SIAM J. SCI. COMPUT.
Vol. 39, No. 5, pp. S610–S632

WEIGHTED INNER PRODUCTS FOR GMRES AND GMRES-DR*

MARK EMBREE†, RONALD B. MORGAN‡, AND HUY V. NGUYEN§

Abstract. The convergence of the restarted GMRES method can be significantly improved, for some problems, by using a weighted inner product that changes at each restart. How does this weighting affect convergence, and when is it useful? We show that weighted inner products can help in two distinct ways: when the coefficient matrix has localized eigenvectors, weighting can allow restarted GMRES to focus on eigenvalues that otherwise cause slow convergence; for general problems, weighting can break the cyclic convergence pattern into which restarted GMRES often settles. The eigenvectors of matrices derived from differential equations are often not localized, thus limiting the impact of weighting. For such problems, incorporating the discrete cosine transform into the inner product can significantly improve GMRES convergence, giving a method we call W-GMRES-DCT. Integrating weighting with eigenvalue deflation via GMRES-DR can also give effective solutions.

Key words. linear equations, localized eigenvectors, weighted inner product, restarted GMRES, GMRES-DR, W-GMRES-DCT, deflation

AMS subject classifications. 65F10, 15A06

DOI. 10.1137/16M1082615

1. Introduction.

We seek to solve large nonsymmetric systems of linear equations using variants of the GMRES method [40] that change the inner product each time the method is restarted.

To solve $Ax = b$ for the unknown x, GMRES computes the approximate solution x_k from the Krylov subspace $\mathcal{K}_k(A, b) = \text{span}\{b, Ab, \ldots, A^{k-1}b\}$ that minimizes the norm of the residual $r_k := b - Ax_k$. This residual is usually optimized in the (Euclidean) 2-norm, but the algorithm can be implemented in any norm induced by an inner product. This generality has been studied in theory (see, e.g., [16, 17]) and in practice. The inner product affects neither the spectrum of A nor the subspace $\mathcal{K}_k(A, b)$ (assuming the method is not restarted), but it can significantly alter the departure of A from normality and the conditioning of the eigenvalues. Numerous authors have proposed inner products in which certain preconditioned saddle point systems are self-adjoint, enabling optimal short-recurrence Krylov methods in place of the long recurrences required for the 2-norm [7, 18, 26, 27, 45]. Recently Pestana and Wathen used nonstandard inner products to inform preconditioner design [36].

For any inner product $\langle \cdot, \cdot \rangle_W$ on $\mathbb{R}^n$, there exists a matrix $W \in \mathbb{R}^{n \times n}$ for which $\langle u, v \rangle_W = v^* W u$ for all $u, v \in \mathbb{R}^n$, and W is Hermitian positive definite in the Euclidean inner product. (Throughout, * denotes the conjugate transpose.) With the W-inner product is associated the norm $\|u\|_W = \sqrt{\langle u, u \rangle_W} = \sqrt{u^* W u}$. To implement

*Received by the editors July 1, 2016; accepted for publication (in revised form) April 10, 2017; published electronically October 26, 2017.

http://www.siam.org/journals/sisc/39-5/M108261.html

Funding: The first author's research was supported in part by National Science Foundation grant DGE-1545362. The second author's research was supported in part by the National Science Foundation Computational Mathematics Program under grant DMS-1418677.

†Department of Mathematics and Computational Modeling and Data Analytics Division, Academy of Integrated Science, Virginia Tech, Blacksburg, VA 24061 (embree@vt.edu).

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

§Department of Mathematics and Computer Science, Austin College, Sherman, TX 75090 (hnguyen@austincollege.edu).

GMRES in the W-inner product, simply replace all norms and inner products in the standard GMRES algorithm. Any advantage gained by this new geometry must be balanced against the extra expense of computing the inner products: essentially one matrix-vector product with W for each inner product and norm evaluation.

In 1998 Essai proposed an intriguing way to utilize the inner product in restarted GMRES [14]: take W to be diagonal and *change the inner product at each restart*. His computational experiments show that this "weighting" can significantly improve the performance of restarted GMRES, especially for difficult problems and frequent restarts. Subsequent work has investigated how weighted GMRES performs on large examples [9, 23, 29, 33, 42]. We offer a new perspective on why weighting helps (or not), justified by simple analysis and careful computational experiments.

Section 2 describes Essai's algorithm. In section 3, we provide some basic analysis for weighted GMRES and illustrative experiments, arguing that the algorithm performs well when the eigenvectors are *localized*. In section 4 we propose a new variant that combines weighting with the discrete cosine transform, which localizes eigenvectors in many common situations. Section 5 describes a variant with deflated restarting (GMRES-DR [31]) that both solves linear equations and computes eigenvectors. The eigenvectors improve the convergence of the solver for the linear equations, and a weighted inner product can help further.

2. Weighted GMRES. The standard restarted GMRES algorithm [40] computes the residual-minimizing approximation from a degree-m Krylov subspace and then uses the residual associated with this solution estimate to generate a new Krylov subspace. We call this algorithm *GMRES(m)* and every set of m iterations after a restart a *cycle*. Essai's variant [14], which we call *W-GMRES(m)*, changes the inner product at the beginning of each cycle of restarted GMRES, with weights determined by the residual vector obtained from the last cycle. Let $r_{km} = b - Ax_{km}$ denote the GMRES(m) residual after k cycles. Then the weight matrix defining the inner product for the next cycle is $W = \mathrm{diag}(w_1, \ldots, w_n)$, where

$$w_j = \max\left\{ \frac{|(r_{km})_j|}{\|r_{km}\|_\infty}, 10^{-10} \right\}, \tag{1}$$

where $(r_{km})_j$ denotes the jth entry of r_{km}.[1] The weighted inner product is then $\langle u, v\rangle_W = v^* W u$ giving the norm $\|u\|_W = \sqrt{\langle u, u\rangle_W}$. At the end of a cycle, the residual vector r_{km} is computed and from it the inner product for the next cycle is built. We have found it helpful to restrict the diagonal entries in (1) to be at least 10^{-10}, thus limiting the condition number of W to $\|W\|_2\|W^{-1}\|_2 \le 10^{10}$. Finite precision arithmetic influences GMRES in nontrivial ways; see, e.g., [28, sect. 5.7]. Weighted inner products can further complicate this behavior; a careful study of this phenomenon is beyond the scope of our study, but note that Rozložník et al. [37] have studied orthogonalization algorithms in general inner products. (Although reorthogonalization is generally not needed for GMRES [34], all tests here use one step of full reorthogonalization for extra stability. Running the experiments in section 2 without this step yields qualitatively similar results, but the iteration counts differ. All computations were performed using MATLAB 2015b. Different MATLAB versions yield somewhat different results, especially for experiments that require many iterations.)

For diagonal W, the inner product takes $3n$ operations (additions and multiplications), instead of $2n$ for the Euclidean case. This increase of roughly 25% in

[1] Essai scales these entries differently, but, as Cao and Yu [9] note, scaling the inner product by a constant does not affect the residuals produced by the algorithm.

the Gram–Schmidt step in GMRES can be significant for very sparse A, but if the matrix-vector product is expensive, the cost of weighting is negligible.

How can weighting help? Suppose we restart GMRES every m iterations. GMRES minimizes a residual of the form $\varphi(A)r$, where φ is a polynomial of degree no greater than m with $\varphi(0) = 1$. When A is nearly normal, one learns much from the magnitude of φ on the spectrum of A. Loosely speaking, since GMRES must minimize the norm of the residual, the optimal φ cannot target isolated eigenvalues near the origin, since φ would then be large at other eigenvalues. This constraint can cause uniformly slow convergence from cycle to cycle; the restarted algorithm fails to match the "superlinear" convergence enabled by higher degree polynomials in full GMRES [48]. (Higher degree polynomials can target a few critical eigenvalues with a few roots, while having sufficiently many roots remaining to be uniformly small on the rest of the spectrum.) Weighting can skew the geometry to favor the troublesome eigenvalues that impede convergence, leaving easy components to be eliminated quickly in subsequent cycles. In aggregate, weighting can produce a string of iterates, each suboptimal in the 2-norm, that collectively give faster convergence than standard restarted GMRES, with its locally optimal (but globally suboptimal) iterates.

Essai's strategy for the inner product relies on the intuition that one should emphasize those components of the residual vector that have the largest magnitude. Intuitively speaking, these components have thus far been neglected by the method and may benefit from some preferential treatment. By analyzing the spectral properties of the weighted GMRES algorithm in section 3, we explain when this intuition is valid, show how it can go wrong, and offer a possible remedy.

Essai showed that his weighting can give significantly faster convergence for some problems but did not provide guidance on matrix properties for which this was the case. Later, Saberi Najafi and Zareamoghaddam [42] showed that diagonal W with entries from a *random* uniform distribution on $[0.5, 1.5]$ can sometimes be effective, too. Weighting appears to be particularly useful when the GMRES(m) restart parameter m is small, though it can help for larger m when the problem is difficult.

Example 2.1. Consider the Add20 example from the Matrix Market collection [6], a nonsymmetric matrix of dimension $n = 2395$ used by Essai [14]. (Apparently our "cycles" correspond to Essai's "iterations.") Here b is a random Normal(0,1) vector. Figure 1 compares GMRES(m) and W-GMRES(m) for $m = 6$ and 20. In terms of matrix-vector products with A, W-GMRES(6) converges about 12.5 times faster than GMRES(6); the relative improvement for $m = 20$ is less extreme, but W-GMRES is still better by more than a factor of two. (Full GMRES converges to $\|r_k\|_2/\|r_0\|_2 \leq 10^{-10}$ in 509 iterations.) Note that $\|r_k\|_2$ does not converge monotonically for W-GMRES, since the residuals are minimized in the weighted norm, not the 2-norm.

Is weighting typically this effective? Essai shows several other tests where weighting helps, though not as much as for Add20. Cao and Yu [9] give evidence that suggests weighting is not as effective for preconditioned problems, though it can still help. Güttel and Pestana arrive at a similar conclusion after conducting extensive numerical experiments on 109 ILU-preconditioned test matrices: "weighted GMRES may outperform unweighted GMRES for some problems, but more often this method is not competitive with other Krylov subspace methods..." [23, p. 733] and "we believe that WGMRES should not be used in combination with preconditioners, although we are aware that for some examples it may perform satisfactorily" [23, p. 750]. Yet weighting is fairly inexpensive and can improve convergence for some difficult problems. When weighting works, what makes it work?

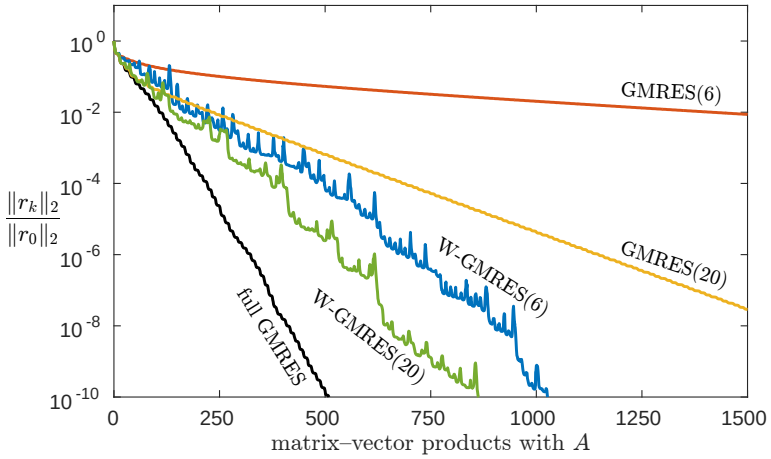


FIG. 1. *Convergence of full GMRES, GMRES(m), and W-GMRES(m) with* $m = 6$ *and* $m = 20$ *for the Add*20 *matrix. In this and all subsequent illustrations, the residual norm is measured in the standard* 2*-norm for all algorithms. Full GMRES converges to this tolerance in about* 509 *iterations.*

3. Analysis of weighted GMRES. Analyzing W-GMRES(m) must be no easier than analyzing standard restarted GMRES, an algorithm for which results are quite limited. (One knows conditions under which GMRES(m) converges for all initial residuals [15] and about cyclic behavior for GMRES($n-1$) for Hermitian (or skew-Hermitian) A [4]. Few other rigorous results are known.) Restarted GMRES is a nonlinear dynamical system involving the entries of A and b, and experiments suggest its convergence can depend sensitively on b [12]. Aware of these challenges, we seek some insight by first studying a setting for which W-GMRES is ideally suited.

Let $W = S^*S$ denote a positive definite matrix that induces the inner product $\langle u, v\rangle_W = v^*Wu$ on $\mathbb{R}^n$, with induced norm

$$\|u\|_W = \sqrt{\langle u, u\rangle_W} = \sqrt{u^*Wu} = \sqrt{u^*S^*Su} = \|Su\|_2.$$

Given the initial guess x_0 and residual $r_0 = b - Ax_0$, at the kth step GMRES in the W-inner product computes the residual r_k that satisfies

$$\begin{aligned}\|r_k\|_W = \min_{\substack{\varphi\in\mathcal{P}_k\\ \varphi(0)=1}} \|\varphi(A)r_0\|_W &= \min_{\substack{\varphi\in\mathcal{P}_k\\ \varphi(0)=1}} \|S\varphi(A)S^{-1}Sr_0\|_2 \\ &= \min_{\substack{\varphi\in\mathcal{P}_k\\ \varphi(0)=1}} \|\varphi(SAS^{-1})Sr_0\|_2.\end{aligned}$$

(Here $\mathcal{P}_k$ denotes the set of polynomials of degree k or less.) Thus, before any restarts are performed, W-GMRES applied to (A, r_0) is equivalent to 2-norm GMRES applied to (SAS^{-1}, Sr_0). (From this observation Güttel and Pestana propose an alternative implementation of W-GMRES, their Algorithm 2 [23, p. 744].) This connection between W-GMRES and standard GMRES was first noted in an abstract by Gutknecht and Loher [22], who considered "preconditioning by similarity transformation." Since A and SAS^{-1} have a common spectrum, one expects full GMRES in the 2-norm and W-norm to have similar *asymptotic* convergence behavior. However, the choice of S can significantly affect the departure of A from normality. For example, if A is diagonalizable with $A = V\Lambda V^{-1}$, then taking $W = V^{-*}V^{-1}$ with $S = V^{-1}$ renders $SAS^{-1} = \Lambda$ a diagonal (and hence normal) matrix.

The W-norms of the residuals produced by a cycle of W-GMRES will decrease monotonically, but this need not be true of the 2-norms, as seen in Figure 1. In fact,

$$\|r_k\|_W = \|Sr_k\|_2 \le \|S\|_2 \|r_k\|_2,$$

and similarly

$$\|r_k\|_2 = \|S^{-1}Sr_k\|_2 \le \|S^{-1}\|_2 \|Sr_k\|_2 = \|S^{-1}\|_2 \|r_k\|_W,$$

so

$$\frac{1}{\|S\|_2}\|r_k\|_W \le \|r_k\|_2 \le \|S^{-1}\|_2 \|r_k\|_W.$$

Pestana and Wathen [36, Thm. 4] show that the same bound holds when the residual in the middle of this inequality is replaced by the optimal 2-norm GMRES residual. That is, if $r_k^{(W)}$ and $r_k^{(2)}$ denote the kth residuals from GMRES in the W-norm and 2-norm (both before any restart is performed, $k \le m$), then

$$\frac{1}{\|S\|_2}\|r_k^{(W)}\|_W \le \|r_k^{(2)}\|_2 \le \|S^{-1}\|_2 \|r_k^{(W)}\|_W.$$

Thus for W-norm GMRES to depart significantly from 2-norm GMRES *over the course of one cycle*, S must be (at least somewhat) ill-conditioned. However, restarting with a new inner product can change the dynamics across cycles significantly.

3.1. The ideal setting for W-GMRES. For diagonal W, as in [14], let

$$S = \operatorname{diag}(s_1, \ldots, s_n) := W^{1/2}.$$

When the coefficient matrix A is diagonal,

$$A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

Essai's weighting is well motivated and its effects can be readily understood.[2] In this case, the eigenvectors are columns of the identity matrix, $V = I$, and $SA = AS$. Consider a cycle of W-GMRES(m) starting with $r_0 = b = [b_1, \ldots, b_n]^T$. For $k \le m$,

$$\text{(2a)}\quad \|r_k\|_W^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|\varphi(A)b\|_W^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|S\varphi(A)b\|_2^2$$

$$\text{(2b)}\quad = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|\varphi(SAS^{-1})Sb\|_2^2$$

$$\text{(2c)}\quad = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|\varphi(A)Sb\|_2^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \sum_{j=1}^{n} |\varphi(\lambda_j)|^2 |s_j b_j|^2.$$

This last formula reveals that the weight s_j affects GMRES in the same way as the component b_j of the right-hand side (which is the component of b in the jth eigenvector direction for this A), *so the weights tune how much GMRES will emphasize any given eigenvalue*. Up to scaling, Essai's proposal amounts to $s_j = \sqrt{|b_j|}$, so

$$\text{(3)}\quad \|r_k\|_W^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \sum_{j=1}^{n} |\varphi(\lambda_j)|^2 |b_j|^3.$$

[2] Examples 1 and 2 of Güttel and Pestana [23] use diagonal A, and the analysis here can help explain their experimental results.

This approach favors large components in b over smaller ones. Small entries in b (reduced at a previous cycle) were "easy" for GMRES to reduce because, for example, they correspond to eigenvalues far from the origin. Large entries in b have been neglected by previous cycles, most likely because they correspond to eigenvalues near the origin. A low-degree polynomial that is small at these points is likely to be large at large magnitude eigenvalues and so will not be picked as the optimal polynomial by restarted GMRES. Weighting tips the scales to favor these neglected eigenvalues near the origin for one cycle; this preference will usually *increase* the residual in "easy" components, but such an increase can be quickly remedied at the next cycle. Consider how standard GMRES(1) handles the scenario

$$0 < \lambda_1 \ll 1 \leq \lambda_2 \approx \cdots \approx \lambda_n.$$

Putting the root ζ of the residual polynomial $\varphi(z) = 1 - z/\zeta$ near the cluster of eigenvalues $\lambda_2 \approx \cdots \approx \lambda_n$ will significantly diminish the residual vector in components $2, \ldots, n$, while not increasing the first component, since $|\varphi(\lambda_1)| < 1$. On the other hand, placing ζ near λ_1 would give $|\varphi(\lambda_j)| \gg 1$ for $j = 2, \ldots, n$, increasing the overall residual norm, but in a manner that could be remedied at the next cycle if GMRES(1) could take a locally suboptimal cycle to accelerate the overall convergence. Essai's weighting enables GMRES to target that small magnitude eigenvalue for one cycle, potentially at an increase to the 2-norm of the residual that can be corrected at the next step without undoing the reduced component associated with λ_1.

3.1.1. The diagonal case: Targeting small eigenvalues.

Example 3.1. Consider the diagonal matrix

$$A = \text{diag}(0.01,\ 0.1,\ 3,\ 4,\ 5,\ \ldots,\ 9,\ 10).$$

Here $b = r_0$ is a Normal(0,1) vector (MATLAB command `randn` with `rng(1319)`), scaled so $\|r_0\|_2 = 1$. Figure 2 compares W-GMRES(m) to GMRES(m) for $m = 3$ and $m = 6$: in both cases, weighting gives a big improvement. Figure 3 compares convergence of GMRES(3) and W-GMRES(3) for 5000 Normal(0,1) right-hand-side vectors (also scaled so $\|r_0\|_2 = 1$). As in Figure 2, W-GMRES(3) usually substantially improves convergence, but it also adds considerable variation in performance. (We have noticed this tendency overall: weighting seems to add more variation, depending on the right-hand side and finite-precision effects.)

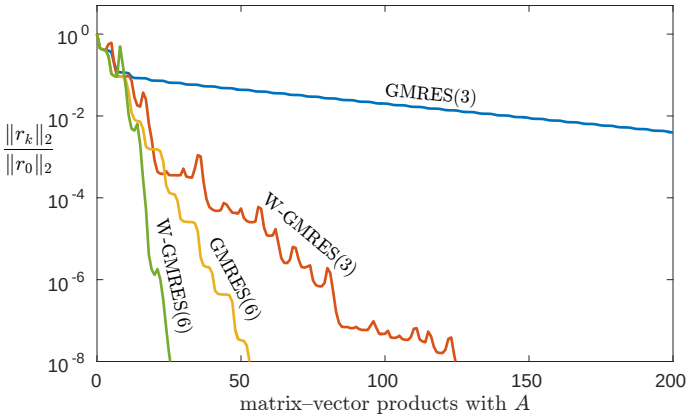


FIG. 2. *GMRES(m) and W-GMRES(m) with* $m = 3, 6$ *for the diagonal test case, Example* 3.1.

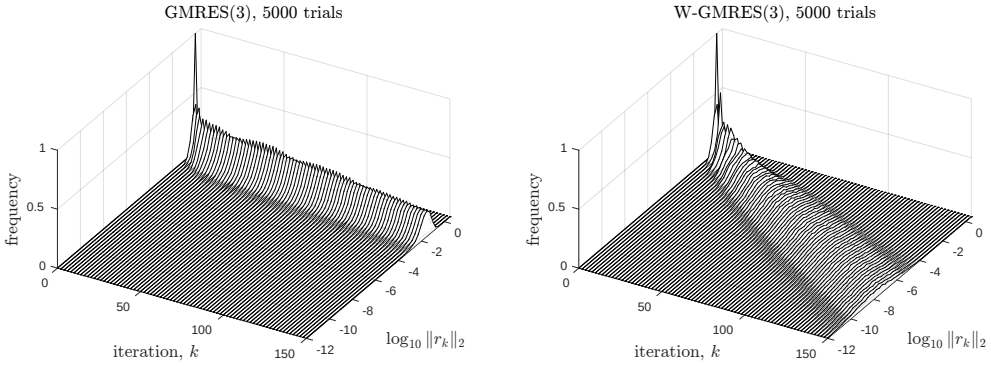


FIG. 3. *Histograms of GMRES(3) and W-GMRES(3) convergence for Example* 3.1 *with* 5000 *random unit vectors* $b = r_0$ *(the same for both methods). For each* k*, these plots show the distribution of* $\|r_k\|_2$*; while W-GMRES(3) converges faster in nearly all cases, the convergence is more variable.*

Focus on $m = 3$ in Figure 2. GMRES(3) converges slowly: the small eigenvalues of A make this problem difficult. Figure 4 shows the residual components in each eigenvector after cycles 2, 3, and 4: GMRES(3) makes little progress in the smallest eigencomponent. Contrast this with W-GMRES(3): after cycles 2 and 3 all eigencomponents have been reduced *except the first one, corresponding to* $\lambda = 0.01$. Cycle 4 emphasizes this neglected eigenvalue, reducing the corresponding eigencomponent in the residual by two orders of magnitude (while increasing some of the other eigencomponents). Without weighting, GMRES(3) does not target this eigenvalue.

Figure 5 examines the GMRES residual polynomials φ_ℓ that advance GMRES(m) and W-GMRES(m) from cycle ℓ to cycle $\ell + 1$ (i.e., $r_{(\ell+1)m} = \varphi_\ell(A) r_{\ell m}$ with $\varphi_\ell(0) = 1$) for $\ell = 2, \ldots, 5$. GMRES(3) shows a repeating pattern, with every other polynomial very similar; such behavior has been noted before [4, 50]. None of these four polynomials is small at $\lambda = 0.01$; see the zoomed-in image on the right of Figure 5. (At cycle 4, $\varphi_4(0.01) \approx 0.908$.) The W-GMRES(3) polynomials at cycles 2, 3, and 5 are much like those for GMRES(3); however, the polynomial for cycle 4 breaks the pattern, causing a strong reduction in the first eigencomponent: $\varphi_4(0.01) \approx 0.00461$. Note that φ_4 is also small at $\lambda = 5$ and $\lambda = 10$. These are the other eigencomponents that are reduced at this cycle; the others all increase, as seen in Figure 4. While the weighted norm decreases from 0.0906 to 0.0061 during cycle 4, the 2-norm shows a more modest reduction, from 0.0913 to 0.0734. Indeed, experiments with different initial residuals often produce similar results, though the cycle that targets $\lambda = 0.01$ can vary, and the 2-norm of the W-GMRES residual can increase at that cycle.

The next result makes such observations precise for 2×2 matrices.

THEOREM 3.2. *Let* $A = \mathrm{diag}(\lambda,\ 1)$ *with real* $\lambda \neq 0$*, and suppose the residual vector at the start of a cycle of GMRES is* $b = [\, b_1 \ \ b_2 \,]^T$ *for* $b_1, b_2 \in \mathbb{R}$ *with* $b_1 \neq 0$*. Define* $\beta := b_2/b_1$*. Then the GMRES(1) polynomial has the root*

$$\frac{\lambda^2 + \beta^2}{\lambda + \beta^2},$$

while the W-GMRES(1) polynomial has the root

$$\frac{\lambda^2 + |\beta|^3}{\lambda + |\beta|^3}.$$

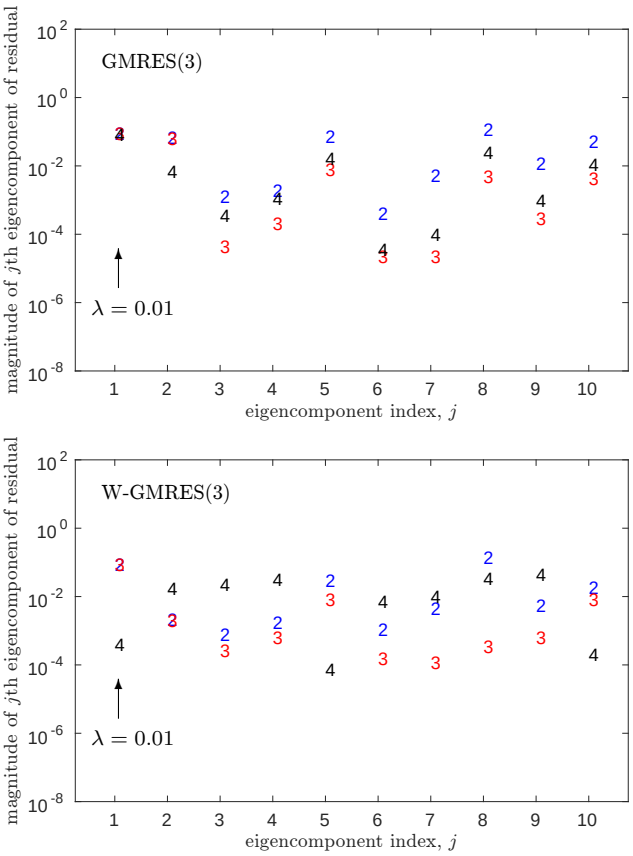


FIG. 4. *Eigencomponents of the residual at the end of cycles $\ell = 2$, 3, and 4 for standard GMRES(3) (top) and W-GMRES(3) (bottom) applied to the matrix in Example* 3.1. *Cycle* 4 *of W-GMRES(3) reduces the eigencomponent corresponding to $\lambda_1 = 0.01$ by two orders of magnitude.*

The proof is a straightforward calculation. To appreciate this theorem, take $\lambda = \beta = 0.1$, giving the roots 0.1818 for GMRES(1) and 0.1089 for W-GMRES(1). These polynomials cut the residual component corresponding to $\lambda = 0.1$ by a factor 0.450 for GMRES(1) and 0.0818 for W-GMRES(1), the latter causing W-GMRES(1) to increase the component for $\lambda = 1$ by a factor of 8.18. This increase is worthwhile, given the reduction in the tough small component; later cycles will handle the easier component. For another example, take $\lambda = \beta = 0.01$: GMRES(1) reduces the $\lambda = 0.01$ component by 0.495, while W-GMRES(1) reduces it by 0.0098, two orders of magnitude.

3.1.2. The diagonal case: Breaking patterns in residual polynomials. Beyond the potential to target small eigenvalues, W-GMRES can also break the cyclic pattern into which GMRES(m) residual polynomials often lapse. Baker, Jessup, and Manteuffel [4, Thm. 2] prove that if A is an $n \times n$ symmetric (or skew-symmetric) matrix, then the GMRES($n-1$) residual vectors exactly alternate in direction, i.e., the residual vector at the end of a cycle is an exact multiple of the residual two cycles before; thus the GMRES residual polynomial at the end of each cycle repeats the same polynomial found two cycles before. Baker, Jessup, and Manteuffel observe the same qualitative behavior for more frequent restarts and suggest that disrupting

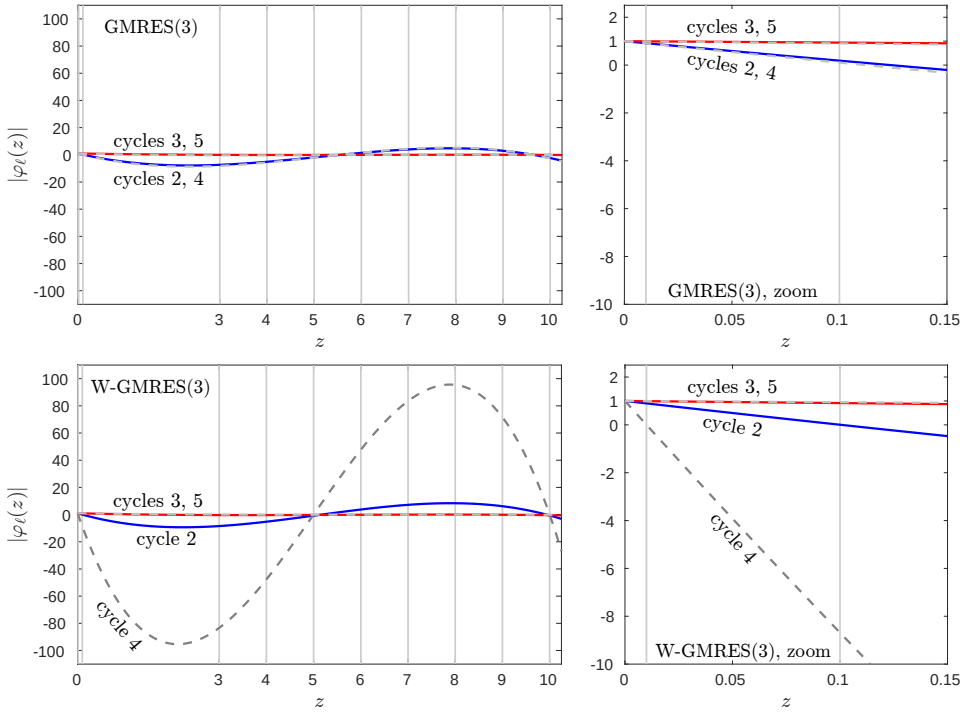


FIG. 5. *Residual polynomials for cycles* 2–5 *of standard GMRES*(3) *(top) and W-GMRES*(3) *(bottom) for Example* 3.1. *The gray vertical lines show the eigenvalues of* A. *At cycle* 4, *the weighted inner product allows W-GMRES*(3) *to target the first eigencomponent, increasing other components.*

this pattern (by varying the restart parameter [3] or augmenting the subspace [4]) can improve convergence. (Longer cyclic patterns can emerge for nonnormal A [50].) Changing the inner product can have a similar effect, with the added advantage of targeting difficult eigenvalues if the corresponding eigenvectors are well-disposed. While Essai's residual-based weighting can break cyclic patterns, other schemes (such as random weighting [23, 42]; see subsection 3.5) can have a similar effect; however, such arbitrary inner products take no advantage of eigenvector structure. The next example cleanly illustrates how W-GMRES can break patterns.

Example 3.3. Apply GMRES(1) to $A = \mathrm{diag}(2, 1)$ with $b = [1\ \ 1]^T$. The roots of the (linear) GMRES(1) residual polynomials alternate between $5/3$ and $4/3$ (exactly), as given by Theorem 3.2. Therefore the GMRES(1) residual polynomials for the cycles alternate between $\varphi_k(z) = 1 - (3/5)z$ and $\varphi_k(z) = 1 - (3/4)z$. GMRES(1) converges ($\|r_k\|_2/\|r_0\|_2 \le 10^{-8}$) in 16 iterations. W-GMRES(1) takes only 7 iterations. As Figure 6 shows, weighting breaks the cycle of polynomial roots, giving 1.667, 1.200, 1.941, 1.0039, 1.999985, 1.0000000002, and ≈ 2: the roots move out toward the eigenvalues, alternating which eigenvalue they favor and reducing $\|r_k\|_2$ much faster.

3.2. Measuring eigenvector localization. The justification for W-GMRES in (2c) required diagonal A, so the eigenvectors are columns of the identity matrix. When A is not diagonal, the motivation for W-GMRES is less compelling. Yet in many applications A is not close to diagonal but still has *localized* eigenvectors: only a few entries are large, and the eigenvectors resemble columns of the identity matrix. Computational evidence suggests that W-GMRES can still be effective, particularly

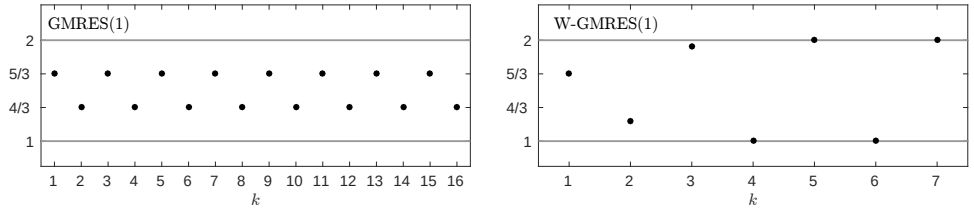
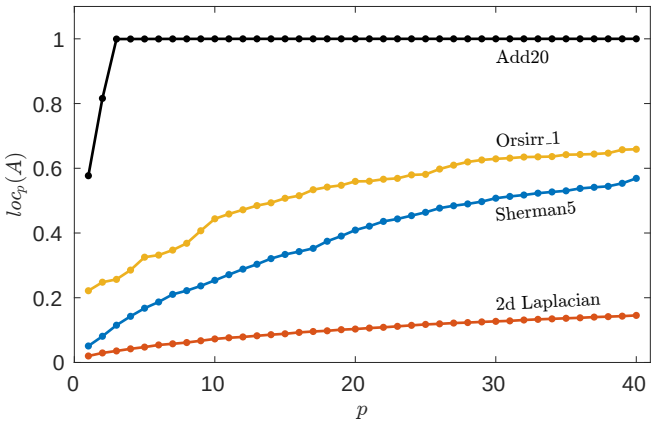


FIG. 6. *Roots of the GMRES(1) and W-GMRES(1) residual polynomials for* $A = \mathrm{diag}(2,1)$ *with* $b = [1,1]^T$*, as a function of the restarted GMRES cycle,* k*. The GMRES(1) roots occur in a repeating pair, and the method takes* 16 *iterations to converge to* $\|r_k\|_2/\|r_0\|_2 \le 10^{-8}$*. In contrast, the W-GMRES(1) roots are quickly attracted to the eigenvalues, giving convergence in just* 7 *steps.*



matrix	n	$loc_5(A)$
Add20	2395	0.9996
Orsirr_1	1030	0.3253
Sherman5	3312	0.1676
2d Laplacian	9801	0.0475

FIG. 7. *The localization measures* $\mathrm{loc}_p(A)$ *for* $p = 1,\ldots,40$ *for four matrices that are used in our W-GMRES experiments. If* $\mathrm{loc}_p(A)$ *is close to* 1*, the eigenvectors associated with smallest magnitude eigenvalues are highly localized.*

when the eigenvectors associated with the small magnitude eigenvalues are localized. The Add20 matrix in Example 2.1 is such an example.

We gauge the localization of the p smallest magnitude eigenvectors by measuring how much they are concentrated in their p largest magnitude entries.[3] For diagonalizable A, label the eigenvalues in increasing magnitude, $|\lambda_1| \le |\lambda_2| \le \cdots \le |\lambda_n|$, with associated unit 2-norm eigenvectors $v_1,\ldots,v_n$. For any $y \in \mathbb{C}^n$, let $y^{(p)} \in \mathbb{C}^p$ denote the subvector containing the p largest magnitude entries of y (in any order). Then

$$loc_p(A) := \frac{1}{\sqrt{p}} \left(\sum_{j=1}^{p} \|v_j^{(p)}\|_2^2 \right)^{1/2}. \tag{4}$$

Notice that $loc_p(A) \in (0,1]$, and $loc_n(A) = 1$. If $loc_p(A) \approx 1$ for some $p \ll n$, the eigenvectors associated with the p smallest magnitude eigenvalues are localized within p positions. This measure is imperfect, for it includes neither the magnitude of the eigenvalues nor the potential nonorthogonality of the eigenvectors. Still $loc_p(A)$ can be a helpful instrument for assessing localization. Figure 7 shows $loc_p(A)$ for

[3]Certain applications motivate more specialized measures of localization based on the rate of exponential decay of eigenvector entries about some central entry.

$p = 1, \ldots, 40$ for four matrices we use in our experiments.[4] The Add20 matrix has strongly localized eigenvectors; the discrete 2d Laplacian has eigenvectors that are far from localized; i.e., they are *global*. The Orsirr_1 and Sherman5 matrices have eigenvectors with intermediate localization.

3.3. The nondiagonal case. When A is not diagonal, the justification for W-GMRES in (2c) is lost; the effect of weighting becomes more subtle. The weights can still break cyclic patterns in restarted GMRES, but the interplay between the weights and the eigenvectors is more difficult to understand. Suppose A is diagonalizable, $A = V\Lambda V^{-1}$, still with $W = S^*S$. Unlike the diagonal case, S will not generally commute with V and Λ, so (2) is replaced by

$$\|r_k\|_W = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|\varphi(SAS^{-1})Sb\|_2 \tag{5a}$$

$$= \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \|(SV)\varphi(\Lambda)(SV)^{-1}Sb\|_2. \tag{5b}$$

The matrix S *transforms the eigenvectors of* A [22]. Suppose, impractically, that V^{-1} were known, allowing the *nondiagonal* weight

$$S = \operatorname{diag}(s_1, \ldots, s_n)V^{-1}.$$

In this case (5b) reduces to

$$\|r_k\|_W^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \sum_{j=1}^{n} |\varphi(\lambda_j)|^2 |s_j b_j|^2 \quad \text{(eigenvector weighting)} \tag{6}$$

a perfect analogue of the diagonal case (2c) that could appropriately target the small magnitude eigenvalues that delay convergence. One might *approximate* this eigenvector weighting by using, as a proxy for V^{-1}, some estimate of the left eigenvectors of A associated with the smallest magnitude eigenvectors. (The rows of V^{-1} are the left eigenvectors of A, since $V^{-1}A = \Lambda V^{-1}$.) We do not pursue this idea here, instead using eigenvector information via deflated restarting in section 5.

The conventional W-GMRES(m) algorithm instead uses diagonal W (and S). In this case, SV scales the *rows* of V, effectively emphasizing certain entries of the eigenvectors at the expense of others. Write out the right and left eigenvectors,

$$V = [v_1\ v_2\ \cdots\ v_n] \in \mathbb{C}^{n\times n}, \qquad V^{-1} = \begin{bmatrix} \widehat{v}_1^* \\ \widehat{v}_2^* \\ \vdots \\ \widehat{v}_n^* \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

Let $c := V^{-1}b$ denote the expansion coefficients for b in the eigenvectors of A: $b = Vc$. Then one can also render (5b) in the form

$$\|r_k\|_W = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \left\| \sum_{j=1}^{n} c_j\, \varphi(\lambda_j) S v_j \right\|_2 \quad \text{(diagonal weighting)} \tag{7}$$

[4] Two of these matrices, Sherman5 and the 2d Laplacian, have repeated (derogatory) eigenvalues; according to the principle of *reachable invariant subspaces* (see, e.g., [5, 28]), the GMRES process only acts upon one eigenvector in the invariant subspace, given by the component of the initial residual in that space. Thus the localization measures in Figure 7 include only one eigenvector for the derogatory eigenvalues, specified by the initial residual vectors used in our experiments.

a different analogue of (2c). Denote the ℓth entry of v_j by $(v_j)_\ell$. In most cases V will be dense. Suppose that only the qth entry of r_0 is large, and that $(v_j)_q \neq 0$ for $j = 1, \ldots, n$. Then Essai's weighting makes $|(Sv_j)_q| \gg |(Sv_j)_\ell|$ for all $\ell \neq q$: all the vectors Sv_j form a small angle with the qth column of the identity. Thus SV is ill-conditioned and SAS^{-1} will have a large departure from normality, often a troublesome case for GMRES; see, e.g., [47, Chap. 26]. It is not evident how such weighting could help W-GMRES(m) focus on small eigenvalues, as it does for diagonal A, suggesting an explanation for the mixed performance of W-GMRES(m) [9, 23].

3.3.1. Stagnation of W-GMRES(m) when GMRES(m) converges. We show that by transforming the eigenvectors, diagonal weighting can even prevent W-GMRES(m) from converging at all. Faber et al. [15] prove that GMRES(m) converges for all initial residuals provided there exists no $v \in \mathbb{C}^n$ such that

$$v^* A^k v = 0 \quad \text{for all } k = 1, \ldots, m.$$

Thus GMRES(m) will converge for all $m \geq 1$ and r_0 if the field of values of A,

$$F(A) := \{v^* A v : v \in \mathbb{C}^n, \|v\|_2 = 1\},$$

does not contain the origin. Weighting effectively applies GMRES to the transformed matrix SAS^{-1}, and now it is possible that $0 \in F(SAS^{-1})$ even though $0 \notin F(A)$. In extreme cases this means that W-GMRES(m) can stagnate even when GMRES(m) converges, as shown by the next example. (In contrast, for A and S diagonal, $SAS^{-1} = A$, so $F(A) = F(SAS^{-1})$.)

Example 3.4 (W-GMRES(1) stagnates while GMRES(1) converges). Consider

$$A = \begin{bmatrix} 1 & -4 \\ 0 & 5 \end{bmatrix}, \qquad r_0 = \begin{bmatrix} 1 \\ \frac{1}{10}(5+\sqrt{5}) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.72360\ldots \end{bmatrix}. \tag{8}$$

Figure 8 shows $F(A)$, an ellipse in the complex plane. From the extreme eigenvalues of the Hermitian and skew-Hermitian parts of A, one can bound $F(A)$ within a rectangle in $\mathbb{C}$ that does not contain the origin,

$$\mathrm{Re}(F(A)) = [3 - 2\sqrt{2}, 3 + 2\sqrt{2}] \approx [0.17157, 5.82843], \qquad \mathrm{Im}(F(A)) = [-2\mathrm{i}, 2\mathrm{i}],$$

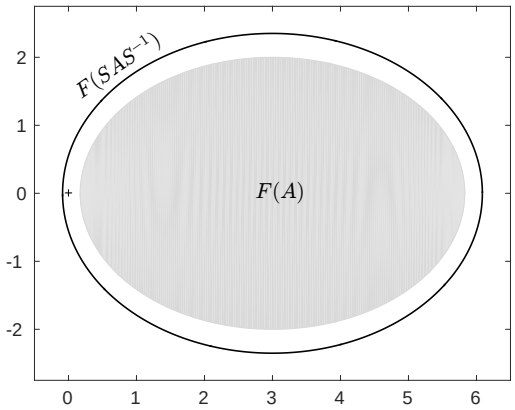


FIG. 8. *The fields of values $F(A)$ (gray region) and $F(SAS^{-1})$ (boundary is a black line) in $\mathbb{C}$ for A and r_0 in (8); the cross (+) marks the origin. Since $0 \notin F(A)$, GMRES(m) converges for any r_0; however, $0 \in F(SAS^{-1})$ and $(Sr_0)^*(SAS^{-1})(Sr_0) = 0$, so W-GMRES(1) completely stagnates.*

and hence GMRES(1) converges for all initial residuals. Now consider the specific r_0 given in (8). Using the scaling $S = \mathrm{diag}(\sqrt{|(r_0)_1|/\|r_0\|_\infty}, \sqrt{|(r_0)_2|/\|r_0\|_\infty})$ at the first step gives the W-GMRES(1) polynomial (see, e.g., [12, eq. (2.2)])

$$r_1 = \varphi(A)r_0 = r_0 - \frac{(Sr_0)^*(SAS^{-1})(Sr_0)}{\|(SAS^{-1})(Sr_0)\|_2^2}(SAS^{-1})(Sr_0) = r_0 - \frac{r_0^* S^2 A r_0}{\|SAr_0\|_2^2} SAr_0.$$

This example has been engineered so that $0 \in F(SAS^{-1})$; indeed,

$$\mathrm{Re}(F(SAS^{-1})) = \left[3 - \sqrt{14 - 2\sqrt{5}}, 3 + \sqrt{14 - 2\sqrt{5}}\right] \approx [-0.086724, 6.08672],$$

as seen in Figure 8. Worse still, this r_0 gives $r_0^* S^2 A r_0 = 0$, so $r_1 = r_0$: W-GMRES(1) makes no progress, and the same weight is chosen for the next cycle. The weighted algorithm completely stagnates, even though GMRES(1) converges for any r_0.

3.3.2. Poor performance when small eigenvalues have global eigenvectors. The last example shows one way GMRES(m) can outperform W-GMRES(m) when A is not diagonal. Next we show that global eigenvectors associated with small magnitude eigenvalues can also contribute to poor W-GMRES(m) convergence.

Example 3.5 (W-GMRES(m) worse than GMRES(m): Global eigenvectors). We revisit Example 3.1, again with dimension $n = 10$ and eigenvalues

$$\Lambda = \mathrm{diag}(0.01, 0.1, 3, 4, 5, \ldots, 9, 10),$$

but now let the eigenvectors of A equal those of the symmetric tridiagonal matrix $\mathrm{tridiag}(-1, 0, -1)$, ordering the eigenvalues from smallest to largest. (The jth column of V has entries $(v_j)_\ell = \sin(j\ell\pi/(n+1))/\|v_j\|_2$; see, e.g., [44].) The resulting $A = V\Lambda V^*$ is symmetric, unitarily similar to Λ. Let b be the same Normal(0,1) vector used in Example 3.1. Since V is unitary, GMRES(m) applied to (Λ, b) produces the same residual norms as GMRES(m) applied to $(V\Lambda V^*, Vb)$. However, weighted GMRES can behave very differently for the two problems: when applied to Λ (as in Example 3.1), W-GMRES(6) vastly outperforms GMRES(6). For A with its nonlocalized eigenvectors, W-GMRES(6) converges much more slowly, as seen in Figure 9.

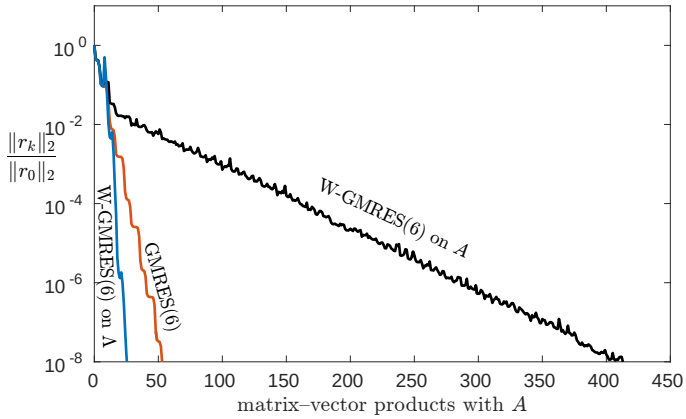


FIG. 9. *GMRES(6) and W-GMRES(6) for Example* 3.5. *Applying GMRES(6) to* (Λ, b) *and* (A, Vb) *produces identical residual* 2*-norms. When applied to* (Λ, b)*, W-GMRES(6) converges faster than GMRES(6) (seen previously in Figure* 2*); when applied to* (A, Vb) *the convergence is much slower (A having nonlocalized eigenvectors).*

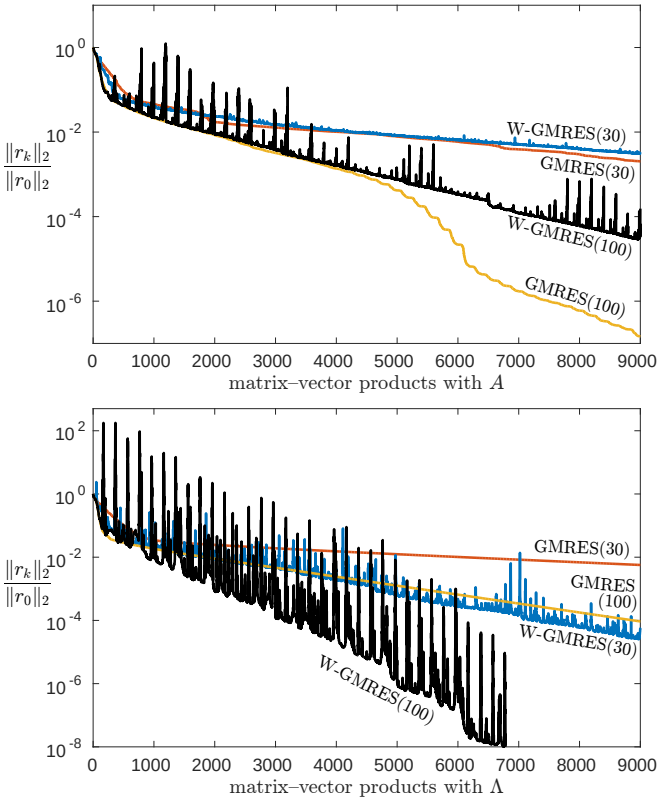


FIG. 10. *GMRES(m) and W-GMRES(m) for the Sherman5 matrix. On the top, the methods are applied to $Ax = b$, and weighting slows convergence. On the bottom, the same methods are applied to $\Lambda x = b$, where Λ is diagonal with the same eigenvalues of A; W-GMRES converges better. (The GMRES(*100*) curve is yellow. Color is available online only.) For both problems, taking $m = 100$ steps with each weighted inner product gives large growth in $\|r_k\|_2$ between restarts.*

The next example shows similar behavior for a larger matrix: eigenvalues on both sides of the origin and global eigenvectors give poor W-GMRES(m) performance.

Example 3.6 (W-GMRES(m) slower than GMRES(m): Global eigenvectors). Consider the nonsymmetric Sherman5 matrix from Matrix Market [6], with $n = 3312$ and with r_0 a Normal(0,1) vector. Figure 10 shows that weighting slows restarted GMRES, particularly for the restart $m = 100$. Figure 7 shows that the eigenvectors are poorly localized; e.g., in the unit eigenvector v_1 corresponding to the smallest eigenvalue, the 100 largest components range only from 0.0509 to 0.0415 in magnitude. This matrix has eigenvalues to the left and right of the origin, and the eigenvectors are not orthogonal.[5] The second plot in Figure 10 shows how the performance of W-GMRES improves when A is replaced by a diagonal matrix with the same spectrum.

3.3.3. Weighting sometimes helps despite global eigenvectors. The past three examples suggest that W-GMRES performs poorly for matrices whose eigenvectors are not localized, but the algorithm is more nuanced than that. The next example

[5]The eigenvalue $\lambda = 1$ has multiplicity 1674, and the corresponding eigenvectors are columns of the identity. The eigenvalues closest to the origin are well-conditioned, while some eigenvalues far from the origin are relatively ill-conditioned. (For the diagonalization computed by the MATLAB command `eig`, $\|V\|_2\|V^{-1}\|_2 \approx 3.237 \times 10^6$.) The pseudospectra of this matrix are well-behaved near the origin; the effect of the eigenvalue conditioning on GMRES convergence is mild [47, Chap. 26].

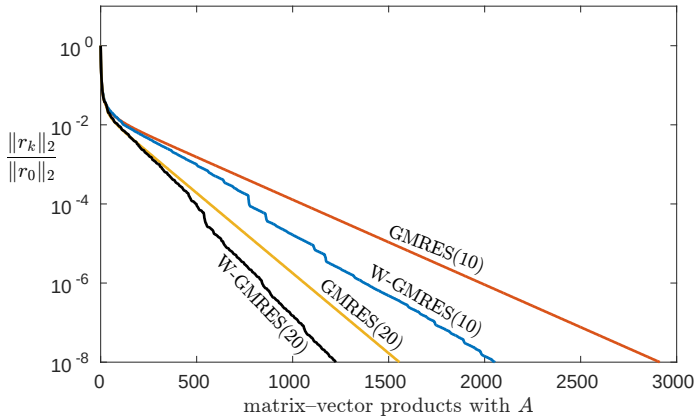


FIG. 11. *GMRES(m) and W-GMRES(m) for a discretization of the Dirichlet Laplacian on the unit square. The eigenvectors are not localized, but weighting still accelerates convergence.*

is characteristic of behavior we have often observed in our experiments. We attribute the improvement over standard restarted GMRES to the tendency of weighting to break cyclic patterns in the GMRES residual polynomials.

Example 3.7 (W-GMRES(m) better than GMRES(m): Global eigenvectors). Consider the standard 5-point finite difference discretization of the Laplacian on the unit square in two dimensions with Dirichlet boundary conditions. The uniform grid spacing $h = 1/100$ in both directions gives a matrix of order $n = 99^2 = 9801$; the initial residual is a random Normal(0,1) vector. The small $loc_p(A)$ values shown in Figure 7 confirm that the eigenvectors are not localized; in fact, they are less localized than for the Sherman5 matrix in the last example. Despite this, Figure 11 shows that W-GMRES(m) outperforms GMRES(m) for $m = 10$ and $m = 20$. In this case, all the eigenvalues are positive, and none is particularly close to the origin.

3.4. Extra weighting. If weighting can improve GMRES, especially when the smallest eigenvalues have localized eigenvectors, might more extreme weighting help even more? Recall from (2) that for diagonal A,

$$\|r_k\|_W^2 = \min_{\substack{\varphi\in\mathcal{P}_k \\ \varphi(0)=1}} \sum_{j=1}^{n} |\varphi(\lambda_j)|^2 |s_j b_j|^2. \tag{9}$$

Essai's standard weighting uses $s_j^2 = w_j = |b_j|$. Suppose one instead takes $s_j^2 = w_j = |b_j|^p$ for $p \geq 0$. For large p we find it particularly important to limit the minimal weight, as in (1); we require $w_j \in [10^{-10}, 1]$.

For most cases we have studied, extra weighting has little qualitative impact. For diagonal A, $p > 1$ can accelerate convergence; for the general case, even with localized eigenvectors, results are mixed. The next example shows a case where it helps.

Example 3.8 (extra weighting, localized eigenvectors). The Orsirr_1 matrix from Matrix Market [6] is a nonsymmetric matrix of dimension $n = 1030$. Figure 7 shows the eigenvectors to be fairly well localized; all the eigenvalues are real and negative. Table 1 reports the matrix-vectors products (with A) required by W-GMRES(m) with different weighting powers p to satisfy the convergence criterion $\|r_k\|_2/\|r_0\|_2 \leq 10^{-8}$. (Here r_0 is a random Normal(0,1) vector.) When m is small (e.g., $m = 5$ and 10),

TABLE 1
W-GMRES(m) for the Orsirr_1 matrix, with extra weighting by the power p. The table shows the matrix-vector products with A required to reduce $\|r_k\|_2/\|r_0\|_2$ by a factor of 10^{-8}.

p	0	1	2	3	4	5	6	8	10
$m = 5$	20000	20000	13854	8126	6641	7640	4043	3947	4512
$m = 10$	16299	13883	5009	5901	3810	3198	3053	3129	3064
$m = 20$	13653	2934	2569	2134	2221	2154	2303	2294	2377
$m = 30$	3750	2572	2224	2012	2075	2009	1998	1961	2094

extra weighting improves convergence significantly. For example, when $m = 10$, the value $p = 6$ gives the best result among the reported p. For larger m, extra weighting has less impact.

3.5. Random weighting. We have argued that when A has local eigenvectors, residual-based weighting can encourage convergence in certain eigenvector directions and break repetitive patterns in GMRES(m). In contrast, when the eigenvectors are predominantly global, weighting can still sometimes help by breaking patterns in GMRES(m), but there is no compelling connection to the eigenvectors. Thus if the eigenvectors of A are global, residual-based weighting should fare no better than the random diagonal weighting proposed by Saberi Najafi and Zareamoghaddam [42]. In contrast, for localized eigenvectors, residual-based weighting should have a significant advantage over random weighting. The next two examples show this behavior.

Example 3.9 (random weights for a matrix with local eigenvectors). The Add20 matrix from Example 2.1 has localized eigenvectors, and W-GMRES(m) improved convergence for $m = 3$ and $m = 6$. Table 2 compares this performance to two random weighting schemes: W is a diagonal matrix with uniform random entries chosen in the interval $[0.5, 1.5]$ (first method) and $[0, 1]$ (second method), with fresh weights generated at each restart. All experiments use the same r_0 as in Figure 1; since performance changes for each run with random weights, we average the matrix-vector products with A (iteration counts) over 10 trials each. All three weighted methods beat GMRES(m), but the residual-based weighting has a significant advantage.

Example 3.10 (random weights for a matrix with global eigenvectors). The results are very different for the discretization of the Dirichlet Laplacian on the unit square (Example 3.7). Table 3 shows that residual-based weighting is inferior to most of the random weighting experiments, with particularly extreme differences for small m. There appears to be no advantage to correlating weighting to residual components. (This example shows a "tortoise and hare" effect [12], where the random weighting with $m = 1$ often converges in fewer iterations than the larger m here.)

4. W-GMRES-DCT: Localizing eigenvectors with Fourier transforms. Equation (6) suggests that, for diagonalizable $A = V\Lambda V^{-1}$, the ideal weighting scheme should incorporate eigenvector information. It will be convenient to now express W in the form $W = Z^*Z$ with $Z = SV^{-1}$ for $S = \mathrm{diag}(s_1, \ldots, s_n)$, giving

$$\|r_k\|_W^2 = \min_{\substack{\varphi \in \mathcal{P}_k \\ \varphi(0)=1}} \sum_{j=1}^{n} |\varphi(\lambda_j)|^2 |s_j b_j|^2.$$

Incorporating V^{-1} into the weighting diagonalizes the problem and makes residual weighting compelling, but such an inner product is clearly impractical. However, we

TABLE 2

For the Add20 matrix, the matrix-vector products (with A) required to converge to $\|r_k\|_2/\|r_0\|_2 \le 10^{-8}$ for GMRES(m) and W-GMRES(m), and random weights chosen uniformly in $[0.5, 1.5]$ and $[0, 1]$. The random counts are averaged over 10 trials (same r_0). W-GMRES(m) does much better than the random schemes for this matrix with local eigenvectors.

m	GMRES(m)	W-GMRES(m)	rand(.5,1.5)	rand(0,1)
1	> 20000	1785	15845.5	4077.1
2	> 20000	2882	10506.7	8190.2
3	> 20000	4580	8664.0	7237.1
6	10032	1022	4270.7	3382.7
10	4220	762	1945.4	1331.9
15	2469	716	1531.6	1126.8
20	1605	650	1109.1	1022.5

TABLE 3

For the discrete Laplacian on the unit square, the matrix-vector products (with A) required to converge to $\|r_k\|_2/\|r_0\|_2 \le 10^{-8}$ for GMRES(m), W-GMRES(m), and two random schemes: weights chosen uniformly in $[0.5, 1.5]$ and $[0, 1]$. The random counts are averaged over 10 trials, all with the same r_0. For smaller values of m, the residual-based weighting of W-GMRES(m) performs significantly worse than the random weighting for this matrix with global eigenvectors.

m	GMRES(m)	W-GMRES(m)	rand(.5,1.5)	rand(0,1)
1	> 20000	> 20000	1446.0	1059.4
2	14449	9958	2499.8	2318.4
3	9687	6450	2670.6	2329.8
6	4867	3704	2267.1	1804.0
10	2912	2054	1726.3	1545.7
15	2017	1650	1742.0	1549.3
20	1556	1225	1404.1	1338.2

have seen that weighting can improve GMRES convergence if the eigenvectors are merely localized. This suggests a new weighting strategy that uses $W = Z^*Z$ with

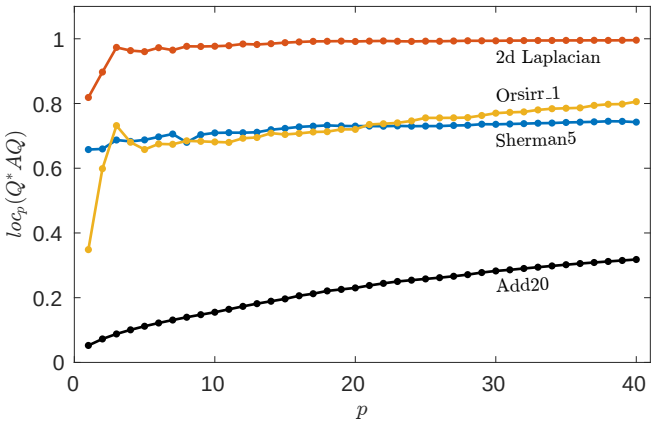
$$Z = SQ^*, \qquad S = \mathrm{diag}(s_1, \ldots, s_n),$$

where Q is unitary and the weights $s_1, \ldots, s_n$ are derived from the residual as before. Since GMRES on (A, b) in this W-inner product amounts to standard 2-norm GMRES on $(ZAZ^{-1}, Zb) = (SQ^*AQS^{-1}, SQ^*b)$, we seek Q in which *the eigenvectors of Q^*AQ are localized*. When A is a discretization of a constant-coefficient differential operator, a natural choice for Q^* is the discrete cosine transform matrix [46],[6] since for many such problems the eigenvectors associated with the smallest-magnitude eigenvalues are *low in frequency*. We call the resulting algorithm W-GMRES-DCT(m). In MATLAB, one could compute $Q^* =$ `dct(eye(n))`, a unitary but dense matrix that should never be formed; instead, it can be applied to vectors in $O(n \log n)$ time using the fast Fourier transform. In MATLAB code, the inner product becomes

$$\langle x, y\rangle_W = y^*QS^*SQ^*x = (SQ^*y)^*(SQ^*x) = \texttt{(s.*dct(y))'*(s.*dct(x))},$$

where $\texttt{s} = [s_1, \ldots, s_n]^T$. Alternatively, Q^* could be the discrete Fourier transform matrix (`fft(eye(n))`, which uses complex arithmetic), a discrete wavelet transform, or any other transformation that localizes (i.e., *sparsifies*) the eigenvectors of A (informed by the motivating application). Figure 12 shows how the DCT completely alters the localization measure loc_p (4) for the matrices shown in Figure 7.

[6] In Strang's parlance [46], the MATLAB `dct` routine gives a DCT-3, scaled so the matrix is unitary.



matrix	n	$loc_5(Q^*AQ)$
Add20	2395	0.1118
Orsirr_1	1030	0.6579
Sherman5	3312	0.6876
2d Laplacian	9801	0.9603

FIG. 12. *Eigenvector localization measures* $loc_p(Q^*AQ)$ *for the coefficient matrices transformed by the DCT matrix* $Q^* =$ `dct(eye(n))`. *The transformation reduces the localization of the Add*20 *eigenvectors, while significantly enhancing localization for the three other matrices.*

Presuming Q is unitary, one can expedite the Arnoldi process by only storing the vectors Q^*v_j, rather than the usual Arnoldi vectors v_j. For example, in the W-inner product, the kth step of the conventional modified Gram–Schmidt Arnoldi process,

$$\begin{aligned}
&\widehat{v} = Av_k \\
&\text{for } j = 1, \ldots, k \\
&\qquad h_{j,k} = \langle \widehat{v}, v_j \rangle_W = v_j^* Q S^* S Q^* \widehat{v} \\
&\qquad \widehat{v} = \widehat{v} - h_{j,k} v_j \\
&\text{end},
\end{aligned}$$

can be sped up by storing only the vectors $u_j := Q^*v_j$, giving

$$\begin{aligned}
&\widehat{u} = Q^*Av_k = Q^*AQ\,u_k \\
&\text{for } j = 1, \ldots, k \\
&\qquad h_{j,k} = \langle \widehat{v}, v_j \rangle_W = u_j^* S^* S \widehat{u} \\
&\qquad \widehat{u} = \widehat{u} - h_{j,k} u_j \\
&\text{end}.
\end{aligned}$$

This latter formulation, which we use in our experiments, removes all applications of Q from the innermost loop. It is mathematically equivalent to the conventional version and differs only through unitary transformations, but over many cycles one can observe small drift in the numerical behavior of the two implementations.

Example 4.1 (Sherman5, revisited with discrete cosine transform). How does the DCT affect the Sherman5 system (Example 3.6)? Compare Figure 13 to the top of Figure 10; the DCT localizes the eigenvectors well, and now W-GMRES-DCT(30) slightly outperforms GMRES(30), while W-GMRES-DCT(100) converges in fewer iterations than GMRES(100). The DCT does not localize the eigenvectors as effectively as diagonalization (bottom of Figure 10), but it is practical.

Example 4.2 (Discrete Laplacian and convection-diffusion). In Example 3.7 we argued that W-GMRES(m) converged better than GMRES(m) despite global eigenvectors because the weighted inner product disrupts the cyclic pattern of GMRES(m). The DCT inner product localizes the eigenvectors, and, as Figure 14 shows, convergence is improved. Indeed, W-GMRES-DCT(20) requires roughly half as many iterations as W-GMRES(20), and both beat GMRES(20).

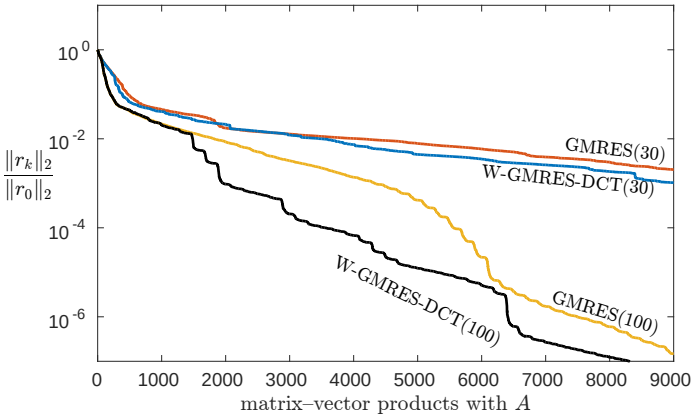


FIG. 13. *GMRES(m) and W-GMRES-DCT(m) for the nonsymmetric Sherman5 matrix. The DCT helps to localize the eigenvectors, resulting in significantly better convergence than seen for W-GMRES(m) in the top plot of Figure* 10.

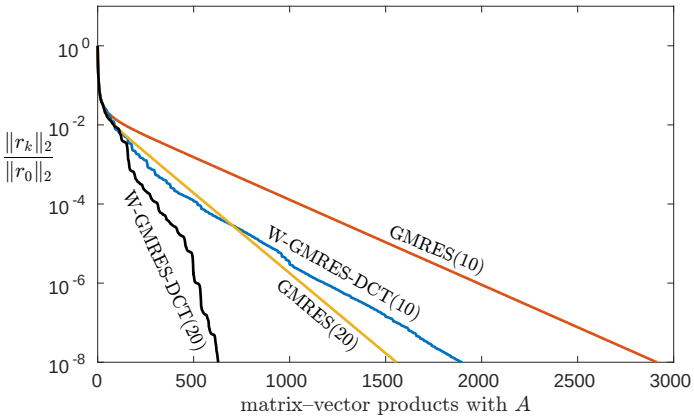


FIG. 14. *GMRES(m) and W-GMRES-DCT(m) for the* 2*d Dirichlet Laplacian (Example* 3.7*). The DCT localizes the eigenvectors, improving the convergence of W-GMRES(m) in Figure* 11.

The eigenvectors of the discrete Laplacian are 2d sinusoids, so the efficacy of the DCT is no surprise. The success of the method holds up when some nonsymmetry is added. Figure 15 compares GMRES(10), W-GMRES(10), and W-GMRES-DCT(10) for finite difference discretizations of the convection-diffusion operator $Lu = -(u_{xx} + u_{yy}) + u_x$, again with Dirichlet boundary conditions and mesh size $h = 1/100$ ($n = 99^2 = 9801$). Due to the convection term u_x the eigenvectors are no longer orthogonal, yet the DCT-enhanced weighting remains quite effective.

5. Weighted GMRES-DR. To improve convergence, many modifications to restarted GMRES have been proposed; see, e.g., [3, 4, 11, 24, 38, 43, 49]. Deflated GMRES methods use approximate eigenvectors in an attempt to remove some troublesome eigenvalues from the spectrum; see [1, 2, 8, 10, 13, 19, 20, 25, 30, 31, 35, 39] and references in [21]. Inner-product weighting has been incorporated into deflated GMRES in [29, 33]. Here we apply weighting to the efficient GMRES-DR method [31].

As detailed in [31], GMRES-DR computes approximate eigenvectors while solving linear equations. We show how to restart with a change of inner product; cf. [14,

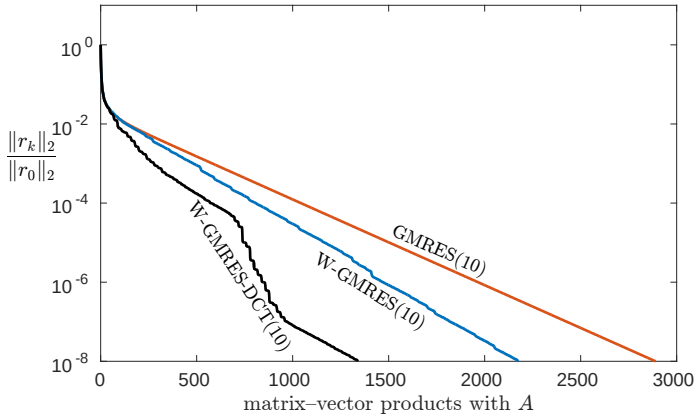


FIG. 15. *GMRES(*10*), W-GMRES(*10*), and W-GMRES-DCT(*10*) for the finite difference discretization of a* 2*d convection-diffusion problem with Dirichlet boundary conditions.*

Prop. 1]. The end of a GMRES-DR cycle in the W-inner product gives

$$AV_k = V_{k+1}H_{k+1,k}, \tag{10}$$

where V_k is an $n \times k$ matrix whose columns span a subspace of approximate eigenvectors, V_{k+1} is the same but with an extra column appended, and $H_{k+1,k}$ is a full $(k+1) \times k$ matrix. Since this recurrence is generated with the W-inner product, the columns of V_{k+1} are W-orthogonal, $V_{k+1}^* W V_{k+1} = I$, and $H_{k+1,k} = V_{k+1}^* W A V_k$. To convert (10) to a new inner product defined by $\widehat{W}$, compute the Cholesky factorization

$$V_{k+1}^* \widehat{W} V_{k+1} = R^* R,$$

where R is a $(k+1) \times (k+1)$ upper triangular matrix. Let R_k denote the leading $k \times k$ submatrix of R, and set

$$\widehat{V}_{k+1} := V_{k+1} R^{-1}, \qquad \widehat{H}_{k+1,k} := R H_{k+1,k} R_k^{-1}.$$

Then we have the new Arnoldi-like recurrence in the $\widehat{W}$-norm:

$$A\widehat{V}_k = \widehat{V}_{k+1}\widehat{H}_{k+1,k}$$

with $\widehat{V}_{k+1}^* \widehat{W} \widehat{V}_{k+1} = I$ and $\widehat{H}_{k+1,k} = \widehat{V}_{k+1}^* \widehat{W} A \widehat{V}_k$. Building on this decomposition, W-GMRES-DR extends the approximation subspace via the Arnoldi process and computes from this subspace the iterate that minimizes the $\widehat{W}$-norm of the residual.

Example 5.1. We apply W-GMRES-DR to the matrix Add20 from Example 2.1. Figure 16 shows convergence for GMRES(20) and GMRES-DR(20,5), with and without weighting. Weighting improves convergence more than deflating eigenvalues; combining the two strategies in W-GMRES-DR modestly improves convergence.

Example 5.2. For the matrix Sherman5, weighting did not help restarted GMRES in Example 3.6. Figure 17 adds GMRES-DR to the picture: deflating eigenvalues is very important, and now adding weighting to GMRES-DR(40,5) gives convergence in 306 fewer matrix-vector products with A.

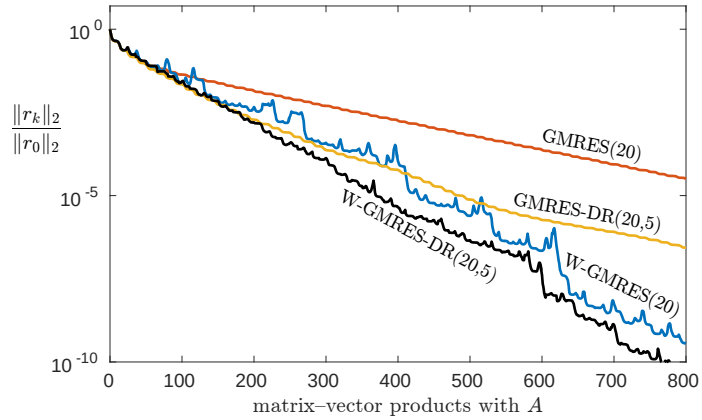


FIG. 16. *Comparison of restarted GMRES, with and without weighting and deflation, for Add*20.

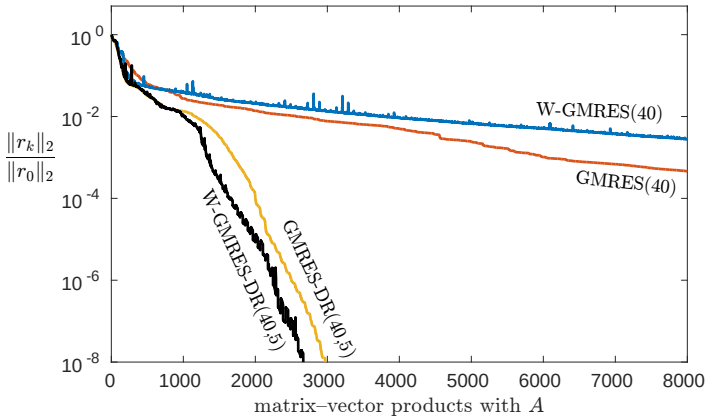


FIG. 17. *Comparison of methods with and without weighting and deflation for Sherman*5.

6. Conclusions. Since its introduction by Essai [14], W-GMRES has been an intriguing method whose adoption has been impeded by an incomplete understanding of its convergence properties. We have argued, through some simple analysis and numerous experiments, that its advantage is due to (1) its ability to break repetitive cycles in restarted GMRES, and (2) its ability to target small-magnitude eigenvalues that slow GMRES convergence, *provided the corresponding eigenvectors are localized*. In cases where those vectors are not localized the proposed W-GMRES-DCT algorithm, which combines residual weighting with a fast transform to sparsify eigenvectors, can potentially help. Weighted inner products can also be applied to deflated GMRES methods, such as GMRES-DR. While the improvement appears to be less significant than for standard GMRES, weighting can sometimes reduce the number of matrix-vector products even further than with the deflation alone or weighting alone.

Many related topics merit further investigation, such as additional sparsifying transformations or more general nondiagonal weighting matrices, application of related ideas to Arnoldi eigenvalue computations (cf. [41]), and the use of weighting to avoid near-breakdown in the nonsymmetric Lanczos method [32].

Acknowledgments. We thank Jörg Liesen, Jennifer Pestana, Andy Wathen, and the two anonymous referees for helpful comments about this work. ME is grateful

for the support of the Einstein Stiftung Berlin, which allowed him to visit the Technical University of Berlin while working on this project. RBM appreciates the support of the Baylor University Sabbatical Program.

REFERENCES

[1] A. M. Abdel-Rehim, R. B. Morgan, D. A. Nicely, and W. Wilcox, *Deflated and restarted symmetric Lanczos methods for eigenvalues and linear equations with multiple right-hand sides*, SIAM J. Sci. Comput., 32 (2010), pp. 129–149, https://doi.org/10.1137/080727361.

[2] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel, *Adaptively preconditioned GMRES algorithms*, SIAM J. Sci. Comput., 20 (1998), pp. 243–269, https://doi.org/10.1137/S1064827596305258.

[3] A. H. Baker, E. R. Jessup, and T. V. Kolev, *A simple strategy for varying the restart parameter in GMRES(m)*, J. Comput. Appl. Math., 230 (2009), pp. 751–761.

[4] A. H. Baker, E. R. Jessup, and T. Manteuffel, *A technique for accelerating the convergence of restarted GMRES*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 962–984, https://doi.org/10.1137/S0895479803422014.

[5] C. Beattie, M. Embree, and J. Rossi, *Convergence of restarted Krylov subspaces to invariant subspaces*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 1074–1109, https://doi.org/10.1137/S0895479801398608.

[6] R. Boisvert, R. Pozo, K. Remington, B. Miller, and R. Lipman, *Matrix Market*, http://math.nist.gov/MatrixMarket, 2004.

[7] J. H. Bramble and J. E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–17.

[8] K. Burrage and J. Erhel, *On the performance of various adaptive preconditioned GMRES strategies*, Numer. Linear Algebra Appl., 5 (1998), pp. 101–121.

[9] Z. Cao and X. Yu, *A note on weighted FOM and GMRES for solving nonsymmetric linear systems*, Appl. Math. Comput., 151 (2004), pp. 719–727.

[10] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1997), pp. 43–66.

[11] E. De Sturler, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp. 864–889, https://doi.org/10.1137/S0036142997315950.

[12] M. Embree, *The tortoise and the hare restart GMRES*, SIAM Rev., 45 (2003), pp. 259–266, https://doi.org/10.1137/S003614450139961.

[13] J. Erhel, K. Burrage, and B. Pohl, *Restarted GMRES preconditioned by deflation*, J. Comput. Appl. Math., 69 (1996), pp. 303–318.

[14] A. Essai, *Weighted FOM and GMRES for solving nonsymmetric linear systems*, Numer. Algorithms, 18 (1998), pp. 277–292.

[15] V. Faber, W. Joubert, E. Knill, and T. A. Manteuffel, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729, https://doi.org/10.1137/S0895479895286748.

[16] V. Faber, J. Liesen, and P. Tichý, *The Faber–Manteuffel theorem for linear operators*, SIAM J. Numer. Anal., 46 (2008), pp. 1323–1337, https://doi.org/10.1137/060678087.

[17] V. Faber and T. Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362, https://doi.org/10.1137/0721026.

[18] B. Fischer, A. Ramage, D. J. Silvester, and A. J. Wathen, *Minimum residual methods for augmented systems*, BIT, 38 (1998), pp. 527–543.

[19] J. Frank and C. Vuik, *On the construction of deflation-based preconditioners*, SIAM J. Sci. Comput., 23 (2001), pp. 442–462, https://doi.org/10.1137/S1064827500373231.

[20] L. Giraud, S. Gratton, X. Pinel, and X. Vasseur, *Numerical experiments on a flexible variant of GMRES-DR*, Proc. Appl. Math. Mech., 7 (2008), pp. 1020701–1020702.

[21] M. H. Gutknecht, *Spectral deflation in Krylov solvers: A theory of coordinate space based methods*, Electron. Trans. Numer. Anal., 39 (2012), pp. 156–185.

[22] M. H. Gutknecht and D. Loher, *Preconditioning by similarity transformations: Another valid option?*, abstract, GAMM Workshop on Numerical Linear Algebra, 2001.

[23] S. Güttel and J. Pestana, *Some observations on weighted GMRES*, Numer. Algorithms, 67 (2014), pp. 733–752.

[24] W. Joubert, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, Numer. Linear Algebra Appl., 1 (1994), pp. 427–447.

[25] 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.
[26] A. Klawonn, *Block-triangular preconditioners for saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 172–184, https://doi.org/10.1137/S1064827596303624.
[27] J. Liesen and B. N. Parlett, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, Numer. Math., 108 (2008), pp. 605–624.
[28] J. Liesen and Z. Strakoš, *Krylov Subspace Methods: Principles and Analysis*, Oxford University Press, Oxford, UK, 2013.
[29] B. Meng, *A weighted GMRES method augmented with eigenvectors*, Far East J. Math. Sci., 34 (2009), pp. 165–176.
[30] R. B. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171, https://doi.org/10.1137/S0895479893253975.
[31] R. B. Morgan, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37, https://doi.org/10.1137/S1064827599364659.
[32] R. B. Morgan and D. A. Nicely, *Restarting the nonsymmetric Lanczos algorithm for eigenvalues and linear equations including multiple right-hand sides*, SIAM J. Sci. Comput., 33 (2011), pp. 3037–3056, https://doi.org/10.1137/100799265.
[33] Q. Niu, L. Lu, and J. Zhou, *Accelerate weighted GMRES by augmenting error approximations*, Int. J. Comput. Math., 87 (2010), pp. 2101–2112.
[34] C. C. Paige, M. Rozložník, and Z. Strakoš, *Modified Gram–Schmidt (MGS), least squares, and backward stability of MGS-GMRES*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 264–284, https://doi.org/10.1137/050630416.
[35] 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, https://doi.org/10.1137/040607277.
[36] J. Pestana and A. J. Wathen, *On choice of preconditioner for minimum residual methods for non-Hermitian matrices*, J. Comput. Appl. Math., 249 (2013), pp. 57–68.
[37] M. Rozložník, M. Tůma, A. Smoktunowicz, and J. Kopal, *Numerical stability of orthogonalization methods with a non-standard inner product*, BIT, 52 (2012), pp. 1035–1058.
[38] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469, https://doi.org/10.1137/0914028.
[39] Y. Saad, *Analysis of augmented Krylov subspace techniques*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 435–449, https://doi.org/10.1137/S0895479895294289.
[40] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Comput., 7 (1986), pp. 856–869, https://doi.org/10.1137/0907058.
[41] H. Saberi Najafi and H. Ghazvini, *Weighted restarting method in the weighted Arnoldi algorithm for computing eigenvalues of a nonsymmetric matrix*, Appl. Math. Comput., 175 (2006), pp. 1276–1287.
[42] H. Saberi Najafi and H. Zareamoghaddam, *A new computational GMRES method*, Appl. Math. Comput., 199 (2008), pp. 527–534.
[43] V. Simoncini, *A new variant of restarted GMRES*, Numer. Linear Algebra Appl., 6 (1999), pp. 61–77.
[44] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd ed., Oxford University Press, Oxford, UK, 1985.
[45] M. Stoll and A. Wathen, *Combination preconditioning and the Bramble–Pasciak$^+$ preconditioner*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 582–608, https://doi.org/10.1137/070688961.
[46] G. Strang, *The discrete cosine transform*, SIAM Rev., 41 (1999), pp. 135–147, https://doi.org/10.1137/S0036144598336745.
[47] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[48] H. A. van der Vorst and C. Vuik, *The superlinear convergence of GMRES*, J. Comput. Appl. Math., 48 (1993), pp. 327–341.
[49] H. A. van der Vorst and C. Vuik, *GMRESR: A family of nested GMRES methods*, Numer. Linear Algebra Appl., 1 (1994), pp. 369–386.
[50] B. Zhong and R. B. Morgan, *Complementary cycles of restarted GMRES*, Numer. Linear Algebra Appl., 15 (2008), pp. 559–571.