PSEUDOEIGENVECTOR BASES AND DEFLATED GMRES FOR HIGHLY NONNORMAL MATRICES*

RONALD B. MORGAN†, ZHAO YANG‡, AND BAOJIANG ZHONG§

Abstract. Pseudoeigenvalues have been extensively studied for highly nonnormal matrices. This paper focuses on the corresponding pseudoeigenvectors. The properties and uses of pseudoeigenvector bases are investigated. It is shown that pseudoeigenvector bases can be much better conditioned than eigenvector bases. We look at the stability and the varying quality of pseudoeigenvector bases. Then applications are considered including the exponential of a matrix. Several aspects of GMRES convergence are looked at, including why using approximate eigenvectors to deflate eigenvalues can be effective even when eigenvectors are not well defined.

Key words. linear equations, iterative methods, GMRES, deflation, eigenvalues, pseudoeigenvalues, nonnormal matrices

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

1. Introduction.

Pseudoeigenvalues have been studied extensively (see for example [28, 22, 19, 23] and references in [30]) for nonnormal matrices and for stability of differential equations and other applications. Here we study the associated pseudoeigenvectors. They have been looked at before [18, 29, 4], but not as extensively studied as pseudoeigenvalues. In particular, pseudoeigenvectors have not been used to analyze GMRES convergence.

Pseudoeigenvectors of a matrix can be thought of as the eigenvectors of a nearby matrix. More specifically, pseudoeigenvectors are defined as follows:

DEFINITION 1.1. *Let A be an n by n matrix. Let E be any matrix such that $||E|| = \epsilon$. Then the eigenvectors of $A + E$ are defined to be ϵ-pseudoeigenvectors of A.*

We observe that with this definition, there are many ϵ-pseudoeigenvectors even for a fixed value of ϵ. There is a different set for each choice of E. For a particular E, the set of ϵ-pseudoeigenvectors will usually have n linearly independent elements. There is an associated set of n ϵ-pseudoeigenvalues. The ϵ-pseudospectrum is made up of the eigenvalues of $A + E$ for all E of norm less than or equal to ϵ.

When the pseudoeigenvectors from a specific perturbation E form a basis for R^n (as they generally will), we call this a pseudoeigenvector basis. Such a basis can be useful when there is not a good basis made up of eigenvectors, as is the case when the eigenvector basis is highly ill-conditioned or when the Jordan canonical form of A is non-diagonal. In this paper we will analyze and apply pseudoeigenvector bases.

One important application is to the GMRES [25] method for solving large nonsymmetric systems of linear equations. Convergence of GMRES for highly nonnormal matrices is a difficult topic. Some aspects have been studied, see for example [12, 11, 10, 5, 24, 14, 27, 31]. Pseudoeigenvalues have been used to analyze GMRES convergence [28, 20, 30, 26]. Here, we analyze with pseudoeigenvector bases.

*The first author was partially supported by the National Science Foundation, Computational Mathematics Program under grant DMS-1418677.

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

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

§School of Computer Science and Technology, Soochow University, Suzhou 215006, China (bjzhong@suda.edu.cn).

Deflated GMRES methods (see for example [15, 13, 3, 1, 2, 7, 16, 21, 9, 8]) can be an improvement over regular restarted GMRES when the matrix has small eigenvalues that slow the convergence. One particular deflated GMRES method called GMRES-DR [16] is a restarted GMRES method that augments the Krylov subspace with approximate eigenvectors. The convergence can often be improved because the approximate eigenvectors essentially remove or deflate out the eigenvalues once they are accurate enough. It is not clear that such a deflated method will be helpful for highly nonnormal matrices because removing some eigenvalues is of dubious value when the eigenvalues do not control the convergence. We will give examples showing that deflated GMRES can nevertheless be effective for highly nonnormal matrices and will offer an explanation.

Section 2 shows that pseudoeigenvector bases can be better conditioned than the original eigenvector basis. It is also observed that the quality of pseudoeigenvector bases can vary and will give an example with what appears to be an optimal basis. Section 3 considers stability of pseudoeigenvector bases. The rest of the paper has some examples of applications of pseudoeigenvector bases. Section 4 has the exponential of a matrix. Tortoise and the Hare convergence of GMRES [5] is in Section 5. Application to general GMRES convergence is in Section 6, while Section 7 has deflated GMRES.

2. Conditioning of Pseudoeigenvector Bases.

2.1. A pseudoeigenvector basis can be better conditioned. In this section, we demonstrate that for a nonnormal matrix A, pseudoeigenvector bases can be better conditioned than the original eigenvector basis. It is certainly possible for a perturbation to not improve conditioning of eigenvectors. However, the structure of many matrices of interest causes their eigenvectors to be very ill-conditioned and allows perturbations to improve them. Davies [4] has an investigation into this same topic, including a formula for evaluating the accuracy of an approximate diagonalization of the form $A + E = XDX^{-1}$.

An obvious case for improvement is when the matrix is not diagonalizable and so there is not a basis of eigenvectors. We use such a matrix for the first example.

Example 1. As one of the test matrices in this paper, we let A be the bidiagonal matrix with 1's in all the diagonal and superdiagonal positions. So the matrix is made up of one large Jordan block. Figure 3.1 has plots of two sets of 10^{-2}-pseudoeigenvalues for this matrix of size $n = 100$ and real perturbations. However, here we have $n = 500$, and we use complex perturbation matrices E of norm ϵ. They are formed by filling an n by n matrix N with entries that have real part generated randomly with normal(0,1) distribution and imaginary part generated the same way. Then E is set to $(\frac{\epsilon}{||N||}) * N$, and the eigenvectors of $A + E$ are computed. These are an ϵ-pseudoeigenvector basis for A. With a perturbation matrix of norm 10^{-8}, the condition number for the pseudoeigenvector basis is $2.1 * 10^{10}$, while with $\epsilon = 10^{-4}$, the condition number is $1.6 * 10^{6}$. There is a trade-off: as the perturbation becomes larger, the conditioning of the pseudoeigenvector basis gets better, but this basis has eigenvectors for a matrix further from the original one.

Next, we consider highly nonnormal matrices that are diagonalizable. In our testing, the conditioning of the eigenvectors can usually be improved with a perturbation. We give a Toeplitz matrix example. Experiments with some other Toeplitz matrices give similar results; some have even more highly ill-conditioned eigenvectors and some less.

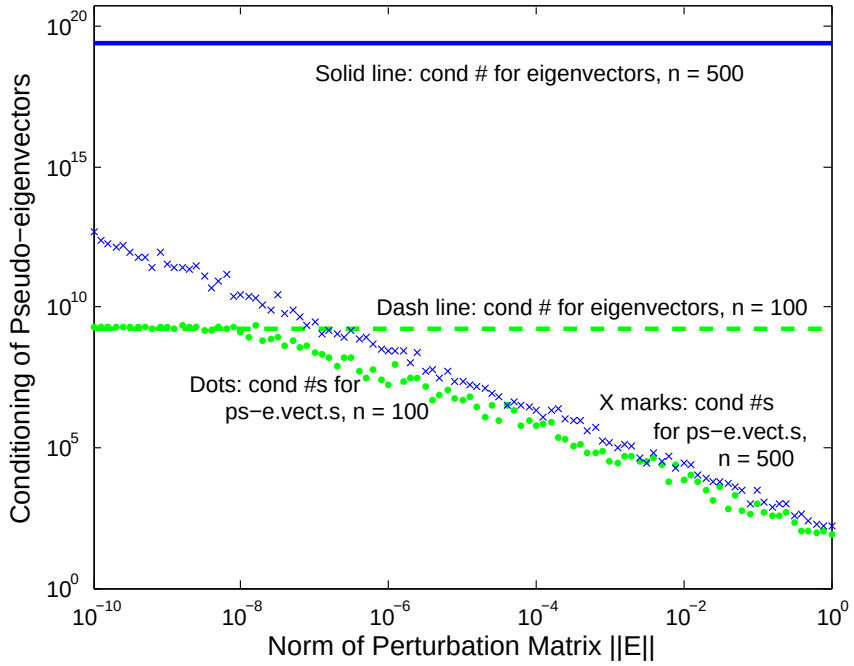

FIG. 2.1. *Conditioning of Pseudoeigenvector Bases for the Butterfly matrix.*

Example 2. The Butterfly matrix from [30] has zeros on the main diagonal, superdiagonal with all -i's, the diagonal above that has 1's, the subdiagonal has i's, and the diagonal below has -1's. We let $n = 100$ and 500. For $n = 100$, the eigenvector basis for the matrix is ill-conditioned with condition number equal to $1.8 * 10^9$. This is shown in Figure 2.1 as the horizontal dashed line. The dots in the figure show the conditioning of the pseudoeigenvector bases that result from perturbing the matrix with random, complex matrices of increasing norm. We see that perturbing the matrix improves the conditioning of the eigenvectors starting with perturbation matrices of norm 10^{-8}. The pseudoeigenvector basis is significantly better conditioned for larger perturbation matrices such as with norm 10^{-4} or 10^{-2}. Next, the Butterfly matrix of size $n = 500$ is considered. The condition number for the matrix of eigenvectors is now $2.5 * 10^{19}$. This is shown on Figure 2.1 with the horizontal solid line. Even with a small perturbations, the conditioning improves dramatically. It is interesting that the relationship on the log-log plot between the norm of the perturbation matrix and the condition number of the basis of pseudoeigenvectors is almost linear. If the perturbation matrices are instead just scaled versions of the same random matrix, then the dots fall even closer to a straight line. Davies' [4] formula for evaluating a pseudoeigenvector basis is $\sigma = \kappa * \epsilon_0 + ||E||$, where κ is the conditioning of the pseudoeigenvector basis and ϵ_0 is machine epsilon. Lower σ indicates a better basis. For this example with $n = 500$, the lowest σ is about 10^{-7} at $\epsilon = ||E|| = 10^{-7}$.

2.2. Quality of pseudoeigenvector bases. In this subsection, we show that even for a fixed ϵ, it is possible for some ϵ-pseudoeigenvector bases to be significantly better than others. We will concentrate on Jordan block matrices such as the one used in Example 1, and will give a conjecture on the optimal pseudoeigenpairs for this matrix.

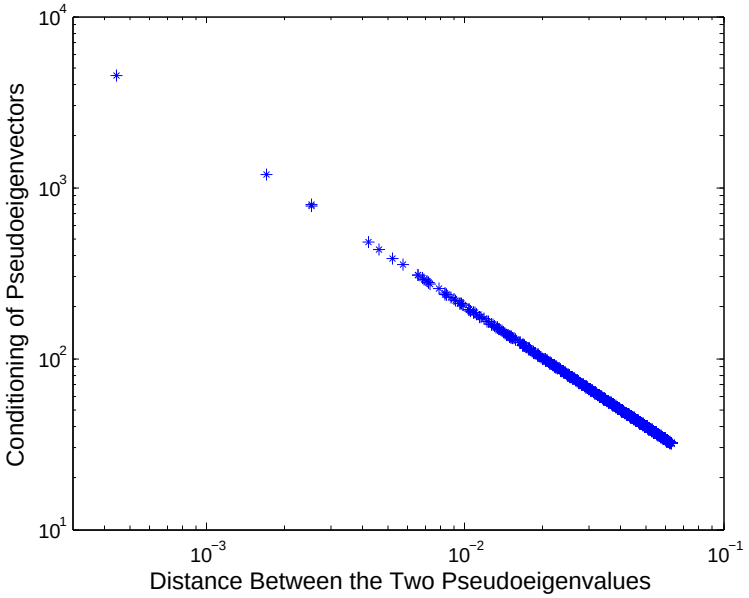

FIG. 2.2. *Condition of pseudoeigenvector basis as a function of distance between pseudoeigenvalues for 2 by 2 matrix.*

Example 3. We consider the 2 by 2 version of the Jordan block matrix in Example 1. We generate 1000 different real perturbation matrices of norm 10^{-3} and look at the conditioning of the resulting pseudoeigenvector bases. It is surprising how much this conditioning varies. There is over two orders of magnitude difference between the highest condition number $4.5 * 10^3$ and the lowest of $3.2 * 10^1$. Figure 2.2 has a log-log plot of the condition numbers versus the distance between the two pseudoeigenvalues for each perturbation. The better conditioned pseudoeigenvector bases occur when the pseudoeigenvalues are more separated. It is interesting how consistent this relationship is; in fact it is linear on this log-log plot. We also note that if instead complex perturbations are used, then there is not as much variance in the conditioning. The range is from $2.4 * 10^2$ down to $3.2 * 10^1$. However, if these complex perturbation results are plotted as in Figure 2.2, the points fall on the same line. It has been noted before that complex perturbations can give a wider spread on the pseudoeigenvalues [30], and as we have seen, here bigger spread means better conditioning.

We next consider the 7 by 7 version of this matrix. Again we use perturbations of norm 10^{-3}. Figure 2.3 has the conditioning of the pseudoeigenvectors plotted against the harmonic distance between the pseudoeigenvalues (this is the reciprocal of the sum of the reciprocals of all of the distances). This is done for 1000 real perturbations (asterisks) and 1000 complex perturbations (circles). The points are not linear in this log-log plot as they were for the 2 by 2 case, but most of the values do fall in a band. The complex perturbations follow the same pattern as the real ones, but do not have the higher conditioned outliers. We also tried 100 by 100 matrices, but did not see a relationship between conditioning and harmonic distance.

We next look at structured perturbations. These have been considered before for the pseudoeigenvalues; see [30] and its references.

Example 4. We consider a specific perturbation for the Jordan block matrix. Let E be the matrix that is all zeros except for a positive entry of size ϵ in the $(n, 1)$

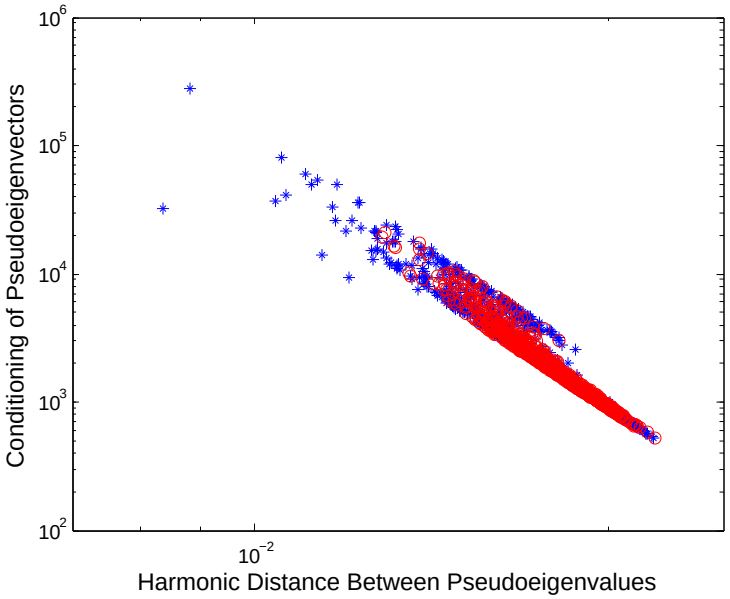

FIG. 2.3. *Condition of pseudoeigenvector basis as a function of distances between pseudoeigenvalues for 7 by 7 matrix.*

position. Then the eigenvalues of the perturbed matrix are

$$1 - \sqrt[n]{\epsilon}\, exp(i\frac{2k\pi}{n}),\ k = 0, 1, \ldots, n-1.$$

A similar formula with + replacing − can be given for perturbation $-\epsilon$ in place of ϵ. In both cases, the pseudoeigenvalues are on a circle of radius $\sqrt[n]{\epsilon}$ around the point $(1, 0)$. So the pseudoeigenvalues are well spaced. The condition number of the pseudoeigenvector basis appears to be $\epsilon^{\frac{1}{n}-1}$. For the case of the 7 by 7 matrix with $\epsilon = 10^{-3}$, this condition number is $3.7 * 10^2$. This compares to values between $5 * 10^2$ and $3 * 10^5$ for the 2000 real and complex perturbations in the previous example. For a Jordan block matrix of size $n = 500$, with 100 random real perturbations and 100 random complex, the minimum pseudoeigenvector bases condition number is $5.0 * 10^4$ and the maximum is $2.1 * 10^6$. Meanwhile, for the special (n,1) entry perturbation, the condition number is $9.8 * 10^2$. So this gives a significant improvement in the conditioning of the pseudoeigenvector basis. We conjecture that this is the optimal perturbation for the Jordan block matrix. Perhaps it will help in finding good pseudoeigenvector bases for other structured matrices.

Example 5. Consider the 3 by 3 matrix with ones on the main diagonal, a 10 in the (1,2) position and a 1 in the (1,3) position and zeros elsewhere. Experimentally, the best perturbation appears to be an ϵ in the (2,1) position.

For some other 3 by 3 matrices, we tested two approaches to finding a good perturbation. One way is to test the 18 perturbation that add plus or minus ϵ to an entry. The second way tests thousands of random perturbations. The two approaches give about the same best conditioned pseudoeigenvector bases. So a simple way of finding a near optimal basis may be possible in some cases.

3. Stability of Pseudoeigenvector Bases. For a basis of pseudoeigenvectors to be useful, it would help if there is some stability relative to the choice of E. This

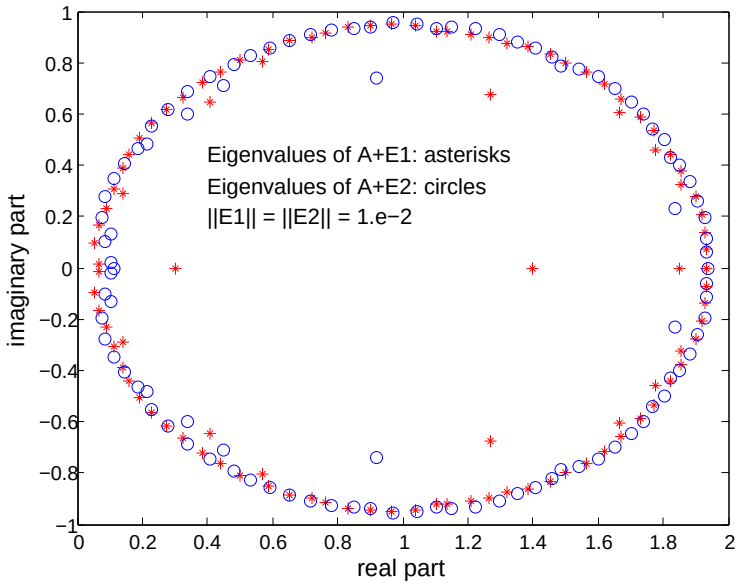

FIG. 3.1. *Pseudoeigenvalues of A from two different perturbations A+E.*

means that using a different E of similar norm will not change everything about the basis. We know there are limits to this stability, because we have just observed in the previous section that pseudoeigenvector bases can vary significantly in quality. However, we now investigate what can be shown.

Example 6. We again use the Jordan block matrix from Example 1. Here we have $n = 100$ and let $\epsilon = 10^{-2}$. The perturbation matrices E are formed by filling an n by n matrix N with real entries generated randomly with normal(0,1) distribution, then multiplying by $\frac{\epsilon}{||N||}$. Figure 3.1 has the pseudoeigenvalues of A that come from two different choices of E, called $E1$ and $E2$. There is a similarity in the general appearance, but some variation in the exact placement of the pseudoeigenvalues. We next look at how much a pseudoeigenvector corresponding to the smallest pseudoeigenvalue changes with a different perturbation. Figure 3.2 gives the cosines of the angles between the eigenvector of $A + E1$ associated with the smallest eigenvalue and the first 10 vectors in an orthogonal basis for the eigenvectors of $A + E2$. The eigenvectors are ordered by distance from the smallest eigenvalue of $A + E1$ and then a QR factorization of the matrix of eigenvectors is used to orthogonalize them. The smallest eigenvector of $A + E1$ is mostly in the direction of the smallest eigenvector of $A + E2$ and then has decreasing components in the directions of the others.

The following theorem relates small singular values of shifted matrices to pseudoeigenvalues. It shows that if $A - \gamma I$ has a small singular value σ, then A has γ as an ϵ-pseudoeigenvalue for a small ϵ, specifically $\epsilon = \sigma$. The right singular vector is the pseudoeigenvector. It follows a standard result (such as in [30], Chapter 2).

THEOREM 3.1. *If (σ, v, u) is a singular triplet of $A - \gamma I$, then v is an σ-pseudoeigenvector of A, with associated σ-pseudoeigenvalue γ.*

Proof. Since (σ, v, u) is a singular triplet of $A - \gamma I$, we have that for unit vectors u and v that

$$\begin{aligned}(A - \gamma I)v &= \sigma u \\ &= \sigma u v^T v.\end{aligned}$$

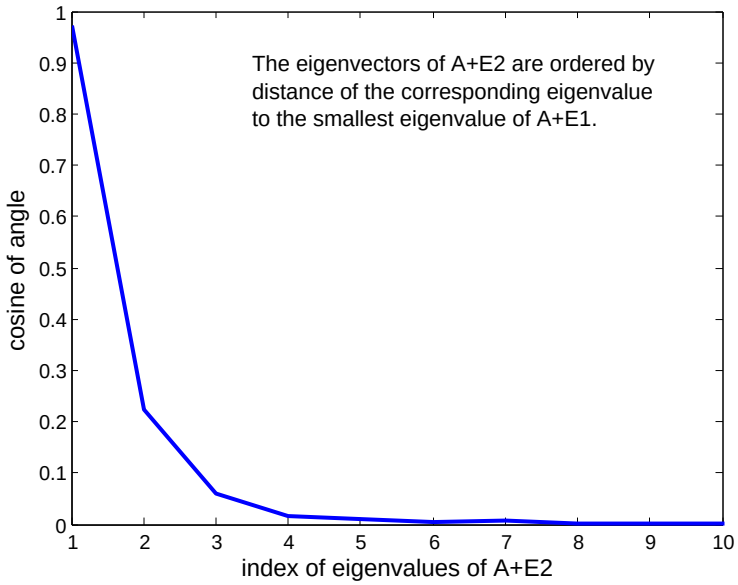

FIG. 3.2. *Cosines of angles between smallest eigenvector of A+E1 and orthogonalized basis for eigenvectors of A+E2.*

So

$$(A - \sigma uv^T)v = \gamma v,$$

and if we set $E = -\sigma uv^T$, then we have

$$(A + E)v = \gamma v,$$

with $||E|| = \sigma$. □

Somewhat conversely, we can show that an ϵ-pseudoeigenvector is composed mostly of singular vectors corresponding to small singular values of a shifted matrix. This indicates that the pseudoeigenvector is somewhat stable. For this pseudoeigenvalue, a corresponding pseudoeigenvector is composed mostly of these singular vectors regardless of the E.

THEOREM 3.2. *Assume that $\{\zeta, x\}$ is an ϵ-pseudoeigenpair of A, and let E be the associated perturbation matrix so that $(A + E)x = \zeta x$ with $||E|| = \epsilon$. Let the singular value decomposition of $A - \zeta I$ be $(A - \zeta I)V = U\Sigma$. Let $x = Vg$, so that the elements of g give the coefficients of the expansion of x in terms of the basis of right singular vectors. Then*

$$||\Sigma g|| \leq \epsilon.$$

Proof.

$$\begin{aligned} ||\Sigma g|| &= ||U\Sigma g|| \\ &= ||(A - \zeta I)Vg|| \\ &= ||((A + E - \zeta I)x - Ex|| \\ &\leq ||Ex|| \\ &\leq \epsilon. \end{aligned}$$

□

Note that $||\Sigma g||$ is the norm of the vector with ith component $\sigma_i g_i$. If ϵ is small, then for each i, either the ith singular value is small or the component of x in that direction is small. Thus x is made up of mostly small singular vectors.

The next theorem has a similar purpose. it gives a condition so that if two ϵ-pseudoeigenvalues are equal but come from different choices of E, then the associated ϵ-pseudoeigenvectors are close to each other.

THEOREM 3.3. *Suppose we have two choices of E, say $E1$ and $E2$, both of norm equal to ϵ and they both give rise to the same ϵ-pseudoeigenvalue, say ζ. Let the associated unit length ϵ-pseudoeigenvectors be $x1$ and $x2$ (so $(A+E1)x1 = \zeta x1$, etc.). Consider the singular values of the matrix $A - \zeta I + E1$, and let γ be the gap between the singular value at zero and the next closest singular value. Then*

$$sin\angle(x1, x2) \leq \frac{2\epsilon}{\gamma}.$$

Proof. Since $(A + E1 - \zeta I)x1 = 0$ and $(A + E2 - \zeta I)x2 = 0$,

$$(A + E1 - \zeta I)d = f, \tag{3.1}$$

where $d = x1 - x2$ and $f = (E2 - E1)x2$. Note that $||f|| \leq 2\epsilon$.

Using the singular value decomposition, $A + E1 - \zeta I = U\Sigma V^T$, where U and V and orthogonal and the singular values are ordered with σ_1 being smallest. So $\sigma_1 = 0$ and $v_1 = x1$. Expanding $x2$ in terms of the basis of right singular vectors, $x2 = \sum_{i=1}^{n} \beta_i v_i$. Then $cos\angle(x1, x2) = \beta_1$ and

$$sin\angle(x1, x2) = \sqrt{\sum_{i=2}^{n} \beta_i^2}. \tag{3.2}$$

Next, $d = (1 - \beta_1)v_1 - \sum_{i=2}^{n} \beta_i v_i$. From (3.1), $f = U\Sigma V^T d = 0 - \sum_{i=2}^{n} \beta_i \sigma_i u_i$. Then $||f|| \geq \sigma_2 \sqrt{\sum_{i=2}^{n} \beta_i}$, and using Equation (3.2), $sin\angle(x1, x2) \leq \frac{||f||}{\sigma_2} \leq \frac{2\epsilon}{\sigma_2} = \frac{2\epsilon}{\gamma}$. □

So if there is some separation of the small singular values of $A - \zeta I + E1$, and ϵ is small, then the pseudoeigenvectors for the two different perturbations of A are close to each other.

We next move to applications of pseudoeigenvector bases.

4. Exponential of a Matrix. A well known way of computing the exponential of a matrix is to find the eigenvalue decomposition $A = Z\Lambda Z^{-1}$. Then

$$exp(A) = Zexp(\Lambda)Z^{-1}. \tag{4.1}$$

However, this does not work well when the eigenvectors are ill-conditioned. A pseudoeigenvector decomposition can be substituted. We are not advocating this as a good way to compute the matrix exponential, but we will show that the pseudoeigenvector basis can give better results that the eigenvector basis. We note that in [4] Davies does consider some matrix functions, though not the exponential.

Example 7. We again use the butterfly matrix from Example 2 and let the size be $n = 500$. We first compute the eigendecomposition using the Matlab 'eig' command and use it in formula (4.1). We compare with the answer Matlab gives with its 'expm'

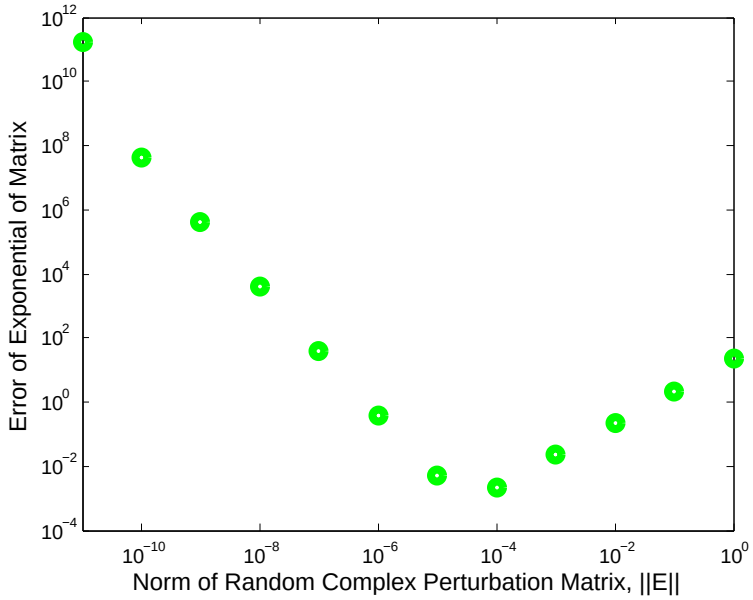

FIG. 4.1. *Difference between exponential of matrix using a pseudoeigenvalue decomposition and the Matlab result.*

command. The eigenvalue decomposition formula differs an enormous amount of $6.2 * 10^{17}$. Next we use the pseudoeigenvalue decomposition. The matrix is perturbed with a random complex matrix of varying norm and then the eigendecomposition is computed and used in formula (4.1). Figure 4.1 has a plot of the difference between the results and Matlab's 'expm' answer. The least difference is $2.2 * 10^{-3}$ with an ϵ of 10^{-4}, but every ϵ gives a big improvement over using the unperturbed eigenvalue decomposition.

5. Tortoise and the Hare Convergence of Restarted GMRES. We now give an example showing how a pseudoeigenvector basis can be of assistance in analyzing convergence of restarted GMRES. Normally with restarted GMRES, the more frequent the restarts, the slower the convergence. Tortoise and the Hare [5] refers to the phenomenon of faster convergence with more frequent restarts. The pseudoeigenvector basis will help in explaning this behavior. It also helps show why there is superlinear convergence.

Example 8. The matrix A is the same as in Example 1, except the size is $n = 20$. The right-hand side is set to all ones and then normed to one. We apply both GMRES(1) and GMRES(3) and the (1,1) panel of Figure 5.1 shows the convergence. Since GMRES(1) converges faster than GMRES(3), this is an example of Tortoise and the Hare convergence. We perturb the matrix with an random real matrix of norm $\epsilon = 10^{-4}$ and compute the eigenvectors to give a pseudoeigenvector basis for the original matrix. The (1,2) panel shows the corresponding pseudoeigenvalues numbered in order of the magnitude of the associated pseudoeigenvalues. The two lower panels show the components of the residual vectors at the beginning of each GMRES cycle expanded in terms of the pseudoeigenvectors. GMRES(1) takes 54 iterations and thus 54 cycles to reach residual norm below 10^{-6}. So there are 54 curves shown in the (2,1) panel for the residual components of GMRES(1), and they move mostly downward as the iteration proceeds. GMRES(3) takes 69 iterations or 23 cycles, and there are 23

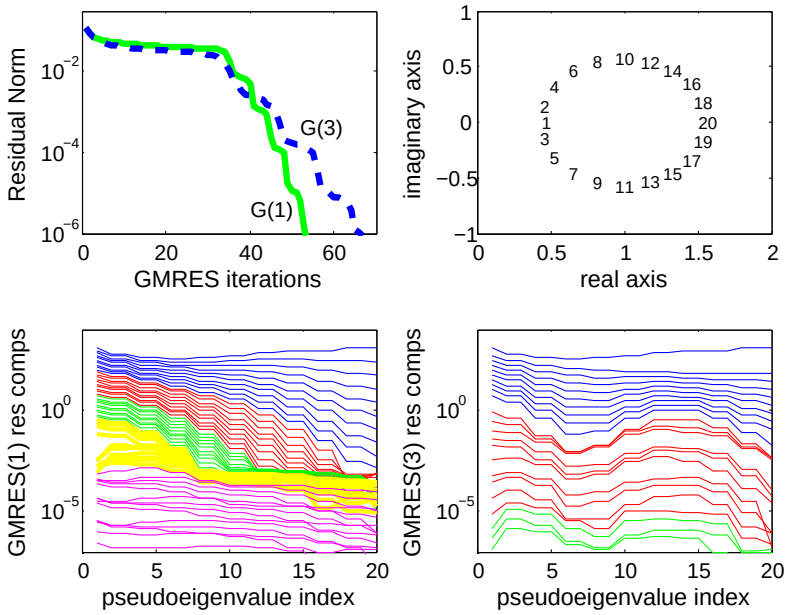

FIG. 5.1. *Tortoise and the Hare GMRES convergence with pseudoeigenvector components.*

curves shown in the (2,2) panel of the figure. We see that GMRES(1) works first on the larger components and meanwhile the overall residual norm plateaus. GMRES(1) next works on the middle-sized components. Then when it finally gets around to reducing the small components, the convergence is rapid. This shows why there is superlinear convergence for GMRES(1): the method is working all along, but only after all regions of the pseudospectrum have been dealt with does the residual norm convergence show this. Now with GMRES(3), the cubic polynomials tend to concentrate in a couple of regions and the residual norm components take on a somewhat cubic shape. GMRES(3) is working too hard in some areas and not enough in others. Results are similar with other right-hand sides and other perturbations, so it seems that GMRES(3) is prone to this cubic-shape behavior. This analysis does not completely explain the convergence, but is does show that using a pseudoeigenvector basis gives some insight.

6. Convergence of GMRES for highly nonnormal matrices. The standard convergence bound for GMRES is not always meaningful for highly nonnormal matrices because of a term involving the condition number of an eigenvector basis. We will use a basis of pseudoeigenvectors to establish a different convergence theorem for GMRES, which may sometimes be more useful. In this theorem, there is a tradeoff between the value of ϵ used and the conditioning of the basis. For small ϵ, the basis may be ill-conditioned. For larger ϵ, another factor in the bound grows.

THEOREM 6.1. *Assume there is a basis for R^n made up of a set of ϵ-pseudoeigenvectors of A. Let the pseudoeigenvectors be $w_1, \ldots, w_n$ and let W be the matrix with these vectors as columns. Let the associated ϵ-pseudoeigenvalues be $\zeta_1, \ldots, \zeta_n$. If the current problem is $A(x - x_0) = r_0$, then after a cycle of restarted GMRES(m), the new*

residual vector satisfies

$$||r|| \leq \big(\kappa(W) \min_{q(0)=1, deg(q)\leq m} \max_{i=1,\ldots,n} q(\zeta_i) + O(\epsilon)\big)||r_0||$$

Proof. Let Λ be the diagonal matrix with the ζ_i's on the diagonal. Then $A+E = W\Lambda W^{-1}$. Next, $q(A) = q(A+E)+O(\epsilon) = q(W\Lambda W^{-1})+O(\epsilon) = Wq(\Lambda)W^{-1}+O(\epsilon)$. Therefore,

$$\begin{aligned}
||r|| &= \min_{q(0)=1, deg(q)\leq m} ||q(A)r_0|| \\
&= \min_{q(0)=1, deg(q)\leq m} ||Wq(\Lambda)W^{-1}r_0|| + O(\epsilon)||r_0|| \\
&\leq \kappa(W) \min_{q(0)=1, deg(q)\leq m} ||q(\Lambda)||||r_0|| + O(\epsilon)||r_0|| \\
&\leq \big(\kappa(W) \min_{q(0)=1, deg(q)\leq m} \max_{i=1,\ldots,n} q(\zeta_i) + O(\epsilon)\big)||r_0||
\end{aligned}$$

□

For a matrix with no basis of eigenvectors or highly ill-conditioned basis of eigenvectors, using a pseudoeigenvector basis and the associated pseudoeigenvalues may be more instructive about convergence.

Example 9. We use two matrices similar to the one from Example 1, but the entries on the main diagonal are now all 1.02 for the first matrix and 1.10 for the second one. Both are size $n = 500$. The right-hand sides are the same random vector. For the case of diagonal elements 1.02, non-restarted GMRES improves the residual norm by a factor of $2.1 * 10^{-4}$ in 300 iterations. For the other matrix, GMRES converges by a factor of $1.3 * 10^{-9}$ in 200 iterations. We consider the convergence estimate $\kappa(Z)(a/c)^{300}$ [24], where $\kappa(Z)$ is the condition number for the pseudoeigenvector basis and where a is the radius and c is the center of a circle containing the pseudospectrum. We use a random real perturbation of norm ϵ for finding pseudoeigenvalues and the pseudoeigenvector basis. The smallest pseudoeigenvalue is used for the estimate of the radius a. See Tables 6.1 and 6.2 for the conditioning of the pseudoeigenvector basis and for the convergence estimate. We notice that we get better convergence estimates with larger values of ϵ, at least up to 10^{-2}. This shows that for this example the basis of pseudoeigenvectors is more useful for larger ϵ. The basis becomes better conditioned as ϵ increases. Next we do the same calculation with the perturbation of only the (n,1) entry that is mentioned in Subsection 2.2. The same calculations are done and only the convergence estimate is given in the last column. This gives estimates that come close to the actual convergence for the larger values of ϵ. However, the estimate may then be losing validity due to a possibly large $O(\epsilon)$ term.

7. Deflated GMRES. Deflated GMRES methods improve the convergence of restarted GMRES by using approximate eigenvectors to remove the effect of some eigenvalues (usually small ones). The eigenvectors can be used to build a preconditioner or they can be put into the subspace. We will consider the latter approach, specifically the method GMRES with deflated restarting (GMRES-DR) [16] which both solves linear equations and computes eigenvalues and eigenvectors. The GMRES-DR method deflates eigenvalues by computing eigenvectors while it is solving linear equations. At the conclusion of a cycle, it uses its subspace to both update the linear equations solution and compute new approximate eigenvectors. Then for the next

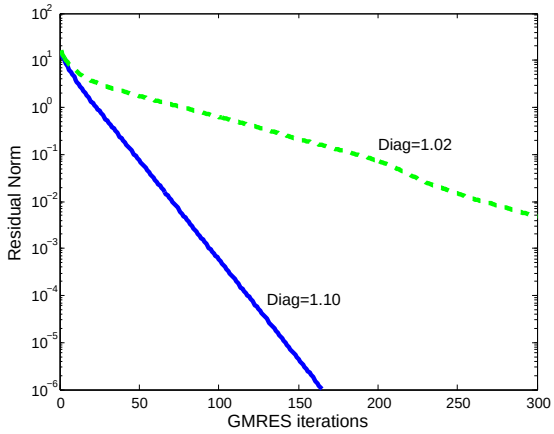

FIG. 6.1. *Convergence of full GMRES for the two bidiagonal matrices.*

TABLE 6.1
Breakdown of a GMRES convergence bound. Matrix with Diag = 1.02. *Residual improves by factor of 2.1e-4 in 300 GMRES iterations.*

ϵ	$\kappa(Z)$	$(a/c)^{300}$	conv. estimate	improved conv. estimate
10^{-8}	1.3e+10	8.5e-9	1.1e+2	4.0
10^{-6}	1.2e+8	1.4e-7	1.6e+1	0.64
10^{-4}	1.0e+6	1.8e-6	1.9	0.10
10^{-3}	9.4e+4	7.6e-6	0.72	0.041
10^{-2}	8.2e+3	9.5e-5	0.78	0.016
10^{-1}	6.8e+2	1.7e-3	1.2	0.0066

cycle it builds a subspace that is spanned by the approximate eigenvectors and new Krylov vectors. This subspace is

$$Span\{\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_k, r_0, Ar_0, A^2r_0, A^3r_0, \ldots, A^{m-k-1}r_0\}, \tag{7.1}$$

where the $\tilde{y}_i$'s are harmonic Ritz vectors generated by the previous cycle and r_0 is the linear equations residual vector from the previous cycle. GMRES-DR(m,k) has maximum size of subspace m and saves k approximate eigenvectors at the restart. The presence of approximate eigenvectors in the subspace essentially removes or deflates the corresponding eigenvalues. This can vastly improve convergence compared to regular GMRES when there are small eigenvalues.

In this section, we show that deflated GMRES can work for highly nonnormal matrices. We will give examples where it is not clear that there are small eigenvalues and the corresponding eigenvectors that can deflated. First a theorem is given that indicates how deflation can work in this situation.

The following theorem shows that the approximate eigenvectors generated by GMRES-DR can be viewed as pseudoeigenvectors and also that the convergence is roughly as if the corresponding pseudoeigenvalues are deflated from the spectrum.

THEOREM 7.1. *Let $y_1, \ldots, y_k$ be normalized approximate eigenvectors of A with associated approximate eigenvalues $\theta_1, \ldots, \theta_k$ and residual vectors $r_i = Ay_i - \theta_i y_i$. Let $Span\{y_1, \ldots, y_k, r_0, Ar_0, \ldots, A^{m-k-1}r_0\}$ be the subspace for a cycle of an augmented*

TABLE 6.2
Breakdown of a GMRES convergence bound. Matrix with Diag = 1.10. Residual improves by factor of 1.3e-9 in 200 GMRES iterations.

ϵ	$\kappa(Z)$	$(a/c)^{200}$	conv. estimate	improved conv. estimate
10^{-6}	1.2e+8	7.3e-12	8.8e-4	2.0e-5
10^{-4}	1.0e+6	4.1e-11	4.3e-5	1.2e-6
10^{-3}	9.4e+4	1.1e-10	1.1e-5	3.3e-7
10^{-2}	8.2e+3	5.7e-10	4.7e-6	8.3e-8
10^{-1}	6.8e+2	4.0e-9	2.7e-6	2.1e-8

minimum residual Krylov method (such as GMRES-DR). Let $Y = [y_1 \dots y_k]$ *and* $R = [r_1 \dots r_k]$. *Let a QR factorization of* Y *be* $Y = QU$ *with* Q n *by* k *orthonormal and* U k *by* k *upper triangular. Let* $E = -RU^{-1}Q^T$ *and* $\epsilon = ||E||$. *Then we first conclude that* $A + E$ *has eigenpairs* (θ_i, y_i) *for* $1 \le i \le k$. *Let the other eigenpairs be* (ξ_i, w_i) *for* $k+1 \le i \le n$, *and let* W *be the matrix with* $y_1, \dots, y_k$ *as the first columns and* $w_{k+1}, \dots w_n$ *as the other columns. Let the current problem be* $A(x - x_0) = r_0$. *Then after a cycle, the new residual vector satisfies*

$$||r|| \le \kappa(W)\Big(\min_{q(0)=1, deg(q)\le m-k} \max_{i=k+1,\dots,n} q(\xi_i) + O(\epsilon)\Big)||r_0||, \tag{7.2}$$

with $\epsilon \le ||R||\kappa(U)$.

Proof. It is quick to show that the (θ_i, y_i)'s are pseudoeigenpairs. Let e_i denote the ith coordinate vector. Then, for $1 \le i \le k$,

$$\begin{aligned}(A+E)y_i &= Ay_i - RU^{-1}Q^T y_i \\ &= Ay_i - RU^{-1}Q^T Y e_i \\ &= \theta_i y_i + r_i - RU^{-1}Q^T QU e_i \\ &= \theta_i y_i + r_i - r_i \\ &= \theta_i y_i.\end{aligned}$$

Next we prove the bound on ϵ, using that Y has normalized columns so the norm of U is bounded below by 1:

$$\epsilon = ||E|| = ||RU^{-1}Q^T|| = ||RU^{-1}|| \le ||R||\,||U^{-1}|| \le ||R||\,||U^{-1}||\,||U|| = ||R||\kappa(U).$$

Finally, let $\xi_i = \theta_i$ for $1 \le i \le k$, and let Ξ be the diagonal matrix with the ξ_i's on the diagonal. Let $r_0 = \Sigma_{i=1}^k \beta_i y_i + \Sigma_{i=k+1}^n \beta_i w_i$. After a cycle of GMRES-DR, let the approximate solution be $\hat{x} = \Sigma_{i=1}^k \gamma_i y_i + p(A)r_0$, for p a polynomial of degree $m-k-1$. Then the new residual vector is $r = r_0 - A\hat{x} = r_0 - \Sigma_{i=1}^k \gamma_i A y_i - Ap(A)r_0 = -\Sigma_{i=1}^k \gamma_i A y_i + q(A)r_0$, where $q(\alpha) \equiv 1 - \alpha p(\alpha)$.

Next, as in the proof of Theorem 6.1, $q(A) = Wq(\Xi)W^{-1} + O(\epsilon)$. In the next equation, the minimum is over all choices of γ_i's and of polynomials q satisfying the given conditions.

$$\begin{aligned}||r|| &= \min_{q(0)=1, deg(q)\le m-k} || - \Sigma_{i=1}^k \gamma_i A y_i + q(A)r_0|| \\ &= \min_{q(0)=1, deg(q)\le m-k} || - \Sigma_{i=1}^k \gamma_i(\theta_i y_i - E y_i) + \Sigma_{i=1}^k \beta_i q(\xi_i) y_i + \Sigma_{i=k+1}^n \beta_i q(\xi_i) w_i + O(\epsilon) r_0||.\end{aligned}$$

We can choose the γ_i's to be any values we want, because the minimum residual will always be as good. Letting $\gamma_i = -\beta_i q(\xi_i)/\theta_i$,

$$||r|| \leq \min_{q(0)=1, deg(q)\leq m-k} || - \Sigma_{i=1}^{k} \frac{\beta_i}{\theta_i} q(\xi_i) E y_i + \Sigma_{i=k+1}^{n} \beta_i q(\xi_i) w_i + O(\epsilon) r_0 ||$$

The β_i's are bounded by $\kappa(W)||r_0||$ and the E terms go into the $O(\epsilon)$:

$$||r|| \leq \kappa(W) \Big(\min_{q(0)=1, deg(q)\leq m-k} \max_{i=k+1,\ldots,n} q(\xi_i) + O(\epsilon) \Big) ||r_0||$$

□

Note the bound in Equation (7.2) has a max over the polynomial at eigenvalues not including k of them. So these k are deflated from the bound.

The $\kappa(U)$ term in the bound on the size of ϵ is one in the symmetric case. Then the size of ϵ depends on the accuracy of the approximate eigenvectors.

Example 10. Let A be the Jordan block matrix from Example 1. Let $n = 500$ and let the right-hand side have random Normal(0,1) entries. This example is extreme, because there is only one eigenvector and all of the eigenvalues are 1. It is not at all clear that deflating eigenvalues will be helpful, because there are no small eigenvalues. Also, with only one eigenvector, using approximate eigenvectors for deflation is questionable. However, as shown in Figure 7.1, GMRES-DR is able to effectively deflate eigenvalues and give faster convergence than restarted GMRES. GMRES-DR(40,20) uses 20 approximate eigenvectors to deflate and a Krylov portion of dimension 20. It converges faster than both GMRES(40) and GMRES(20). Notice there is a Tortoise and the Hare effect [5] here for restarted GMRES as $m = 20$ converges faster than $m = 40$.

GMRES-DR computes approximate eigenvectors that are also pseudoeigenvectors and deflates the corresponding pseudoeigenvalues, as indicated by Theorem 7.1. Figure 7.2 shows the residual norms in GMRES-DR(30,10) for the harmonic Ritz pairs which are the approximate eigenpairs. They seem to be converging well. However, the harmonic Ritz values behave differently than for most matrices. They do not settle down. Figure 7.3 shows that they move to the right in the complex plane as they converge. Since the only eigenvalue is at 1.0, it makes sense that approximate eigenvalues would move towards this value as they converge. Figure 7.4 shows all of the harmonic Ritz values for GMRES-DR(30,10) after cycle 30 and also the values for GMRES(20) for cycle 30. GMRES-DR(30,10) has small pseudoeigenvalues that are deflated out and improve the convergence.

Next we look at applying a portion of Theorem 7.1 to this situation. We look at the perturbed matrix $A+E$ for $E = -RU^{-1}Q^T$, where the quantities are defined in the theorem. The theorem claims that GMRES-DR convergence goes mostly according to the spectrum of this matrix with k eigenvalues deflated out. Figure 7.5 shows the smallest eigenvalues of this perturbed matrix after 10 and after 30 cycles of GMRES-DR(40,20). For after 10 cycles, the dots are the eigenvalues of $A+E$, and the asterisks are the 20 smallest harmonic Ritz values. The eigenvalues of $A + E$ that have an asterisk around them are deflated from the spectrum. However, there are some dots close to the origin without asterisks, so the convergence at 10 cycles is still slow (there are 220 matrix-vector products after 10 cycles). After 30 cycles and 620 matrix-vector products, the eigenvalues of $A + E$ are shown with pluses and all of the small ones correspond to harmonic Ritz values, shown with circles. At that point, convergence is very fast because the small eigenvalues of the perturbed matrix are being deflated.

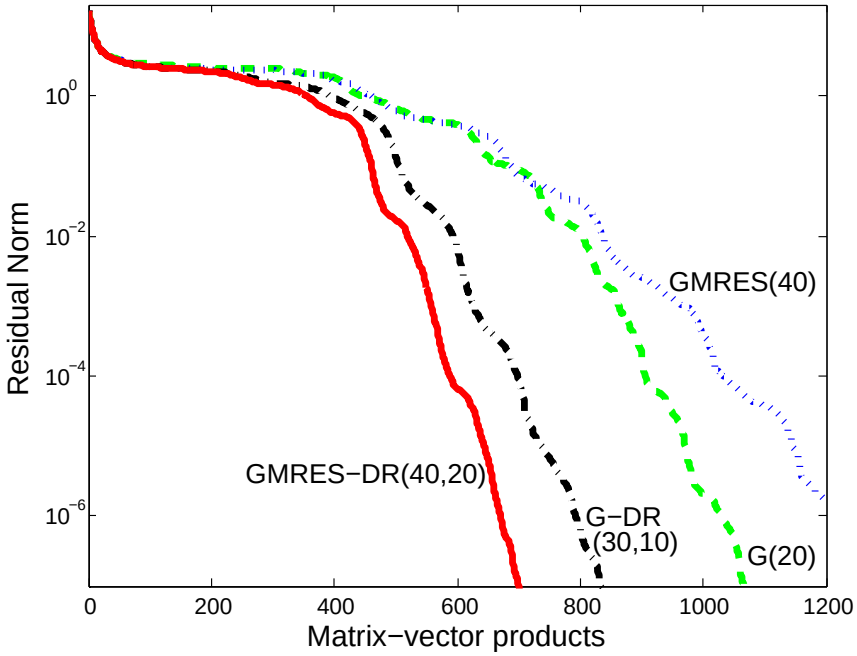

FIG. 7.1. *GMRES and GMRES-DR for the Jordan block matrix. Deflation of eigenvalues is effective even though there are no small eigenvalues.*

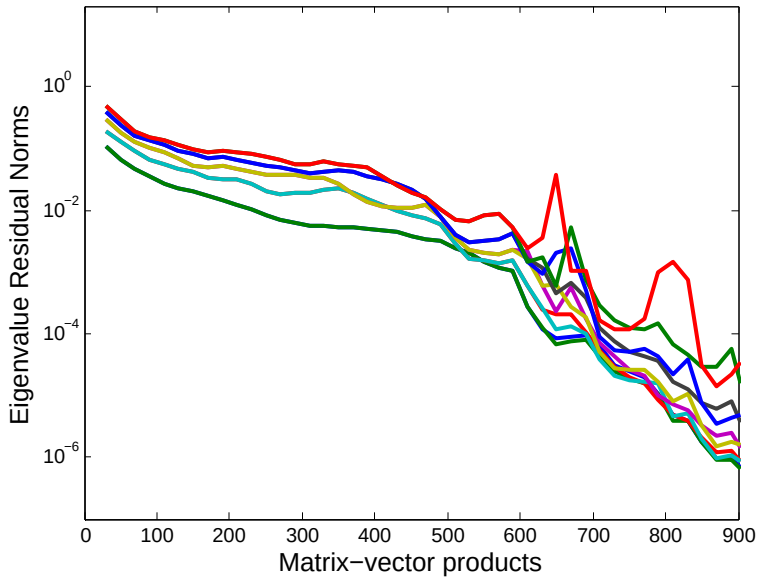

FIG. 7.2. *Convergence of harmonic Ritz pairs for GMRES-DR(30,10).*

We now run a test that shows that the approximate eigenvectors generated by GMRES-DR are useful for deflating eigenvalues for another system of equation with the same matrix and different right-hand side. We take the harmonic Ritz vectors from the end of a run of GMRES-DR(40,20) that goes for 36 cycles or 700 matrix-vector products. Then the GMRES(20)-Proj(20) method [17] is applied. It alternates

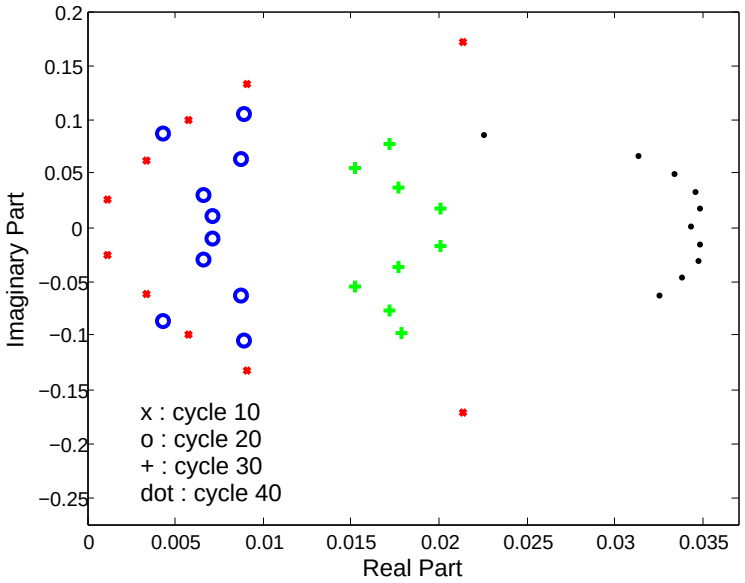

FIG. 7.3. *Smallest harmonic Ritz values for GMRES-DR(30,10) after cycles 10, 20, 30 and 40.*

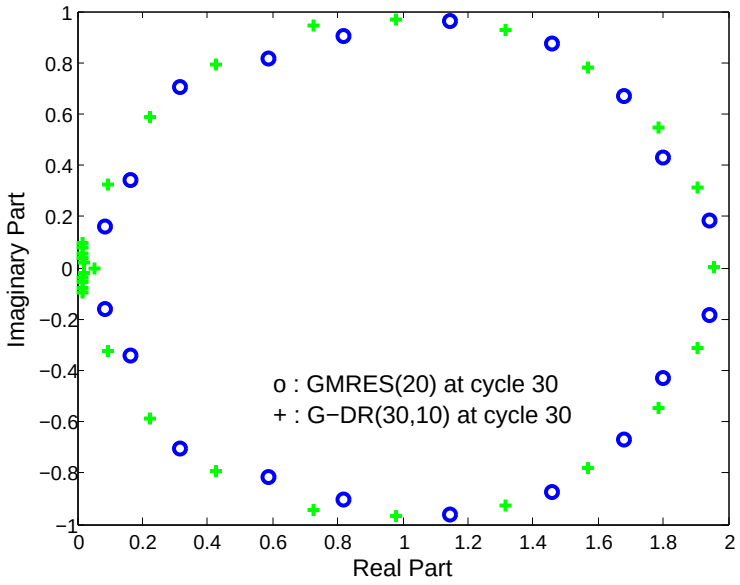

FIG. 7.4. *Harmonic Ritz values for GMRES(20) and GMRES-DR(30,10) after cycle 30.*

minimum residual projection over the 20 harmonic Ritz vectors with cycles of GMRES(20). Figure 7.6 shows the convergence versus regular GMRES(20). The approximate eigenvalues are useful even though they are fixed vectors with fixed harmonic Ritz values instead of the continually moving values used by GMRES-DR. These pseudoeigenvectors are not always as effective in this second right-hand side situation; see the tests in Example 12.

Example 11. We now shift the Jordan block matrix in the previous example to make it more difficult. We slightly change all of the diagonal elements to 0.999

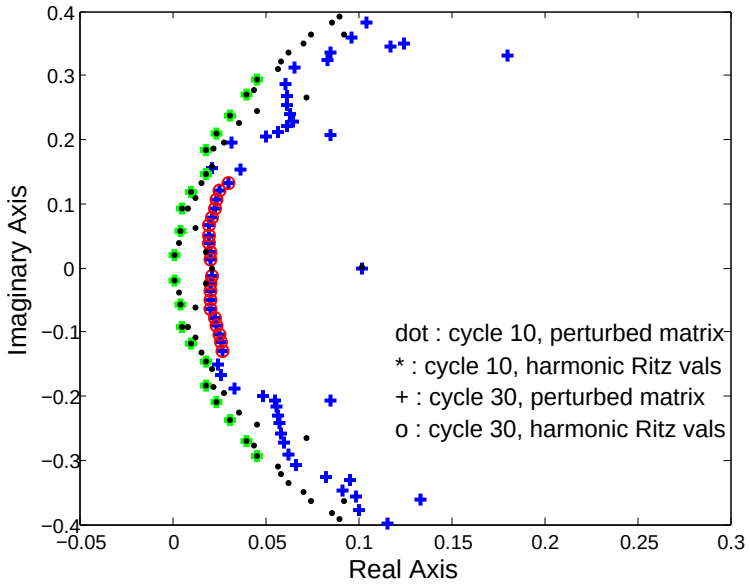

FIG. 7.5. *Eigenvalues of the perturbed matrix from Theorem 7.1*

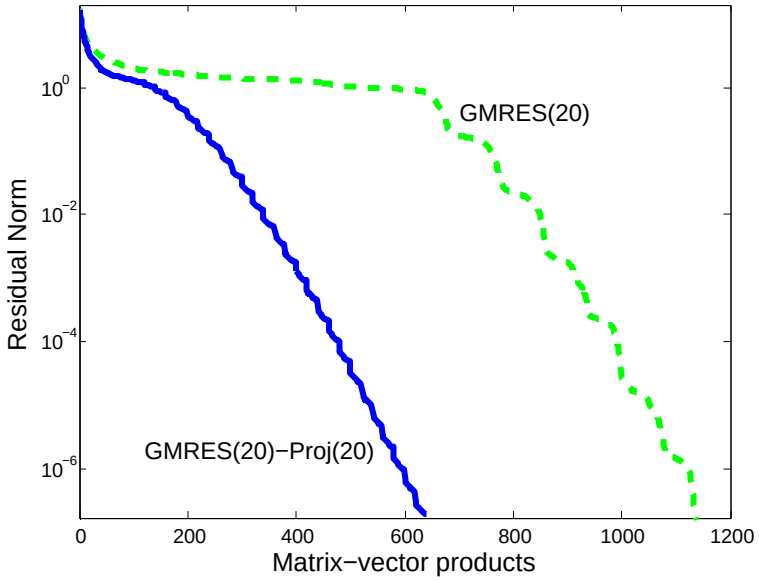

FIG. 7.6. *Deflation of pseudoeigenvalues for a second right-hand sided.*

and leave the superdiagonal elements at 1.0. Figure 7.7 shows that this shift makes the problem more difficult for restarted GMRES. The small pseudoeigenvalues are shifted closer to the origin by 0.001. With a further shift of diagonal elements to 0.998, the problem is too difficult for GMRES(20) and GMRES(40). They do not converge. Meanwhile GMRES-DR(40,20) works about the same as before. In fact, even with the matrix having diagonal elements of 0.992, GMRES-DR(40,20) converges to relative residual tolerance of 10^{-10} in 920 matrix-vector products versus 760 for the first case of diagonal elements equal to 1.0. However, with diagonal elements of 0.990,

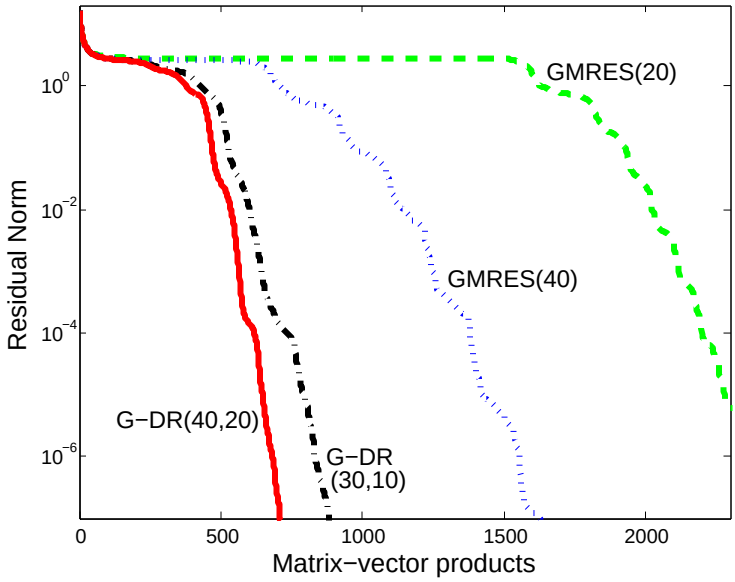

FIG. 7.7. *GMRES and GMRES-DR for shifted Jordan block matrix with diagonal elements 0.999.*

TABLE 7.1
List of number of pseudoeigenvectors needed for deflation with different shifts.

	n = 500	n = 1000
shift	k needed	k needed
0	0	0
0.005	10	45
0.01	15	50
0.02	25	55
0.04	35	80
0.08	45	90
0.16	50	100
0.32	60	120
0.64	80	165

GMRES-DR(40,20) also stagnates (interestingly, GMRES-DR(35,15) does converge).

Next, we use different shifts of the original Jordan block matrix and look at how many pseudoeigenvalues need to be deflated in order for GMRES-DR to converge. We use GMRES-DR(20+k,k) for values of k that are multiples of five. We start with a shift of 0.005 that gives diagonal elements of 0.995 and then keep doubling the shift. The size of k needed is given in Table 7.1 for each shift. This number of pseudoeigenvalues needed for deflation increases as the pseudospectrum is pushed to the left. Results are also given for the matrix of twice the size and then many more pseudoeigenvectors need to be deflated. For the larger shifts, about twice as many are needed. It makes sense that the pseudoeigenvalues are packed in about twice as thick for the larger matrix, so more need to be deflated.

Example 12. We next test with the Butterfly matrix from Example 2. The size is $n = 500$, and the diagonal elements of the matrix are changed from 0 to 1.9. This

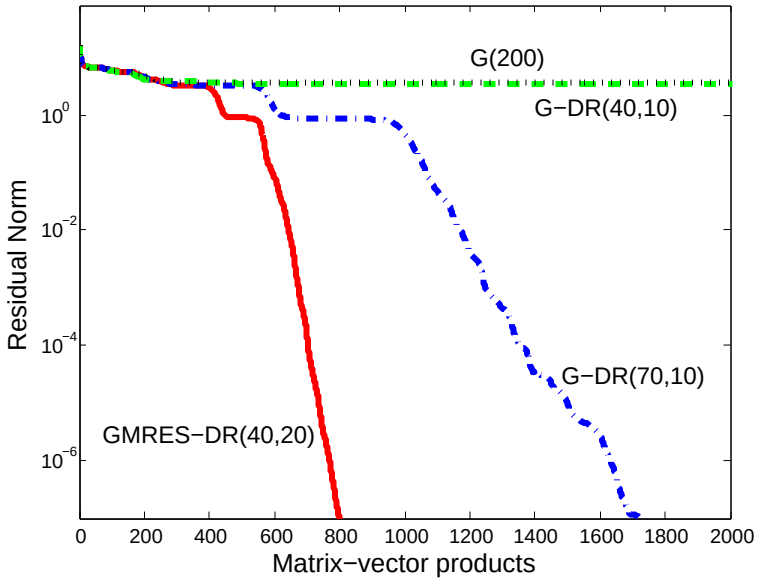

FIG. 7.8. *GMRES and GMRES-DR for the Butterfly matrix.*

shifts the pseudospectrum from surrounding the origin in the complex plane to just right of the origin (depending on the ϵ). Deflation of eigenvalues is necessary for convergence. Figure 7.8 shows that deflating 20 pseudoeigenvalues is very effective. Figure 7.9 has eigenvalues of the perturbed matrix $A + E$ from Theorem 7.1 at the end of 10 cycles and 30 cycles of GMRES-DR(40,20). The closeup in Figure 7.10 shows the perturbed matrix has some small eigenvalues that do not correspond to any of the 20 harmonic Ritz values and so are not being deflated. However, by cycle 30, all of the smallest eigenvalues of the perturbed matrix (denoted by pluses) do correspond to harmonic Ritz values. So at that point, there is fast convergence for GMRES-DR(40,20).

Next, we again use pseudoeigenvectors from solving one system to deflate pseudoeigenvalues for a second system with a different right-hand side. We apply GMRES-DR(40,20) for solving the first right-hand side, then deflate the 20 pseudoeigenvectors for GMRES(20)-Proj(20) on the second right-hand side. However, this method does not converge; see Figure 7.11. Using 30 pseudoeigenvectors is effective. It seems that GMRES-Proj may need more pseudoeigenvectors than GMRES-DR.

Example 13. The next test matrix is a SUPG matrix from [6] with $n = 2500$ and $\nu = .01$. Figure 7.12 shows the eigenvalues of the matrix and the pseudoeigenvalues using perturbations with three different values of ϵ. Looking only at the eigenvalues, the problem appears easy, because the eigenvalues are clumped together well away from the origin. However, the pseudoeigenvalues come close to the origin for the larger values of ϵ, so the problem is not so simple for restarted GMRES. Figure 7.13 shows that deflating pseudoeigenvectors is very helpful. GMRES-DR(40,20) converges about three times faster than the restarted GMRES runs. The harmonic Ritz values for GMRES(20) and GMRES-DR(30,10) after cycles 3 and 5 are shown in Figure 7.14.

8. Conclusion. This paper shows that pseudoeigenvectors can be used to assist understanding of GMRES convergence. In particular, it is explained why deflating

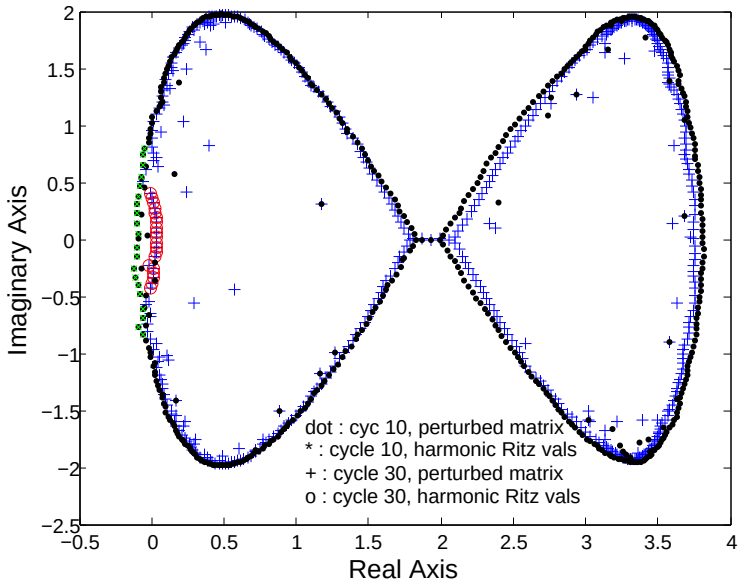

FIG. 7.9. *Eigenvalues of the perturbed matrix from Theorem 7.1 for the Butterfly matrix.*

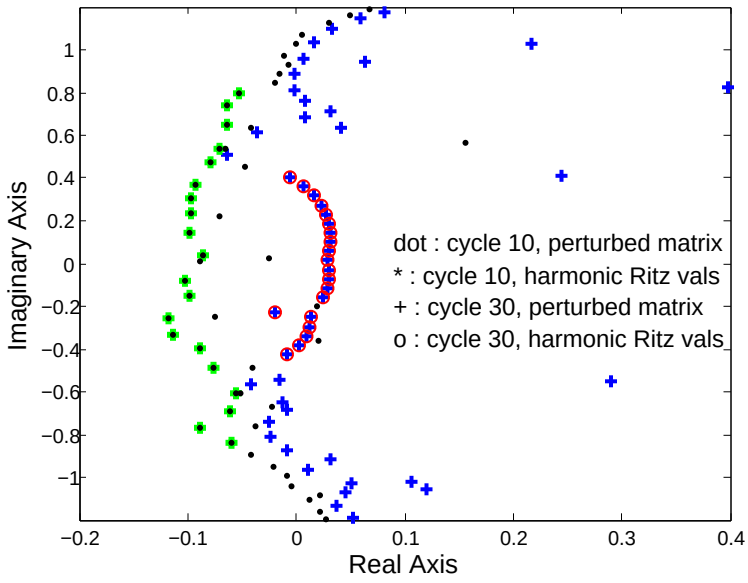

FIG. 7.10. *Closeup of eigenvalues of the perturbed matrix from Thm 7.1 for the Butterfly matrix.*

eigenvalues can be effective even when the eigenvectors are poorly defined. There is still much to be studied with GMRES convergence and pseudoeigenvectors are one possible tool.

Pseudoeigenvectors are also potentially useful for computations, such as for matrix exponential. They can perform much better than eigenvectors. They may not be competitive method with leading approaches for small matrices. However, it is possible that for large matrices some pseudoeigenvectors can be used for an approximation. These can come from either eigenvectors of a perturbed matrix or from inaccurately

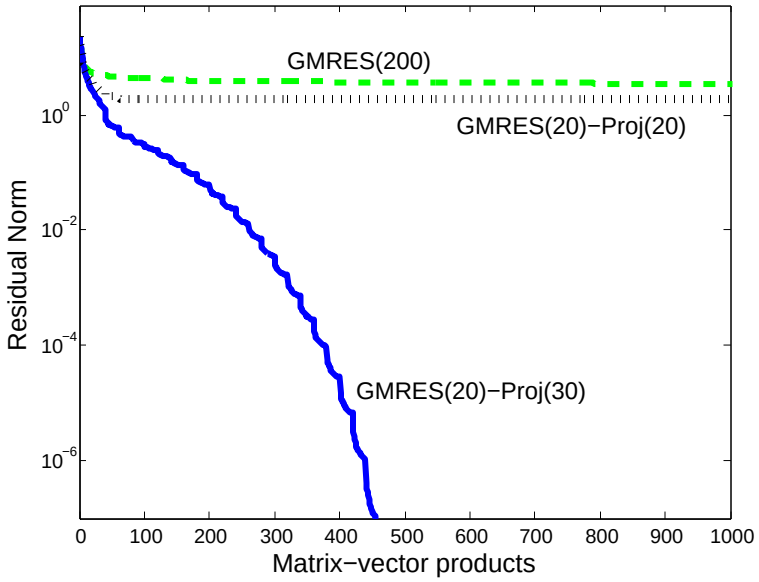

FIG. 7.11. *Deflation of pseudoeigenvalues for a second right-hand sided.*

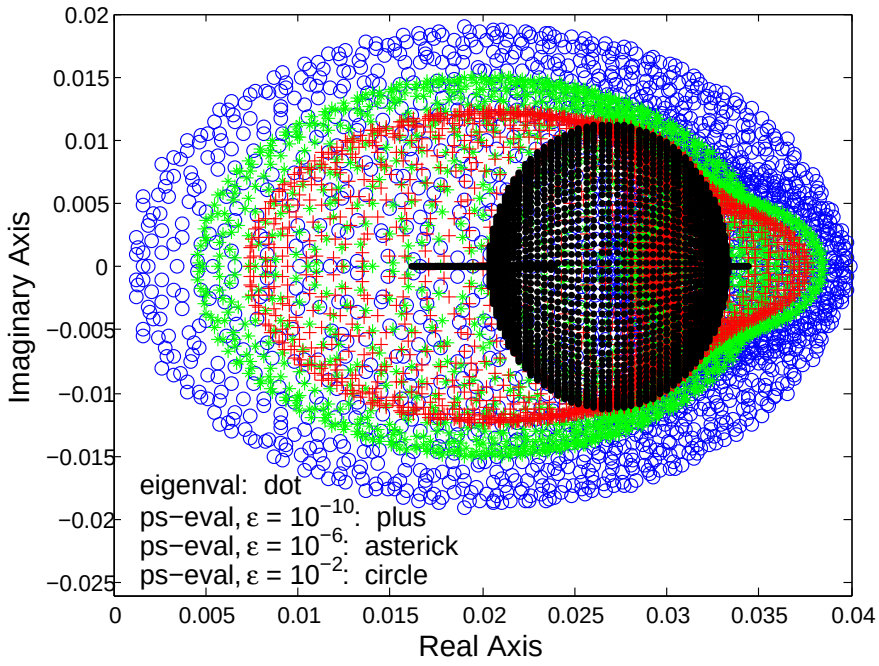

FIG. 7.12. *Eigenvalues and some pseudoeigenvalues of SUPG matrix.*

computing eigenvectors of the original matrix.

This paper also looks at aspects of stability and variability of pseudoeigenvector bases and compares them to eigenvector bases for highly non-normal matrices. While pseudoeigenvector bases for a fixed matrix do vary and some are better than others, some stability can be shown.

Perhaps in the future pseudoeigenvectors will receive more attention like their

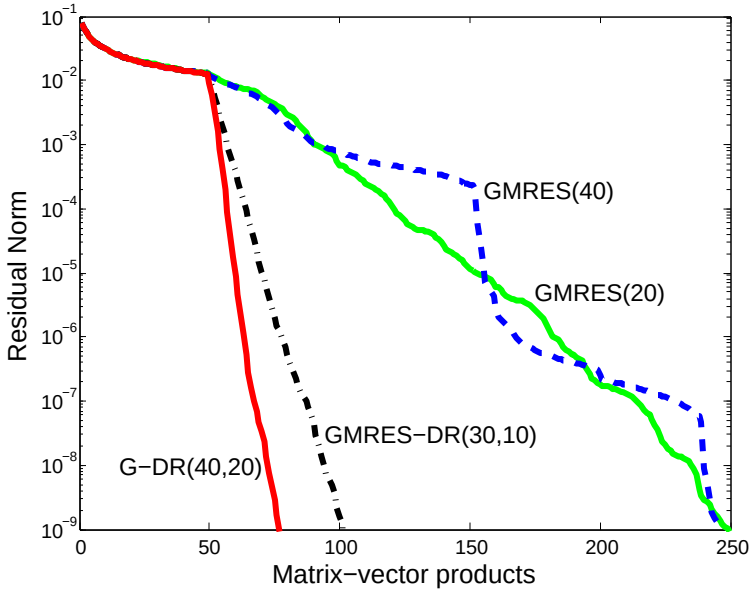

FIG. 7.13. *GMRES and GMRES-DR for the SUPG matrix. Deflation of eigenvalues is effective even though there are no small eigenvalues.*

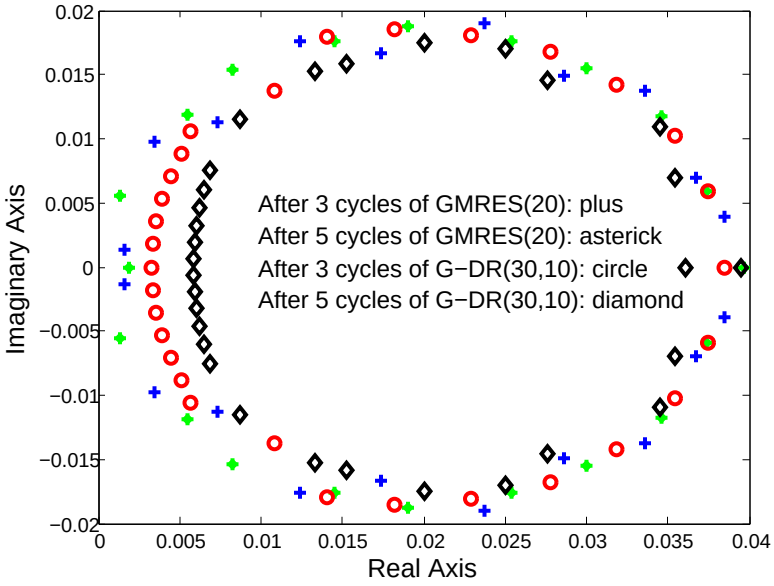

FIG. 7.14. *Harmonic Ritz values for GMRES(20) and GMRES-DR(30,10) after cycles 3 and 5.*

companion pseudoeigenvalues.

Acknowledgments. The authors would like to thank Mark Embree for many discussions and helpful comments.

REFERENCES

[1] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel. Adaptively preconditioned GMRES algorithms. *SIAM J. Sci. Comput.*, 20:243–269, 1998.
[2] K. Burrage and J. Erhel. On the performance of various adaptive preconditioned GMRES strategies. *Num. Lin. Alg. with Appl.*, 5:101–121, 1998.
[3] A. Chapman and Y. Saad. Deflated and augmented Krylov subspace techniques. *Num. Lin. Alg. with Appl.*, 4:43–66, 1997.
[4] E. B. Davies. Approximate diagonalization. *SIAM J. Matrix Anal. Appl.*, 29:1051–1064, 2007.
[5] M. Embree. The tortoise and the hare restart GMRES. *SIAM Review*, 45:259–266, 2003.
[6] B. Fischer, A. Ramage, D. J. Silvester, and A. J. Wathen. On parameter choice and iterative convergence for stabilised discretisations of advection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 179:185–202, 1999.
[7] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. *SIAM J. Sci. Statist. Comput.*, 23:442–462, 2001.
[8] A. Gaul, M. H. Gutknecht, J. Liesen, and R. Nabben. A framework for deflated and augmented Krylov subspace methods. *SIAM J. Matrix Anal. Appl.*, 34:495–518, 2013.
[9] L. Giraud, S. Gratton, X. Pinel, and X. Vasseur. Flexible GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 32:1858–1878, 2010.
[10] A. Greenbaum, V. Ptak, and Z. Strakos. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, 17:465–469, 1996.
[11] A. Greenbaum and L. N. Trefethen. GMRES and Arnoldi as matrix approximation problems. *SIAM J. Sci. Comput.*, 15:359–367, 1994.
[12] W. Joubert. On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems. *Num. Lin. Alg. with Appl.*, 1:427–447, 1994.
[13] S. A. Kharchenko and A. Y. Yeremin. Eigenvalue translation based preconditioners for the GMRES(k) method. *Num. Lin. Alg. with Appl.*, 2:51–77, 1995.
[14] J. Liesen and Z. Strakos. Convergence of GMRES for tridiagonal Toeplitz matrices. *SIAM J. Matrix Anal. Appl.*, 26:233–251, 2004.
[15] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16:1154–1171, 1995.
[16] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 24:20–37, 2002.
[17] R. B. Morgan and W. Wilcox. Deflated iterative methods for linear equations with multiple right-hand sides. arXiv:math-ph/0405053v2, 2004.
[18] J. Moro, J. V. Burke, and M. L. Overton. On the Lidskii–Vishik–Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. *SIAM J. Matrix Anal. Appl.*, 18:793–817, 1997.
[19] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.*, 13:778–795, 1992.
[20] N. M. Nachtigal, L. Reichel, and L. N. Trefethen. A hybrid GMRES algorithm for nonsymmetric linear systems. *SIAM J. Matrix Anal. Appl.*, 13:796–825, 1992.
[21] 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.*, 28:1651–1674, 2006.
[22] S. C. Reddy and L. N. Trefethen. Lax-stability of fully discrete spectral methods via stability regions and pseudo-eigenvalues. *Comput. Methods Appl. Mech. Engrg.*, 80:147–164, 1990.
[23] L. Reichel and L. N. Trefethen. Eigenvalues and pseudo-eigenvalues of toeplitz matrices. *Linear Algebra Appl.*, 30:153–185, 1992.
[24] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd Edition.* SIAM, Philadelphia, PA, 2003.
[25] Y. Saad and M. H. Schultz. GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7:856–869, 1986.
[26] J. A. Sifuentes, M. Embree, and R. B. Morgan. GMRES convergence for perturbed coefficient matrices, with application to approximate deflation preconditioning. *SIAM J. Matrix Anal. Appl.*, 34:1066–1088, 2013.
[27] P. Tichý, J. Liesen, and V. Faber. On worst-case GMRES, ideal GMRES, and the polynomial numerical hull of a Jordan block. *ETNA*, 26:453–473, 2007.
[28] L. N. Trefethen. Approximation theory and numerical linear algebra. In J. C. Mason and M. G. Cox, editors, *Chapman and Hall.* Springer, London, 1990.
[29] L. N. Trefethen and S. J. Chapman. Wave packet pseudomodes of twisted Toeplitz matrices. *Comm. Pure Appl. Math.*, 57:1233–1264, 2004.
[30] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators.* Princeton Univ. Press, Princeton, NJ, 2005.
[31] B. Zhong and R. B. Morgan. Complementary cycles of restarted GMRES. *Numer. Linear Algebra Appl.*, 15:559–571, 2008.