

DEFLATED AND RESTARTED SYMMETRIC LANCZOS METHODS FOR EIGENVALUES AND LINEAR EQUATIONS WITH MULTIPLE RIGHT-HAND SIDES*

ABDOU M. ABDEL-REHIM[†], RONALD B. MORGAN[‡], DYWAYNE A. NICELY[§], AND
WALTER WILCOX[¶]

Abstract. A deflated restarted Lanczos algorithm is given for both solving symmetric linear equations and computing eigenvalues and eigenvectors. The restarting limits the storage so that finding eigenvectors is practical. Meanwhile, the deflating from the presence of the eigenvectors allows the linear equations to generally have good convergence in spite of the restarting. Some reorthogonalization is necessary to control roundoff error, and several approaches are discussed. The eigenvectors generated while solving the linear equations can be used to help solve systems with multiple right-hand sides. Experiments are given with large matrices from quantum chromodynamics that have many right-hand sides.

Key words. linear equations, deflation, eigenvalues, Lanczos, conjugate gradient, QCD, multiple right-hand sides, symmetric, Hermitian

AMS subject classifications. 65F10, 65F15, 15A06, 15A18

DOI. 10.1137/080727361

1. Introduction. We consider a large matrix A that is either real symmetric or complex Hermitian. We are interested in solving the system of equations $Ax = b$, possibly with multiple right-hand sides, and in solving the associated eigenvalue problem. Both eigenvalues and eigenvectors are desired. Symmetric and Hermitian problems can take advantage of fast algorithms such as the conjugate gradient (CG) method [21, 52] for linear equations and the related Lanczos algorithm [24, 46] for eigenvalues. However, regular CG can be improved upon for the case of multiple right-hand sides, and Lanczos may have storage and accuracy issues while computing eigenvectors. We give new methods for these problems.

An approach is presented called Lanczos with deflated restarting (Lan-DR). It simultaneously solves the linear equations and computes the eigenvalues and eigenvectors. A restarted Krylov subspace approach is used for the linear equations, but it also saves approximate eigenvectors at the restart and uses them in the subspace for the next cycle. The restarting of the Lanczos algorithm does not slow down the convergence as it normally would, because once the approximate eigenvectors are accurate enough, they essentially remove or deflate the associated eigenvalues from the problem. The eigenvalue portion of Lan-DR has already been presented in [67], but

*Received by the editors June 15, 2008; accepted for publication (in revised form) March 20, 2009; published electronically February 5, 2010. This work was partially supported by the National Science Foundation, by Computational Mathematics Program grant 0310573, and by the National Computational Science Alliance. It utilized the Baylor High-Performance Computing Cluster. The second author was also supported by the Baylor University Sabbatical Program.

<http://www.siam.org/journals/sisc/32-1/72736.html>

[†]Department of Computer Science, College of William and Mary, Williamsburg, VA 23187-8795 (amrehim@cs.wm.edu).

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

[§]Department of Natural Science and Mathematics, University of Charleston, 2300 MacCorkle Avenue SE, Charleston, WV 25304 (dywaynenicely@ucwv.edu).

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

as mentioned, we add on the solution of linear equations. Also, some reorthogonalization is necessary to control roundoff error. We give some new approaches for this reorthogonalization. We also give a minimal residual (MINRES)/harmonic version of the algorithm.

An important example where both linear equations need to be solved and eigenvectors are desired is the case of multiple right-hand sides. The eigenvector information generated for one right-hand side can be used to improve the convergence of the systems with subsequent right-hand sides. We give a method called deflated conjugate gradients (D-CG) that implements such an approach. For the second and subsequent right-hand sides, there is first a projection over the approximate eigenvectors generated by Lan-DR on the first right-hand side. Then the CG iteration is applied. We will apply this to large Hermitian systems of linear equations in quantum chromodynamics (QCD).

Section 2 has a review of methods that this paper builds on. The Lan-DR method for eigenvalues and linear equations is presented in section 3. Section 4 has comparison of reorthogonalization approaches. Multiple right-hand sides are considered in section 5. Then section 6 has MINRES-DR and deflated MINRES, which are of interest particularly for indefinite systems and interior eigenvalues.

2. Review.

2.1. Restarted methods for eigenvalue problems. Restarted Krylov methods for nonsymmetric eigenvalue problems took a jump forward with the implicitly restarted Arnoldi method (IRAM) by Sorensen [62]. IRAM restarts with several vectors, so it not only computes many eigenvectors simultaneously, but it also has improved convergence. At the time of a restart, let the Ritz vectors be $\{y_1, y_2, \dots, y_k\}$, and let r be a multiple of the residual vectors for these Ritz vectors (the residuals are parallel). Then the next cycle of IRAM builds the subspace

$$(2.1) \quad \text{Span}\{y_1, y_2, \dots, y_k, r, Ar, A^2r, A^3r, \dots, A^{m-k-1}r\}.$$

This subspace is equivalent [29] to

$$(2.2) \quad \text{Span}\{y_1, y_2, \dots, y_k, Ay_i, A^2y_i, A^3y_i, \dots, A^{m-k}y_i\}.$$

This last form helps to show why IRAM is effective. The subspace contains a Krylov subspace with each approximate eigenvector as the starting vector.

Versions of IRAM for symmetric problems are given in [6, 4]. A mathematically equivalent method to IRAM that does not use the implicit restarting is in [29] (this approach is equivalent at the end of each cycle). Wu and Simon present a mathematically equivalent approach for symmetric problems called thick restarted Lanczos (TRLAN) [67]. They put the approximate eigenvectors at the beginning of a new subspace instead of at the end as was done in [29]. Stewart gives a framework for restarted Krylov methods in [64]. See [68] for a symmetric block version. Nonsymmetric and harmonic versions of restarted Arnoldi following the TRLAN approach are in [35]. For block Arnoldi methods, see [2] and its references.

2.2. Deflated restarted methods for linear equations. Deflated Krylov methods for linear equations compute eigenvectors and use them to deflate eigenvalues and thus improve convergence of the linear equations solution. For problems that have slow convergence due to small eigenvalues, deflation can make a big difference. For nonsymmetric problems, approximate eigenvectors are formed during GMRES [54] in

[28, 22, 14, 8, 53, 3, 5, 25, 11, 30, 31]. Methods in [11] are related in that they save information during restarted GMRES in order to improve convergence.

Now for symmetric problems, it is assumed in [38] that there is a way to get approximate eigenvectors which are used for a deflated CG. Eigenvectors are formed during CG in [55].

We now look at a particular deflated GMRES method that will be used in this paper. GMRES-DR(m, k) [31] uses the subspace

$$(2.3) \quad \text{Span}\{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_k, r_0, Ar_0, A^2r_0, A^3r_0, \dots, A^{m-k-1}r_0\},$$

where r_0 is the initial residual for the linear equations at the start of the new cycle, and $\{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_k\}$ are the harmonic Ritz vectors corresponding to the smallest harmonic Ritz values. The dimension of the whole subspace is m , including the k approximate eigenvectors (the harmonic Ritz vectors). This can be viewed as a Krylov subspace generated with starting vector r_0 augmented with approximate eigenvectors. Remarkably, the whole subspace turns out to be a Krylov subspace itself (though not with r_0 as starting vector) [30]. FOM-DR [31] is a version that, as in the full orthogonalization method (FOM) [49], uses a Galerkin projection instead of MINRES projection. FOM-DR also needs regular Ritz vectors instead of harmonic.

2.3. Multiple right-hand sides. Systems with multiple right-hand sides occur in many applications (see [17] for some examples). Block methods are a standard way to solve systems with multiple right-hand sides (see, for example, [40, 52, 17, 32, 20]). However, block methods are not ideal for every circumstance. Other approaches for multiple right-hand sides use information from the solution of the first right-hand side (and possibly others) to assist subsequent right-hand sides. Seed methods [61, 7, 23, 45, 51, 66, 15] project over entire subspaces generated while solving previous right-hand sides. Simoncini and Gallopoulos [59, 60] suggest methods including using seeding, using blocks, and using Richardson iteration with a polynomial generated from GMRES applied to the first right-hand side. In [33] a small subspace is generated with GMRES-DR applied to the first right-hand side that contains important information about approximate eigenvectors, and this is used to improve the subsequent right-hand sides. See [44] for a method for multiple right-hand sides that can also handle a changing matrix.

3. Lanczos with deflated restarting. We propose a restarted Lanczos method that both solves linear equations and computes eigenvalues and eigenvectors. It is called Lanczos with deflated restarting (Lan-DR). The Lan-DR method is a version of FOM-DR [31] for symmetric and Hermitian problems and is closely related to GMRES-DR [31]. As mentioned earlier, the eigenvalue portion of Lan-DR is TRLAN [67] and is mathematically equivalent to the IRAM [62].

For Lan-DR, the number of desired eigenvectors k must be chosen, along with which eigenvalues are to be targeted. Normally the eigenvalues nearest the origin are the most important ones for deflation purposes, but other eigenpairs can be computed. In particular, deflating large outstanding eigenvalues may help convergence of the linear equations solution and may be useful for stability.

At the time of a restart, let r_0 be the residual vector for the linear equations and let the Ritz vectors from the previous cycle be $\{y_1, y_2, \dots, y_k\}$. Then the next cycle of Lan-DR builds the subspace

$$(3.1) \quad \text{Span}\{y_1, y_2, \dots, y_k, r_0, Ar_0, A^2r_0, A^3r_0, \dots, A^{m-k-1}r_0\}.$$

Lan-DR generates the recurrence formula

$$(3.2) \quad AV_m = V_{m+1}\bar{T}_m,$$

where V_m is an n by m matrix whose columns span the subspace (3.1) and V_{m+1} is the same except for an extra column. Also \bar{T}_m is an $m+1$ by m matrix that is tridiagonal except for the $k+1$ by $k+1$ leading portion. This portion is nonzero only on the main diagonal (which has the Ritz values) and in the $k+1$ row and column. A part of recurrence (3.2) can be separated out to give

$$(3.3) \quad AV_k = V_{k+1}\bar{T}_k,$$

where V_k is an n by k matrix whose columns span the subspace of Ritz vectors, V_{k+1} is the same except for an extra column, and \bar{T}_k is the leading $k+1$ by k portion of T_m . This recurrence allows access to both the approximate eigenvectors (the Ritz vectors) and their products with A while requiring storage of only $k+1$ vectors of length n . The approximate eigenvectors in Lan-DR span a small Krylov subspace of dimension k [31].

It is necessary to maintain some degree of orthogonality of the columns of V_{m+1} . We mention here an approach to reorthogonalization that we call k -selective reorthogonalization (k -SO). For cycles after the first one, all new Lanczos vectors are reorthogonalized against the k Ritz vectors. This uses Parlett and Scott's idea of selective reorthogonalization [47], but is more natural in this setting, because we are already computing the approximate eigenvectors. Also, because of the restarting, there is no need to store a large subspace. Simon's partial reorthogonalization [58, 67] is also a possibility, as is periodic reorthogonalization [19]. See section 4 for more on this, including hybrid approaches. Full reorthogonalization or partial reorthogonalization can be used for the first cycle, but this may not be necessary. By Paige's theorem [41, 46, 65], orthogonality is lost only when some eigenvectors begin to converge. This may not happen in the first cycle if there are no outstanding eigenvalues. In our tests, we do reorthogonalize for the first cycle.

Lan-DR(m, k)

1. *Start.* Choose m , the maximum size of the subspace, and k , the desired number of approximate eigenvectors. If there is an initial guess, x_0 , then the problem becomes $A(x - x_0) = r_0$.
2. *First cycle.* Apply m iterations of the standard symmetric Lanczos algorithm. This computes the matrix V_{m+1} that has the Lanczos vectors as columns and the $m+1$ by m tridiagonal matrix \bar{T}_m . In addition, fully reorthogonalize all the Lanczos vectors.
3. *Eigenvector computation.* Compute the k smallest (or others, if desired) eigenpairs, (θ_i, g_i) , of T_m , the m by m portion of \bar{T}_m . The g_i should be normalized. For $i = 1, \dots, k$, form $y_i = V_m g_i$. The shortcut residual norm formula is $\|r_i\| = |t_{m+1,m} g_{m,i}|$, where $g_{m,i}$ is the last element of g_i .
4. *Linear equations.* Let $c_m = V_k^T r_0$ (for the first cycle, this is all zeros except for $\|r_0\|$ in the first position; for other cycles it is all zeros except for $\|r_0\|$ in the $k+1$ position). Solve $T_m d = c_m$, and set $\tilde{x} = x_0 + V_m d$. Then $r = r_0 - A\tilde{x} = r_0 - V_{m+1}\bar{T}_m d$. If satisfied with convergence of the linear equations and the eigenvalues, stop. If not, let the new $x_0 = \tilde{x}$ and $r_0 = r$ and continue.
5. *Restart.* For $i = 1, \dots, k$, reassign $v_i = y_i$. Set v_{k+1} to be the previous v_{m+1} . Set the k by k portion of \bar{T}_k^{new} to be the diagonal matrix with the Ritz

values θ_i as diagonal entries. For $i = 1, \dots, k$, the $k+1, i$ element of \bar{T}_k^{new} is computed by $t_{k+1,i}^{new} = t_{m+1,m} g_{m,i}$. Set $\bar{T}_k = \bar{T}_k^{new}$.

6. *Main iteration.* First we compute the v_{k+2} vector. Compute $w = Av_{k+1} - \sum_{i=1}^k t_{k+1,i} v_i$, then $t_{k+1,k+1} = v_{k+1}^T w$ and $w = w - t_{k+1,k+1} v_{k+1}$. Let $t_{k+2,k+1} = \|w\|$. Set $v_{k+2} = w/\|w\|$. Next, compute the other v_i vectors, for $i = k+3, \dots, m+1$, using the standard Lanczos iteration. \bar{T}_m is formed at the same time. Also reorthogonalize the v_i vectors as desired (see the next section). Go to step 3.

We give the expense for Lan-DR(m, k) without reorthogonalization (see the next section for reorthogonalization expenses). The cost is $m - k$ matrix-vector products per cycle and about $6(m - k) + (k + 2)m$ length n vector operations (such as dot products and daxpy's) per cycle. For k small relative to m , Lan-DR uses one matrix-vector product and roughly $k + 8$ vector operations per iteration. For k near $m/2$, there are about $2k + 10$ vector operations per iteration. This compares to about 5 length n vector operations per iteration for CG. Of course, CG does not calculate eigenvectors like Lan-DR does.

Example 1. We use a test matrix that has many small eigenvalues. It is a diagonal matrix of dimension $n = 5000$ whose diagonal elements are $0.1, 0.2, 0.3, \dots, 9.8, 9.9, 10, 11, 12, \dots, 4909, 4910$. The right-hand side is a random normal vector. We apply the method Lan-DR(100,40), which at each restart keeps the 40 Ritz vectors corresponding to the smallest Ritz values and then builds a subspace of dimension 100 (including the 40 approximate eigenvectors). With k -SO, at every iteration of all of the cycles except the first, there is a reorthogonalization against 40 vectors. We first look at how well Lan-DR computes the eigenvalues. Figure 1 shows the residual norms for the 40 Ritz vectors. The desired number of eigenvalues is 30 and the desired residual tolerance is 10^{-8} . It takes 57 cycles for the first 30 eigenvalues to reach this level. Since this is a fairly difficult problem with eigenvalues clustered together, it takes a while for eigenvalues to start converging. However, from that point, eigenvalues converge regularly, and it can be seen that many eigenvalues and their eigenvectors can be computed accurately. The orthogonalization costs are significantly less than for fully reorthogonalized IRAM (about 84 vector operations for orthogonalization per Lan-DR iteration versus an average of 280 for IRAM). For a matrix that is fairly sparse so that the matrix-vector product is inexpensive (and also has a cheap preconditioner, if there is one), the difference in orthogonalization is significant.

We continue the example by comparing Lan-DR(100,40) to unrestarted Lanczos. Figure 2 has the residual norms for the smallest and 30th eigenvalues with each method. The results are very similar for the first eigenvalue, in spite of the fact that Lan-DR is restarted. The presence of the approximate eigenvectors corresponding to the nearby eigenvalues essentially deflates them and thus gives good convergence for Lan-DR. For eigenvalue 30, Lan-DR trails unrestarted Lanczos but is still competitive. This is significant, since Lan-DR(100,40) requires storage of only about 100 vectors compared to nearly 3000 for unrestarted Lanczos.

Next, we look at the solution of the linear equations. Figure 3 shows Lan-DR with three choices of m and k and also shows CG. The convergence of Lan-DR(100,40) is very close to that of CG. The deflation of eigenvalues in Lan-DR(100,40) allows it to compete with an unrestarted method such as CG. Lan-DR(100,10) is not too far behind CG, but Lan-DR(30,10) restarts too frequently and is much slower.

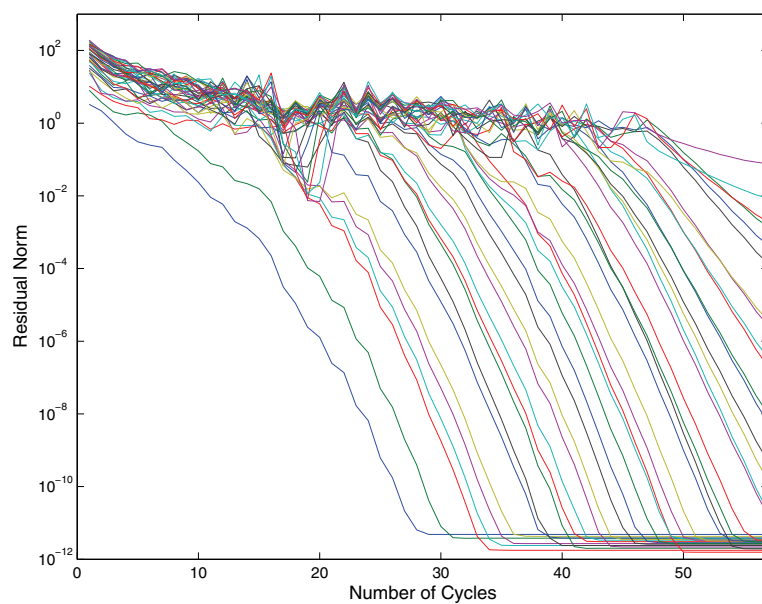


FIG. 1. Computing many eigenvalues of a matrix with small eigenvalues.

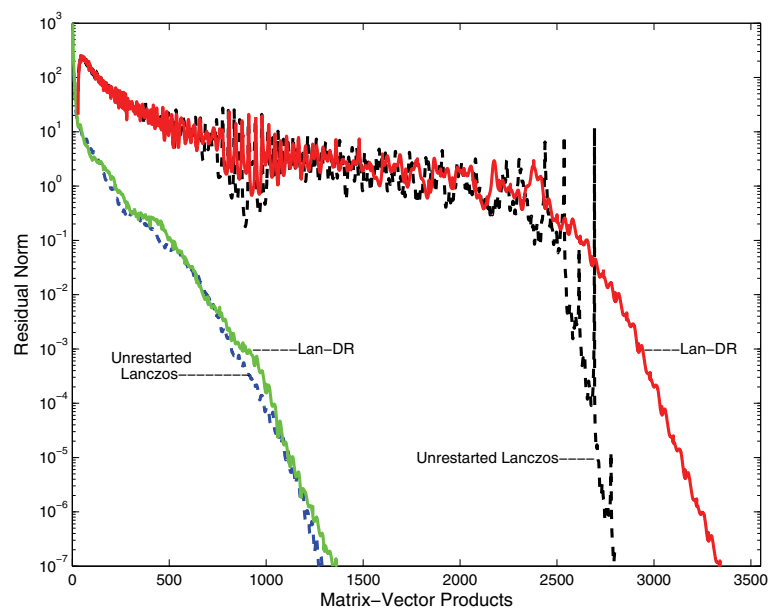


FIG. 2. Comparison of Lan-DR(100,40) with unrestarted Lanczos for first and 30th eigenvalues.

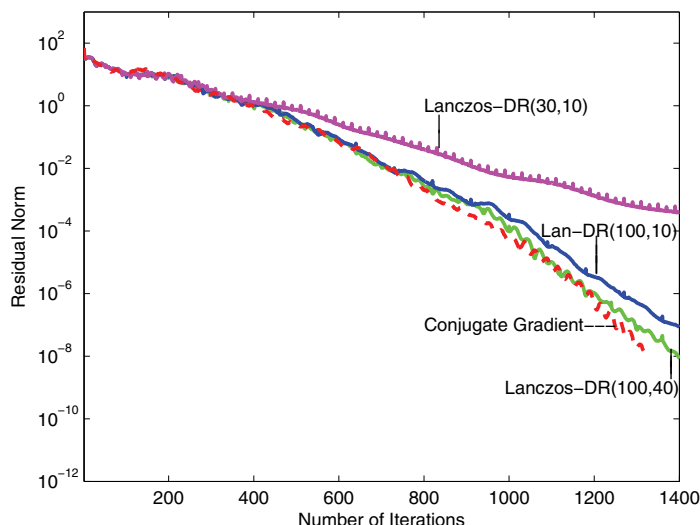


FIG. 3. Comparison of *Lan-DR* with *CG* for solving linear equations.

4. Reorthogonalization. It was mentioned earlier that some reorthogonalization is necessary to control roundoff error. We now look at this in more detail and give several possible approaches. The first is full reorthogonalization, which takes every Lanczos vector formed by the three-term recurrence and reorthogonalizes it against every previous vector. The expense for this in terms of vector operations of length n varies from about $2k$ to $2m$ per iteration. The second approach is periodic reorthogonalization [19]. For this, we always reorthogonalize the v_{k+1} and v_{k+2} vectors, and then at regular intervals reorthogonalize two consecutive vectors (see [67, 65] for why consecutive vectors need to be reorthogonalized). The cost for this varies, as with full reorthogonalization when it is applied, but it saves considerably if the reorthogonalization is not frequently needed. The next approach is partial reorthogonalization (PRO) [58, 57, 67, 65], which monitors loss of orthogonality and thus determines when to reorthogonalize. As suggested in [67], we use “global orthogonalization,” so when we reorthogonalize, it is against all previous vectors. As with the periodic approach, we always reorthogonalize the v_{k+1} and v_{k+2} vectors and always reorthogonalize two consecutive vectors. This approach can be cheaper than periodic reorthogonalization, because it waits until reorthogonalization is needed.

The next three reorthogonalization methods are related to the three above, but reorthogonalization is done only against the first k vectors. Here we are using the idea from selective reorthogonalization (SO) [47] that orthogonality is lost only in the direction of converged or converging Ritz vectors [41, 46]. Since restarting is used, normally only the k Ritz vectors that are kept at the restart have an opportunity to converge. There can be exceptions, as will be discussed in Examples 3 and 4. The first SO-type approach has been used in the previous section. It is called k -SO and has reorthogonalization at every iteration against the k Ritz vectors. This requires about $2k$ vector operations per iteration versus an average of about $k + m$ vector operations for full reorthogonalization. The last two methods are k -periodic and k -PRO. These are the same as the periodic approach and PRO, except they reorthogonalize against only the first k vectors.

TABLE 1
Full reorthogonalization versus reorthogonalize only at the restart.

	Orthogonality	rn 1	rn 30
$k+1, k+2$ vectors only	2.2×10^{-12}	5.6×10^{-12}	9.9×10^{-9}
Full reorthog.	1.2×10^{-14}	6.7×10^{-12}	9.9×10^{-9}

Example 2. We consider again Lan-DR for the matrix of Example 1. For this problem, loss of orthogonality is controlled by the restarting and the reorthogonalization of the v_{k+1} and v_{k+2} vectors at the restart. No further reorthogonalization is needed. Table 1 shows a comparison with full reorthogonalization. The second column gives the loss of orthogonality as measured by $\|V_m^T V_m - I_{m \times m}\|$ at the end of 57 cycles. The next two columns have the residual norms of the first and 30th Ritz pairs. While full reorthogonalization gives greater orthogonality of the Lanczos vectors, the Ritz vectors end up with similar accuracy. The 30th eigenvector continues to converge beyond cycle 57 and eventually reaches a residual norm of 3.7×10^{-12} even with the approach of reorthogonalizing only at the restart. For this example, the Ritz vectors converge slowly enough that we do not have a Ritz vector appear and significantly converge in one cycle (see Figure 1). So before an eigenvector has converged, an approximation to it is among the group of Ritz vectors that v_{k+1} and v_{k+2} are reorthogonalized against. This explains why reorthogonalizing at restarts is all that is needed.

This example points out that restarting can make reorthogonalization easier. Reorthogonalization against only 40 vectors is done for two vectors every 60 iterations. If we compare this to unrestarted Lanczos using PRO with global reorthogonalization and with a tolerance on loss of orthogonality of the square root of machine epsilon, we see that the number of vectors reorthogonalized is similar. However, as the unrestarted Lanczos iteration proceeds, there are many previous vectors to reorthogonalize against. Also unrestarted Lanczos with PRO gives converged eigenvectors with residual norms of just below 10^{-6} compared to well below 10^{-11} for Lan-DR(100,40) with reorthogonalization only at the restart. We note, however, that the PRO tolerance can be adjusted for more frequent reorthogonalization and greater accuracy.

The next matrix is designed so that Lan-DR needs more reorthogonalization. We compare approaches and look at some potential problems.

Example 3. Let the matrix be diagonal with dimension $n = 5000$ and diagonal elements $1, 2, 3, \dots, 9, 10, 100, 101, 102, \dots, 5088, 5089$. The right-hand side is a random normal vector. We use 12 cycles of Lan-DR(120,40) with the reorthogonalization approaches described at the beginning of this section. More reorthogonalization is needed than in Example 2, because eigenvectors converge more quickly. Table 2 shows the results. Two different tolerances on the loss of orthogonality for the PRO methods are used: square root of machine epsilon and three-quarters power. The second column of the table gives the frequency of reorthogonalization (of two vectors) for periodic methods. The next three columns are the same as the columns in Table 1. The last column has another measure of the effect of the roundoff error. It gives the number of iterations needed to solve a second right-hand side using the deflated CG method, which will be given in the next section. We see that k -periodic reorthogonalizing of two vectors every 40 iterations gives fairly good results. Accuracy drops as the reorthogonalization is done less frequently. With a frequency of 75, Lan-DR is still able to compute some eigenvectors accurately, but unlike with a

TABLE 2
Comparison of reorthogonalization methods.

	Reor. freq.	Orthogonality	rn 1	rn 30	2nd rhs it's
Full	1	1.3×10^{-14}	7.5×10^{-12}	1.7×10^{-10}	57
k -SO	1	1.8×10^{-11}	7.5×10^{-12}	1.8×10^{-10}	57
k -periodic	40	6.7×10^{-10}	5.8×10^{-12}	1.2×10^{-8}	57
	60	1.7×10^{-7}	4.8×10^{-12}	6.6×10^{-7}	57
	70	4.0×10^{-5}	5.9×10^{-12}	3.7×10^{-5}	70
	75	9.0	4.6×10^{-12}	-	218
Periodic	40	3.1×10^{-10}	5.8×10^{-12}	2.0×10^{-9}	57
	70	2.8×10^{-6}	5.9×10^{-12}	2.0×10^{-5}	57
	75	2.4×10^{-5}	4.6×10^{-12}	5.5×10^{-5}	57
	80	5.5	5.5×10^{-12}	8.0×10^{-3}	213
k -PRO, ϵ_0^5	-	8.4×10^{-9}	4.9×10^{-12}	1.5×10^{-7}	57
k -PRO, ϵ_0^{75}	-	1.0×10^{-11}	4.9×10^{-12}	2.1×10^{-10}	57
PRO, ϵ_0^5	-	7.6×10^{-9}	4.9×10^{-12}	1.1×10^{-7}	57
PRO, ϵ_0^{75}	-	4.8×10^{-12}	4.9×10^{-12}	1.7×10^{-10}	57

frequency of 70, multiple copies of eigenvalues appear. Also, Lan-DR is no longer helpful for the solution of the second right-hand side. PRO gives results somewhat similar to periodic restarting without the need to select the frequency ahead of time. For k -PRO with ϵ_0^{75} , a total of 82 vectors are reorthogonalized and, with ϵ_0^5 , there are 44 reorthogonalizations. This compares to 50 for k -periodic with a frequency of 40.

Example 4. We continue with the same matrix. Because it has fairly rapidly converging eigenvalues, it can be used to demonstrate a possible problem with reorthogonalizing against only the k saved Ritz vectors. Table 3 shows results for Lan-DR with k -SO for $k = 40$ and changing values of m . Each test is run for the number of cycles needed so that the total number of matrix-vector products is 1000 or just over. Even though k -SO reorthogonalizes at every iteration, there is increasing loss of orthogonality as m increases, until a high level of orthogonality comes back at $m = 200$. We explain first what happens when $m = 140$. Recall that this matrix has 10 well-separated eigenvalues. Also recall we fully reorthogonalize in the first cycle. After that first cycle of Lan-DR(140,40) with k -SO, there are only seven Ritz values below 10. After the second cycle, all 10 small eigenvalues have converged to a high degree. Some orthogonality is lost in that cycle because, by the end of the cycle, there are converged Ritz vectors in the subspace that are not reorthogonalized against. The effect is more extreme with $m = 180$, because then nine of the 10 small eigenvalues converge part way in the first cycle while one eigenvalue is missing (the one at 7.0). That missing one converges very early in the second cycle, and orthogonality is lost after that. So it is dangerous to use k -SO if some eigenvectors converge very rapidly (within one cycle). On the other hand, k -SO and the other k -methods have been successful for other problems, such as in Example 2 and for the large QCD matrices that are in later experiments. The next example shows another possible problem with k -SO.

Example 5. For matrices with outstanding eigenvalues other than the small ones that can converge in one cycle, care must be taken. For k -SO to work, Ritz vectors corresponding to those eigenvalues need to be included in the Ritz vectors that are saved at the restart. We use the same matrix as in the previous example, except the

TABLE 3
An effect of rapidly converging eigenvectors.

m, k	k -SO		Full reorthogonalization	
	Orthogonality	rn30	Orthogonality	rn 30
100, 40	4.8×10^{-12}	1.2×10^{-10}	1.2×10^{-14}	6.3×10^{-11}
120, 40	1.8×10^{-11}	1.8×10^{-10}	7.5×10^{-12}	1.8×10^{-10}
140, 40	1.1×10^{-8}	3.6×10^{-7}	5.8×10^{-12}	1.2×10^{-8}
160, 40	1.3×10^{-3}	6.3×10^{-5}	4.8×10^{-12}	6.6×10^{-7}
180, 40	1.0	1.4	5.9×10^{-12}	3.7×10^{-5}
200, 40	1.2×10^{-12}	3.4×10^{-11}	4.6×10^{-12}	3.4×10^{-11}

largest eigenvalue is changed from 5089 to 5400. This eigenvalue is now outstanding enough to converge rapidly. Lan-DR(120,40) with k -SO loses orthogonality of its basis. After 12 cycles, the orthogonality level is 3.4×10^{-3} , and it is actually worse (near 1) at earlier cycles. As mentioned, this can be fixed by including the large eigenpair among those selected to be saved for the next cycle. Another option is to use the reorthogonalization against all previous vectors instead of just the first k .

5. Multiple right-hand sides. Next, solution of systems with multiple right-hand sides is considered. We suggest a simple approach that uses the eigenvectors generated during the solution of the first right-hand side to deflate eigenvalues from the solution of the second right-hand side. First, a projection is done over the Ritz vectors that were formed by Lan-DR while solving the first right-hand side. Then the standard CG method is used. We call this approach deflated CG (D-CG) since it deflates out eigenvalues before applying CG. It is similar to the init-CG approach [15] that also does a projection before CG. However, init-CG projects over the entire Krylov subspace generated by CG while solving the first right-hand side, while D-CG uses a projection over a compact space that has the important eigen-information. In the algorithm that follows, this projection is in Step 3.

D-CG

1. After applying the initial guess x_0 , let the system of equations be $A(x - x_0) = r_0$.
2. If it is known that the right-hand sides are closely related, project over the previous computed solution vectors. For example, if we project over one solution at a time with a Galerkin projection, then for each system $Ax^i = b^i$ that has already been solved, compute $\delta = (x^i)^T r_0 / ((x^i)^T b^i)$, then form $x_0 = x_0 + \delta x^i$ and $r_0 = r_0 - \delta b^i$.
3. Apply the Galerkin projection for V_k . This uses the V_{k+1} and \bar{T}_k matrices from (3.3) that were developed while solving the first right-hand side with Lan-DR. Specifically, solve $T_k d = V_k^T r_0$, where T_k is the diagonal k by k portion of \bar{T}_k , and let the new approximate solution be $x_0 = x_0 + V_k d$ and the new residual be $r_0 = r_0 - AV_k d = r_0 - V_{k+1} \bar{T}_k d$.
4. Apply CG to $A(x - x_0) = r_0$ until satisfied with convergence.

5.1. Examples.

Example 6. We consider the same matrix as in Example 1, and we consider solution of the second right-hand side. The first right-hand side system has been solved with Lan-DR(100,40). We first illustrate how increasing the accuracy of the approximate eigenvectors helps the convergence of the second right-hand side. Figure 4 shows

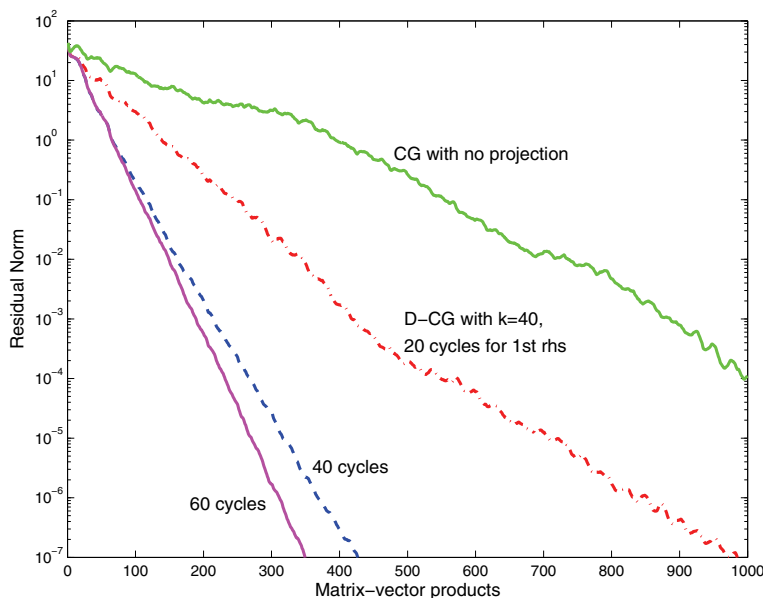


FIG. 4. Solving the first right-hand side to different levels of accuracy.

convergence curves for the second right-hand side with D-CG when Lan-DR has been run for different numbers of cycles. CG is also given and it lags behind all the D-CG curves. The Lan-DR linear equations relative residual converges at 20 cycles. However, D-CG is not as effective as it can be if Lan-DR has only been run 20 cycles. The eigenvectors are not all accurate enough to deflate out the eigencomponents from the residual of the second right-hand side, and eventually CG has to deal with these components, and this slows convergence. D-CG after 40 cycles converges rapidly and does not slow down as it proceeds. Using 60 cycles of Lan-DR is only a little better. Figure 5 shows the components in the directions of the eigenvectors corresponding to the 80 smallest eigenvalues of the residual vector for D-CG on the second right-hand side after the projection over the approximate eigenvectors that come from the first right-hand side. These components improve significantly as Lan-DR on the first right-hand side is run more cycles.

We next consider varying the number of approximate eigenvectors that are used for deflation. For the first right-hand side, $\text{Lan-DR}(m, k)$ is run for 50 cycles with changing k and with $m = k + 60$. Figure 6 shows the convergence results for then applying D-CG to the second right-hand side. With $k = 10$ eigenvectors, D-CG is already significantly better than regular CG, but deflating more eigenvalues is even better. For this example there is a significant jump upon going from 80 to 120 eigenvectors. This happens because having 120 eigenvalues allows us to get past the 100 clustered eigenvalues and pushes well into the rest of the spectrum.

Figure 7 shows that Lan-DR/D-CG can come out ahead of regular CG in terms of matrix-vector products even if we spend more on Lan-DR for the first right-hand side. We use 10 right-hand sides. The first is solved with 44 cycles of $\text{Lan-DR}(180, 120)$. Then the other nine use D-CG with relative residual tolerance of 10^{-8} . This all takes about as many matrix-vector products as solving three systems with CG.

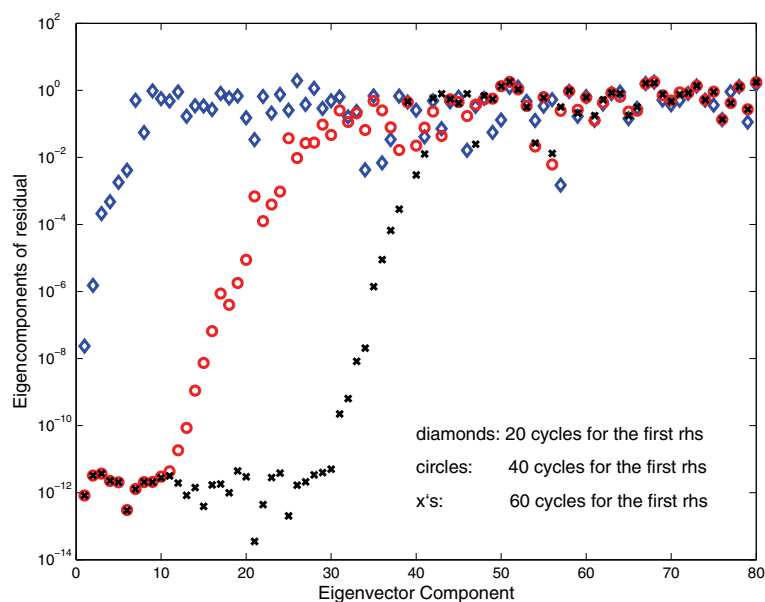


FIG. 5. *Eigencomponents for residual for the second right-hand side after solving the first right-hand side to different levels of accuracy.*

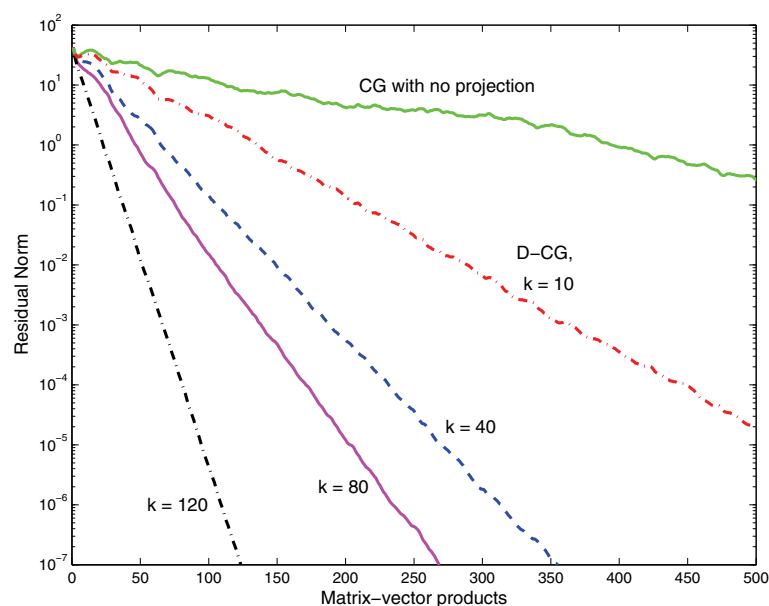


FIG. 6. *Deflating different numbers of eigenvalues.*

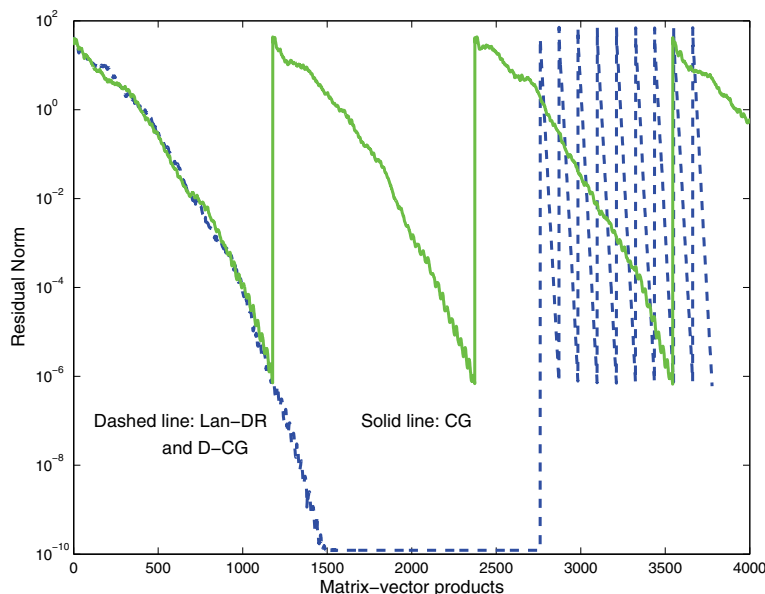


FIG. 7. Lan-DR and D-CG versus regular CG for 10 right-hand sides.

5.2. Comparison with block-CG. Block-CG [40, 39, 56] generates a Krylov subspace, with each right-hand side as a starting vector, then combines them all together into one large subspace. This large subspace can generally develop approximations to the eigenvectors corresponding to the small eigenvalues, so block-CG naturally deflates eigenvalues as it goes along. As a result, block-CG can be very efficient in terms of matrix-vector products.

Block methods require all the right-hand sides to be available at the same time, while Lan-DR/D-CG needs only one right-hand side at a time. Block methods have extra orthogonalization expense compared to nonblock methods. Simple block-CG can be unstable, particularly if the right-hand sides are related to each other. This can be controlled by the somewhat complicated process of removing (“deflating”) right-hand sides [39].

As mentioned, block-CG can converge quickly. For instance, with the matrix from Example 1 and 20 random right-hand sides, block-CG needs only 3100 matrix-vector products. Lan-DR(180,120)/D-CG uses 4909. The next example shows that Lan-DR/D-CG can be competitive with block-CG.

Example 7. We use the same matrix as in Example 3, but with $n = 10,000$ (so the largest element is 10,089). We compare the Lan-DR/D-CG approach with block-CG for 20 random right-hand sides. Lan-DR has $m = 100$ and $k = 15$ and runs for four cycles. Figure 8 shows that the two approaches converge at almost the same number of matrix-vector products. However, Lan-DR has less orthogonalization expense. Lan-DR with k -periodic reorthogonalization of two vectors every 40 iterations followed by D-CG for the remaining 19 right-hand sides uses 14,000 vector operations of length n . Block-CG needs 177,000.

5.3. Related right-hand sides.

Example 8. We use the same matrix as in the previous example. There are again 20 right-hand sides, but this time the first is chosen randomly and the others

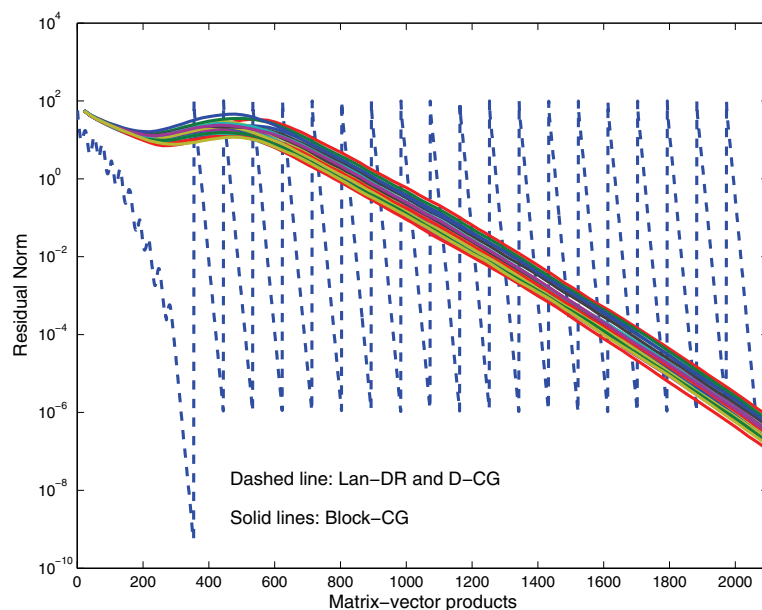


FIG. 8. Comparison with block-CG for 20 right-hand sides.

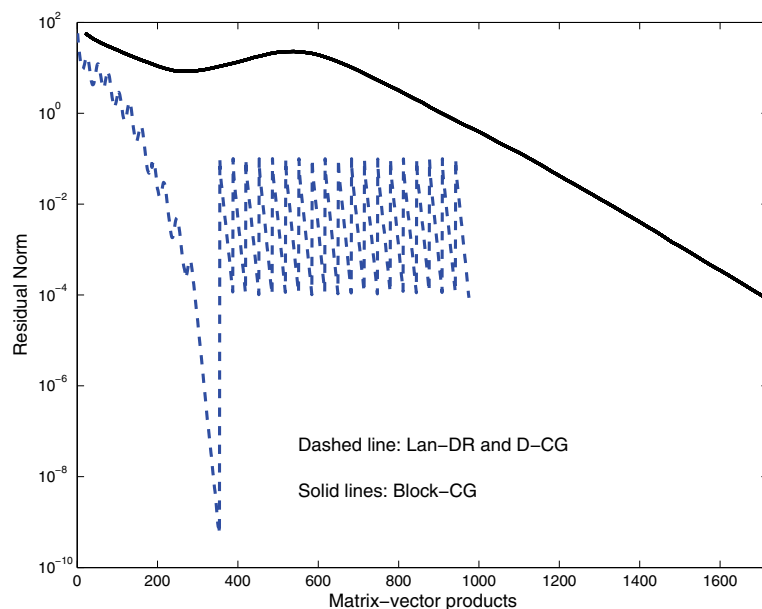


FIG. 9. Comparison with block-CG for 20 related right-hand sides.

are chosen as $b^{(i)} = b^{(1)} + 10^{-3} * \text{ran}^{(i)}$, where $\text{ran}^{(i)}$ is a random vector (elements chosen randomly with normal(0,1) distribution). The convergence tolerance is moved to a relative residual below 10^{-6} , because block-CG has instability after that point. Figure 9 shows that Lan-DR does a better job of taking advantage of the related right-hand sides. However, as mentioned earlier, block-CG can be improved by removing right-hand sides once they become linearly dependent.

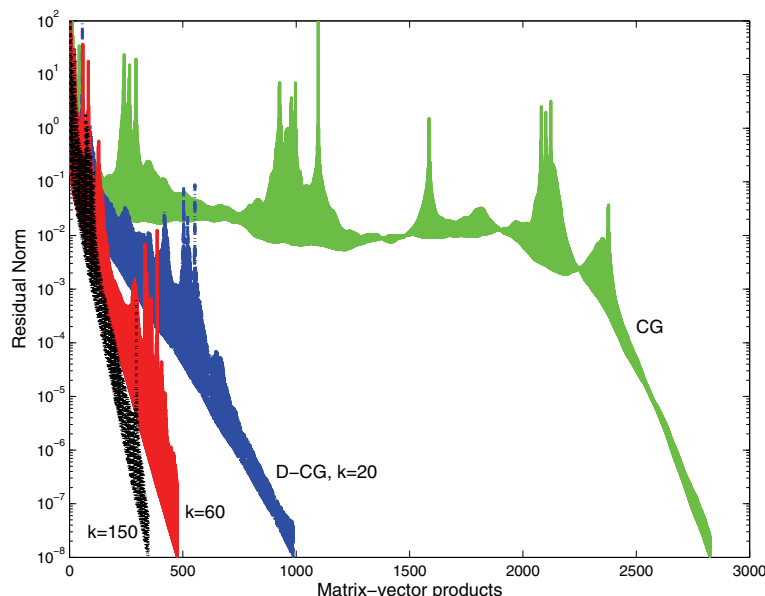


FIG. 10. Comparison of CG to D-CG with varying numbers of eigenvalues deflated for a large QCD matrix.

5.4. Example from QCD. Many problems in lattice quantum chromodynamics (lattice QCD) have large systems of equations with multiple right-hand sides. For example, the Wilson–Dirac formulation [18, 9] and overlap fermion [36, 1] computations both lead to such problems. Deflation of eigenvalues is used for QCD problems, for example, in [10, 13, 12, 37, 9, 33, 63, 26]. QCD matrices have complex entries. They generally are non-Hermitian, but it may be possible to change them into Hermitian form. This can be done by multiplying the system by γ_5 [18] or multiplying by the complex conjugate of the QCD matrix. We consider the first of these options, which gives an indefinite matrix.

Example 9. We choose a large QCD matrix of size $n = 1.5$ million. The κ value is set near to κ -critical, which makes it a difficult problem. There are at least a dozen right-hand sides for each matrix, and sometimes over 100. The first right-hand side is solved using Lan-DR(200, k) with several values of k . The linear equations solution does not converge past a residual norm of 0.036. This shows Lan-DR may not be stable for an indefinite problem and helps motivate the development of MINRES methods in the next section. However, Lan-DR does generate useful eigenvectors that can be used to solve the other right-hand sides. Figure 10 shows the convergence for the solution of the second right-hand side system using D-CG with three choices of k . These are compared to CG. We note that deflating 20 eigenvalues gives a big improvement over regular CG. Using 150 eigenvectors is almost an order of magnitude improvement over CG.

The results in Figure 10 are fairly typical. Figure 11 shows the even iterations for only five configurations (matrices). D-CG with $k = 100$ eigenvalues deflated is compared with CG. There is some variance in CG, but with 100 small eigenvalues taken out, the convergence of D-CG is almost identical for all matrices.

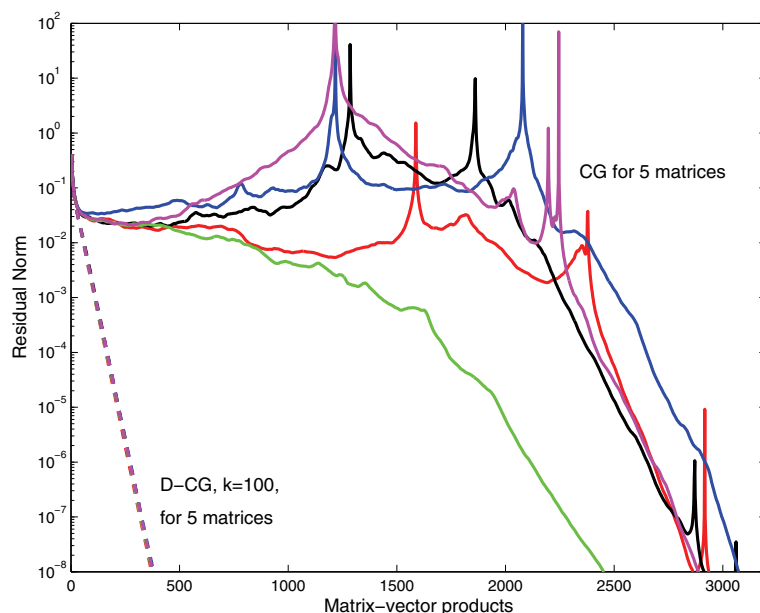


FIG. 11. Comparison of CG to D-CG with $k = 100$ eigenvalues deflated for five QCD matrices.

6. MINRES-DR and deflated MINRES. We now consider indefinite problems. For stability, we need MINRES [43, 50, 52] versions of our methods. Instead of Lan-DR, a MINRES-DR method can be used. It solves the linear equations problem with a MINRES projection, and it computes harmonic Ritz vectors [27, 16, 42, 34, 65, 35] instead of regular Ritz vectors. Harmonic Ritz approximations are more reliable for interior eigenvalues. We next give only the steps that are changed from Lan-DR to MINRES-DR.

MINRES-DR(m, k)

3. *Eigenvector computation.* Compute the k smallest (or others, if desired) harmonic Ritz values, θ_i , and let g_i be the corresponding vectors of unit length. See [35] for more details, including residual norm formulas.
4. *Linear equations.* Let $c_{m+1} = V_{k+1}^T r_0$. Solve the least squares problem $\min \|c_{m+1} - \bar{T}_m d\|$ for d and set $\tilde{x} = x_0 + V_m d$. Then $r = r_0 - A\tilde{x} = r_0 - V_{m+1} \bar{T}_m d$. If satisfied with convergence of the linear equations and the eigenvalues, stop. If not, let the new $x_0 = \tilde{x}$ and $r_0 = r$, and continue.
5. *Restart.* Let P be the $m+1$ by $k+1$ matrix whose first k columns come from orthonormalizing the g_i vectors (and adding zeros for the $m+1$ row). Let e_m be the m th coordinate vector of length m . Then the $k+1$ column of P is the vector $[-t_{m+1,m} T_m^{-T} e_m, 1]^T$ [48] orthonormalized against the previous columns. The new V and T matrices are formed from the old ones: $V_{k+1}^{new} = V_{m+1} P$ and $\bar{T}_k^{new} = P^T \bar{T}_m P_{m,k}$, where $P_{m,k}$ has the first m columns and k rows of P . Set $V_{k+1} = V_{k+1}^{new}$ and $\bar{T}_k = \bar{T}_k^{new}$.

For the second and subsequent right-hand sides, D-MINRES (deflated MINRES) can be used. Like D-CG, it has a projection over the approximate eigenvectors, but this is followed by MINRES [43].

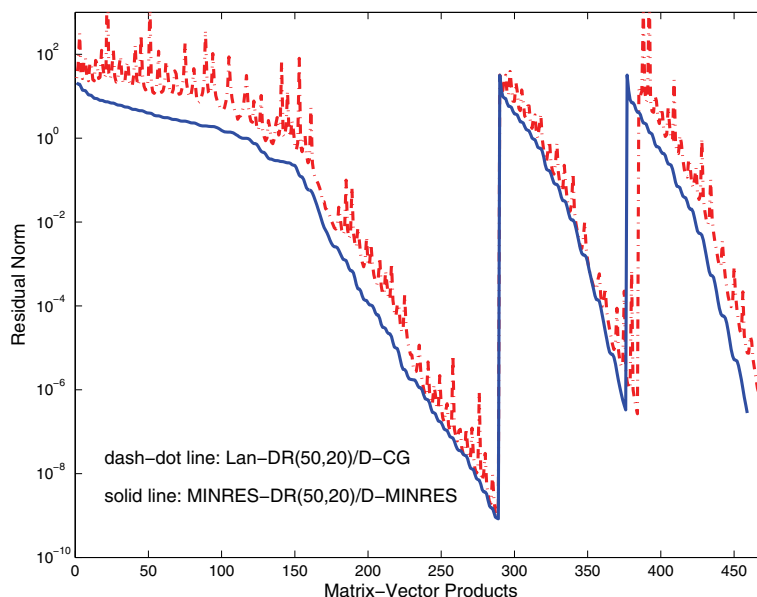


FIG. 12. MINRES-DR versus Lan-DR for the first right-hand side with D-MINRES for the second and third right-hand sides.

Example 10. For an indefinite matrix, we use a diagonal matrix of dimension $n = 1000$ whose diagonal entries are generated with random numbers distributed normal(0,1) that are shifted 2.0 to the right. Then there are 22 negatives among the 1000 eigenvalues. The five closest to the origin are -0.015 , -0.033 , 0.041 , -0.53 , and -0.57 . Indefinite problems can be difficult if there are many eigenvalues on both sides of the origin, unless the eigenvalues are well separated from the origin. Figure 12 shows a comparison of MINRES methods with Galerkin methods for a solution of three right-hand sides. So we compare MINRES-DR followed by D-MINRES for the second and third right-hand sides with Lan-DR followed by D-CG. MINRES-DR converges more smoothly than Lan-DR, and D-MINRES is slightly better than D-CG.

Example 11. The MINRES methods are next applied to the first large QCD matrix from Example 9. Figure 13 shows MINRES-DR(200,60) as diamonds at the end of each cycle. MINRES-DR converges almost as fast as standard unrestarted MINRES. We also see that D-MINRES gives greater improvement as the number of approximate eigenvectors is increased. Finally, D-MINRES converges a little faster than D-CG does in Example 9. For instance, following Lan-DR(200,150), D-CG takes 348 matrix-vector products. Meanwhile, D-MINRES needs only 310 when following MINRES-DR(200,150).

7. Conclusion. Lan-DR is a restarted Lanczos method that solves a symmetric/Hermitian positive definite system of linear equations and simultaneously computes eigenvalues and eigenvectors. The restarting allows the computation of eigenvectors when there are limits on storage. The presence of the approximate eigenvectors in the subspace helps the convergence. In fact, the convergence is often close to that of unrestarted Lanczos for both the linear equations and the eigenvalues, in spite of the restarting.

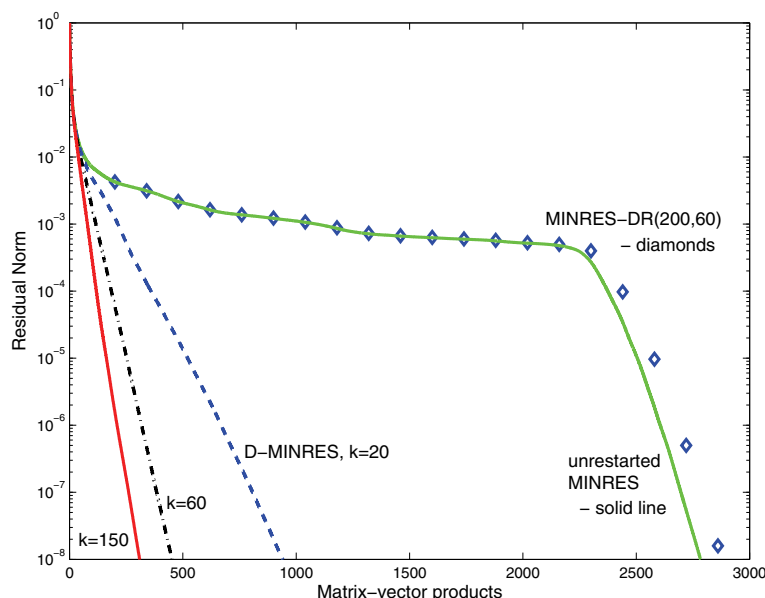


FIG. 13. MINRES-DR and D-MINRES for the large QCD matrix.

There are several options for reorthogonalizing. Included are methods that re-orthogonalize only against the k saved Ritz vectors. For these methods, there may be trouble if there are rapidly converging eigenvalues that converge in one cycle. Restarting generally keeps reorthogonalization costs down.

For subsequent right-hand sides, D-CG first uses a projection over the eigenvectors that were computed by Lan-DR while solving the first right-hand side system, then applies CG. For difficult problems, the convergence of CG can be much faster after the small eigenvalues are deflated out. Experiments on large problems from QCD back this up.

MINRES-DR is a more stable version of Lan-DR for indefinite problems. For indefinite systems with multiple right-hand sides, a deflated MINRES method is given.

Acknowledgments. We would like to thank the referees for their suggestions and corrections. A. Abdel-Rehim would like to thank the Departments of Physics and Computer Science, College of William and Mary, as well as Thomas Jefferson National Laboratory for support during the writing of this paper.

REFERENCES

- [1] G. ARNOLD, N. CUNDY, J. VAN DEN ESHOF, A. FROMMER, S. KRIEG, TH. LIPPERT, AND K. SCHÄFER, *Numerical Methods for the QCD Overlap Operator: II. Sign-Function and Error Bounds*, Technical Report BUW-SC 2003/2, Bergische Universität GH Wuppertal, Wuppertal, Germany, 2003.
- [2] J. BAGLAMA, *Augmented block Householder Arnoldi method*, Linear Algebra Appl., 429 (2008), pp. 2315–2334.
- [3] J. BAGLAMA, D. CALVETTI, G. H. GOLUB, AND L. REICHEL, *Adaptively preconditioned GMRES algorithms*, SIAM J. Sci. Comput., 20 (1998), pp. 243–269.
- [4] J. BAGLAMA, D. CALVETTI, AND L. REICHEL, *Iterative methods for the computation of a few eigenvalues of a large symmetric matrix*, BIT, 36 (1996), pp. 400–421.

- [5] K. BURRAGE AND J. ERHEL, *On the performance of various adaptive preconditioned GMRES strategies*, Numer. Linear Algebra Appl., 5 (1998), pp. 101–121.
- [6] D. CALVETTI, L. REICHEL, AND D. SORESENSEN, *An implicitly restarted Lanczos method for large symmetric eigenvalue problems*, Electron. Trans. Numer. Anal., 2 (1994), pp. 1–21.
- [7] T. F. CHAN AND W. L. WAN, *Analysis of projection methods for solving linear systems with multiple right-hand sides*, SIAM J. Sci. Comput., 18 (1997), pp. 1698–1721.
- [8] A. CHAPMAN AND Y. SAAD, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1997), pp. 43–66.
- [9] D. DARNELL, R. B. MORGAN, AND W. WILCOX, *Deflation of eigenvalues for iterative methods in lattice QCD*, Nucl. Phys. B (Proc. Suppl.), 129 (2004), pp. 856–858.
- [10] P. DE FORCRAND, *Progress on lattice QCD algorithms*, Nucl. Phys. B (Proc. Suppl.), 47 (1996), pp. 228–235.
- [11] E. DE STURLER, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp. 864–889.
- [12] S. J. DONG, F. X. LEE, K. F. LIU, AND J. B. ZHANG, *Chiral symmetry, quark mass, and scaling of the overlap fermions*, Phys. Rev. Lett., 85 (2000), pp. 5051–5054.
- [13] R. G. EDWARDS, U. M. HELLER, AND R. NARAYANAN, *Study of chiral symmetry in quenched QCD using the overlap Dirac operator*, Phys. Rev. D, 59 (1999), pp. 0945101–0945108.
- [14] J. ERHEL, K. BURRAGE, AND B. POHL, *Restarted GMRES preconditioned by deflation*, J. Comput. Appl. Math., 69 (1996), pp. 303–318.
- [15] J. ERHEL AND F. GUYOMARCH, *An augmented conjugate gradient method for solving consecutive symmetric positive definite linear systems*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1279–1299.
- [16] R. W. FREUND, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, J. Comput. Appl. Math., 43 (1992), pp. 135–158.
- [17] R. W. FREUND AND M. MALHOTRA, *A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides*, Linear Algebra Appl., 254 (1997), pp. 119–157.
- [18] A. FROMMER, *Linear systems solvers—recent developments and implications for lattice computations*, Nucl. Phys. B (Proc. Suppl.), 53 (1997), pp. 120–126.
- [19] J. GRGAR, *Analyses of the Lanczos Algorithm and of the Approximation Problem in Richardson's Method*, Ph.D. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL, 1981.
- [20] M. H. GUTKNECHT, *Block Krylov space methods for linear systems with multiple right-hand sides: An introduction*, in Modern Mathematical Models, Methods and Algorithms for Real World Systems, A. H. Siddiqi, I. S. Duff, and O. Christensen, eds., Anamaya Publishers, New Delhi, India, 2007, pp. 420–447.
- [21] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.
- [22] 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.
- [23] M. KILMER, E. MILLER, AND C. RAPPAPORT, *QMR-based projection techniques for the solution of non-Hermitian systems with multiple right-hand sides*, SIAM J. Sci. Comput., 23 (2001), pp. 761–780.
- [24] 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.
- [25] C. LE CALVEZ AND B. MOLINA, *Implicitly restarted and deflated GMRES*, Numer. Algorithms, 21 (1999), pp. 261–285.
- [26] M. LÜSCHER, *Local coherence and deflation of the low quark modes in lattice QCD*, J. High Energy Phys., 2007, no. 7, paper 081 (electronic).
- [27] R. B. MORGAN, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154/156 (1991), pp. 289–309.
- [28] R. B. MORGAN, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.
- [29] R. B. MORGAN, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Math. Comp., 65 (1996), pp. 1213–1230.
- [30] R. B. MORGAN, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1112–1135.
- [31] R. B. MORGAN, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
- [32] R. B. MORGAN, *Restarted block-GMRES with deflation of eigenvalues*, Appl. Numer. Math., 54 (2005), pp. 222–236.
- [33] R. B. MORGAN AND W. WILCOX, *Deflated Iterative Methods for Linear Equations with Multiple Right-Hand Sides*, preprint, 2004. Available online at <http://www.arxiv.org/abs/math-ph/0405053>.

- [34] R. B. MORGAN AND M. ZENG, *Harmonic projection methods for large non-symmetric eigenvalue problems*, Numer. Linear Algebra Appl., 5 (1998), pp. 33–55.
- [35] R. B. MORGAN AND M. ZENG, *A harmonic restarted Arnoldi for calculating eigenvalues and determining multiplicity*, Linear Algebra Appl., 415 (2006), pp. 96–113.
- [36] R. NARAYANAN AND H. NEUBERGER, *Alternative to domain wall fermions*, Phys. Rev. D, 62 (2000), paper 074504.
- [37] H. NEFF, N. EICKER, TH. LIPPERT, J. W. NEGELE, AND K. SCHILLING, *Low fermionic eigenmode dominance in QCD on the lattice*, Phys. Rev. D, 64 (2001), paper 114509.
- [38] R. A. NICOLAIDES, *Deflation of conjugate gradients with applications to boundary value problems*, SIAM J. Numer. Anal., 24 (1987), pp. 355–365.
- [39] A. A. NIKISHIN AND A. YU. YEREMIN, *Variable block CG algorithms for solving large sparse symmetric positive definite linear systems on parallel computers, I: General iterative scheme*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1135–1153.
- [40] D. P. O’LEARY, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.
- [41] C. C. PAIGE, *The Computation of Eigenvectors and Eigenvalues of Very Large Sparse Matrices*, Ph.D. Thesis, University of London, 1971.
- [42] 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.
- [43] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
- [44] 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.
- [45] B. N. PARLETT, *A new look at the Lanczos algorithm for solving symmetric systems of linear equations*, Linear Algebra Appl., 29 (1980), pp. 323–346.
- [46] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [47] B. N. PARLETT AND D. S. SCOTT, *The Lanczos algorithm with selective orthogonalization*, Math. Comp., 33 (1979), pp. 217–238.
- [48] S. ROLLIN AND W. FICHTNER, *Improving the accuracy of GMRES with deflated restarting*, SIAM J. Sci. Comput., 30 (2007), pp. 232–245.
- [49] Y. SAAD, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp., 37 (1981), pp. 105–126.
- [50] Y. SAAD, *Practical use of some Krylov subspace methods for solving indefinite and nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 203–228.
- [51] Y. SAAD, *On the Lanczos method for solving symmetric linear systems with several right-hand sides*, Math. Comp., 48 (1987), pp. 651–662.
- [52] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
- [53] Y. SAAD, *Analysis of augmented Krylov subspace techniques*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 435–449.
- [54] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [55] Y. SAAD, M. YEUNG, J. ERHEL, AND F. GUYOMARC’H, *A deflated version of the conjugate gradient algorithm*, SIAM J. Sci. Comput., 21 (2000), pp. 1909–1926.
- [56] T. SCHMELZER, *Block Krylov Methods for Hermitian Linear Systems*, Diploma thesis, Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany, 2004.
- [57] H. D. SIMON, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, Linear Algebra Appl., 61 (1984), pp. 101–132.
- [58] H. D. SIMON, *The Lanczos algorithm with partial reorthogonalization*, Math. Comp., 42 (1984), pp. 115–136.
- [59] V. SIMONCINI AND E. GALLOPOULOS, *An iterative method for nonsymmetric systems with multiple right-hand sides*, SIAM J. Sci. Comput., 16 (1995), pp. 917–933.
- [60] V. SIMONCINI AND E. GALLOPOULOS, *A hybrid block GMRES method for nonsymmetric systems with multiple right-hand sides*, J. Comput. Appl. Math., 66 (1996), pp. 457–469.
- [61] C. SMITH, A. PETERSON, AND R. MITTRA, *A conjugate gradient algorithm for the treatment of multiple incident electromagnetic fields*, IEEE Trans. Antennas and Propagation, 37 (1989), pp. 1490–1493.
- [62] D. C. SORESENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
- [63] A. STATHOPOULOS AND K. ORGINOS, *Computing and Deflating Eigenvalues While Solving Multiple Right Hand Side Linear Systems in Quantum Chromodynamics*, preprint, 2008. Available online at <http://www.arxiv.org/abs/0707.013/>.

- [64] G. W. STEWART, *A Krylov-Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614.
- [65] G. W. STEWART, *Matrix Algorithms. Volume II: Eigensystems*, SIAM, Philadelphia, 2001.
- [66] H. A. VAN DER VORST, *An iterative method for solving $f(A)x = b$ using Krylov subspace information obtained for the symmetric positive definite matrix A* , J. Comput. Appl. Math., 18 (1987), pp. 249–263.
- [67] K. WU AND H. SIMON, *Thick-restart Lanczos method for large symmetric eigenvalue problems*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 602–616.
- [68] Y. ZHOU AND Y. SAAD, *Block Krylov-Schur Method for Large Symmetric Eigenvalue Problems*, Technical Report umsi-2004-215, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN, 2004.