Some observations on weighted GMRES*

Stefan Güttel · Jennifer Pestana

Abstract We investigate the convergence of the weighted GMRES method for solving linear systems. Two different weighting variants are compared with unweighted GMRES for three model problems, giving a phenomenological explanation of cases where weighting improves convergence, and a case where weighting has no effect on the convergence. We also present a new alternative implementation of the weighted Arnoldi algorithm which under known circumstances will be favourable in terms of computational complexity. These implementations of weighted GMRES are compared for a large number of examples. We find that weighted GMRES may outperform unweighted GMRES for some problems, but more often this method is not competitive with other Krylov subspace methods like GMRES with deflated restarting or BICGSTAB, in particular when a preconditioner is used.

S. Güttel
Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK.
Current address: School of Mathematics, The University of Manchester, Alan Turing Building, Manchester M13 9PL, UK, E-mail: stefan.guettel@manchester.ac.uk

J. Pestana
Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK,
E-mail: pestana@maths.ox.ac.uk

1 Introduction

The GMRES method of Saad and Schultz [28] is one of the most popular
Krylov subspace methods for solving a non-Hermitian system of linear equa-
tions $A\boldsymbol{x} = \boldsymbol{b}$, where $A \in \mathbb{C}^{N \times N}$ is invertible and $\boldsymbol{b} \in \mathbb{C}^N$. Given an initial
guess $\boldsymbol{x}^{(0)}$, GMRES computes successive iterates $\boldsymbol{x}^{(k)}$, $k = 1, 2, \ldots$, so that

$$\|\boldsymbol{r}^{(k)}\|_2 = \min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(A)\boldsymbol{r}^{(0)}\|_2,$$

where $\mathcal{P}_k$ denotes the linear space of polynomials of degree at most k, and
$\boldsymbol{r}^{(k)} = \boldsymbol{b} - A\boldsymbol{x}^{(k)}$ is the k-th residual.

Since GMRES uses the Arnoldi algorithm, its computational cost increases
with each iteration. An alternative is to restart GMRES after m iterations [28],
taking the last computed residual as the next initial residual. We call the
original method *full* GMRES and the latter *restarted* GMRES or GMRES(m).
The set of m Arnoldi iterations between successive restarts will be called a
cycle.

Although in exact arithmetic full GMRES is guaranteed to terminate with
the exact solution in at most N steps, the restarted version may stagnate [5,
12,28,34] or converge slowly [4,35,36]. The behaviour of restarted GMRES
has been well studied and a number of remedies for slow convergence have
been proposed [11,22,27,30–32].

One such remedy is the *weighted* GMRES method of Essai [13], shortly de-
noted as WGMRES(m), that aims to improve the convergence of GMRES(m)
by using a weighted inner product, which we call a *D-inner product*, that
changes at each cycle. This D-inner product, and associated D-norm, are
defined for any Hermitian positive definite $D \in \mathbb{C}^{N \times N}$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ as
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_D = \boldsymbol{y}^H D \boldsymbol{x}$, $\|\boldsymbol{x}\|_D = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_D}$, where $\boldsymbol{y}^H$ represents the Hermitian
conjugate of $\boldsymbol{y}$.

The WGMRES(m) method also starts from an initial guess $\boldsymbol{x}^{(0)}$ and com-
putes successive approximations $\boldsymbol{x}^{(k)}$ at each cycle $k = 1, 2, \ldots$, such that at
the end of the k-th cycle

$$\|\boldsymbol{r}^{(k)}\|_D = \min_{\substack{p \in \mathcal{P}_m \\ p(0)=1}} \|p(A)\boldsymbol{r}^{(k-1)}\|_D.$$

For further details we refer to Essai [13].

The essential ingredient of weighted GMRES is the weighted Arnoldi algo-
rithm [13] that, after m iterations, generates basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ of the
Krylov space $\mathcal{K}_m(A, \boldsymbol{r}) = \mathrm{span}\{\boldsymbol{r}, A\boldsymbol{r}, \ldots, A^{m-1}\boldsymbol{r}\}$. If one collects the Krylov
basis vectors in a matrix $V_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}] \in \mathbb{C}^{N \times m}$, one can write down
an Arnoldi decomposition

$$AV_m = V_{m+1}\underline{H_m} = V_m H_m + \boldsymbol{v}_{m+1} h_{m+1,m} \boldsymbol{e}_m^T, \tag{1}$$

where $\underline{H_m} \in \mathbb{C}^{(m+1)\times m}$ is the upper Hessenberg matrix

$$\underline{H_m} = \begin{bmatrix} H_m \\ h_{m+1,m}\boldsymbol{e}_m^T \end{bmatrix},$$

and $\boldsymbol{e}_m \in \mathbb{R}^m$ is the m-th canonical unit vector. The matrix V_{m+1} is D-orthonormal, i.e., $V_{m+1}^H D V_{m+1} = I_{m+1}$, the identity matrix of dimension $m+1$. The weighted Arnoldi algorithm requires more computation per iteration than standard Arnoldi in the Euclidean inner product and, consequently, one cycle of WGMRES(m) is computationally more expensive than one cycle of GMRES(m). However, convergence may occur more quickly.

We would like to emphasize that *weighting* differs from *preconditioning*. Left preconditioning, for example, solves $P^{-1}A\boldsymbol{x} = P^{-1}\boldsymbol{b}$, thereby seeking a solution from $\mathcal{K}_{km+1}(P^{-1}A, P^{-1}\boldsymbol{r}^{(0)})$. One expects that a Krylov subspace method that uses this space converges faster, and typically this means that the eigenvalues of $P^{-1}A$ are clustered. (Right preconditioning has an analogous effect.) Weighting, on the other hand, does not change the Krylov space at all, instead affecting the inner product that is used to extract an approximation from the Krylov space built with the original matrix A.

Essai [13] considered the particular weight matrix

$$D = \frac{1}{\sqrt{N}\|\boldsymbol{r}^{(k-1)}\|_2} \operatorname{diag}\big(|r_1^{(k-1)}|, |r_2^{(k-1)}|, \ldots, |r_N^{(k-1)}|\big), \qquad (2)$$

where the $r_j^{(k-1)}$ are the entries of the residual vector $\boldsymbol{r}^{(k-1)}$, so that greater emphasis is given to large components of the residual at each cycle. Note that $D = D^{(k)}$ changes at each cycle, but to keep notation simple, we typically omit the superindex k. The matrix D may be poorly conditioned if the diagonal entries vary too much in magnitude. In such cases, adding a small multiple of the identity will improve the conditioning of D.

For a number of test problems, WGMRES(m) with the weight matrix (2) required fewer cycles and less CPU time than the standard GMRES(m) method [13]. Application of WGMRES(m) to systems left-preconditioned by ILU(0) [21] also resulted in a slight reduction in the number of cycles required for convergence when compared with GMRES(m) [6]. However, the CPU time for WGMRES(m) was greater as a consequence of the computation of nonstandard inner products and norms. The weighted GMRES method has also been used to solve shifted linear systems [19], and systems with multiple right-hand sides [17]. We remark that Niu *et al.* [25] showed that WGMRES(m) can be accelerated by augmenting the Krylov space at cycle k with the ℓ most recent error approximations $\boldsymbol{z}^{(i)}$, $i = k-\ell, \ldots, k-1$, where $\boldsymbol{z}^{(i)} = \boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i-1)}$ when $i > 0$ and $\boldsymbol{0}$ otherwise.

Although intuitively it seems sensible to emphasize those entries of the residual vector that are large in magnitude, the convergence behaviour of WGMRES(m) is not well understood. We attempt to remedy this here by examining the harmonic Ritz values associated with WGMRES(m). In view of the limited understanding even of the convergence of full GMRES, it seems

unlikely at this stage that a simple and complete convergence theory for WGM-RES can be developed. However, some insight can be gained by studying several model problems. We also propose a new implementation of the weighted Arnoldi algorithm and compare its cost with the original.

The outline of this paper is as follows. An analysis of the harmonic Ritz values associated with GMRES(m) and WGMRES(m) is given in Section 2. In Section 3 we compare Essai's implementation of the weighted Arnoldi algorithm with an alternative. Finally, in Section 4, the different implementations are tested on a number of problems and compared with standard GMRES(m), GMRES(m) with deflated restarting, and BICGSTAB.

2 Harmonic Ritz values and the convergence of weighted GMRES

In this section we try to shed some light on the convergence behaviour of weighted GMRES and explain why this method may converge faster than unweighted GMRES in some cases, or why weighting may have no effect on the convergence. It should be emphasized that GMRES(m) and WGMRES(m) after k cycles yield residuals $\boldsymbol{r}^{(k)}$ from the *same* Krylov space $\mathcal{K}_{km+1}(A, \boldsymbol{r}^{(0)})$ but the harmonic Ritz values that uniquely determine the residual polynomials may exhibit considerably different behaviour. In other words, the approximation spaces of both methods are the same but the extractions from these spaces may be different. This very property makes weighting quite different from what is typically achieved by a preconditioner P.

The convergence of GMRES (and its restarted and weighted variants) is generally very difficult to analyse, if not impossible, as in theory any nonincreasing convergence curve can be obtained with any choice of eigenvalues and Ritz values [3,9,16]. Additionally, restarted GMRES may exhibit any admissible cycle-convergence behaviour, where the two admissible situations are that the residuals decrease strictly monotonically at each cycle or that there is complete stagnation [32]. Nevertheless, we still consider it instructive to make clear the relations between the unweighted and weighted (harmonic) Ritz values in the following. At the end of this section we will study three (unpreconditioned) model problems. As no set of examples can be exhaustive, our primary aim must be to illustrate and analyse some effects that may cause the difference in the convergence of GMRES(m) and WGMRES(m) observed in practical examples.

Facts about harmonic Ritz values. Let us start by collecting some well-known facts about harmonic Ritz values, see [15,33,37]. First of all, the weighted harmonic Ritz values θ_j with corresponding Ritz vectors $\boldsymbol{u}_j = V_m \boldsymbol{z}_j$ satisfy

$$\left(H_m + |h_{m+1,m}|^2 \boldsymbol{f}_m \boldsymbol{e}_m^T \right) \boldsymbol{z}_j = \theta_j \boldsymbol{z}_j, \tag{3}$$

where $\boldsymbol{f}_m = H_m^{-H} \boldsymbol{e}_m$. It is also well known that the harmonic Ritz values $\theta_1^{(k)}, \ldots, \theta_m^{(k)}$ associated with cycle k are the zeros of the residual polynomial

$p_m^{(k)} \in \mathcal{P}_m$, $p_m^{(k)}(0) = 1$, which is uniquely determined by the condition

$$\begin{aligned}
\|\boldsymbol{r}^{(k)}\|_{D^{(k)}} &= \|p_m^{(k)}(A)p_m^{(k-1)}(A)\cdots p_m^{(1)}(A)\boldsymbol{r}^{(0)}\|_{D^{(k)}} \\
&= \min_{\substack{p\in\mathcal{P}_m \\ p(0)=1}} \|p(A)p_m^{(k-1)}(A)\cdots p_m^{(1)}(A)\boldsymbol{r}^{(0)}\|_{D^{(k)}}.
\end{aligned}$$

If A is normal, then with $\widetilde{p}^{(k)} := p_m^{(k)}p_m^{(k-1)}\cdots p_m^{(1)}$ we have

$$\|\boldsymbol{r}^{(k)}\|_{D^{(k)}} \leq \|\boldsymbol{r}^{(0)}\|_{D^{(k)}} \max_{\lambda\in\Lambda(A)} |\widetilde{p}^{(k)}(\lambda)|,$$

so that the convergence of (restarted) GMRES in the 2-norm can be understood in terms of the uniform convergence of residual polynomials on the discrete set of eigenvalues $\Lambda(A)$.

Relationship between Ritz values with and without weighting. Starting from a given vector, the Arnoldi method in the D-inner product builds D-orthogonal vectors that satisfy (1). With the same starting vector, the Arnoldi method in the Euclidean inner product computes orthogonal vectors $\widehat{\boldsymbol{v}}_i$ such that

$$A\widehat{V}_m = \widehat{V}_{m+1}\underline{\widehat{H}_m}, \quad \widehat{V}_m = [\widehat{\boldsymbol{v}}_1,\ldots,\widehat{\boldsymbol{v}}_m], \quad \underline{\widehat{H}_m} = \begin{bmatrix} \widehat{H}_m \\ \widehat{h}_{m+1,m}\boldsymbol{e}_m^T \end{bmatrix}, \qquad (4)$$

where again $\underline{\widehat{H}_m}$ is an upper Hessenberg matrix. Additionally, the matrices V_m and $\widehat{V}_m$, and $\underline{H_m}$ and $\underline{\widehat{H}_m}$ are linked by

$$V_m = \widehat{V}_m S_m, \; \underline{H_m} = S_{m+1}^{-1}\underline{\widehat{H}_m}S_m, \qquad (5)$$

where S_m is upper triangular and is nonsingular in the absence of breakdown. From (5) it follows that[1] [13, Proposition 2, Corollary 1]

$$\begin{aligned}
\widehat{H}_m &= S_m H_m S_m^{-1} + \frac{\widehat{h}_{m+1,m}}{s_{m+1,m+1}}\boldsymbol{s}_{m+1}\boldsymbol{e}_m^T, \\
H_m &= S_m^{-1}\widehat{H}_m S_m + \widehat{h}_{m+1,m}s_{m,m}\boldsymbol{g}_{m+1}\boldsymbol{e}_m^T,
\end{aligned} \qquad (6)$$

where $\boldsymbol{s}_{m+1}$ and $\boldsymbol{g}_{m+1}$ are the vectors obtained from the first m elements of the last column of S_{m+1} and S_{m+1}^{-1}, respectively, and $s_{m+1,m+1} = \boldsymbol{e}_{m+1}^H S_{m+1}^{-1}\boldsymbol{e}_{m+1}$.

We wish to relate the weighted harmonic Ritz values, defined by (3), and the unweighted harmonic Ritz values, defined by

$$(\widehat{H}_m + |\widehat{h}_{m+1,m}|^2\widehat{\boldsymbol{f}}_m\boldsymbol{e}_m^T)\widehat{\boldsymbol{y}}_m = \widehat{\theta}_m\widehat{\boldsymbol{y}}_m,$$

where $\widehat{\boldsymbol{f}}_m = \widehat{H}_m^{-H}\boldsymbol{e_m}$. Using (6), we find that

$$\widehat{H}_m + |\widehat{h}_{m+1,m}|^2\widehat{\boldsymbol{f}}_m\boldsymbol{e}_m^T = S_m\left(H_m + |h_{m+1,m}|^2\boldsymbol{f}_m\boldsymbol{e}_m^T\right)S_m^{-1} + \boldsymbol{z}_m\boldsymbol{e}_m^T, \qquad (7)$$

---

[1] Note that the formula in Corollary 1 in [13] is incorrect and should be as in (6).

where

$$\boldsymbol{z}_m = \frac{\widehat{h}_{m+1,m}}{s_{m+1,m+1}}\boldsymbol{s}_{m+1} + |\widehat{h}_{m+1,m}|^2\widehat{\boldsymbol{f}}_m - |h_{m+1,m}|^2 g_{m,m}S_m\boldsymbol{f}_m.$$

If $g_{m+1,m+1} = \boldsymbol{e}_{m+1}^T S_{m+1}^{-1}\boldsymbol{e}_m$ then, using (5), we have that $h_{m+1,m} = \boldsymbol{e}_{m+1}^T\underline{H}_m\boldsymbol{e}_m = g_{m+1,m+1}\widehat{h}_{m+1,m}s_{m,m}$. Additionally, by (6) and the Sherman–Morrison formula [18, page 19],

$$\boldsymbol{f}_m = H_m^{-H}\boldsymbol{e}_m = \frac{1}{1 + \widehat{h}_{m+1,m}^* s_{m,m}^* \boldsymbol{g}_{m+1}^H T_m^{-H}\boldsymbol{e}_m}T_m^{-H}\boldsymbol{e}_m,$$

where $T_m = S_m^{-1}\widehat{H}_m S_m$. Now, $T_m^{-H}\boldsymbol{e}_m = g_{m,m}^*S_m^H\widehat{\boldsymbol{f}}_m$, and, from the proof of Corollary 2 in [13], $S_m\boldsymbol{g}_{m+1} = -g_{m+1,m+1}\boldsymbol{s}_{m+1}$. Thus,

$$\boldsymbol{f}_m = \frac{g_{m,m}^*}{1 - \widehat{h}_{m+1,m}^* g_{m+1,m+1}^* \boldsymbol{s}_{m+1}^H\widehat{\boldsymbol{f}}_m}S_m^H\widehat{\boldsymbol{f}}_m$$

from which it follows that

$$\begin{aligned}
\boldsymbol{z}_m =&|\widehat{h}_{m+1,m}|^2\left(I - \frac{|g_{m+1,m+1}|^2}{1 - (g_{m+1,m+1}\widehat{h}_{m+1,m})^*\boldsymbol{s}_{m+1}^H\widehat{\boldsymbol{f}}_m}S_mS_m^H\right)\widehat{\boldsymbol{f}}_m \\
&+ \frac{\widehat{h}_{m+1,m}}{s_{m+1,m+1}}\boldsymbol{s}_{m+1}.
\end{aligned} \tag{8}$$

Applying the Bauer–Fike theorem [18, Theorem 6.3.2] to (7) with (8), gives

$$\min_j |\theta_j - \widehat{\theta}_i| \le \kappa(S_m^{-1}X_m)\|\boldsymbol{z}_m\|_2, \ i = 1,\dots,m,$$

where X_m is an eigenvector matrix of $H_m + |h_{m+1,m}|^2\boldsymbol{f}_m\boldsymbol{e}_m^T$ and $\kappa(S_m^{-1}X_m)$ is the 2-norm condition number of $S_m^{-1}X_m$. Although the influence of the weighting matrix D is not obvious from the above inequality, we can obtain a (typically pessimistic) bound that displays the effect of weighting more clearly since, by (5) and the D-orthogonality of V_m, $S_m^{-H}S_m^{-1} = \widehat{V}_m^H D\widehat{V}_m$. Thus, it follows that $\kappa(S_m^{-1})^2 = \kappa(\widehat{V}_m^H D\widehat{V}_m) \le \kappa(D)$ and that

$$\min_j |\theta_j - \widehat{\theta}_i| \le \sqrt{\kappa(D)}\kappa(X_m)\|\boldsymbol{z}_m\|_2, \ i = 1,\dots,m.$$

This shows that the difference between the weighted and unweighted harmonic Ritz values depends on the nonnormality of $H_m + |h_{m+1,m}|^2\boldsymbol{f}_m\boldsymbol{e}_m^T$, through X_m, and the conditioning of the change of basis matrix S_m. The latter term is bounded by the condition number of D, which in the case of Essai's weighting (2) is given by the ratio of the largest and smallest values of $\boldsymbol{r}^{(k-1)}$ in magnitude. Consequently, we obtain a smaller bound when the entries of the residual vector have similar magnitudes.

The nearness of the harmonic Ritz values additionally depends on $\|\boldsymbol{z}_m\|_2$ in (8). In GMRES, $|\widehat{h}_{m+1,m}|$ is small in magnitude when the residual norm is

small [28, Proposition 1]. When the angle between $\boldsymbol{v}_{m+1}$ and $\widehat{\boldsymbol{v}}_{m+1}$ is small, $|s_{m+1,m+1}|$ is large and $|g_{m+1,m+1}| = 1/|s_{m+1,m+1}|$ is small relative to the remaining entries in the $(m + 1)$-th column of S_{m+1}, i.e., the entries of $\boldsymbol{s}_{m+1}$. The size of $\boldsymbol{z}_m$ also depends on the conditioning of $\widehat{H}_m$, through $\widehat{\boldsymbol{f}}_m$, and on the norm of S_m, which can be bounded by the norm of D^{-1}. Thus, when the entries of D are of the same magnitude, and $H_m + |h_{m+1,m}|^2 \boldsymbol{f}_m \boldsymbol{e}_m^T$ has a well conditioned eigenvector matrix, we might expect the weighted and unweighted harmonic Ritz values to be close; otherwise they may differ significantly, which may be advantageous for weighted GMRES.

Relationship between WGMRES and GMRES residuals. If the GMRES and WGMRES residuals coincide at the $(k - 1)$-th cycle then, since WGMRES is a quasi-minimal residual method [10, Section 4.3], it cannot have a smaller residual in the Euclidean norm than GMRES at the end of the k-th cycle. Properties of quasi-minimal residual methods are described in Section 4.3 of Eiermann and Ernst [10], where, for example, it is shown that the difference between the GMRES and WGMRES residuals is bounded by [10, Theorem 4.8]

$$\|\boldsymbol{r}^{(k)}_{GMRES}\|_2 \leq \|\boldsymbol{r}^{(k)}_{WGMRES}\|_2 \leq \sqrt{\kappa(D^{(k)})}\|\boldsymbol{r}^{(k)}_{GMRES}\|_2.$$

The difference is small if all elements of $\boldsymbol{r}^{(k-1)}$ are of approximately equal magnitude, but may be larger when the sizes of these elements vary. Thus, if WGMRES performs better than GMRES this must be due to a better choice of starting residual for the next cycle. We explore the effect of changing $\boldsymbol{r}^{(k)}$ on convergence and on the harmonic Ritz values in the rest of this section.

Three model problems. We now compare the harmonic Ritz values generated by GMRES(m) with the weighted harmonic Ritz values of WGMRES(m) for three model problems. As well as Essai's weight matrix (2) we consider an alternative, proposed by Najafi and Zareamoghaddam [24], who were concerned that as the magnitudes of the entries of the residual became smaller it would be difficult to compute with (2); D_{rand} is a diagonal matrix with random uniformly distributed entries in (0.5,1.5). We use reorthogonalization to minimize the effect of finite precision on our results. Although the following three examples have little practical relevance, we believe that they serve the purpose of giving insight into how weighting can possibly improve the convergence of restarted GMRES (Examples 1 and 3), or how weighting can have absolutely no effect on the convergence (Example 2).

Example 1 (interval) *We examine the harmonic Ritz values of a diagonal matrix with diagonal entries (and, hence, eigenvalues) $1, 2, \ldots, 100$. The right-hand side $\boldsymbol{b}$ is a vector of all ones, scaled to unit length. As one can see in Figure 1 (a), it appears that for unweighted GMRES(m) with $m = 5$ the harmonic Ritz values of every second cycle have m accumulation points, giving*

asymptotically $2m$ accumulation points $\theta_1^*, \ldots, \theta_{2m}^*$ in total. In this example these accumulation points are approximately

$$3.348, \; 22.208, \; 51.510, \; 79.318, \; 96.908,$$
$$3.453, \; 20.616, \; 49.477, \; 79.784, \; 98.155.$$

It should be noted that a similar 2-cyclic behaviour has been observed and analysed for the so-called optimum gradient method in the 1950's [14, 2] (this method can be interpreted as restarted FOM), and in the context of matrix function approximations in [1]. This restarted GMRES behaviour may also be related to the asymptotic orthogonality of successive initial residuals $\boldsymbol{r}^{(k-1)}$ and $\boldsymbol{r}^{(k)}$, proven for the case $m = N-1$ in [4, Theorem 2]. A detailed investigation of this phenomenon is beyond the scope of this paper, but we expect that tools similar to those used in the mentioned papers can be applied.

In Figure 1 (b) we show the level lines of the modulus of the nodal polynomial $q_{2m}(z) = \prod_{j=1}^{2m}(z - \theta_j^*)$. These level lines are also known as lemniscates, *see also the discussion in [1]. One can read off from this plot that the level line $10^{14.535}$ is the smallest one containing $\Lambda(A)$ in its interior, and the level line $10^{14.909}$ passes through the origin. By the normalization condition of residual polynomials, the modulus of $q_{2m}(\lambda)/q_{2m}(0)$ is at most $10^{-0.3740} \approx 0.4227$ for all $\lambda \in \Lambda(A)$. This residual polynomial is the result of two restart cycles, hence the expected convergence rate of restarted unweighted GMRES(m) in this example is approximately $\sqrt{0.4227} \approx 0.6502$. This rate is shown in Figure 1 (c) as the black dashed line, and it coincides well with the observed linear convergence of unweighted GMRES(m) (black curve with + markers).*

The convergence of the weighted GMRES(m) variants under consideration appears much less regular. The harmonic Ritz values associated with Essai's weighting appear to cover the spectral interval of A more evenly, and this is also indicated by the histogram in Figure 1 (d), which shows the distribution of harmonic Ritz values over the spectral interval of A. This "randomization" of interpolation nodes causes the method to converge faster than linearly. A similar effect is achieved by random weighting with D_{rand}. The fact that these harmonic Ritz values are spread out over the spectral interval of A makes visually clear that weighting does not attempt to cluster the spectrum (and thereby the harmonic Ritz values) as we might expect a preconditioner to do. We have not attempted to plot the lemniscates associated with the harmonic Ritz values produced by the weighted GMRES variants as, due to the observed irregular behaviour, these lemniscates cannot be described by just a few accumulation points. Therefore the evaluation of the residual polynomials at zero would not give more information than the computed residual norms.

Example 2 (circle) *Our second example is a diagonal matrix with $N = 100$ diagonal elements (eigenvalues) $\beta \cdot e^{2i\pi j/N} + 1$ on a circle of radius $\beta = 0.9$ centered at $z = 1$, $j = 1, 2, \ldots, N$. The right-hand side $\boldsymbol{b}$ is a vector of all ones, scaled to unit length, and the restart length is $m = 5$. As can be seen from Figure 2 (a), the harmonic Ritz values "spiral" towards the point $z = 1$, being almost evenly spaced on concentric circles. This effect appears for all types*
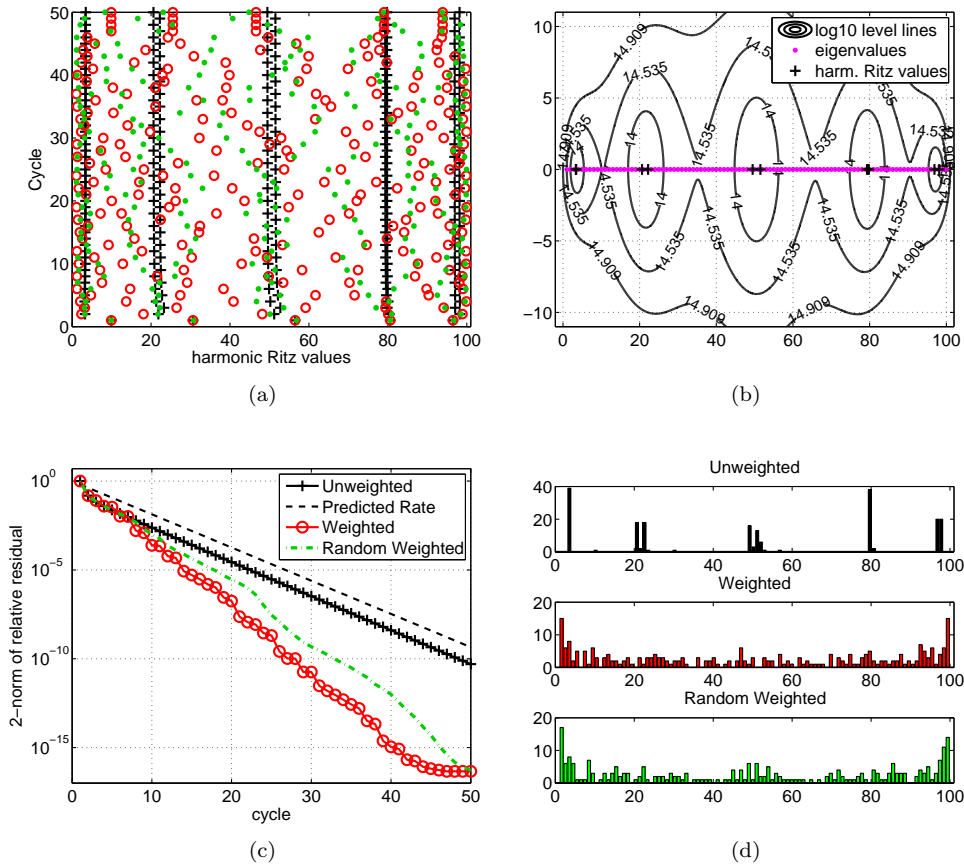
(a)

(b)

(c)

(d)

Fig. 1: (a) Harmonic Ritz values for GMRES(5) (black +), WGMRES(5) with (2) (red ∘), and WGMRES(5) with D_{rand} (green dots) for a diagonal matrix with equispaced eigenvalues on $[1, 100]$. The harmonic Ritz values are shown at the end of each of 50 cycles. (b) Lemniscates associated with unweighted GM-RES(5). (c) Relative 2-norm residuals for the considered GMRES(5) variants. (d) Histogram indicating the distribution of harmonic Ritz values.

of weighting under investigation, and the convergence shown in Figure 2 (b) seems to be unaffected by whatever weighting method we use. To explain this observation, assume that at some cycle the harmonic Ritz values are unit roots of order m shifted and scaled to a circle of radius $\alpha < \beta$ centered at $z = 1$. The corresponding nodal polynomial is $q_m(z) = (z-1)^m - \alpha^m$. As can be verified easily, the maximal modulus of q_m on the circle of radius β centered at $z = 1$ is attained at points "in the middle" of two neighboring eigenvalues, for example, at the point $z^* = 1 + \beta e^{\pi i/N}$. Hence, the modulus of the residual
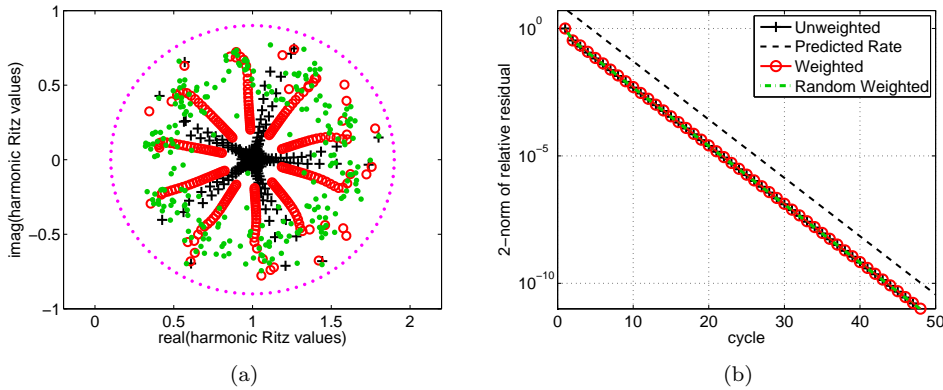
Fig. 2: (a) Harmonic Ritz values for GMRES(5) (black +), WGMRES(5) with (2) (red ∘), and WGMRES(5) with D_{rand} (green dots) for a 100×100 matrix with eigenvalues distributed on the shifted unit circle. The harmonic Ritz values are shown at the end of each of 50 cycles. (b) Relative 2-norm residuals for the considered GMRES(5) variants. (All three convergence curves are visually indistinguishable.)

polynomial is bounded by

$$\left| \frac{q_m(z^*)}{q_m(0)} \right| = \left| \frac{q_m(1 + \beta e^{\pi i/N})}{q_m(0)} \right| = \left| \frac{\beta^m e^{\pi i m/N} - \alpha^m}{(-1)^m - \alpha^m} \right| \approx \beta^m$$

for sufficiently small α. This explains why we see convergence with rate β in Figure 2 (b), indicated by the black dashed line, and weighting has essentially no effect on the convergence here.

Example 3 (Jordan block) *Our next example is an upper triangular Jordan block J of size $N = 100$ with eigenvalue 1. The right-hand side $\boldsymbol{b}$ is a vector of all ones, scaled to unit length. As one can see in Figure 3 (b), the unweighted GMRES(5) method will stagnate except in the first cycle, where a little progress is made. The corresponding harmonic Ritz values reappear at 10 points on a circle of radius one around the eigenvalue 1, with the real point $\theta = 2$ being counted twice due to symmetry, see Figure 3 (a). The harmonic Ritz values associated with Essai's weighted GMRES(5) method move closer towards the eigenvalue 1 with each cycle. After 23 cycles, weighted GMRES(5) has found the exact solution of $J\boldsymbol{x} = \boldsymbol{b}$. The random weighting matrix D_{rand} leads to stagnation just as unweighted GMRES(5). As opposed to the previous two examples, the harmonic Ritz values shown in Figure 3 (a) do not explain the convergence curves in Figure 3 (b) for this (highly) nonnormal example.*

To still give some insight into the different behaviour of unweighted GMRES and Essai's weighted GMRES, we visualize in Figure 3 (c) and (d) the

entries of the residual vector after each cycle. The special structure of the Jordan matrix results in large residual entries being shifted up the vector with each cycle. With unweighted GMRES(5) the residual vector is initially largest in its last entries. This phenomenon is fully described by Theorem 2.1 in [20], which is stated in terms of the transpose of the Jordan block, J^T, but is also valid for J. Some intuition is gained by observing that if $S = \begin{bmatrix} \mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_{N-1} \end{bmatrix}$ is the noncircular shift matrix, then $\mathcal{K}_m(J, \mathbf{r}) = \mathcal{K}_m(S, \mathbf{r})$ for any vector $\mathbf{r}$. It follows that the first two basis vectors, $\mathbf{r}$ and $S\mathbf{r}$ differ only in the last component and this affects the weight of the residual. More generally, any two basis vectors $S^j\mathbf{r}$ and $S^{j+1}\mathbf{r}$ differ in the $(N-j)$-th component only. Thus the "support" of nonzero entries in the residual vector at the end of the first cycle is in the last five components of the residual. At later cycles, this "support" forms a band that gets wider with each cycle, eventually polluting all entries of the residual vector and causing the method to stagnate. WGMRES(5) with Essai's weighting initially has largest entries at the bottom of the residual vector, although the other components also have some weight. Weighted GMRES then "cleans up" the entries of the residual vector which were large in the previous cycle because more weight is placed at those entries. Eventually, this WGMRES(5) variant finds the exact solution in the 24-th cycle.

Note that this Jordan example also explains why the Krylov space and hence the residual at the end of a cycle may depend sensitively on the initial residual for that cycle: instead of working with the matrix J and right-hand side vector $\mathbf{b}$ we could as well run restarted weighted or unweighted GMRES with $A = XJX^{-1}$ and $\widetilde{\mathbf{b}} = X\mathbf{b}$, where $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ is an arbitrary invertible matrix. Since $\mathcal{K}_m(A, \widetilde{\mathbf{b}}) = X\mathcal{K}_m(J, \mathbf{b})$, each column of Figure 3 (c) and (d) can now be interpreted as the components of a residual vector in the basis of generalized eigenvectors of A. If a zero component r_j of a residual vector $\mathbf{r} = [r_1, \ldots, r_N]^T$ is altered from 0 to $\epsilon > 0$, for example by finite precision arithmetic, then this can cause a change of an eigenvector component in $A\mathbf{r}$ (and the following Krylov subspace vectors) of order $\epsilon\|\mathbf{x}_j\|$, which can be arbitrarily large depending on $\|\mathbf{x}_j\|$.

We conclude this section by remarking on the observations of Cao and Yu [6], who compared WGMRES(m) with GMRES(m) on ILU-preconditioned problems. They found that although the number of WGMRES(m) cycles was usually slightly lower, the CPU time for WGMRES(m) was greater unless the restart length was short. Since ILU preconditioners often cluster eigenvalues, weighting—the effect of which is typically to shift the harmonic Ritz values from those obtained by GMRES(m)—appears to offer little benefit in this case. Our own experiments (see Example 5) confirm that for many systems right-preconditioned by ILU(0), WGMRES(m) requires a number of cycles that is similar to, or slightly greater than, that required by GMRES(m). However, WGMRES(m) may be of some benefit when the restart length is short or when the spectrum of the preconditioned matrix is not nicely clustered.
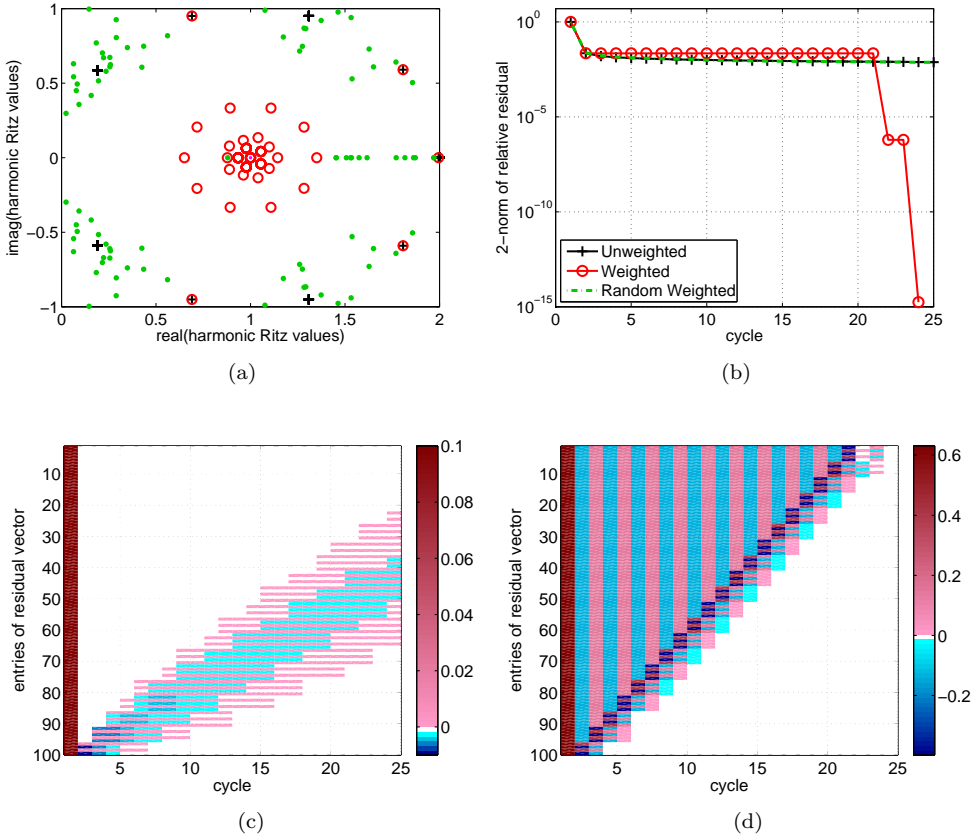
Fig. 3: (a) Harmonic Ritz values for GMRES(5) (black +), WGMRES(5) with (2) (red ○), and WGMRES(5) with D_{rand} (green dots) for a Jordan block with eigenvalue 1. The harmonic Ritz values are shown at the end of each of the 25 cycles. The eigenvalue at 1 is plotted as an orange dot. (b) Relative 2-norm residuals for the considered GMRES(5) variants. (The convergences curves of WGMRES(5) with D_{rand} and GMRES(5) are visually hard to distinguish.) (c) Entries of the residual vectors after each cycle of unweighted GMRES(5). (d) Entries of the residual vectors after each cycle of GMRES(5) with Essai's weighting.

3 The weighted Arnoldi algorithm

In this section we discuss two mathematically equivalent variants of the weighted Arnoldi algorithm for constructing the D-orthonormal basis required by WGMRES. Algorithm 2 is, to our knowledge, a new implementation.

3.1 Variants of the algorithm

The most straightforward implementation of the weighted Arnoldi algorithm replaces Euclidean inner products in a standard Arnoldi algorithm by D-inner products (see, e.g., [13,29]). With modified Gram–Schmidt orthogonalization (MGS)[2] the j-th iteration requires the computation of j D-inner products (see Algorithm 1). If D is a diagonal matrix like (2), the inner products can be efficiently implemented as $(\boldsymbol{v} \circ \boldsymbol{d})^H \boldsymbol{u}$ or $\boldsymbol{v}^H (\boldsymbol{d} \circ \boldsymbol{u})$, where $\circ$ represents the Hadamard product and $\boldsymbol{d}$ the vector of diagonal elements of D. Each step of Algorithm 1 is more expensive than a step of the Arnoldi algorithm in the Euclidean inner product because of the D-inner products and D-norms. However, Algorithm 1 can be used with a preconditioner in a straightforward manner.

As an alternative to computing D-inner products at each step of the weighted Arnoldi algorithm, we can apply the Arnoldi algorithm in the Euclidean inner product to the transformed matrix $\widetilde{A} = D^{\frac{1}{2}} A D^{-\frac{1}{2}}$ and starting vector $\widetilde{\boldsymbol{r}} = D^{\frac{1}{2}} \boldsymbol{r}$. Doing so gives matrices $\widetilde{V}_m$ and $\underline{\widetilde{H}_m}$ that satisfy

$$\widetilde{A}\widetilde{V}_m = \widetilde{V}_{m+1}\underline{\widetilde{H}_m}, \tag{9}$$

where $\underline{\widetilde{H}_m}$ is an upper Hessenberg matrix, $\widetilde{V}_m^H \widetilde{V}_m = I_m$, and the columns of $\widetilde{V}_m$ form a basis of $\mathcal{K}_m(\widetilde{A}, \widetilde{\boldsymbol{r}})$. Premultiplying both sides of (9) by $D^{-\frac{1}{2}}$ gives the Arnoldi decomposition $AV_m = V_{m+1}\underline{H_m}$, where $V_m = D^{-\frac{1}{2}}\widetilde{V}_m$ and $\underline{H_m} = \underline{\widetilde{H}_m}$. The columns of V_m are D-orthonormal and span $\mathcal{K}_m(A, \boldsymbol{r})$ since the columns of $\widetilde{V}_m$ span $\mathcal{K}_m(D^{\frac{1}{2}} A D^{-\frac{1}{2}}, D^{\frac{1}{2}}\boldsymbol{r})$.

The resulting algorithm with MGS orthogonalization is given in Algorithm 2. We note that a similar orthogonalization strategy was considered in the context of rounding error analysis in [26]. Additionally, Heyouni and Essai [17] considered the use of matrix square roots for enforcing D-orthogonality when solving systems with multiple right-hand sides, although they still computed D-inner products at each iteration of their weighted Arnoldi algorithm.

It does not seem feasible to use Algorithm 2 in this form with a preconditioner P. However, a (right) preconditioner can be incorporated by replacing the two-sided scaling for $\widetilde{A}$ in line 1 of this algorithm with a one-sided scaling $\widetilde{A}_1 = D^{\frac{1}{2}} A$, and line 5 with $\boldsymbol{w} = \widetilde{A}_1 P^{-1} D^{-1/2} \widetilde{\boldsymbol{v}}_k$ (and similarly for left preconditioning).

In the remainder of this manuscript, we use Algorithms 1 and 2 to refer to both the weighted Arnoldi variants and the corresponding WGMRES(m) methods. The meaning will be clear from the context. Although it is easiest

---

[2] When classical Gram–Schmidt orthogonalization (CGS) is used instead, nonstandard inner products can be replaced by Euclidean inner products at the expense of two matrix-vector products with D per cycle. One step of such a CGS version may be much cheaper than Algorithm 1, but without reorthogonalization there may be considerable differences in the convergence behavior compared to MGS. We therefore decided to stick with the MGS version in this paper.

Variants of the weighted Arnoldi algorithm:

Inputs: Matrix $A \in \mathbb{C}^{N \times N}$, diagonal positive definite weight matrix $D \in \mathbb{C}^{N \times N}$, vector $\boldsymbol{r} \in \mathbb{C}^N$, number of Arnoldi iterations m

Outputs: D-orthonormal Arnoldi vectors $\{\boldsymbol{v}_1 \ldots, \boldsymbol{v}_m\}$ of $\mathcal{K}_m(A, \boldsymbol{r})$ and upper Hessenberg matrix $\underline{H_m} = [h_{ij}] \in \mathbb{C}^{(m+1) \times m}$

Algorithm 1: Explicit D-inner products and MGS orthogonalization

1 $\boldsymbol{v}_1 = \boldsymbol{r}/\|\boldsymbol{r}\|_D$
2 **for** $k = 1, 2, \ldots, m$ **do**
3       $\boldsymbol{w} = A\boldsymbol{v}_k$
4   **for** $j = 1, 2, \ldots, k$ **do**
5            $h_{jk} = \boldsymbol{v}_j^* D\boldsymbol{w}$
6            $\boldsymbol{w} = \boldsymbol{w} - \boldsymbol{v}_j h_{jk}$
7   **end**
8       $h_{k+1,k} = \|\boldsymbol{w}\|_D$
9   **if** $h_{k+1,k} = 0$ **then**
10   Stop
11   **end**
12      $\boldsymbol{v}_{k+1} = \boldsymbol{w}/h_{k+1,k}$
13 **end**

Algorithm 2: Implicit D-inner products and MGS orthogonalization

1 $\widetilde{A} = D^{\frac{1}{2}} A D^{-\frac{1}{2}}$
2 $\boldsymbol{w} = D^{\frac{1}{2}} \boldsymbol{r}$
3 $\widetilde{\boldsymbol{v}}_1 = \boldsymbol{w}/\|\boldsymbol{w}\|_2$
4 **for** $k = 1, 2, \ldots, m$ **do**
5       $\boldsymbol{w} = \widetilde{A}\widetilde{\boldsymbol{v}}_k$
6   **for** $j = 1, 2, \ldots, k$ **do**
7            $h_{jk} = \widetilde{\boldsymbol{v}}_j^* \boldsymbol{w}$
8            $\boldsymbol{w} = \boldsymbol{w} - \widetilde{\boldsymbol{v}}_j h_{jk}$
9   **end**
10      $h_{k+1,k} = \|\boldsymbol{w}\|_2$
11   **if** $h_{k+1,k} = 0$ **then**
12   Stop
13   **end**
14      $\widetilde{\boldsymbol{v}}_{k+1} = \boldsymbol{w}/h_{k+1,k}$
15 **end**
16 $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m] = D^{-\frac{1}{2}} [\widetilde{\boldsymbol{v}}_1, \ldots, \widetilde{\boldsymbol{v}}_m]$

to monitor $\|\boldsymbol{r}^{(k)}\|_{D^{(k)}}$ in WGMRES(m), we measure instead the reduction of the residual $\|\boldsymbol{r}^{(k)}\|_2$ for fair comparison with GMRES(m).

3.2 Operation counts

In both variants of the Arnoldi algorithm discussed here, the number of matrix-vector products is m, and their computation requires $2m \times$ Nnz arithmetic operations, where Nnz is the number of nonzero elements in the matrix A or $\widetilde{A}$, respectively. Furthermore, the successive orthogonalization of m Krylov basis vectors requires $m(m + 1)/2$ inner products and vector updates of the form $\boldsymbol{w} = \boldsymbol{w} - \boldsymbol{v}_j h_{jk}$. One such vector update requires $2N$ arithmetic operations. Computing a single inner product requires $2N$ or $3N$ arithmetic operations in the unweighted or weighted case, respectively. The row and column[3] scaling

---

[3] As pointed out by one of the referees, the column scaling in Algorithm 2 can be eliminated, at the expense of storing m additional vectors, by computing each Arnoldi vector $\boldsymbol{v}_k = D^{-\frac{1}{2}} \widetilde{\boldsymbol{v}}_k$ as soon as $\widetilde{\boldsymbol{v}}_k$ is available. In this case, $\boldsymbol{w} = D^{-\frac{1}{2}} A D^{\frac{1}{2}} \widetilde{\boldsymbol{v}}_k = D^{-\frac{1}{2}} A \boldsymbol{v}_k$.

Table 1: Operation counts for a cycle of Algorithms 1 and 2 and GMRES-DR.

Algorithm 1	$2m \times \texttt{Nnz} + \frac{5}{2}Nm^2$
Algorithm 2	$2m \times \texttt{Nnz} + N(2m^2 + m + 2p + 3)$
GMRES-DR	$2m \times \texttt{Nnz} + N(2m^2 + 2m\ell)$

of the matrix A to form $\widetilde{A}$ and vector $\boldsymbol{r}$ to form $\widetilde{\boldsymbol{r}}$ in Algorithm 2 requires $3N + 2 \times \texttt{Nnz}$ operations, counting the computation of a square root or a division as a single arithmetic operation. Algorithm 2 also requires Nm multiplications to scale the basis vectors at the end of each cycle.

In Section 4 we will compare different variants of WGMRES(m) with GMRES-DR(m, ℓ) [23] on various numerical examples, but it easy to discuss the computational cost theoretically as well. In each cycle, GMRES-DR(m, ℓ) augments the Krylov basis computed by GMRES(m) with the Schur vectors corresponding to the ℓ smallest harmonic Ritz values. These Schur vectors, and the initial residual of the new cycle k, form the columns of a matrix $P_{\ell+1} \in \mathbb{C}^{(m+1) \times (\ell+1)}$, from which the first $\ell + 1$ basis vectors of the new cycle are obtained via $V_{\ell+1}^{(k)} = V_{m+1}^{(k-1)} P_{\ell+1}$. The dense matrix-matrix multiplication to obtain $V_{\ell+1}^{(k)}$ requires $2N(m+1)(\ell+1)$ arithmetic operations and so dominates the cost of augmenting the Krylov basis. The remaining columns of $V_{\ell+m}^{(k)}$ are then computed by the Arnoldi algorithm. Depending on the restart length m and the number of Schur vectors ℓ, one cycle of GMRES-DR(m, ℓ) may be more costly than a cycle of WGMRES(m).

We summarize the operation counts for a single cycle of Algorithms 1 and 2 and GMRES-DR in Table 1, letting $p = \texttt{Nnz}/N$ be the average number of nonzeros per row. Although the performance of each algorithm is machine dependent Algorithm 2, for which the cost of the weighted inner product is independent of the restart length, can become more efficient than Algorithm 1 when $2p < m^2/2 - m - 3$, i.e., when the average number of nonzeros per row is sufficiently small and the restart length is sufficiently large (or when reorthogonalization is used).

4 Numerical experiments

In this section we compare the two variants of WGMRES(m) with each other and with more established Krylov subspace methods. The first example comprises the five problems for which Essai [13] compared WGMRES(m) and GMRES(m) and are available from the University of Florida Sparse Matrix Collection [8]. Essai found that WGMRES(m) converged faster than GMRES(m) for all problems and so we report only results for Algorithms 1 and 2 and the more competitive GMRES-DR(m, ℓ) here. The motivations for choosing GMRES-DR(m, ℓ) for comparison are as follows: first, this method is very close in spirit to GMRES(m) and WGMRES(m); in particular, all these methods extract their approximations from the same Krylov spaces. Second,

Table 2: Size of, and number of nonzeros in, the matrices in Essai's problems.

Problem	N	Nnz
add20	2 395	13 151
bfwa782	782	7 514
fs_541_2	541	4 282
memplus	17 758	99 147
orsirr_1	1 030	6 858

the storage needs and orthogonalization costs of these algorithms are comparable. The last example compares Algorithm 1, unweighted GMRES(m), GMRES-DR(m, ℓ) and BICGSTAB over a large set of right-preconditioned test problems that is described below.

In the first example the right-hand side $\boldsymbol{b}$ is a random vector, to be consistent with Essai, while in the second $\boldsymbol{b}$ either comes with the matrix, or is generated randomly. We set $\boldsymbol{x}^{(0)} = \boldsymbol{0}$ and stop a method when $\|\boldsymbol{r}_j^{(k)}\|_2/\|\boldsymbol{b}\|_2$ falls below 10^{-10} or when 100 cycles are performed; the latter case is denoted by '—'. Iteration counts are given in the form $it_{\mathrm{out}}(it_{\mathrm{in}})$, where it_{out} is the number of cycles and it_{in} is the number of steps in the last cycle. In our notation GMRES-DR(m, ℓ) augments a Krylov subspace of dimension m with ℓ approximate Schur vectors so that that all GMRES variants require exactly m matrix-vector products with A in a cycle. We choose $\ell = 5, 10$ so that the cost of augmenting with Schur vectors is not too high.

Example 4 (Essai problems) *The first example comprises the five problems considered by Essai [13] (see Table 2 for problem details). We see from Table 3 that GMRES-DR converges in fewer cycles than WGMRES(m) for* add20, bfwa782 *and* fs_541_2. *Indeed, for* fs_541_2 *WGMRES with* $m = 40, 80$ *fails to converge within 100 cycles. For* memplus, *on the other hand, WGMRES(m) requires fewer cycles than GMRES-DR($m, 5$), except when* $m = 80$, *but more than GMRES-DR($m, 10$), except when* $m = 50$. *When applied to* orsirr_1, *WGMRES(m) converges faster than GMRES-DR($m, 5$) when* $m = 40, 50, 80$, *while GMRES-DR($m, 10$) does not converge within 100 cycles for any* m.

Essai found that for these problems WGMRES(m) consistently outperformed GMRES(m). However, WGMRES does not always perform as well as GMRES-DR, which also alters the harmonic Ritz values via deflation. Nevertheless, WGMRES is still competitive for problems like memplus *and* orsirr_1, *for which GMRES-DR can stagnate or converge slowly if the number of approximate Schur vectors* ℓ *in not well chosen. We note that in practice the optimal* ℓ *is usually unknown, while WGMRES has no parameters to choose.*

Comparing the two WGMRES variants, we observe that Algorithms 1 and 2 behave fairly similarly, in particular for add20 *and* bfwa782, *and that neither is consistently better. On this evidence it appears that efficiency and preconditioning should determine the most appropriate variant (since preconditioning with Algorithm 2 is not straightforward).*

Table 3: Number of cycles for Essai's problems.

Problem	m	GMRES-DR$(m, 5)$	GMRES-DR$(m, 10)$	Alg 1	Alg 2
add20	40	20(23)	18(20)	21(35)	21(35)
	50	14(44)	14(5)	15(32)	15(32)
	60	11(49)	11(26)	12(43)	12(43)
	70	10(36)	10(15)	11(12)	11(12)
	80	9(12)	8(69)	10(5)	10(5)
bfwa782	40	17(13)	9(2)	65(34)	69(24)
	50	11(34)	7(23)	34(26)	34(27)
	60	8(33)	6(18)	28(58)	28(58)
	70	7(42)	5(36)	22(5)	22(5)
	80	6(46)	4(70)	14(56)	14(56)
fs_541_2	40	35(27)	29(21)	—	—
	50	22(18)	20(49)	23(47)	23(42)
	60	18(49)	16(53)	25(60)	25(60)
	70	13(48)	12(62)	21(65)	22(9)
	80	10(78)	9(76)	—	—
memplus	40	60(40)	48(32)	50(24)	54(29)
	50	42(27)	33(41)	35(33)	33(2)
	60	33(9)	25(54)	28(31)	27(60)
	70	25(51)	21(50)	26(26)	26(37)
	80	20(51)	18(2)	20(49)	20(9)
orsirr_1	40	70(9)	—	61(35)	58(2)
	50	55(13)	—	46(46)	48(11)
	60	34(55)	—	35(41)	34(32)
	70	25(12)	—	28(39)	28(18)
	80	—	—	22(32)	23(8)

Example 5 (performance profile) *To get a general idea of whether WGM-RES is a practical method when preconditioners are used, we compare GMRES and Algorithm 1, as well as GMRES-DR and BICGSTAB, on a large number of problems from the University of Florida Sparse Matrix Collection. In contrast to Cao and Yu [6] we use right preconditioning which minimizes the residual.*

Our method of comparison is as follows. We first retrieve the 220 non-symmetric matrices A of sizes between 10^4 and 10^6 with at most 15 nonzero elements per row on average. We then apply sparse reverse Cuthill–McKee reordering as implemented in Matlab's **symrcm**. *Next we scale the columns of A to have unit Euclidean norm, followed by a scaling of the rows of A to unit norm. Our aim is to compute an ILU preconditioner with thresholding and pivoting via Matlab's* **ilu**. *For stability reasons we compute an ILU factorization of $A + \sigma I$, where $\sigma = 10^{-12}$ if all diagonal elements of A are zero, or $\sigma = 10^{-12} \max\{|a_{ii}|\}$ if some but not all diagonal elements a_{ii} of A zero, or $\sigma = 0$ otherwise. This procedure follows recommendations in [7]. We successively use a drop tolerance of $10^{-3}, 10^{-4}, \ldots, 10^{-8}$, and stop when the U factor of the factorization has a condition number below 10^{15}, so that it can be assumed numerically nonsingular. If this condition is not satisfiable with a drop tolerance of 10^{-8}, then the matrix A is skipped.*

We now run the Krylov methods GMRES(10), WGMRES(10), GMRES-DR(10,5), BICGSTAB on the test matrices. A method is marked as failed if more than 50 restart cycles or 250 BICGSTAB iterations are required to obtain a relative residual norm of 10^{-8}. Note that in our notation GMRES-DR(10,5) requires 10 matrix-vector products (MVP) per cycle, exactly like GMRES(10) and WGMRES(10), and BICGSTAB requires 2 MVP per iteration, so that the maximal number of MVP is 500 for all methods. If all methods fail on a matrix it is excluded from the test. Since the collection contains many singular matrices, related to eigenvalue or least-squares problems, only 109 out of the 220 retrieved matrices are finally included in our test.

The performance profile in Figure 4 allows us to compare the number of MVP needed for each method to converge across all test problems. More specifically, if for each linear system the performance ratio measures the number of MVP for the k-th method to converge to the number of MVP for the best performing method to converge, then the function $f_k(\alpha)$ measures the fraction of problems in the test set for which the performance ratio of method k is less than or equal to α. Thus, $\alpha = 1$ shows the fraction of problems for which the k-th method requires the fewest MVP of all methods. Also, $\lim_{\alpha \to \infty} f_k(\alpha)$ indicates the number of failures.

It is somewhat disappointing that WGMRES(10) is generally outperformed by GMRES(10). We note that GMRES(10) and WGMRES(10) fail on at least 16 matrices from our test set; this failure rate is considerably higher than for GMRES-DR and BICGSTAB. Overall, GMRES-DR(10,5) requires fewest MVP in general, and thereby outperforms all other methods under consideration. It is, however, less robust than BICGSTAB, the latter of which fails for 3 matrices only but typically requires the most MVP. To summarize, we believe that WGMRES should not be used in combination with preconditioners, although we are aware that for some examples it may perform satisfactorily.

5 Conclusions

The weighted GMRES variant presented by Essai has recently gained interest for solving linear systems. This method is justified by a heuristic that emphasizes large residual components via a weighted inner product. With the help of simple model problems we have given insight into how weighting affects the distribution of harmonic Ritz values, or how it affects entries in the residual vector after each cycle. For example, in one case where the harmonic Ritz values appeared in cyclic pairs on the spectral interval of a matrix, weighting had the effect of "randomizing" these harmonic Ritz values, thereby covering the spectral interval more evenly. This led to an improved convergence of WGMRES compared to the linear convergence observed for GMRES on the same example.

We presented two different implementations of weighted GMRES and compared their cost. Our numerical results suggest that the variants converge similarly, but an alternative to Essai's may require fewer arithmetic operations.
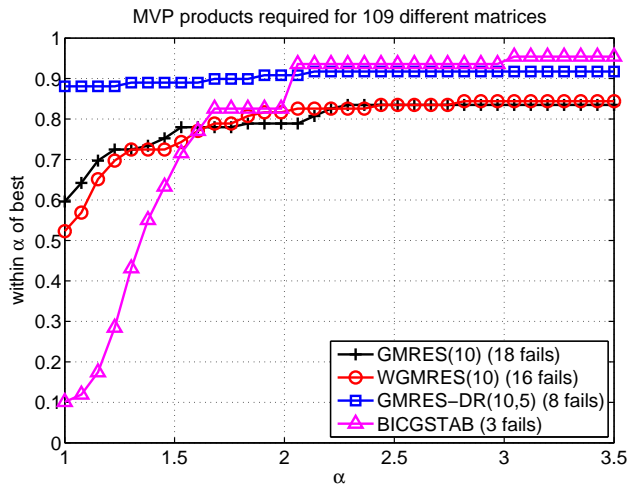
Fig. 4: Performance profile of matrix vector products required by various Krylov methods applied to 109 matrices from the University of Florida Sparse Matrix Collection with an ILU preconditioner.

When applied to unpreconditioned problems, WGMRES(m) can outperform GMRES(m). However, a test run with many matrices from the University of Florida Sparse Matrix Collection revealed, similarly to observations in [6], that weighted GMRES is typically outperformed by GMRES if a preconditioner is used. In addition, we compared these methods with other state-of-the-art Krylov methods like GMRES-DR (GMRES with deflated restarting) and BICGSTAB. GMRES-DR required fewest matrix-vector products, whereas BICGSTAB appeared to be the most robust method in our test, at the cost of requiring the most matrix-vector products. One advantage of WGMRES(m) over GMRES-DR(m,ℓ) is that there is no parameter ℓ to be chosen.

We find that, although weighted GMRES may outperform unweighted GMRES for some examples, in general this method is not competitive with other Krylov subspace methods like BICGSTAB or deflated GMRES, in particular when preconditioners are used.

References

1. Afanasjew, M., Eiermann, M., Ernst, O.G., Güttel, S.: A generalization of the steepest descent method for matrix functions. Electron. Trans. Numer. Anal. **28**, 206–222 (2008)

2. Akaike, H.: On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. Ann. Inst. Statist. Math. **11**, 1–16 (1959)
3. Arioli, M., Pták, V., Strakoš, Z.: Krylov sequences of maximal length and convergence of GMRES. BIT **38**, 636–643 (1998)
4. Baker, A.H., Jessup, E.R., Manteuffel, T.: A technique for accelerating the convergence of restarted GMRES. SIAM J. Matrix Anal. Appl. **26**, 962–984 (2005)
5. Brown, P.: A theoretical comparison of the Arnoldi and GMRES algorithms. SIAM J. Sci. Stat. Comput. **12**, 58–78 (1991)
6. Cao, Z.H., Yu, X.Y.: A note on weighted FOM and GMRES for solving nonsymmetric linear systems. Appl. Math. Comput. **151**, 719–727 (2004)
7. Chow, E., Saad, Y.: Experimental study of ILU preconditioners for indefininte matrices. J. Comput. Appl. Math. **86**, 387–414 (1997)
8. Davis, T.A., Hu, Y.: The University of Florida Sparse Matrix Collection. ACM Trans. Math. Softw. **38**, 1:1–1:25 (2011)
9. Duintjer Tebbens, J., Meurant, G.: Any Ritz value behavior is possible for Arnoldi and for GMRES with any convergence curve. SIAM J. Matrix Anal. Appl. **33**(3), 958–978 (2012)
10. Eiermann, M., Ernst, O.G.: Geometric aspects in the theory of Krylov subspace methods. Acta Numer. **10**, 251–312 (2001)
11. Eiermann, M., Ernst, O.G., Schneider, O.: Analysis of acceleration strategies for restarted minimal residual methods. J. Comput. Appl. Math. **123**, 261–292 (2000)
12. Embree, M.: The tortoise and the hare restart GMRES. SIAM Rev. **45**, 259–266 (2003)
13. Essai, A.: Weighted FOM and GMRES for solving nonsymmetric linear systems. Numer. Algorithms **18**, 277–292 (1998)
14. Forsythe, G.E., Motzkin, T.S.: Asymptotic properties of the optimum gradient method (abstract). Bull. Amer. Math. Soc. **57**, 183 (1951)
15. Goossens, S., Roose, D.: Ritz and harmonic Ritz values and the convergence of FOM and GMRES. Numer. Linear Algebra Appl. **6**, 281–293 (1997)
16. Greenbaum, A., Pták, V., Strakoš, Z.: Any nonincreasing convergence curve is possible for GMRES. SIAM J. Matrix Anal. Appl. **17**, 465–469 (1996)
17. Heyouni, M., Essai, A.: Matrix Krylov subspace methods for linear systems with multiple right-hand sides. Numer. Algorithms **40**, 137–156 (2005)
18. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, New York, NY (1990)
19. Jing, Y.F., Huang, T.Z.: Restarted weighted full orthogonalization method for shifted linear systems. Comput. Math. Appl. **57**, 1583–1591 (2009)
20. Liesen, J., Strakoš, Z.: Convergence of GMRES for tridiagonal Toeplitz matrices. SIAM J. Matrix Anal. Appl. **26**, 233–251 (2004)
21. Meijerink, J.A., van der Vorst, H.A.: An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Math. Comput. **31**, 148–162 (1977)
22. Morgan, R.B.: A restarted GMRES method augmented with eigenvectors. SIAM J. Matrix Anal. Appl. **16**, 1154–1171 (1995)
23. Morgan, R.B.: GMRES with deflated restarting. SIAM J. Sci. Comput. **24**, 20–37 (2002)
24. Najafi, H.S., Zareamoghaddam, H.: A new computational GMRES method. Appl. Math. Comput. **199**, 527–534 (2008)
25. Niu, Q., Lu, L., Zhou, J.: Accelerate weighted GMRES by augmenting error approximations. Int. J. Comput. Math. **87**, 2101–2112 (2010)
26. Rozložník, M., Tůma, M., Smoktunowicz, A., Kopal, J.: Numerical stability of orthogonalization methods with a non-standard inner product. BIT Numer. Math. **52**, 1035–1058 (2012)
27. Saad, Y.: Analysis of augmented Krylov subspace methods. SIAM J. Matrix Anal. Appl. **18**, 435–449 (1997)
28. Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**, 856–869 (1986)
29. Sarkis, M., Szyld, D.B.: Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. Comput. Methods Appl. Mech. Eng. **196**, 1612–1621 (2007)

30. Simoncini, V., Szyld, D.B.: On the occurence of superlinear convergence of exact and inexact Krylov subspace methods. SIAM Rev. **47**, 247–272 (2005)
31. Vecharynski, E., Langou, J.: The cycle-convergence of restarted GMRES for normal matrices is sublinear. SIAM J. Sci. Comput. **32**, 186–196 (2010)
32. Vecharynski, E., Langou, J.: Any admissible cycle-convergence behavior is possible for restarted GMRES at its initial cycles. Numer. Linear Algebra Appl. **18**, 499–511 (2011)
33. Wang, Z., Qi, J., Liu, C., Li, Y.: Weighted harmonic Arnoldi method for large interior eigenproblems. World Acad. Sci. Eng. Technol. **56**, 1476–1479 (2011)
34. Zavorin, I., O'Leary, D.P., Elman, H.: Complete stagnation of GMRES. Linear Algebra Appl. **367**, 165–183 (2003)
35. Zhong, B., Morgan, R.B.: Complementary cycles of restarted GMRES. Numer. Linear Algebra Appl. **15**, 559–571 (2008)
36. Zhong, B.J.: A product hybrid GMRES algorithm for nonsymmetric linear systems. J. Comput. Math. **23**, 83–92 (2005)
37. Zhong, H.X., Wu, G.: Thick restarting the weighted harmonic Arnoldi algorithm for large interior eigenproblems. Int. J. Comput. Math. **88**, 994–1012 (2011)