

TWO-GRID DEFLATED KRYLOV METHODS FOR LINEAR EQUATIONS*

RONALD B. MORGAN[†], TRAVIS WHYTE[‡], WALTER WILCOX[§], AND ZHAO YANG[†]

Abstract. An approach is proposed for solving large linear systems that combines Krylov methods with the use of two different grid levels. Eigenvectors are computed on the coarse grid and then used to deflate the eigenvalues on the fine grid with an efficient projection. GMRES-type methods are first used on both the coarse and fine grids. Then another approach is given that uses a novel restarted BiCGStab (or IDR) method on the fine grid. While BiCGStab is generally considered to be a non-restarted method, restarting works well in this context. Tests show that this new approach can be very efficient for difficult linear equations problems.

Key words. linear equations, deflation, GMRES, BiCGStab, eigenvalues, two-grid

AMS subject classifications. 65F10, 15A06

1. Introduction. We look at solving large systems of linear equations $Ax = b$ that result from discretizations of partial differential equations. There exists a variety of iterative methods for solving these problems. In particular, multigrid methods [11, 14, 15, 23, 47] are extremely effective under certain conditions. In other situations, simple multigrid methods do not work well. However, they can be used as preconditioners for Krylov methods. Here we propose a new approach for combining multigrid with Krylov subspace methods. We do not compare our approach with sophisticated multigrid methods and Krylov with multigrid methods and do not claim that our approach is better. However, we are giving an alternative that is worth considering.

Convergence of iterative methods is generally affected by the conditioning of the matrix, or more specifically by the presence of small eigenvalues. One reason why multigrid methods are effective is that the eigenvectors corresponding to small eigenvalues generally have a similar shape on different grid levels. Multigrid is able to handle these small eigenvalues on the coarse grids, where the computations are cheaper. Also, the smaller linear systems generally involve a better-conditioned matrix.

Krylov subspace methods may sometimes converge slowly. Work has been done on dealing with the detrimental presence of small eigenvalues for Krylov methods. Restarted methods such as GMRES are particularly sensitive to the presence of small eigenvalues. Deflated GMRES methods [7, 16, 17, 24, 25, 26, 28, 32, 34, 40, 44, 45] compute approximate eigenvectors and use them to remove or deflate the effect of the small eigenvalues. In particular, we will use the method GMRES-DR [34], which both solves linear equations and simultaneously computes eigenvectors.

The computation of approximate eigenvectors can be expensive for difficult problems. The approach in this paper starts with computing eigenvectors on a coarse grid and moving them to the fine grid. Then these approximate eigenvectors are used for a version of deflated GMRES on the fine grid (specifically, GMRES-Proj [36], which will be described later).

*Received May 6, 2020. Accepted October 21, 2024. Published online on February 24, 2025. Recommended by Kirk Soodhalter. The first author was supported by the National Science Foundation under grant DMS-1418677 and from a Baylor University Research Committee Grant.

[†]Department of Mathematics, Baylor University, Waco, TX 76798-7328
(Ronald_Morgan@baylor.edu, zhao_yang@alumni.baylor.edu). Corresponding author: R. Morgan.

[‡]Jülich Supercomputing Centre, 54245 Jülich, Germany (t.whyte@fz-juelich.de).

[§]Department of Physics, Baylor University, Waco, TX 76798-7316 (Walter_Wilcox@baylor.edu).



This work is published by ETNA and licensed under the Creative Commons license [CC BY 4.0](#).

For fairly sparse matrices, the GMRES orthogonalization can be a major expense. We present an approach that substitutes either BiCGStab or IDR in place of GMRES on the fine-grid. Here we restart the BiCGStab and IDR procedures even though they are normally non-restarted methods. This may seem risky because of the often erratic convergence of these nonsymmetric Lanczos methods. However, in our testing, restarted BiCGStab and IDR converge reliably.

We now discuss the contributions of this paper. The main three are: 1) approximate eigenvectors are efficiently computed on a coarse grid and then used to deflate linear equations on the fine grid. Also, the GMRES-DR algorithm is used on the coarse-grid, which allows us to begin solving the linear equations at the same time as the eigenvalues are computed. 2) On the fine grid there is an efficient way of deflating eigenvalues with a projection, which allows many eigenvalues to be dealt with. Other approaches that build a preconditioner with eigenvectors and thus use them at every iterations can become expensive if more than a few eigenvalues are used. 3) A restarted BiCGStab or IDR method is combined with deflation for the fine-grid solution. We are not aware of any similar previous methods. The first two contributions have appeared in some form in other papers, and this will be mentioned in the next paragraphs. We conclude this paragraph with a couple of smaller contributions of this paper. Some analysis and discussion is given for how accurate the eigenvectors need to be on the fine-grid. Mainly it is shown that while high accuracy is not needed, the required amount of accuracy varies with the size of the smallest eigenvalue and possibly other factors. Also, some explanation will be given for why BiCGStab is effective even with restarting.

With respect to the first main contribution from the previous paragraph, there have been works that find eigenvectors on a coarse grid. The papers [4, 39] use approximate eigenvectors from a coarse grid to deflate for a symmetric matrix and the conjugate gradient method. Most similar to this paper is Sifuentes' thesis [42], which involves two-grid deflation but with the Arnoldi method on the coarse grid and with the more expensive approach of building a deflating preconditioner for the GMRES method on the fine grid. Instead of using computed approximate eigenvectors, Erlangga and Nabben [21] deflate using vectors from the interpolation operator that maps from the coarse to the fine grid. Somewhat related are works using coarse-grid eigenvectors for eigenvalue problems; see for example [12, 37, 50, 51]. These type of methods improve the coarse-grid eigenvectors in different ways on the fine grid or on multiple grids. Not so closely related, but worth noting, are methods that combine multigrid with the GMRES method; see for example the paper by Elman, Ernst, and O'Leary [20].

As mentioned above, the second contribution is an efficient deflation. This is used previously in [32, 36] for a deflated GMRES algorithm (see also [46]) but not with two grids. Many works use an approach of building a preconditioner from approximate eigenvectors [3, 5, 7, 16, 24, 28, 31, 38, 40, 43, 44, 45, 45]. The eigenvectors are used to deflate in many different situations. As mentioned, this can be much more expensive, but it is generally more robust. Of particular note amongst these are [3, 5, 31, 38], which deflate from the unrestarted MINRES and BiCGStab method or, as the last one, from IDR. Also, some of those use the GMRES-DR method to compute eigenvectors, as this paper does.

Of the three main contributions in the current paper, no previous works have used more than one of them, and there has been no previous use of the restarted BiCGStab method. So the new procedure in this paper significantly stands out from previous approaches. Section 2 of this paper reviews some of the previous methods that will be used. Section 3 presents the two-grid deflated GMRES method. The two-grid deflated BiCGStab/IDR algorithm is then given in Section 4. Several further examples are presented in Section 5, including a Helmholtz problem with multigrid preconditioning.

2. Review. Next, we quickly describe some of the methods that will be used in the rest of the paper.

2.1. The GMRES-DR method. The GMRES with deflated restarting (GMRES-DR) [34] method uses Krylov subspaces to both solve linear equations and compute the eigenpairs for the smallest eigenvalues. Once the eigenvectors converge well enough, their presence in the subspace can essentially remove or deflate the effect of the small eigenvalues on the linear equations. Usually only moderate accuracy is needed before the approximate eigenvectors have a beneficial effect.

The GMRES-DR(m, k) method saves k approximate eigenvectors at the restart and constructs a subspace of dimension m from these. Specifically, for one restarted cycle it uses the subspace

$$\text{Span} \{y_1, \dots, y_k, r_0, Ar_0, A^2r_0, \dots, A^{m-k-1}r_0\},$$

where the y_i 's are harmonic Ritz vectors from the previous cycle and r_0 is the residual vector at the start of the cycle. For the GMRES-DR and GMRES methods, a "cycle" is the building and use of a Krylov subspace in between restarts. The augmented subspace of GMRES-DR is actually a Krylov subspace itself, and it contains Krylov subspaces with each y_i as starting vector. This makes the eigenvectors generally converge along with the linear equations.

The GMRES-DR method converges faster than the restarted GMRES version for difficult problems with small eigenvalues. It also often converges faster than the BiCGStab method in terms of matrix-vector products but has greater orthogonalization costs per matrix-vector product.

2.2. The GMRES-Proj method. There are situations where approximate eigenvectors are available at the beginning of the computation of solutions of linear equations. For example, if there are multiple right-hand sides, then some eigenvectors could have been computed during the computation for earlier right-hand sides [36, 40, 46]. The method GMRES-Proj [36] uses these approximate eigenvectors to deflate the corresponding eigenvalues while solving linear equations. The GMRES(m)-Proj(k) method assumes that k approximate eigenvectors have been previously computed and alternates projections onto these vectors with cycles of GMRES(m).

Algorithm 1 GMRES(m)-Proj(k).

0. Let k be the number of approximate eigenvectors that are available. Choose m , the dimension of subspaces generated by the restarted GMRES method.
 1. Alternate between A) and B) until convergence:
 - A) Apply a Galerkin projection onto the subspace spanned by the k approximate eigenvectors.
 - B) Apply one cycle of GMRES(m).
-

The algorithm for the projection step is given in Algorithm 2. MinRes projection can be used instead of a Galerkin projection. It is the same as for the Galerkin case except that $H = (AV)^T AV$ and $c = (AV)^T r_0$. We have found the Galerkin projection to be more reliable than the MinRes one. All experiments in this paper use a Galerkin projection.

2.3. The two-grid Arnoldi method. A two-grid method for computing eigenvalues and eigenvectors is given in [37]. Eigenvectors are computed on a coarse grid with a standard Arnoldi method, they are transferred to the fine grid (with spline interpolation), and then

Algorithm 2 Galerkin projection over a set of approximate eigenvectors.

0. Let the current system of linear equations be $A(x - x_0) = r_0$.
 1. Let V be an n by k orthonormal matrix whose columns span the set of approximate eigenvectors.
 2. Form $H = V^T AV$ and $c = V^T r_0$.
 3. Solve $Hd = c$, and let $\hat{x} = Vd$.
 4. The new approximate solution is $x_p = x_0 + \hat{x}$, and the new residual is $r = r_0 - Ax_p = r_0 - AVd$.
-

they are improved on the fine grid with Arnoldi-E [33], a method that can accept initial approximations.

3. The two-grid deflated GMRES method. Our new methods generate approximate eigenvectors from the coarse grid and use them to deflate eigenvalues on the fine grid. In this section, we give a version using GMRES methods. The GMRES-DR procedure is applied on the coarse grid. This generates eigenvectors and solves the coarse-grid linear equations. This solution of the equations on the coarse grid is mapped to the fine grid with spline interpolation or prolongation and used there as the initial guess. Here we use interpolation for all examples except Example 10. The eigenvectors are similarly moved to the fine grid and, if necessary, improved on the fine grid. The GMRES-Proj method is applied on the fine grid using these approximate eigenvectors. This can achieve much faster convergence than using a restarted GMRES method. Compared to running GMRES-DR on the fine grid, it is cheaper to implement and can deflate eigenvalues from the beginning.

Algorithm 3 Two-grid deflated GMRES.

0. Choose m and k for the coarse grid. Pick nev , the number of eigenpairs that are required to converge to an eigenvalue tolerance, say $rtolev$. For the fine grid, pick $rtol$, the linear equations residual tolerance, and pick $m3$, the number of steps of when to restart GMRES.
 1. Apply GMRES-DR(m,k) on the coarse grid. Map the approximate eigenvectors to the fine grid (with spline interpolation or a prolongation operator). Map the solution of the coarse-grid linear equations problem to the fine grid and use it as an initial guess for the fine-grid problem.
 2. (If needed:) Improve the approximate eigenvectors on the fine grid using Arnoldi-E(m,k) (see the two-grid Arnoldi method in [37]).
 3. Apply GMRES($m3$)-Proj(k) on the fine grid.
-

Below, for our first examples, there is no need for phase 2, but it is used in Sections 4.1, 5.2, and 5.3.

EXAMPLE 1. We consider a system of linear equations from a finite difference discretization of the 2-D convection-diffusion equation

$$-e^{5xy}(u_{xx} + u_{yy}) + 40u_x + 40u_y = c \sin(x) \cos(x) e^{xy}$$

on the unit square with zero boundary conditions and c chosen to make the right-hand side have norm one. Central difference formulas are used for all discretizations. The discretization size is $h = \frac{1}{512}$, leading to a matrix of dimension $n = 511^2 = 262,121$. The coarse-grid discretization size is $h = \frac{1}{64}$, giving a matrix of dimension $63^2 = 3969$.

The first phase on the coarse grid uses GMRES-DR(150,100) and runs until 80 eigenpairs have converged to drop the level of the residual norm below 10^{-8} . These residual norms are computed only at the end of the cycles. The eigenvectors are moved to the fine grid and are there accurate enough to be effective in deflating the eigenvalues (after the Rayleigh-Ritz procedure is applied to all 100 vectors that are moved from the coarse grid, the smallest 80 Ritz pairs have residual norms at or below $1.4 \cdot 10^{-3}$). So as mentioned, the second phase is not needed. The third phase of solving the fine-grid linear equations is stopped when the relative residual norm drops below a tolerance of $rtol = 10^{-10}$. This is also tested only at the end of each GMRES cycle but of course could be easily monitored during the GMRES runs.

The top of Figure 3.1 illustrates convergence curves for both the linear equations and the eigenvalues on the coarse grid. The linear equations solution converges well before all 80 eigenpairs become accurate. The linear equations require 19 cycles of GMRES-DR(150,100), which use 1050 matrix-vector products (150 for the first cycle and 50 each for the next 18). The eigenvalues require 107 cycles or 5450 matrix-vector products.

The bottom left part of the figure illustrates the convergence of the linear equations on both grids versus the number of fine-grid-equivalent matrix-vector products. The coarse-grid matrix is about 64 times smaller than the fine-grid matrix, so we scale the number of matrix-vector products by a factor of 64 to get the fine-grid equivalents. The coarse-grid linear equations converge so rapidly that the convergence curve is barely noticeable on the left-hand side of the graph. Then, there is a small gap for both the convergence of the eigenpairs on the coarse grid and the matrix-vector products needed to form the projection matrix H for the fine grid. This gap is visualized more detailed at the bottom right part of the figure, which is a close-up of the left one. Three different values of $m3$ are used for GMRES($m3$)-Proj(100): 50, 100, and 200. While the fastest convergence is achieved with GMRES(200), the least expensive one is for $m3 = 100$. We define the approximate cost as $cost = 5mvp + vops$, where mvp is the number of matrix-vector products; the value 5 comes from the approximate number of non-zeros per row, and $vops$ is the number of length- n vector operations such as dot-products and daxpy's (so $2n$ flops for one unit of cost). This approximate cost includes the entire process, with the cost for the first phase on the coarse grid being scaled according to the length of the coarse-grid vectors relative to the fine-grid vectors. This cost is $cost = 1.10 \cdot 10^6$ for $m3 = 50$, $1.06 \cdot 10^6$ for $m3 = 100$, and finally $1.30 \cdot 10^6$ for $m3 = 200$. The orthogonalization expense for GMRES(200) is especially significant with such a sparse matrix.

Next, we consider different sizes for the coarse grid. Smaller coarse-grid matrices means that less work is needed to find the eigenpairs, however, they may not be as accurate for the fine-grid case. Table 3.1 has results for coarse grids from sizes $65,025 = 255^2$ down to $49 = 7^2$. All of the tests use the same type of coarse-grid computation as before and the GMRES(100)-Proj(100) method on the fine grid. The results show that for this matrix, the fine-grid convergence is fairly robust with respect to the coarse-grid size. The number of fine-grid cycles increases by less than half as the coarse grid goes from size 65,025 down to 255. For the 255-case, the accuracy of 80 eigenpairs on the fine grid is $5.7 \cdot 10^{-3}$ or better, and this is enough to be fairly effective. The convergence is 10 times faster than with no deflation, which is illustrated by the last row of the table. The larger coarse-grid matrices do yield more accurate eigenvectors on the fine grid, for instance with residual norms $4.4 \cdot 10^{-5}$ and below for the size 65,025. However, this level of accuracy is not needed.

The two-grid deflated GMRES method can be better than the BiCGStab method in terms of matrix-vector products. So it can be an effective procedure if the matrix is not very sparse or if an expensive preconditioner is used. Otherwise, the GMRES orthogonalization expense can be significant. This motivates replacing GMRES with a restarted BiCGStab method on the fine grid.

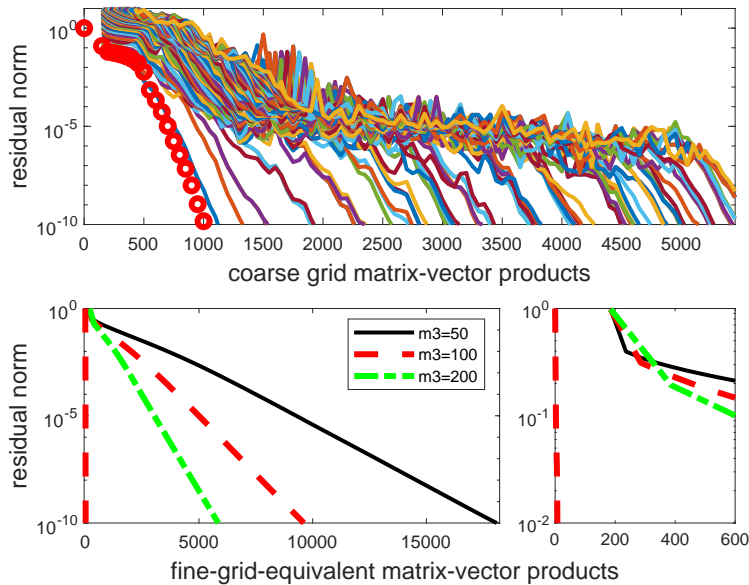


FIG. 3.1. The performance of the two-grid deflated GMRES method is shown for the convection-diffusion example. The fine-grid matrix size is $n = 261,121$, and the coarse-grid matrix size is 3969. The top part corresponds to GMRES-DR(150,100) on the coarse grid; the convergence of the linear equations is shown with circles at the end of each cycle, and the convergence of 80 eigenpairs is shown with lines. The bottom left part has the coarse-grid linear equations solution on the very left, scaled by 64 to correspond to fine-grid matrix-vector products. Then this is followed by GMRES(m_3)-Proj(100) on the fine grid, with $m_3 = 50, 100, 200$. The small bottom right part of the figure is a close-up of the upper left portion of the previous graph.

4. The two-grid deflated BiCGStab method.

4.1. The algorithm.

We wish to use approximate eigenvectors from the coarse grid to deflate eigenvalues from a BiCGStab or IDR method on the fine grid. In [1, 2, 35] a deflated BiCGStab method is given. A single projection is applied before running BiCGStab using both right and left eigenvectors. Here a single projection will not be effective because our eigenvectors are not accurate on the fine grid. So we implement BiCGStab/IDR as a restarted method with projections at each restart. Not only does this allow us to use less accurate eigenvectors, but also it does not require left eigenvectors. We use the same Galerkin projection as for the two-grid deflated GMRES method.

We now give an implementation of this restarted, deflated BiCGStab/IDR method. It replaces phase 3 in the two-grid deflated GMRES algorithm, Algorithm 3, given earlier and thus is part of a two-grid deflated BiCGStab/IDR algorithm. The new BiCGStab(ncyc)-Proj(k) procedure is similar to GMRES(m_3)-Proj(k), but *ncyc* is the total number of BiCGStab cycles, not the length of the cycles (so there are *ncyc* - 1 restarts). As before, *k* gives the number of approximate eigenvectors that are used in the projection step. The stopping criterion for the algorithm uses the minimum of two different quantities. First, we require each cycle to converge to a fraction of the remaining distance to final convergence in terms of orders of magnitude (the relative distance to convergence in orders of magnitude is given by $\frac{rtol \|r_0\|}{\|r\|}$ since $\frac{\|r_0\|}{\|r\|}$ is the current improvement and *rtol* is the required final tolerance). This fraction is determined by the number of remaining cycles (the goal is to equalize the amount of convergence expected of each cycle). Second, for a cycle *icyc* we want to reach a point that is at least the fraction *icyc/ncyc* away from convergence (again in orders of magnitude). The

TABLE 3.1
Effect of coarse-grid size.

coarse grid matrix size	coarse grid cycles	accuracy of 80 eigenpairs on fine grid (max res. norm)	fine grid cycles	cost (millions of fine grid equiv.) vector ops)
$65,025 = 255^2$	1023	$4.4 \cdot 10^{-5}$	85	8.02
$16,129 = 127^2$	270	$3.1 \cdot 10^{-4}$	86	1.40
$3969 = 63^2$	107	$1.4 \cdot 10^{-3}$	94	1.06
$961 = 31^2$	43	$5.8 \cdot 10^{-3}$	106	1.15
$225 = 15^2$	12	$5.7 \cdot 10^{-3}$	120	1.30
$49 = 7^2$	1	$(9.2 \cdot 10^{-3} \text{ for } 40)$	177	1.91
no coarse grid	-	-	1255	13.2

first criteria is usually stronger (asking for further convergence), and so it is the one enforced, but in case the residual norm jumps up during the projection, the second criteria is needed (see Example 8).

Algorithm 4 Two-grid deflated BiCGStab/IDR.

0. Assume k approximate eigenvectors are provided. Let $rtol$ be the specified relative residual tolerance for the linear equations solution. Choose $ncyc$, the requested number of cycles of BiCGStab/IDR.
 1. Apply GMRES-DR(m,k) on the coarse grid. Then transfer the approximate eigenvectors to the fine grid and also transfer the solution of the coarse-grid linear equations as an initial guess for the fine-grid problem.
 2. (If needed:) Improve the eigenvectors on the fine grid using Arnoldi-E (see [37]).
 3. For $icyc = 1 : ncyc$
 - a) Apply a Galerkin projection onto the approximate eigenvectors.
 - b) Let $\|r\|$ be the current residual. Set the relative residual tolerance for this cycle, $rticyc$, to be the minimum (the further convergence point) of $(rtol\|r_0\|/\|r\|)^{\frac{1}{ncyc-icyc+1}}$ and $(\|r_0\|/\|r\|)(rtol)^{\frac{icyc}{ncyc}}$, where r_0 be the initial right-hand side for the fine-grid iteration.
 - c) Run BiCGStab or IDR with a relative residual tolerance of $rticyc$.
 - d) Break out of the loop if $\|r\|$ is already below $rtol$.
-

We will refer to the last phase of these new methods as BiCGStab($ncyc$)-Proj(k) or IDR($ncyc$)-Proj(k). For our tests, the Matlab BiCGStab program is called. For IDR, we use the program described by van Gijzen and Sonneveld in [48], which is available as a MATLAB code from the authors. The default version IDR(4) is called.

EXAMPLE 2. We return to the same convection-diffusion problem as in Example 1 with a dimension $n = 262,121$ and a coarse-grid matrix of size 3969. Figure 4.1 illustrates how the convergence of the restarted, deflated BiCGStab method is effected by the number of cycles. It shows results for BiCGStab($ncyc$)-Proj(100) with $ncyc = 5, 10, \text{ and } 20$. With $ncyc = 5$, the convergence generally slows down during a cycle. The eigenvalue deflation has a good effect initially, but as the cycle proceeds, the small eigencomponents are not small enough, and the iterations must deal with them. For $ncyc = 20$, there are more frequent restarts, so the deflation is performed more frequently. This allows the BiCGStab iterations to not need to deal with small eigencomponents, and so the method converges faster. It is surprising that in spite

TABLE 4.1

Comparing different values of $ncyc$ for BiCGStab($ncyc$)-Proj(100). The number of matrix-vector products is given along with the approximate cost in thousands of vector ops. These are only for the deflated BiCGStab phase. For the overall cost, add 80.2 thousands for the cost before that phase.

$ncyc$	5	10	15	20	30	50	100	150	200
mvp's	9638	7606	6390	5421	5357	5278	(use 99)	(140)	(168)
cost (th's)	116.7	93.3	79.7	69.1	70.3	73.4	83.1	93.4	100.4

of the very jagged behavior of the residual norms of BiCGStab, the overall convergence with $ncyc = 20$ is quite consistent. With even more frequent restarts, the convergence is similar to $ncyc = 20$. These tests are not shown on the figure because they mostly overlie the $ncyc = 20$ curve. However, see Table 4.1 for the number of matrix-vector products needed with some other values of $ncyc$. As $ncyc$ increases, even when the number of matrix-vector products goes down, the cost may go up due to the more projections onto the approximate eigenvectors (see the last row of the table). Note that with large values of $ncyc$, there may be convergence before all cycles are used, and the break out of the loop in part 3d of the algorithm is activated. This is shown in parentheses in the table. For example, only 168 runs are needed when 200 are specified. We also tried restarting BiCGStab after 272 matrix-vector products (the average length for the $ncyc = 20$ test), and the convergence was a little slower than the $ncyc = 20$ method, using 5865 matrix-vector products instead of 5421. Probably the $ncyc = 20$ test has an advantage because it often restarts after an iteration where the residual norm comes down significantly. Next, we try the restarted BiCGStab method with $ncyc = 20$ but no deflation of the eigenvalues. This requires 31,617 total matrix-vector products and shows that deflation is very helpful here.

Figure 4.2 presents several methods. The top portion illustrates convergence with respect to matrix-vector products. The two-grid deflated BiCGStab method uses $ncyc = 20$ and has projection over 100 approximate eigenvectors found from GMRES-DR(150,100) on the coarse grid. This deflated BiCGStab method converges even faster than the two-grid deflated GMRES version. Both deflated methods are much better than regular BiCGStab. IDR does not converge and is not shown. Also given is a two-grid deflated BiCGStab result for solving a second system of linear equations with the same matrix but a different right-hand side (this right-hand side is generated randomly). It converges similarly as for the first right-hand side but does not need the coarse-grid work or a formation of the projection matrix. The bottom of Figure 4.2 provides a plot of convergence versus the approximate cost (calculated as the sum of all length- n vector operations and five times the number of matrix-vector products, since there are about five non-zeros per row of the matrix). We note that the two-grid GMRES method exhibits very slow convergence in this plot because of its orthogonalization. However, the two-grid BiCGStab method converges faster than regular BiCGStab in spite of getting a late start due to the cost of the coarse-grid phase. The second right-hand side does not have this initial cost, and it has less than half the expense of solving the first right-hand side system. The two-grid method presented in this paper is particularly useful for the case of multiple right-hand sides.

The next example at first shows that if regular preconditioning is added to the previous example, then the regular BiCGStab method is less expensive than deflated BiCGStab. So deflation may not be needed for fairly easy problems. Secondly, if the matrix is shifted to give an indefinite problem, then deflation can help again.

EXAMPLE 3. We continue the previous example with several changes: we add incomplete factorization preconditioning, then shift the diagonal to the left to create more difficult

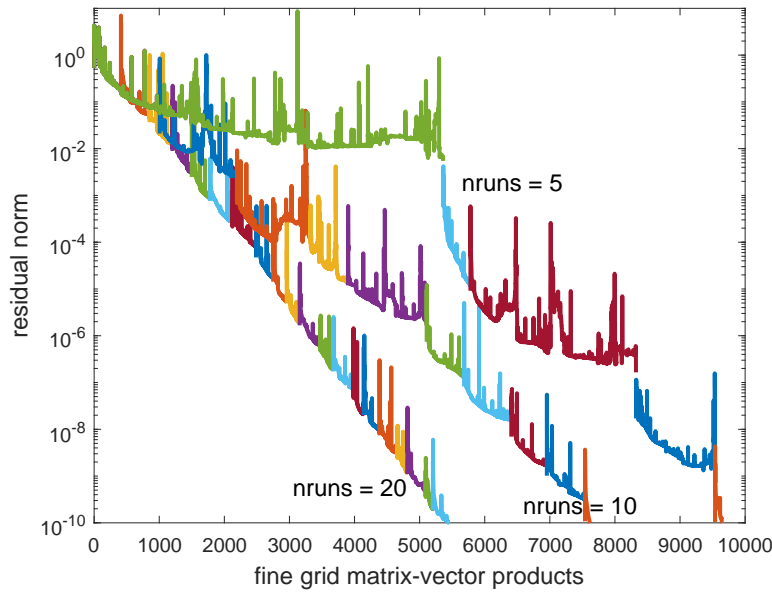


FIG. 4.1. Convergence curves for the two-grid deflated BiCGStab method on the fine grid are shown. GMRES-DR(150,100) is first used on the coarse grid, and then BiCGStab(ncyc)-Proj(100) on the fine grid. The number of cycles for the restarted BiCGStab phase is $ncyc = 5, 10,$ and 20 . The color changes with each new cycle.

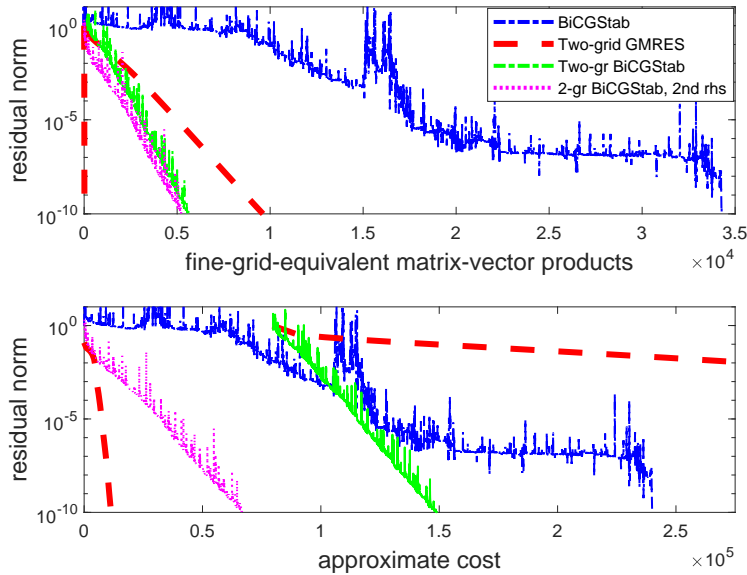


FIG. 4.2. Convergence is shown for the two-grid deflated BiCGStab method, with GMRES-DR(150,100) on the coarse grid and BiCGStab(20)-Proj(100) on the fine grid. This is compared to other methods. The two-grid deflated GMRES method involves GMRES-DR(150,100) on the coarse grid and GMRES(100)-Proj(100) on the fine grid. A test of deflated BiCGStab with a second right-hand side is also shown. In the legend, “Two-grid GMRES” refers to the two-grid deflated GMRES method, “Two-gr BiCGStab” to the two-grid deflated BiCGStab, and “2-gr BiCGStab, 2nd rhs” refers to the two-grid deflated BiCGStab method applied to a second right-hand side but deflated using the approximate eigenvectors from the first right-hand side.

TABLE 4.2

Conv-diff with $nfg = 262,121$ and $ncg = 16,129$. Compare the two-grid method with regular BiCGStab for increasingly more indefinite matrices.

shift	rtol2	Phase 2 cycles	Phase 3 mvps	overall time	BiCGStab mvps	BiCGStab time
0	10^{-2}	0	212	56	765	7.1
10^2	10^{-2}	0	360	59	1842	17
15^2	10^{-2}	0	323	61	12,298	112
20^2	10^{-2}	0	-	-	-	-
20^2	10^{-3}	7	384	216	-	-
25^2	10^{-3}	6	-	-	-	-
25^2	10^{-4}	21	53,949	1031	-	-
25^2	10^{-5}	39	464	970	-	-

problems, and finally solve 10 systems with different right-hand sides. The incomplete factorization ILU(0) is used. Also, the size of the coarse-grid matrix is changed to $ncg = 16,129$. Again, GMRES-DR(150,100) on the coarse-grid develops the approximate eigenvectors. The preconditioning makes the problem much easier, and now the regular BiCGStab method is very effective. The row in Table 4.2 with $shift = 0$ shows that BiCGStab is converging to a residual norm below 10^{-10} with 765 matrix-vector products and in only 7.1 seconds. For the two-grid method, the cost for developing the approximate eigenvectors on the coarse grid and moving them to the fine grid is significant enough that even though less matrix-vector products are needed for the deflated BiCGStab phase on the fine grid, the overall cost is much higher.

We now shift the matrix, first as $A - 10^2 I$. The results are presented in the second row of Table 4.2. The matrix-vector products for the regular BiCGStab method goes up to 1842, but it still runs quickly in 17 seconds (Matlab timing). However, the next row has a shift of 15^2 , and BiCGStab takes much longer. The two-grid approach converges in about half the time. The reduction in matrix-vector products is even greater, down from 12,298 to 323. With further shifting, BiCGStab does not converge. The two-grid method needs some phase 2-improvement of the eigenvectors. For a shift of 25^2 , the eigenvectors need to be improved to accuracy below residual norms of 10^{-5} , and then the deflated BiCGStab phase is very fast. This is expensive because of the time spent in phase 2. Table 4.3 includes some details for the shifts 15^2 and 25^2 (with the eigenvalue improvement just mentioned included for the 25^2 -case). Also, the last two columns have results for nine additional right-hand sides. Unlike the first right-hand side, these are generated randomly from a normal distribution and then normed to one. For these shifts, the two-grid method is better for the first right-hand side and is much faster for additional right-hand sides. With the shift 25^2 , solving an additional right-hand side takes only an average of 6.8 seconds compared to 357 seconds for the regular BiCGStab method. We also note that the two-grid method gives accurate answers for all cases, while the regular BiCGStab method is not able to (only three of the right-hand sides have solutions with residual norms below 10^{-6}).

4.2. Accuracy of the eigenvectors. Here we discuss whether an improvement of the approximate eigenvectors is needed in phase 2 and if so, how accurate the eigenvectors need to be. The main conclusion is negative in the sense that it is difficult to know a priori whether phase 2 is necessary. We have seen this in the previous example where with the shift 15^2 , an accuracy of 10^{-2} is sufficient for the approximate eigenvectors and no phase 2-improvement is needed. Meanwhile, for a shift of 25^2 , a significant improvement is needed so that there is a residual accuracy below 10^{-5} . Next, we show that even for a simple 2 by 2 case, the needed accuracy varies.

TABLE 4.3

Conv-diff with $nfg = 262,121$ and $ncg = 16,129$. Compare two-grid method with regular BiCGStab for indefinite matrices and multiple right-hand sides. The "Total mvps" column give the fine-grid-equivalent matrix-vector products for the whole process, while the columns with "BiCGSt mvps" and "mvps" are for just the deflated BiCGStab or regular BiCGStab method.

shift		First rhs			Average for other 9 rhs's	
		Total mvps	BiCGSt mvps	time	mvps	time
15^2	2-grid	564	323	61	285	5.5
	BiCGSt	12,298	12,298	112	11,401	103
25^2	2-grid	6655	464	970	457	6.8
	BiCGSt	-	-	-	37,741	357

EXAMPLE 4. We consider the diagonal 2 by 2 matrix A with diagonal entries α and 10. The current residual (or right-hand side) for the linear equations is $r = [1 \ 1]^T$. The approximate eigenvector is $y = [1 \ y_2]^T$. We do a projection over the approximate eigenvector and look at how accurate the eigenvector needs to be for the residual to be improved. The goal of the deflation is to reduce the first component of the residual because this is the more difficult one corresponding to the small eigenvalue. We say that the residual is improved if the first component is reduced more in terms of orders of magnitude than the second component is increased. For an improvement with $\alpha = 0.01$, the accuracy of the approximate eigenvector needs to approximately have a residual norm below 0.1. But for $\alpha = 0.001$, the accuracy needs to be below 0.02. Jumping to much smaller α , with $\alpha = 0.000001$, the eigenvector residual norm needs to be below 0.0002. So the point is that with a smaller eigenvalue, a more accurate approximation of the eigenvector is needed for successful deflation. We have seen in past deflation work that eigenvalue residual norms below 10^{-2} are needed for many matrices, but here for some difficult matrices, a higher accuracy is required. An important point is that in the actual solution of large systems of equations, it may not be possible to know ahead of time how accurate the approximate eigenvectors need to be. Some experimentation with the particular linear equations may be necessary.

4.3. Effectiveness of the restarted BiCGStab method. Table 4.1 shows something remarkable for the deflated, restarted BiCGStab method. For $ncyc \geq 20$, the results are fairly close to being invariant of the number of cycles. This is in spite of much smaller subspaces being used for the larger values of $ncyc$. Also, the convergence is at a very consistent pace considering the usual erratic convergence of BiCGStab. The deflated GMRES method in Figure 3.1 is very sensitive to the subspace sizes, with $m3 = 50$ converging three times slower than for $m3 = 200$. The deflated BiCGStab method is not so sensitive: with 20 cycles, it uses an average of 271 matrix-vector products per cycle, while with $ncyc = 100$, an average of 53 per cycle converges slightly faster.

We next investigate the reason for this effectiveness of the restarted BiCGStab method. First, we illustrate that this is not due to the deflation. To show this, the next example does not use deflation and still exhibits a similar phenomenon for the restarted BiCGStab method.

EXAMPLE 5. The matrix is the same as in the previous examples, except the size is $127^2 = 16,129$, and there is no coarse-grid matrix. We use a random right-hand side to try to make the example more general, though usually there is not much influence coming from the right-hand side. We run the restarted BiCGStab method with no deflation between the cycles. Table 4.4 shows results for tests with different numbers of cycles. Surprisingly, even for a large numbers of cycles, the results are similar and often better than for the regular non-restarted BiCGStab method. For example, with $ncyc$ specified to be 400, the method ends up using 198 cycles, 3452 matrix-vector products, and an average number of matrix-vector products per

TABLE 4.4

The restarted BiCGStab method with different numbers of cycles (no deflation is used). The number of matrix-vector products is given.

ncyc	1 (no restart)	5	10	25	50	75 (use 74)	150 (118)	250 (171)	400 (198)
mvp's	3969	3755	4371	3968	3386	3578	4049	3764	3452

cycle of $17\frac{1}{2}$. This compares to 3969 matrix-vector products with the non-restarted BiCGStab method. Also for the GMRES(18) method, 28,838 matrix-vector products are required.

It has been proposed that the GMRES method should involve changing cycle lengths [8]. We suggest that variable cycle lengths is the main reason that the restarted BiCGStab method is so effective even with small subspace sizes. Table 4.5 provides results for the restarted BiCGStab method with average subspace dimensions of approximately 50, 35, and 18 (the lowest dimension we could achieve was about 18 because when a high number of cycles is specified, the algorithm finishes in far fewer cycles than requested). The number of matrix-vector products is similar for each of the three average subspace sizes (these three numbers come from Table 4.4 with *ncyc* to be 75, 150, and 400).

The next column has results for GMRES(*m*) for $m = 50, 35,$ and 18 . As mentioned, the number of matrix-vector products is very large for small m . The third result is for the GMRES method restarted exactly as for BiCGStab (so with the same number of matrix-vector products for each corresponding cycle, though it does not necessarily need all of the cycles). Now the GMRES method uses even less matrix-vector products than BiCGStab. This is a remarkable improvement over the regular restarted GMRES method. The regular GMRES(*m*) method can get stuck in a pattern [9, 52], and changing the sizes of the subspaces can break the pattern.

The final column of the table involves results for a different way of implementing a variable restarting for the GMRES procedure. A maximum cycle length is specified, and the length of each individual cycle is randomly chosen between 1 and the maximal length (for example, a maximal cycle length of 36 is used in the last test because it gives an average of 18). This approach performs much better than the fixed cycle length approach, but it is not as good as using the BiCGStab lengths. We conclude from these tests that our BiCGStab method restarts in a surprisingly effective way. It is important for the new two-grid deflated BiCGStab method that frequent restarts still give an effective method because if only a few restarts are used, then there may not be frequent enough deflations, as seen in Figure 4.1 with the *ncyc* = 5- and 10-cases.

Tables 4.4 and 4.5 make it seem like the sizes of the subspaces used for the BiCGStab and GMRES method, restarted as for BiCGStab, do not have much effect for this example. This would be surprising since it is well known that large subspaces can be an advantage for Krylov methods. And indeed, even though the GMRES procedure does well with these small subspaces, it performs significantly better without restarting, using 1383 matrix-vector products. Since the non-restarted BiCGStab method has 3969 matrix-vector products, we see that here the GMRES method is better capable to take advantage of a large non-restarted Krylov subspace.

5. Further experiments.

5.1. Helmholtz matrix.

EXAMPLE 6. We next test using the Helmholtz matrix from a finite difference discretization of the differential equation

$$-u_{xx} - u_{yy} - 100^2u = f$$

TABLE 4.5
Matrix-vector products for the restarting BiCGStab and GMRES methods.

average cycle length	restarted BiCGStab	GMRES(m), m fixed	GMRES restarted as was BiCGStab	GMRES with random restarts
50	3578	6250	3266	3498
35	4049	11,760	3244	3733
18	3452	28,838	3411	5013

on the unit square with zero boundary conditions. The discretization size is again $h = \frac{1}{512}$ resulting again in $n = 262,121$. The value of $rtol$ is 10^{-10} . The right-hand side is generated randomly from a normal distribution. The coarse-grid discretization size is $h = \frac{1}{128}$, giving a matrix of dimension $127^2 = 16,129$. This problem is difficult because of the significantly indefinite spectrum. See [10, 19, 30, 47] for work on multigrid methods for indefinite problems. Perhaps an easier option is to use multigrid as part of a preconditioner; see the next example.

Figure 5.1 illustrates the convergence of Krylov methods plotted against matrix-vector products. We have found that the IDR method performs better than BiCGStab for indefinite matrices, and only this one is shown in the figure, but even here, IDR does not converge. GMRES-DR(200,150) converges eventually (though not quite to the requested $rtol$), however, it is expensive due to orthogonalization costs. Our two-grid deflated GMRES method first uses GMRES-DR(200,150) on the coarse grid until 120 eigenvalues converge to a residual norm below 10^{-8} . This requires 285 cycles. Then, on the fine grid GMRES(100)-Proj(150) is employed. This method converges faster than GMRES-DR(200,150) and is much less expensive because on the fine grid, GMRES(100) has less orthogonalization. However, two-grid deflated IDR is the best method. It uses IDR(20)-Proj(150) on the fine grid. It converges faster in terms of matrix-vector products and is also much less expensive per matrix-vector product.

We did try deflating only 100 eigenvalues, as was used in the previous examples. However, in this case the deflated IDR method converges more than twice as slow. For this difficult indefinite problem, many eigenvalues are needed for an effective deflation.

Helmholtz problems are fairly complicated. For example, if the wave number is increased and the fine grid uses the same discretization, then the coarse grid may need to be finer than in Example 6 in order to have good enough eigenvector approximations. Our goal here is merely to demonstrate the potential for the new approach; much more work is needed on Helmholtz problems for a thorough study.

EXAMPLE 7. For this example, the problem is the same as in the previous one, but we now use multigrid preconditioning. Since a standard multigrid method does not converge for this matrix, we instead use a preconditioner from the solution of linear equations with the positive shifted Laplacian with the operator $-u_{xx} - u_{yy} + 100^2u$ [29]. With this positive shift, multigrid easily converges and thus can precondition the Helmholtz matrix (the negatively shifted Laplacian). This preconditioning is used on both the fine grid and the coarse grid and makes the problem easier to solve; see Figure 5.2. The problem is still indefinite, and while the BiCGStab method still does not converge, now IDR does. GMRES-DR(150,100) converges in less than half the number of matrix+preconditioner applications than IDR, but it has more orthogonalization expense. The two-grid deflated methods use GMRES-DR(150,100) on the coarse grid and a stopping when 80 eigenpairs have converged to a residual norm 10^{-8} . Then on the fine grid, the deflated GMRES method uses GMRES(100)-Proj(100) and the deflated IDR method uses IDR(5)-Proj(100) (the IDR method converges almost 25% slower with $ncyc = 20$ instead of 5). Both deflated methods converge faster than the other methods, but the deflated IDR method has much less orthogonalization costs.

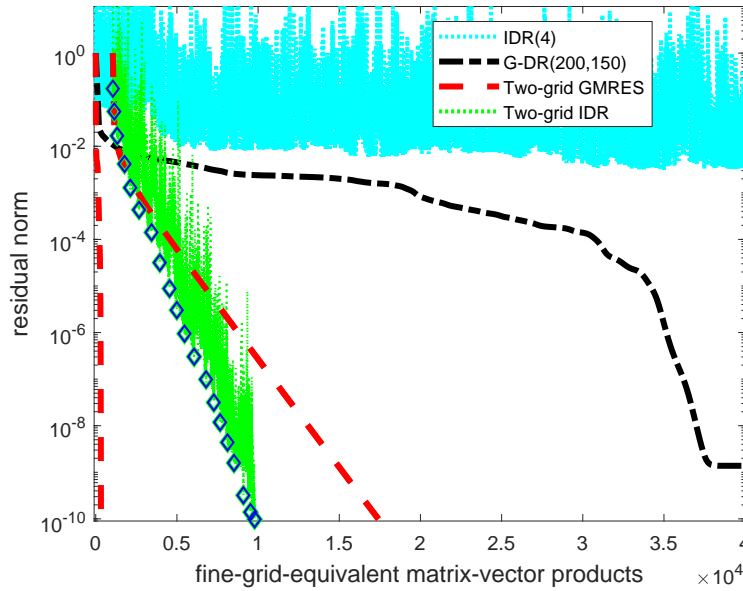


FIG. 5.1. The matrix is from the simple Helmholtz equation with $\kappa = 100$. The two-grid deflated GMRES method uses GMRES-DR(200,150) on the coarse grid and GMRES(100)-Proj(150) on the fine grid. The two-grid deflated IDR method uses GMRES-DR(200,150) on the coarse grid and IDR(20)-Proj(150) on the fine grid; diamonds show the residual norm at the end of each of the 20 cycles. Also compared are IDR and GMRES-DR(200,150). In the legend, “G-DR” refers to GMRES-DR, “Two-grid GMRES” to two-grid deflated GMRES, and “Two-grid IDR” to the two-grid deflated IDR method.

In this experiment, the positively shifted Laplacian linear equations are solved accurately. However, one can relax this and apply the multigrid method only until partial convergence. The methods still work, but the results vary to some extent. We also tried complex shifts for the multigrid preconditioner [22]. Faster convergence is observed at a higher cost per iteration due to the subspaces becoming complex. For a complex Helmholtz problem, this would not be a disadvantage and should be further studied.

5.2. Biharmonic Matrix. We next consider matrices from discretizing a biharmonic differential equation. Matrices from this differential equation quickly become very ill-conditioned as the discretization size gets small. The biharmonic examples demonstrate first that the residual norms can jump up during the projection onto the approximate eigenvectors. The second example illustrates faster convergence if the approximate eigenvectors are improved on the fine grid.

EXAMPLE 8. The partial differential equation is

$$-u_{xxxx} - u_{yyyy} + 40u_{xxx} = f.$$

The matrix sizes are $n = 65,025$ and $ncg = 961$. Central difference formulas are used for discretizing both the third- and fourth-order derivatives. The right-hand side is chosen randomly from a normal distribution. Due to the ill-conditioning, all biharmonic matrix tests have residual tolerance for the linear equations of only 10^{-8} . BiCGStab and IDR do not converge, and GMRES-DR is slow and expensive. We only give results for the deflated BiCGStab method. The top half of Figure 5.3 displays results with three choices of m and k for GMRES-DR(m,k) on the coarse grid. GMRES-DR(100,50) finds 40 eigenpairs with residual

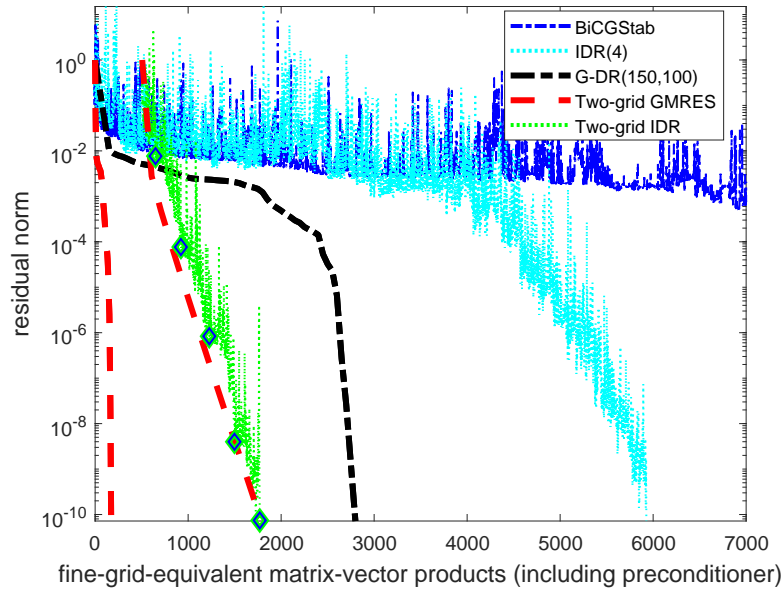


FIG. 5.2. The matrix is from the simple Helmholtz equation with $\kappa = 100$. Now multigrid preconditioning is used. The two-grid deflated GMRES method uses GMRES-DR(150,100) on the coarse grid and GMRES(100)-Proj(100) on the fine grid. The two-grid deflated IDR method uses GMRES-DR(150,100) on the coarse grid and IDR(5)-Proj(100) on the fine grid; diamonds show the residual norm at the end of each of the 5 cycles. These methods are compared to IDR and GMRES-DR(150,100). In the legend, “G-DR” refers to GMRES-DR, “Two-grid GMRES” to the two-grid deflated GMRES, and “Two-grid IDR” to the two-grid deflated IDR method.

norms below 10^{-8} , GMRES-DR(150,100) gets 80 to that tolerance, and GMRES(200,150) stops when 120 have converged. The deflated BiCGStab method is used on the fine grid with $ncyc = 20$ cycles. The convergence is plotted against the cost for both coarse- and fine-grid phases. Here each matrix-vector product is counted as 13 vector operations (the number of non-zeros in most rows). The cost for the coarse-grid phase is higher when more eigenvalues are computed, but then convergence is faster on the fine grid. Using 100 approximate eigenvectors is best overall for this example. If there were multiple right-hand sides, then 150 would be clearly better for subsequent right-hand sides because of the faster fine-grid convergence.

The lower half of Figure 5.3 shows a portion of the fine-grid restarted BiCGStab residual curve with $k = 150$, plotted against matrix-vector products. The residual norm jumps up by a significant amount during each projection onto the approximate eigenvectors. For instance, it increases from $2.5 \cdot 10^{-6}$ to $1.2 \cdot 10^{-5}$ in between the cycles 14 and 15 (at the matrix-vector product 3415). As a result, the second of the two convergence criteria in part 3b of the algorithm is activated. Figure 5.4 displays eigencomponents of the residual vectors for a smaller version of this problem. The matrix is if size $n = 961$, and the coarse-grid matrix is of size 49. Here 10 eigenvectors are computed accurately on the coarse grid and moved to the fine grid. The top of the figure illustrates all 961 eigencomponents during a part of a deflated BiCGStab run on the fine grid. The (red) circles are obtained from the residual vector after the projection at the start of the fifth cycle. Then the (black) squares are the results after BiCGStab has been applied. Finally, the (blue) dots correspond to the results after the next projection at the start of the sixth cycle, and they mostly overlie the circles but are a little better on average. Most of the components increase dramatically with the projection. Then, fortunately they are reduced by the Krylov iteration. The lower part of the figure shows that

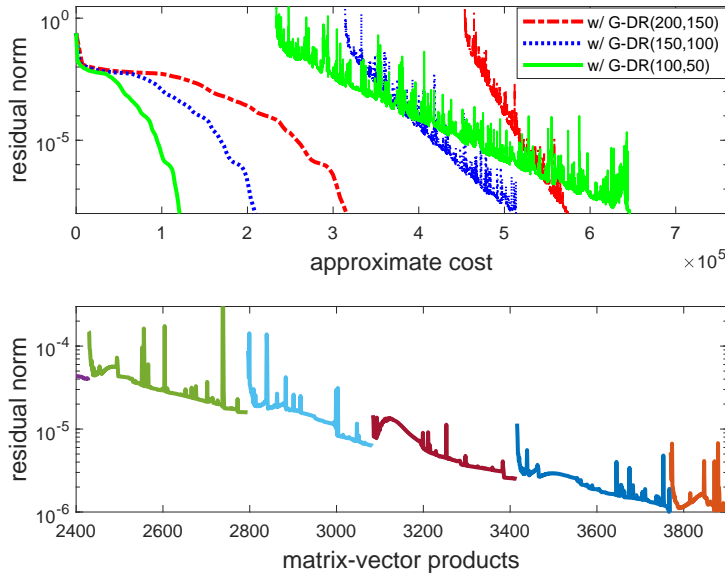


FIG. 5.3. The matrix is from a biharmonic differential equation. The top half corresponds to the deflated BiCGStab method, with convergence shown on both the coarse and fine grids. The coarse grid involves GMRES-DR(m,k) for three choices of m and k . The fine grid involves BiCGStab(20)-Proj(k) for the three values of k . The gap between the two pieces of each of these curves is due to finishing the GMRES-DR iteration for computing the approximate eigenvectors. The bottom half has a view of part of the fine-grid convergence with $k = 150$, showing the jump in the residual norm with each projection. In the legend, “G-DR” is GMRES-DR.

some of the components corresponding to the small eigenvalues are reduced by the projection. This reduction is important because they are for the most part not reduced by the BiCGStab iteration. This essential reduction of the small eigencomponents by the projection makes up for the increase of the other components because overall this deflated method does better than the regular non-restarted BiCGStab method.

We continue the biharmonic example, but now use an incomplete LU factorization. This example also involves improving the eigenvectors on the fine grid.

EXAMPLE 9. We use incomplete factorization preconditioning for the biharmonic differential equation from the previous example. This is done with Matlab’s “ilu” command with no fill, after adding 0.5 to all diagonal elements of the matrix (this is needed to make the preconditioning effective). We choose a finer discretization than in the previous example because the preconditioning allows us to solve a harder problem. The matrix has size $n = 511^2 = 261,121$, and the coarse grid is of size $127^2 = 16,129$. An ILU preconditioner is generated for both the fine and coarse grids. As in the previous example, $rtol$ is set to 10^{-8} .

Table 5.1 illustrates the results of a few tests. The second row of the table (“cycles phase 1”) gives the number of GMRES-DR(150,100) cycles on the coarse grid. The third row gives the cycles of Arnoldi-E on the fine grid used to improve the approximate eigenvectors (phase 2 of the two-grid algorithm). The fourth row gives the number of cycles on the fine grid; this happens to be exactly 200 for GMRES-Proj and is specified to be 50 for the restarted BiCGStab method. The fifth row of the table has the number of applications of matrix-vector product plus preconditioner on the fine grid. Then the next row adds to this the coarse grid mvp’s+preconditionings, scaled by 16 to give the fine-grid-equivalent total. The last row gives the approximate cost, which counts 26 for each matrix-vector product plus preconditioner (13

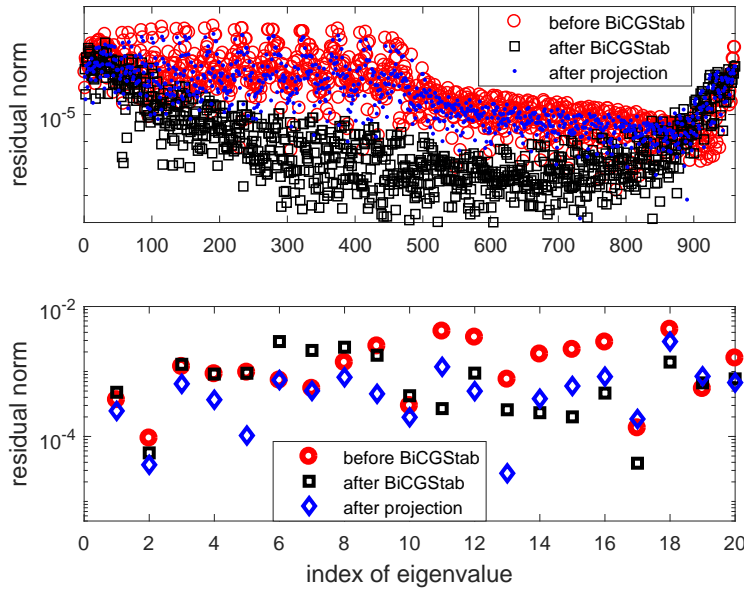


FIG. 5.4. Small matrix from a biharmonic differential equation. The top portion of the figure has eigencomponents of the residual vector before applying BiCGStab in the fifth cycle (red circles), then after the BiCGStab method (black squares), and then after the projection over the approximate eigenvectors (blue dots). The bottom portion has only the first 20 eigencomponents.

for the mvp and 13 for preconditioning), and then adds the vector ops. The fine-grid-equivalent coarse-grid costs are included.

The first result in the table (the second column) is for the regular, non-restarted BiCGStab method. For convergence, this takes 34,907 applications of a matrix-vector product plus a preconditioning. The next column corresponds to the two-grid deflated GMRES method, with GMRES-DR(150,100) on the coarse grid and GMRES(50)-Proj(100) on the fine grid. This method performs much better than BiCGStab in terms of the numbers of applications of matrix-vector product plus preconditioner, using the fine-grid equivalent of 10,203 of them. However, the cost estimate in the last row of the table is only little better (854 thousand compared to 1152 thousand) due to the orthogonalization in the 200 cycles of GMRES(50). If GMRES(100)-Proj(100) is used instead, then the number of cycles on the fine grid reduces to 73, and thus the fine-grid-equivalent matrix-vector product plus preconditioners goes down to 7576. However the cost goes up to 988 thousands. The next column corresponds to the deflated BiCGStab method. The same GMRES-DR run is used on the coarse grid, and then the fine grid uses BiCGStab(50)-Proj(100), with 50 cycles of BiCGStab and projections over the 100 approximate eigenvectors in between. The cost is significantly reduced compared to the other methods. We note that with $ncyc = 20$, there are 12,168 mvp+prec's. For this example, more restarts are needed perhaps because the approximate eigenvectors are not as good and so need to be deflated more often. The next to last column has Arnoldi-E(150,100) applied to the approximate eigenvectors for 10 cycles with the 10 smallest approximate eigenvectors used as starting vectors. Then the last column corresponds to the Arnoldi-E method applied as in [37] until the first 80 approximate eigenvectors have residual norms below 10^{-3} (before this improvement, these 80 have residual norms from $3.3 \cdot 10^{-4}$ to $2.4 \cdot 10^{-2}$). This improvement takes 41 cycles. Both of these raise the cost but reduce the BiCGStab iterations. For cases

TABLE 5.1

Results for the biharmonic matrix preconditioned by ILU factorization (in the first column, “mvp +” refers to a matrix-vector product plus an application of the preconditioner). The first method is the regular non-restarted BiCGStab method. Second is the two-grid deflated GMRES method with GMRES-DR(150,100) on the coarse grid and GMRES(50)-Proj(100) on the fine grid. The last three columns have the same coarse-grid work and then use restarted BiCGStab(50)-Proj(100) on the fine grid. The first of these three tests involve no improvement of the approximate eigenvectors on the fine grid, while the last two have 10 and 41 cycles of improvement, respectively.

fine grid method	BiCGSt	Defl. GMRES	Defl. BiCGSt	Defl. BiCGSt	Defl. BiCGSt
cycles phase 1	-	31	31	31	31
cycles phase 2	-	0	0	10	41
cycles phase 3	1	200	50	50	50
fine grid mvp +	34,907	10,000	7877	6074	3069
total mvp + (fine gr. equi.)	34,907	10,203	8080	6827	5372
cost (in thousands)	1152	854	322	801	2369

with multiple right-hand sides, this eigenvector improvement could be worthwhile. Probably the best way to decide if eigenvectors should be improved is by experimenting.

5.3. QCD problem. An important computational area that especially needs further development of linear solvers is quantum chromodynamics (QCD). Increasingly large and difficult systems of equations are developed in QCD, and so the methods need to increase in effectiveness. Multigrid methods have been developed for QCD [6, 13, 18]. They show potential but are not as efficient as multigrid for standard PDE’s and are much more complicated to implement. Currently, multigrid methods are not standard for QCD but are one possible tool. We wish to add the two-grid deflated BiCGStab method to the pool of possible methods. For this we will use some of the framework that has been previously developed for QCD multigrid. Here we do preliminary testing with a moderately sized problem in the simpler QCD situation of the two-dimensional Schwinger model [41]. It is beyond the scope of this paper to test other models or to give a full comparison with other QCD methods.

EXAMPLE 10. As mentioned above, the matrix comes from the Schwinger model. It is of size $n = 294,912$. A coarse-grid matrix of size $n = 9216$ is constructed using techniques described in [49]. Along with this, prolongation operators are formed for moving vectors from a grid to a finer grid. The original fine-grid matrix has 9 non-zeros per row, while the coarse-grid matrix has 80. There are significant costs involved in forming the coarse-grid matrix, including 1128 matrix-vector products with the fine-grid matrix to create very rough approximations to small eigenvectors that are needed in the process of forming it. There are of course also costs for GMRES-DR on the coarse grid. However, these costs may not be significant if many right-hand sides are solved as is common for QCD problems. Therefore, here we only compare matrix-vector products during the solve phase on the fine grid (phase 3). The matrix is shifted as $A - \sigma I$ by three different σ -values, 0.061, 0.062, and 0.063. What is referred to as “critical mass” is roughly at 0.062 and shifting past that point makes the matrix slightly indefinite and the problem more difficult. See [27] for the physical relevance of such a shift.

The first row of the results in Table 5.2 is for the regular BiCGStab method, which does not converge for the most difficult matrix. It is not unusual for BiCGStab to fail for difficult QCD problems. Then three tests of BiCGStab(20)-Proj(40) are given with increasing computational work on developing the approximate eigenvectors. The second row of results has GMRES-DR(80,40) for 20 cycles on the coarse grid and no phase 2-improvement. This is better than the regular BiCGStab method for the first shift of 0.061, slower for the second

TABLE 5.2

Linear equations are solved for a 2-D QCD matrix. GMRES(80,40) is used on the coarse grid. Fine grid matrix-vector products are given for solving one right-hand side, not including set-up costs for forming the coarse-grid matrix.

Method \ Shift	0.061	0.062	0.063
non-restarted BiCGStab	4880	6187	-
Defl. Bi., 20 cyc phase 1, 0 cyc phase 2	3589	12,284	-
Defl. Bi., 20 cyc phase 1, 15 cyc phase 2	3247	4002	4519
Defl. Bi., 43 cyc phase 1, 15 cyc phase 2	2725	3354	3055

shift, and also does not converge for the third. The third row of results has an added 15 cycles of phase 2-improvement on the fine grid, targeting only the smallest 5 eigenvalues and eigenvectors (this is because in testing, it is more important to have accurate eigenvectors for the eigenvalues near zero). The cost for this improvement is significant for one right-hand side, but it is not as important if many of them are involved. But also, we are now able to achieve convergence for the "difficult" shift. The last row corresponds to the same method except there is more computational work on the coarse grid with 43 phase 2-cycles (this is enough to have the 5 smallest eigenpairs converge to an accuracy of 10^{-8}). These experiments show that the new approach has potential for difficult problems.

6. Conclusion. We have proposed a two-grid method that finds approximate eigenvectors with the coarse grid and uses them to deflate the eigenvalues for linear equations on the fine grid. This includes deflation for the BiCGStab and IDR methods using only approximate right eigenvectors and a novel use of restarting for these normally non-restarted nonsymmetric Lanczos methods. This two-grid deflation is a very efficient way to deflate eigenvalues because the difficult work of finding approximate eigenvectors is done for an easier problem on the coarse grid. This is particularly useful for multiple right-hand side problems because the coarse-grid computation only needs to be done once and then can be applied for all right-hand sides. The new approach can be combined with multigrid preconditioning (see Example 7).

For the deflated, restarted BiCGStab method, one choice that needs to be made is the number of cycles. We have found that 20 cycles is generally a good choice. However, if the approximate eigenvectors are not very accurate, then more cycles with more frequent deflation may help (see Example 9).

Many facets of these two-grid deflated methods could use further investigation. For instance, three-dimensional problems may have greater potential because the coarse-grid matrix can be relatively smaller compared to the fine-grid matrix. QCD problems could use much further investigation, including going to four-dimensional problems. Other possible future work is to use more grid levels; see [37] for a multiple-grid method for computing eigenvalues.

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

REFERENCES

- [1] A. M. ABDEL-REHIM, R. B. MORGAN, AND W. WILCOX, *Deflated BiCGStab for linear equations in QCD problems*, in Proceedings of the XXV International Symposium on Lattice Field Theory—PoS(LATTICE 2007), Proc. Sci. Vol. 48, SISSA, Trieste, 2008, pp. 026/1–026/7.
- [2] A. M. ABDEL-REHIM, A. STATHOPOULOS, AND K. ORGINOS, *Extending the eigCG algorithm to nonsymmetric Lanczos for linear systems with multiple right-hand sides*, Numer. Linear Algebra Appl., 21 (2014), pp. 473–493.

- [3] K. AHUJA, P. BENNER, E. DE STURLER, AND L. FENG, *Recycling BiCGSTAB with an application to parametric model order reduction*, SIAM J. Sci. Comput., 37 (2015), pp. S429–S446.
- [4] O. AKSEL'SON, M. NEICHEVA, AND B. POLMAN, *The bordering method as a preconditioning method*, Vestnik Moskov. Univ. Ser. XV Vychisl. Mat. Kibernet., (1996), pp. 3–25, 81.
- [5] A. AMRITKAR, E. DE STURLER, K. ŚWIRYDOWICZ, D. TAFTI, AND K. AHUJA, *Recycling Krylov subspaces for CFD applications and a new hybrid recycling solver*, J. Comput. Phys., 303 (2015), pp. 222–237.
- [6] R. BABICH, J. BRANNICK, R. C. BROWER, M. A. CLARK, T. A. MANTEUFFEL, S. F. MCCORMICK, J. C. OSBORN, AND C. REBBI, *Adaptive multigrid algorithm for the lattice Wilson-Dirac operator*, Phys. Rev. Lett., 105 (2010), Art. No. 201602, 4 pages.
- [7] J. BAGLAMA, D. CALVETTI, G. H. GOLUB, AND L. REICHEL, *Adaptively preconditioned GMRES algorithms*, SIAM J. Sci. Comput., 20 (1998), pp. 243–269.
- [8] A. H. BAKER, E. R. JESSUP, AND T. V. KOLEV, *A simple strategy for varying the restart parameter in GMRES(m)*, J. Comput. Appl. Math., 230 (2009), pp. 751–761.
- [9] A. H. BAKER, E. R. JESSUP, AND T. MANTEUFFEL, *A technique for accelerating the convergence of restarted GMRES*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 962–984.
- [10] J. H. BRAMBLE, D. Y. KWAK, AND J. E. PASCIAK, *Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1746–1763.
- [11] A. BRANDT, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.
- [12] A. BRANDT, S. MCCORMICK, AND J. RUGE, *Multigrid methods for differential eigenproblems*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 244–260.
- [13] J. BRANNICK, A. FROMMER, K. KAHL, B. LEDER, M. ROTTMANN, AND A. STREBEL, *Multigrid preconditioning for the overlap operator in lattice QCD*, Numer. Math., 132 (2016), pp. 463–490.
- [14] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, New York, 2000.
- [15] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.
- [16] K. BURRAGE AND J. ERHEL, *On the performance of various adaptive preconditioned GMRES strategies*, Numer. Linear Algebra Appl., 5 (1998), pp. 101–121.
- [17] A. CHAPMAN AND Y. SAAD, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1997), pp. 43–66.
- [18] S. D. COHEN, R. C. BROWER, M. A. CLARK, AND J. C. OSBORN, *Multigrid algorithms for domain-wall Fermions*, in Proceedings of XXIX International Symposium on Lattice Field Theory—PoS(Lattice 2011), Proc. Sci. Vol. 139, SISSA, Trieste, 2012, pp. 030/1-030/7.
- [19] S. COOLS, B. REPS, AND W. VANROOSE, *A new level-dependent coarse grid correction scheme for indefinite Helmholtz problems*, Numer. Linear Algebra Appl., 21 (2014), pp. 513–533.
- [20] H. C. ELMAN, O. G. ERNST, AND D. P. O'LEARY, *A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations*, SIAM J. Sci. Comput., 23 (2001), pp. 1291–1315.
- [21] Y. A. ERLANGGA AND R. NABBEN, *Multilevel projection-based nested Krylov iteration for boundary value problems*, SIAM J. Sci. Comput., 30 (2008), pp. 1572–1595.
- [22] Y. A. ERLANGGA, C. W. OOSTERLEE, AND C. VUIK, *A novel multigrid based preconditioner for heterogeneous Helmholtz problems*, SIAM J. Sci. Comput., 27 (2006), pp. 1471–1492.
- [23] R. P. FEDORENKO, *The speed of convergence of one iterative process*, USSR Comput. Math. Math. Phys., 4 (1964), pp. 227–235.
- [24] J. FRANK AND C. VUIK, *On the construction of deflation-based preconditioners*, SIAM J. Sci. Comput., 23 (2001), pp. 442–462.
- [25] A. GAUL, M. H. GUTKNECHT, J. LIESEN, AND R. NABBEN, *A framework for deflated and augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 495–518.
- [26] L. GIRAUD, S. GRATTON, X. PINEL, AND X. VASSEUR, *Flexible GMRES with deflated restarting*, SIAM J. Sci. Comput., 32 (2010), pp. 1858–1878.
- [27] M. GOLTERMAN, S. SHARPE, AND R. SINGLETON, *Effective theory for quenched lattice QCD and the Aoki phase*, Phys. Rev. D, 71 (2005), Art. No. 094503, 17 pages.
- [28] S. A. KHARCHENKO AND A. Y. YEREMIN, *Eigenvalue translation based preconditioners for the GMRES(k) method*, Numer. Linear Algebra Appl., 2 (1995), pp. 51–77.
- [29] A. L. LAIRD AND M. B. GILES, *Preconditioned iterative solution of the 2D Helmholtz equation*, Tech. Rep. 02/12, Oxford Computer Laboratory, Oxford, 2002.
- [30] J. MANDEL, *Multigrid convergence for nonsymmetric, indefinite variational problems and one smoothing step*, Appl. Math. Comput., 19 (1986), pp. 201–216.
- [31] L. A. M. MELLO, E. DE STURLER, G. H. PAULINO, AND E. C. N. SILVA, *Recycling Krylov subspaces for efficient large-scale electrical impedance tomography*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 3101–3110.
- [32] R. B. MORGAN, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.

- [33] ———, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Math. Comp., 65 (1996), pp. 1213–1230.
- [34] ———, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
- [35] R. B. MORGAN AND D. A. NICELY, *Restarting the nonsymmetric Lanczos algorithm for eigenvalues and linear equations including multiple right-hand sides*, SIAM J. Sci. Comput., 33 (2011), pp. 3037–3056.
- [36] R. B. MORGAN AND W. WILCOX, *Deflated iterative methods for linear equations with multiple right-hand sides*, Preprint on arXiv, 2004. <https://arxiv.org/abs/math-ph/0405053>
- [37] R. B. MORGAN AND Z. YANG, *Two-grid and multiple-grid Arnoldi for eigenvalues*, SIAM J. Sci. Comput., 40 (2018), pp. A3470–A3494.
- [38] M. P. NEUENHOFEN AND C. GREIF, *Mstab: stabilized induced dimension reduction for Krylov subspace recycling*, SIAM J. Sci. Comput., 40 (2018), pp. B554–B571.
- [39] A. PADIY, O. AXELSSON, AND B. POLMAN, *Generalized augmented matrix preconditioning approach and its application to iterative solution of ill-conditioned algebraic systems*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 793–818.
- [40] M. L. PARKS, E. DE STURLER, G. MACKEY, D. D. JOHNSON, AND S. MAITI, *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput., 28 (2006), pp. 1651–1674.
- [41] J. S. SCHWINGER, *Gauge invariance and mass. II.*, Phys. Rev., 128 (1962), pp. 2425–2429.
- [42] J. SIFUENTES, *Preconditioned Iterative Methods for Inhomogeneous Acoustic Scattering Applications*, Ph.D Thesis, Rice University, Houston, 2010.
- [43] 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.
- [44] K. M. SOODHALTER, *Block Krylov subspace recycling for shifted systems with unrelated right-hand sides*, SIAM J. Sci. Comput., 38 (2016), pp. A302–A324.
- [45] K. M. SOODHALTER, D. B. SZYLD, AND F. XUE, *Krylov subspace recycling for sequences of shifted linear systems*, Appl. Numer. Math., 81 (2014), pp. 105–118.
- [46] A. STATHOPOULOS AND K. ORGINOS, *Computing and deflating eigenvalues while solving multiple right-hand side linear systems with an application to quantum chromodynamics*, SIAM J. Sci. Comput., 32 (2010), pp. 439–462.
- [47] U. TROTTEBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, San Diego, 2001.
- [48] M. B. VAN GIJZEN AND P. SONNEVELD, *Algorithm 913: an elegant IDR(s) variant that efficiently exploits biorthogonality properties*, ACM Trans. Math. Software, 38 (2011), Art. No. 5, 19 pages.
- [49] T. WHYTE, W. WILCOX, AND R. B. MORGAN, *Deflated GMRES with multigrid for lattice QCD*, Phys. Lett. B, 803 (2020), Art. No. 135281, 4 pages.
- [50] J. XU AND A. ZHOU, *A two-grid discretization scheme for eigenvalue problems*, Math. Comp., 70 (2001), pp. 17–25.
- [51] Y.-K. YONG AND Y. CHO, *Algorithms for eigenvalue problems in piezoelectric finite element analyses*, in 1994 Proceedings of IEEE Ultrasonics Symposium, vol. 2, IEEE Conference Proceedings, Los Alamitos, 1994, pp. 1057–1062.
- [52] B. ZHONG AND R. B. MORGAN, *Complementary cycles of restarted GMRES*, Numer. Linear Algebra Appl., 15 (2008), pp. 559–571.