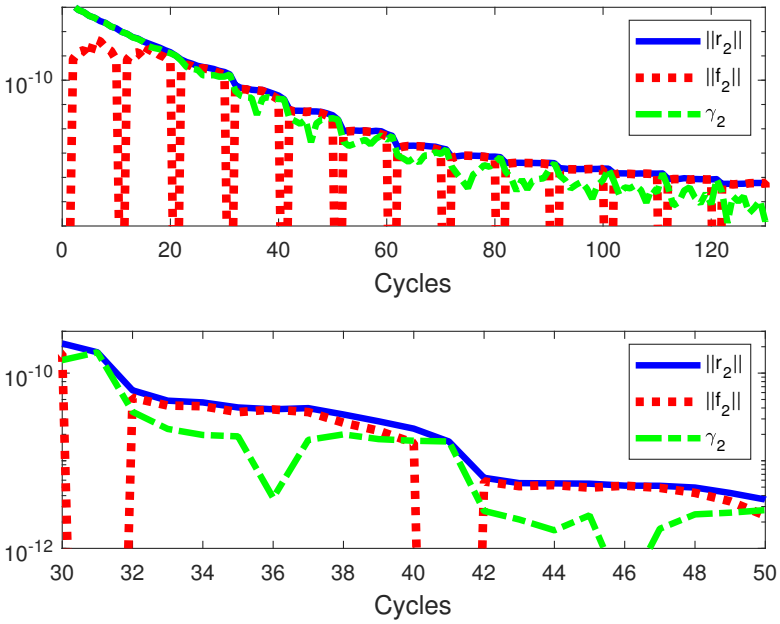

FIG. 4.4. *Breakdown of the residual vector for y_2 with starting vectors for Arnoldi-E cycled through* 10 *Ritz vectors. Convergence slows as the deviation from parallel generally becomes larger relative to the rest of the residual. Matrix is symmetric from the* 1*-D Laplacian. Lower graph is a close-up of a portion of the upper graph.*

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 than γ_2, then it can wash out this correction. When $||f_2||$ is significantly smaller than γ_2, it should not have much effect, and the subspace will have an accurate approximation to Ay_2. In Figure 4.3, the $||r_2||$ curve starts to level out once γ_2 is 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 top of Figure 4.1. However, this time we continue to rotate through all 10 even after some have converged. The residual curve r_2 is shown with a solid line in the top of Figure 4.4, and because of this change, it is concave up while the ones in Figure 4.1 converge fairly consistently. We wish to see 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 onto w be $r_2 = \gamma_2 w + f_2$. So we have

$$Ay_j = \theta_j y_j + \gamma_j w, \quad \text{and} \quad Ay_2 = \theta_2 y_2 + \gamma_2 w + f_2, \quad \text{where } f_2 \perp w. \tag{4.14}$$

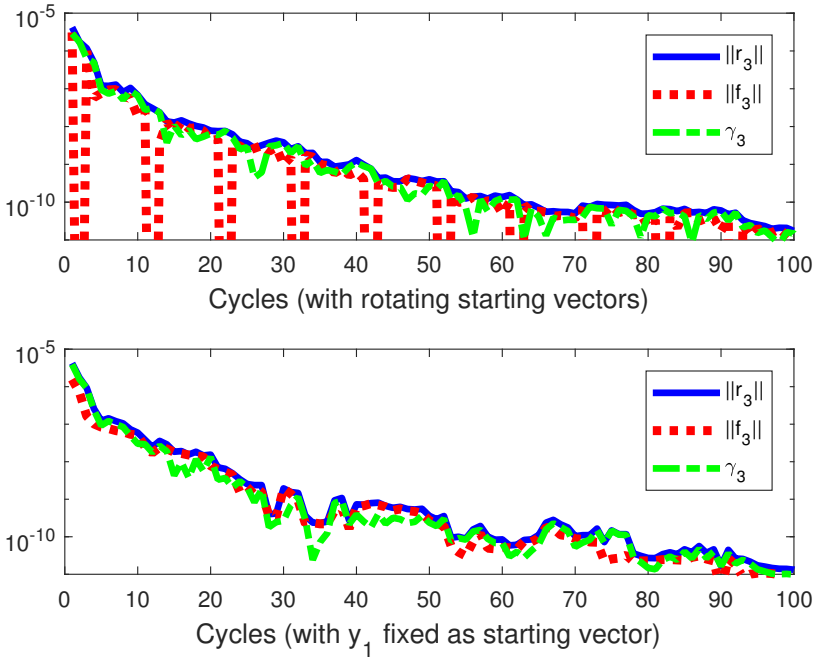

FIG. 4.5. *Residual vector for y_3 with nonsymmetric matrix from the 1-D conv-diff equation, $\beta = 102.4$. Top figure has rotating through 10 Ritz vectors and the bottom keeps y_1 as the starting vector. Even with y_1, the deviation from parallel keeps going down, and the residual for y_3 keeps improving.*

The top part of Figure 4.4 also has γ_2 and $||f_2||$. These quantities are computed at the end of each cycle, so an $||f_2||$ smaller than γ_2 helps on the next cycle. Every 10th cycle, y_2 is the starting vector in Arnoldi-E and so $||f_2||$ is zero. Initially with γ_2 larger than $||f_2||$, y_2 converges rapidly. Here $||r_2||$ does not stall out as it does in Figure 4.3, because every 10th 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 γ_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 4.4 has a close-up from the top part for cycles 30 through 50. When γ_2 comes up near $||f_2||$, the residual improves for the next cycle.

Next let the matrix be nonsymmetric by adding a convection term $\beta = 102.8$. We still have $n_c = 255$ and $n_f = 1023$ but now have 10^{-8} for the first phase *rtol*. Again we cycle through 10 y_j's as starting vectors, even if converged. The convergence of the third residual vector is given in Figure 4.5 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 γ_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, $\|f_3\|$ keeps being reduced and y_3 keeps improving. Perhaps this happens because the eigenvectors are related to each other instead of being orthogonal (see Figure 12.3 of [29]). In fact, the cosine of the angle between y_3 and y_1 for the last cycle of this test is 0.992, so the vectors are closely aligned. Pushing toward y_1 is also partially going toward y_3.

TABLE 5.1
Two-grid Arnoldi versus 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 Arnoldi cycle equiv's	227	41.8	15.6	9.56	11.9	9.86	10.1

5. Multiple grids for Arnoldi. We now investigate using more than two grids. More grid levels make it possible to get from grid to grid with smaller changes in grid size. As before, the problem size is n_f. We let nl be the number of grid levels used, with grid level 1 being the coarsest and grid level nl being the finest corresponding to matrix size n_f. 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 n_f. Choose the grid levels. Let nl be the number of grids ordered from coarsest to finest.
   Choose m and k, the max subspace size and number of Ritz vectors retained, nev = number desired eigenpairs, and $rtol$ = 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* $= 2, \ldots, nl$:
   A. **Move to next finer grid:**
      Move the k current Ritz vectors to the next finer grid (e.g., splines).
   B. **Finer grid computation:**
      Apply step 3 of the two-grid Arnoldi algorithm. Stop if this is the finest grid, otherwise go back to A.

Example 8. This and the next two examples use decreasing sizes of the matrices n_f, $\frac{n_f+1}{2} - 1$, $\frac{n_f+1}{4} - 1$, $\ldots$, $\frac{n_f+1}{2^{nl-1}} - 1$. We use the 1-D convection-diffusion equation with convection of $\beta = 51.2$. The size is $n_f = 4095$. 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. The multiple-grid Arnoldi result with the 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 $n_c = 1023$. With multiple-grid Arnoldi, we can get below 10 fine-grid-equivalents.

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

Example 9. We again have a matrix from the 1-D convection-diffusion equation, but the convection is increased to $\beta = 102.4$ and the size of the matrix is reduced to $n_f = 1023$. We use Arnoldi(30,16) since the matrix is more nonnormal. 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. Multiple-grid Arnoldi beats two-grid on some of the choices, but not by much. Using too small a coarsest grid can make things worse, as with a coarsest grid of size 31. Approximations from a coarse grid to the next are not as accurate with the increased convection.

TABLE 5.2
Two-grid Arnoldi versus 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

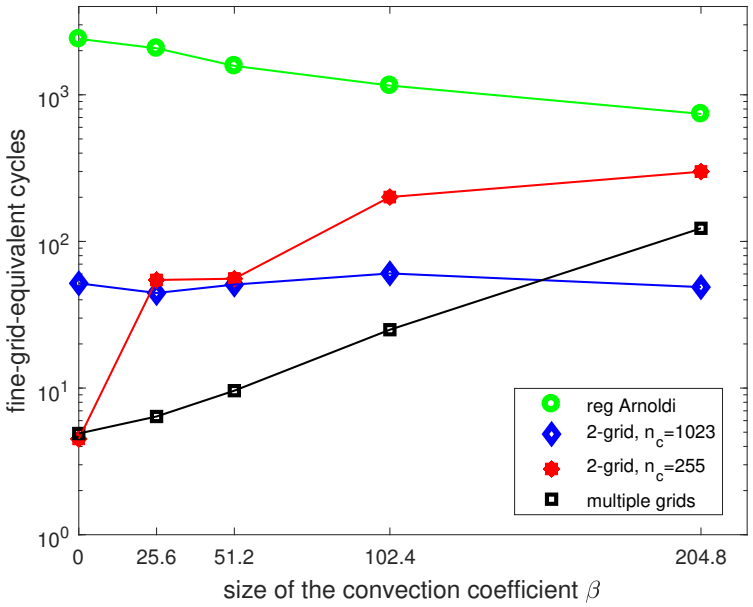

FIG. 5.1. *Multiple-grid Arnoldi with coarsest grid of size* 255 *compared to two-grid Arnoldi and regular Arnoldi. The convection coeffient varies from* 0 *to* 204.8. *Fine grid matrix size is* $n = 4095$.

Example 10. For this example we go back to the larger fine grid matrix of size $n = 4095$. 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 are regular Arnoldi and two types of two-grid Arnoldi, first with a coarse grid matrix of size $n_c = 1023$ and then with $n_c = 255$. For the case of a symmetric matrix, $\beta = 0$, both multiple-grid Arnoldi and two-grid Arnoldi with $n_c = 255$ 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.

Example 11. We next use large matrices and compare multiple-grid Arnoldi to two-grid and also the two methods from Example 2 that use standard multigrid, inverted Arnoldi, and LOBPCG. This time the multigrid method for solving the linear equations uses three Gauss–Seidel relaxations per grid for each V-cycle (twice downward and once upward) using a code from Scott Fulton of Clarkson University. Also, with inverted Arnoldi the tolerance is left at 10^{-6}. Relaxing this tolerance as in Example 2 would give inaccurate results, because the eigenvalue residual norms all drop, then go back up when second copies of double or near double eigenvalues appear. The outer Arnoldi loop is not restarted.

TABLE 5.3
Two-grid and multiple-grid Arnoldi versus inverted Arnoldi and LOBPCG with multigrid. Matrix is dimension $n = 1{,}046{,}529$ from 2-D Laplacian.

Two-grid Arnoldi	Coarse grid	cg cycles	fg cycles	mvp equiv's	Time
	$(127)^2$	149	7	181.8	46.3
Multi-grid Arnoldi	Coarsest grid	# levels	cycles	mvp equiv's	Time
	$(63)^2$	5	45, 10, 9, 2, 0	58.7	11.0
Inverted Arnoldi	rtol for lin. eq's	Arnoldi iter's	mvp equiv's per solve	mvp equiv's total	Time
	1.e-6	44	24	1127	53.6
LOBPCG	rtol for lin. eq's	Precond. calls	mvp equiv's per solve	mvp equiv's total	Time
	1.e-1	142	4.3	770	61.2

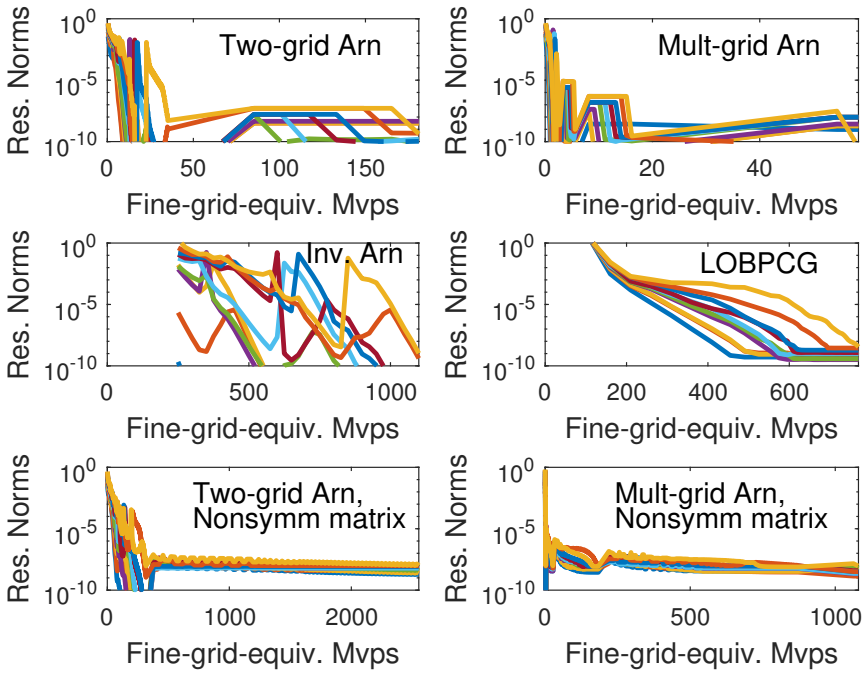

FIG. 5.2. *Two-grid and multiple-grid Arnoldi versus inverted Arnoldi and LOBPCG with multigrid. Matrix is dimension $n = 1{,}046{,}529$. The top four parts of the figure are for the 2-D Laplacian. The lower two parts are with added convection term $10u_x$.*

The matrix is first the 2-D discrete Laplacian on the unit square of size $n = (1023)^2 = 1{,}046{,}529$. Results are in Table 5.3 and in the top four parts of Figure 5.2. Multiple-grid Arnoldi is the best method, using far fewer fine-grid-equivalent matrix-vector products than the two multigrid methods. Also there is less cpu time, 11.0 seconds compared to 46 or more for the other methods. Part of the success of multiple-grid Arnoldi is due to the eigenvectors of the finer grids not changing much as they move to the next grid, so few cycles are needed on the finer grids (2 and 0 on the last two).

The situation is different when a convection term is added so the differential equation is the same as in Example 1. LOBPCG does not apply for this nonsymmetric

problem, but inverted Arnoldi takes 49 seconds while multiple-grid Arnoldi with five levels is much slower with 203 seconds (two-grid Arnoldi is even slower at 829 seconds with coarse grid of size 255^2). Matrix-vector products are almost the same between (multiple-grid Arnoldi (1079) and inverted Arnoldi (1085). The lowest level of Figure 5.2 has convergence for two-grid and multiple-grid (inverted Arnoldi is not included, but it is fairly similar to the plot for the Laplacian in the left middle part of the figure).

Next, with more convection, $25u_x + 15u_y$, the multigrid linear equations solves for inverted Arnoldi fail. Meanwhile, multiple-grid Arnoldi converges; the best is for four levels with a time of 659 seconds (five and three levels are 2 minutes more).

Interestingly, for this last experiment, if the residual tolerance is lower at 10^{-7}, then only 49 seconds are needed. This points out again, as was discussed at the end of Example 1, that the success of both two-grid and multiple-grid Arnoldi depends on both having a coarse grid that provides accurate approximations and how accurate the eigenvectors are needed on the fine grid. Determining the optimal fine grid requires either knowledge of multigrid theory for the particular problem or some experimenting. Additionally, determining if these approaches are better than the multigrid methods that we have been comparing them to is even more difficult. Certainly it appears that the new approaches are more sensitive to the amount of accuracy that is desired. That can be seen especially in the left bottom part of Figure 5.2, where two-grid Arnoldi starts with good accuracy on the fine grid then converges slowly. Multiple-grid Arnoldi seems to be less sensitive to the choice of coarsest grid but still is quite affected by the demanded accuracy, as the first sentence of this paragraph pointed out.

6. Conclusion. Eigenvalue problems from differential equations are easier for smaller grids, because smaller matrices mean less work per iteration, and also the spectrum is easier. Here we gave Krylov methods that use the power of coarse grids. 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 and where eigenvectors from coarse grids are fairly good approximate eigenvectors on finer grids. Compared to traditional multigrid, the use of Krylov subspaces makes this approach more robust in the sense of being applicable to more problems. Effectiveness can be explained with near-Krylov properties of the approximate eigenvectors that are passed from coarse to fine grids. This approach can use two grids or a sequence of increasingly finer grids. The best method should be further studied but depends on the problem, the desired accuracy, and how good approximate eigenvectors are when moved from a coarse to a fine grid.

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

Acknowledgment. The authors would like to thank the referees for their very careful readings and many suggestions and corrections that improved the paper.

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 (2010), pp. 129–149.

[2] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[3] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.
[4] A. Brandt, S. McCormick, and J. Ruge, *Multigrid methods for differential eigenproblems*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 244–260.
[5] S. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, New York, 1994.
[6] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.
[7] R. P. Fedorenko, *The speed of convergence of one iterative process*, USSR Comput. Math. Math. Phys., 4 (1964), pp. 227–235.
[8] A. V. Knyazev and K. Neymeyr, *Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method*, Electron. Trans. Numer. Anal., 15 (2003), pp. 38–55.
[9] C. Lanczos, *An iterative method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.
[10] R. B. Lehoucq and D. C. Sorensen, *Deflation techniques for an implicitly restarted Arnoldi iteration*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 789–821.
[11] Q. Lin and H. Xie, *A multi-level correction scheme for eigenvalue problems*, Math. Comp., 84 (2015), pp. 71–88.
[12] J. Mandel and S. McCormick, *A multilevel variational method for $Au = \lambda Bu$ on composite grids*, J. Comput. Phys., 80 (1989), pp. 442–452.
[13] R. B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154–156 (1991), pp. 289–309.
[14] R. B. Morgan, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Math. Comp., 65 (1996), pp. 1213–1230.
[15] R. B. Morgan, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1112–1135.
[16] R. B. Morgan, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
[17] R. B. Morgan and M. Zeng, *Harmonic projection methods for large non-symmetric eigenvalue problems*, Numer. Linear Algebra Appl., 5 (1998), pp. 33–55.
[18] R. B. Morgan and M. Zeng, *A harmonic restarted Arnoldi algorithm for calculating eigenvalues and determining multiplicity*, Linear Algebra Appl., 415 (2006), pp. 96–113.
[19] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.
[20] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[21] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl., 34 (1980), pp. 269–295.
[22] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, 2nd ed., Classics in Appl. Math. 66, SIAM, Philadelphia, 2011.
[23] 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 (2013), pp. 1066–1088.
[24] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.
[25] D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
[26] A. Stathopoulos, Y. Saad, and K. Wu, *Dynamic thick restarting of the Davidson, and the implicitly restarted Arnoldi methods*, SIAM J. Sci. Comput., 19 (1998), pp. 227–245.
[27] G. W. Stewart, *A Krylov–Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614.
[28] G. W. Stewart, *Backward error bounds for approximate Krylov subspaces*, Linear Algebra Appl., 340 (2002), pp. 81–86.
[29] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[30] U. Trottenberg, C. W. Oosterlee, and A. Schuller, *Multigrid*, Academic Press, London, 2001.
[31] J. van den Eshof and G. L. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153.

[32] K. Wu and H. Simon, *Thick-restart Lanczos method for symmetric eigenvalue problems*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 602–616.
[33] J. Xu and A. Zhou, *A two-grid discretization scheme for eigenvalue problems*, Math. Comp., 70 (2001), pp. 17–25.
[34] Z. Yang, *A Multigrid Krylov Method for Eigenvalue Problems*, Ph.D. thesis, Baylor University, Waco, TX, 2015.