Acceleration Techniques for Restarted GMRES

Manmeet Bhabra

May, 2020

1 Introduction

Iterative methods have become quite popular in solving large sparse system of linear equations of the form

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{A} \in \mathbb{R}^{m\times m}, \quad \boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^m \tag{1}$$

In the class of iterative methods, Krylov subspace methods are some of the most popular and, among them, the generalized minimum residual (GMRES) algorithm is one of the most used method for non-symmetric matrices. A central difficulty with the GMRES method is with scaling in computational work and storage as the size of the Krylov subspace used increases. For large matrices $\boldsymbol{A}$, the standard GMRES method becomes impractical as it becomes infeasible to store the larger Krylov subspace needed to find the solution. Due to the this, the restarted GMRES method (GMRES(l)) is used where, after an l-dimensional Krylov subspace is formed, it is used to find an approximate solution. The method is then restarted using this approximate solution as the new initial guess, and the process is repeated until convergence.

A downside of restarted GMRES is the process of completely discarding old Krylov subspaces on every restart cycle. This loss of information from earlier subspaces can result in slower convergence and, in some cases, stalling, where no drop in the residual is seen after a restart cycle. To tackle this, several GMRES variants have been proposed, with the general idea of retaining more information between restart cycles and, in turn, accelerate convergence. In this work, we investigate the GMRES-E (GMRES with eigenvectors) and LGMRES ("Loose" GMRES) algorithms which were formualted with the central purpose of accelerating the restarted GMRES algorithm.

2 Background

2.1 GMRES

The GMRES algorithm is a projection method where the approximate solution is searched for in a Krylov subspace. Given an initial guess $\boldsymbol{x}_0$ and the residual associated with that guess $\boldsymbol{r}_0$, the l-th Krylov subspace is given as:

Algorithm 1: The Restarted GMRES

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, the initial guess $\boldsymbol{x}_0$ and l (the Krylov subspace dimension)
Result: The approximate solution $\boldsymbol{x}_l$ to the system of equations
1 *Compute* $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$
2 *Generate the Arnoldi basis and matrix* $\tilde{\boldsymbol{H}}_l$ *using Arnoldi's method starting with* $\frac{\boldsymbol{r}_0}{\|\boldsymbol{r}_0\|}$
3 $\boldsymbol{y}_l = \arg\min_{\boldsymbol{y}} \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_1 - \tilde{\boldsymbol{H}}_l \cdot \boldsymbol{y} \right\|_2$
4 $\boldsymbol{x}_l = \boldsymbol{x}_0 + \boldsymbol{V}_l \cdot \boldsymbol{y}_l$
5 *If* $\boldsymbol{r}_l$ *is small, return* $\boldsymbol{x}_l$*. Else, set* $\boldsymbol{x}_0 = \boldsymbol{x}_m$ *and GoTo* 1

$$\mathcal{K}_l(\boldsymbol{A}, \boldsymbol{r}_0) = \langle \boldsymbol{r}_0, \boldsymbol{A}\boldsymbol{r}_0, ..., \boldsymbol{A}^{l-1}\boldsymbol{r}_0 \rangle \tag{2}$$

Arnoldi's method is then used to get an orthogonal basis for the subspace, which we will denote using the matrix $\boldsymbol{V}_l$. We proceed next to search for the solution to the linear system in the affine space $\boldsymbol{x}_0 + \mathcal{K}_l$. The candidate solution that is optimal in the sense that it minimizes the residual norm is given as:

$$\begin{aligned} \boldsymbol{x}_l &= \boldsymbol{x}_0 + \boldsymbol{V}_l \cdot \boldsymbol{y}_l \\ \boldsymbol{y}_l &= \arg\min_{\boldsymbol{y}} \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_1 - \tilde{\boldsymbol{H}}_l \cdot \boldsymbol{y} \right\|_2 \end{aligned} \tag{3}$$

where $\tilde{\boldsymbol{H}}_l \in \mathbb{R}^{(l+1)\times l}$ is an upper Hessenberg matrix.

A key challenge that arises with the standard GMRES method is that of memory use. If m (the size of the matrix $\boldsymbol{A}$) is large, storing the orthonormal basis becomes impractical for large subspace dimensions. One approach to address this issue is through restarting where, after an l-dimensional Krylov subspace is created, a candidate solution is found by minimizing the residual. If this candidate solution has an error that is small enough, then it is returned. If not, the candidate is used as a new initial guess and the process repeats. A high level pseudocode of the restarted GMRES algorithm, adapted from [1], is given in Algorithm (1).

2.2 GMRES-E

The GMRES-E algorithm aims to accelerate convergence by augmenting the subspace in which an approximate solution is searched for. In particular, the central aim of this method is to augment the subspace with approximate eigenvectors that neutralize the effect of small eigenvalues and in turn improve the convergence of the algorithm. For further details about this method, we refer the reader to [2].

For matrices $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ that are nearly normal (with a spectral decomposition $\boldsymbol{A} = \boldsymbol{Z} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{Z}^{-1}$), the convergence of GMRES can be shown to be similar to that of the Conjugate Gradient method and given by the following (in which the initial guess $\boldsymbol{x}_0$ is assumed to be the zero vector):

$$\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq 2 \|\boldsymbol{Z}\| \left\|\boldsymbol{Z}^{-1}\right\| \left(1 - \frac{1}{\sqrt{\kappa} + 1}\right)^m \tag{4}$$

If the eigenvalues are arranged in increasing order (in absolute value), then κ in the above bound is given as $\kappa = \frac{|\lambda_n|}{|\lambda_1|}$. It is clear form equation (4) that the convergence rate deteriorates if the matrix in question has both large and very small eigenvalues resulting in a large κ value.

This convergence bound can change drastically if the subspace used in the residual minimization includes some eigenvectors. Consider, once again, the matrix $\boldsymbol{A}$ from before along with its spectral decomposition. Furthermore, consider the k eigenvectors $\{\boldsymbol{z}_1, ..., \boldsymbol{z}_k\}$ associated with the k *smallest* eigenvalues of $\boldsymbol{A}$. Say now that the minimum residual solution $\hat{\boldsymbol{x}}$ to the linear system of equations is searched for in the subspace $\langle \boldsymbol{b}, \boldsymbol{Ab}, ..., \boldsymbol{A}^{m-1}\boldsymbol{b}, \boldsymbol{z}_1, ..., \boldsymbol{z}_k \rangle$. The residual of the solution can then be shown to converges as:

$$\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq 2 \|\boldsymbol{Z}\| \left\|\boldsymbol{Z}^{-1}\right\| \left(1 - \frac{1}{\sqrt{\kappa_e} + 1}\right)^m \tag{5}$$

where $\kappa_e = \frac{|\lambda_n|}{|\lambda_{k+1}|}$. The result in equation (5) reveals a clear source of improvement to the standard GMRES method. If a matrix has a handful of very small eigenvalues that deteriorate convergence of the standard GMRES algorithm, the influence of these eigenvalues can be removed by augmenting the Krylov subspace with their corresponding eigenvectors.

2.2.1 Implementation Details

The challenge now comes to finding the relevant eigenvectors of $\boldsymbol{A}$, a problem that is hard in and of itself. Fortunately, information concerning the eigenvalues and eigenvectors of $\boldsymbol{A}$ is available during GMRES due to Arnoldi's method when generating the orthogonal vectors to span the Krylov subspace.

The standard GMRES-E restart iteration starts with a guess $\boldsymbol{x}_0$ (which gives an initial residual $\boldsymbol{r}_0$) and a set of k approximate eigenvectors $\{\boldsymbol{z}_k\}$ of $\boldsymbol{A}$ that will be used to augment the l-dimensional Krylov subspace $\mathcal{K}_l(\boldsymbol{r}_0, \boldsymbol{A})$. Thus, the dimension of the subspace in which the approximate solution $\hat{\boldsymbol{x}}$ will be searched for is of dimension $s = l + k$. GMRES-E proceeds just as in the standard GMRES algorithm where Arnoldi's method is used first to compute an orthogonal basis for the subspace. Denote by $\boldsymbol{W} \in \mathbb{R}^{m \times s}$ the vector whose first l vectors are the orthonormal Arnoldi vectors $\boldsymbol{q}_i$ for the l-dimensional Krylov subspace and the subsequent k vectors are the approximate eigenvectors:

$$\boldsymbol{W} = [\boldsymbol{Q}_l \;\; \boldsymbol{Z}_k], \quad \boldsymbol{W} \in \mathbb{R}^{m \times s},\; \boldsymbol{Q}_l \in \mathbb{R}^{m \times l},\; \boldsymbol{Z}_k \in \mathbb{R}^{m \times k} \tag{6}$$

Arnoldi's method is used first to obtain the orthogonal vectors for the Krylov subspace $\mathcal{K}_l(\boldsymbol{r}_0, \boldsymbol{A})$ (just as in the standard GMRES method) to obtain $\boldsymbol{Q}_l$ and a matrix $\boldsymbol{Q}_{l+1} \in \mathbb{R}^{m \times (l+1)}$. In the GMRES-E approach, the orthogonalization process is continued with the vectors $\boldsymbol{A} \cdot \boldsymbol{z}_k$. The net result then of the whole Arnoldi process is the following:

$$\boldsymbol{A} \cdot \boldsymbol{W} = \tilde{\boldsymbol{Q}} \cdot \tilde{\boldsymbol{H}} \tag{7}$$

$$\tilde{\boldsymbol{Q}} = [\boldsymbol{Q}_{l+1} \;\; \boldsymbol{Q}_k], \quad \tilde{\boldsymbol{Q}} \in \mathbb{R}^{m \times (s+1)}, \quad \tilde{\boldsymbol{H}} \in \mathbb{R}^{(s+1) \times s} \tag{8}$$

where $\tilde{\boldsymbol{H}}$ is upper-Hessenberg. Note that in equation (7) the matrix $\boldsymbol{W}$ is not orthonormal. The approximate solution $\hat{\boldsymbol{x}}$ is then searched for in the space range($\boldsymbol{W}$), $\hat{\boldsymbol{x}} = \boldsymbol{W}\boldsymbol{y}$, by computing the optimal $\boldsymbol{y}$ that minimizes the residual norm just as in the standard GMRES method.

The computation of the approximate eigenvectors for $\boldsymbol{A}$ then follows, where a Rayleigh-Ritz procedure is used. We search for approximate eigenvectors in the subspace range($\boldsymbol{W}$) by solving the following:

$$\begin{aligned} \boldsymbol{A} \cdot [\boldsymbol{W}\boldsymbol{g}_i] - \theta_i \cdot [\boldsymbol{W}\boldsymbol{g}_i] &\perp \text{range}(\boldsymbol{W}) \\ \boldsymbol{W}^T\boldsymbol{A}\boldsymbol{W} \cdot \boldsymbol{g}_i &= \theta_i \cdot \boldsymbol{W}^T\boldsymbol{W} \cdot \boldsymbol{g}_i \end{aligned} \tag{9}$$

which is a generalized eigenvalue problem. A modification that can be made to get good approximations for eigenvalues close to zero (and their corresponding eigenvectors) is found by solving the modified problem:

$$\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{W} \cdot \boldsymbol{g}_i = \frac{1}{\theta_i} \cdot \boldsymbol{W}^T\boldsymbol{W} \cdot \boldsymbol{g}_i \tag{10}$$

Solving equation (10) thus yields an approximation for the eigenvectors of the form $\boldsymbol{z}_k = \boldsymbol{W}\boldsymbol{g}_k$. These are subsequently used to augment the Krylov subspace in the next restart iteration.

Therefore, at every restart iteration, GMRES-E retains some information between restarts in the form of the approximate eigenvectors of the matrix. As restart iterations progress, and these approximations become better and better, the faster convergence in the residual starts to be observed as predicted by equation (5). An important point to note however is that GMRES-E works best for certain problems than others. In particular, for problems that are close to normal and where convergence is hampered by a few eigenvalues, GMRES-E will perform well. However, as problems become highly nonnormal or for cases where solving the eigenvalue problem at each step becomes computationally expensive, the benefits of the method deteriorate. An overview of the pseudocode for the GMRES-E algorithm is given in Algorithm (2).

2.3 LGMRES

Consider the standard restarted GMRES method using an l-dimensional Krylov subspace. The residual after i restart cycles can be denoted as $\boldsymbol{r}_i$ (this residual was thus generated using a total of $m \times i$ iterations). The angle between the residual between consecutive restart cycles, $\angle(\boldsymbol{r}_i, \boldsymbol{r}_{i+1})$, is typically referred to as the *sequential* angles, and the angle between every other residual, $\angle(\boldsymbol{r}_{i-1}, \boldsymbol{r}_{i+1})$, can be referred to as the *skip* angles. The process of discarding prior Krlov subspaces between successive restarts in the standard GMRES method has been shown in many cases to slow convergence. This primarily is a result of the fact that orthogonality with the previously discarded subspaces is not maintained, an issue that can result in slow convergence and in some cases even stalling. Moreover, in several cases with slow convergence, the lack of orthogonality with earlier subspaces results in a distinctive pattern being observed where the residual vectors in *every other* restart cycle point in nearly the same direction, that is, the skip angles $\angle(\boldsymbol{r}_{i-1}, \boldsymbol{r}_{i+1})$ are small. LGMRES uses this finding as its primary

Algorithm 2: The GMRES-E Algorithm

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, $\{\boldsymbol{z}_i\}$ (the set of k eigenvector approximations), the initial guess $\boldsymbol{x}_0$ and l (the Krylov subspace dimension)

Result: The approximate solution $\boldsymbol{x}_l$ to the system of equations

1 *Compute* $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$ *and* $\boldsymbol{q}_1 = \frac{\boldsymbol{r}_0}{\|\boldsymbol{r}_0\|}$

/* Arnoldi's Method: */

2 **for** $j = 1, 2, \ldots, l$ **do**

3 $\quad \boldsymbol{w}_j = \boldsymbol{A}\boldsymbol{q}_j$

4 $\quad$ *Use an orthogonalization process (ex. Modified Gram − Schmidt) to get* $\boldsymbol{q}_{j+1}$ *and entries in* $\tilde{\boldsymbol{H}}$ *from* $\boldsymbol{w}_j$

5 **end**

/* Augment the subspace using the approximate eigenvectors */

6 **for** $j = 1, 2, \ldots, k$ **do**

7 $\quad \boldsymbol{w}_j = \boldsymbol{A}\boldsymbol{z}_j$

8 $\quad$ *Use an orthogonalization process to get* $\boldsymbol{q}_j$ *and entries in* $\tilde{\boldsymbol{H}}$ *from* $\boldsymbol{w}_j$

9 **end**

/* Solve the least squares problem using the augmented subspace for the approximate solution */

10 $\boldsymbol{y}_l = \arg\min_{\boldsymbol{y}} \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_1 - \tilde{\boldsymbol{H}} \cdot \boldsymbol{y} \right\|_2$

11 $\boldsymbol{x}_l = \boldsymbol{x}_0 + \boldsymbol{W} \cdot \boldsymbol{y}_l$

/* Form the new approximate eigenvectors */

12 *Solve* $\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{W} \cdot \boldsymbol{g}_i = \frac{1}{\theta_i} \cdot \boldsymbol{W}^T\boldsymbol{W} \cdot \boldsymbol{g}_i$ *for the new approximate eigenvectors* $\boldsymbol{z}_i = \boldsymbol{W}\boldsymbol{g}_i$

/* Return solution or restart if necessary */

13 *If* $\boldsymbol{r}_l$ *is small, return* $\boldsymbol{x}_l$*. Else, set* $\boldsymbol{x}_0 = \boldsymbol{x}_m$ *and GoTo* 1

motivation. In essence, the method is based on the principle that the small angle between every other restart cycle can be seen as an indication that if some form of orthogonality to earlier subspaces were maintained, convergence could be accelerated. For further details about this method beyond what is described next, we refer the reader to [3, 4].

LGMRES looks to try to impose orthogonality to earlier subspaces and, in turn, increase the skip angles $\angle(\boldsymbol{r}_{i-1}, \boldsymbol{r}_{i+1})$ by following along the simple idea from GMRES-E of augmenting the traditional Krylov subspace used in each restart iteration by appending some vectors. Consider the following error at the i^{th} restart iteration

$$\boldsymbol{e}_i = \boldsymbol{x}_e - \boldsymbol{x}_i, \tag{11}$$

where $\boldsymbol{x}_e$ is the exact solution. Suppose now that, at the next restart iteration, the approximate solution is searched for in the following affine space:

$$\boldsymbol{x}_{i+1} \in \boldsymbol{x}_i + \mathcal{M}, \quad \mathcal{M} = \mathcal{K}_l(\boldsymbol{A}, \boldsymbol{r}_i) \cup span(\boldsymbol{e}_i). \tag{12}$$

It is clear that, in this space, the exact solution can be found (thus, if the exact error were to be augmented to the space, the linear system would be exactly solved). LGMRES follows this idea by augmenting the subspace with approximations of the error between restart cycles defined as

$$\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{x}_{i-1} \tag{13}$$

LGMRES uses the k previous error approximations $(\boldsymbol{z}_j, j = i-k+1, i-k, ..., i)$ to augment the subspace used to find the approximate solution at the $i+1$ restart iteration given by $\mathcal{M} = \mathcal{K}_l(\boldsymbol{A}, \boldsymbol{r}_i) \cup span(\{\boldsymbol{z}_j\}_{j=(i-k+1):i})$. It is important to note also that the error approximation $\boldsymbol{z}_i \in \mathcal{K}_l(\boldsymbol{A}, \boldsymbol{r}_{i-1})$. Therefore, including these vectors to augment subspaces with allows in a way to retain information from previous Krylov subspaces when searching for solutions (and some form of orthogonality to earlier subspaces is maintained).

LGMRES follows GMRES-E in a similar approach in how Arnoldi's method is used to get an orthogonal representation of the augmented subspace. Consider the i^{th} restart cycle and let $\boldsymbol{Z}_k \in \mathbb{R}^{m\times k}$ hold the k previous error approximation vectors. Furthermore, let $\boldsymbol{Q}_l \in \mathbb{R}^{m\times l}$ hold the orthonormal Arnoldi vectors $\boldsymbol{q}_i$ for the l-dimensional Krylov subspace $\mathcal{K}_l(\boldsymbol{A}, \boldsymbol{r}_{i-1})$. Then, if $\boldsymbol{W} \in \mathbb{R}^{m\times(l+k)}$, $\boldsymbol{W} = [\boldsymbol{Q}_l \;\; \boldsymbol{Z}_k]$, Arnoldi's method when used in LGMRES produces:

$$\boldsymbol{A} \cdot \boldsymbol{W} = \tilde{\boldsymbol{Q}} \cdot \tilde{\boldsymbol{H}} \tag{14}$$

$$\tilde{\boldsymbol{Q}} = [\boldsymbol{Q}_{l+1} \;\; \boldsymbol{Q}_k], \quad \tilde{\boldsymbol{Q}} \in \mathbb{R}^{m\times(s+1)}, \quad \tilde{\boldsymbol{H}} \in \mathbb{R}^{(s+1)\times s}, \quad s = (l+k) \tag{15}$$

where $\tilde{\boldsymbol{H}}$ is upper Hessenberg. Searching for the optimal solution $\boldsymbol{x}_l \in span(\boldsymbol{W})$ that minimizes the residual then proceeds just as in the standard GMRES method.

We once again note that LGMRES performs well for certain problems. In conditions where skip angles are small, LGMRES provides space for acceleration. In cases where GMRES itself stalls, however, LGMRES typically will be unable to resolve this. Fortunately, LGMRES is able to provide this speed-up with effectively no increase in computational cost relative to the standard restarted GMRES method, making it therefore an attractive algorithm to consider. A general overview of the LGMRES algorithm is provided in Algorithm (3).

3 Numerical Results

In this section, we perform several numerical tests to verify the theoretical properties of the methods presented. In all cases, one of the key metrics used for comparisons is the number of restart cycles needed to convergence. It should be noted that computational time/complexity was not included for comparisons. This was primarily due to the fact that all solvers (including the standard GMRES method) were implemented from scratch and these implementations were not optimized effectively to allow for a fair comparison. For instance, in the GMRES method, several methods perform a running QR factorization as the Krylov subspace increases in order to make the least squares solve more efficient, however this was not done in our case. Similarly, for GMRES-E, the formation of the Rayleigh-Ritz eigenvalue problem can be optimized using upper triangular matrices, however this was not considered as well for simplicity reasons.

Presented first in the following section are separate numerical results for simple cases for the GMRES-E and LGMRES method. Following this, several matrices from the University of Florida Sparse Matrix Collection have been used to see how these methods perform on practical, large-scale, problems. Furthermore, for the standard restarted GMRES method, the notation GMRES(l) has been used to denote a solver in which an l-dimensional subspace is used between restarts. Similarly, GMRES-E(l, k) and LGMRES(l, k) solvers correspond to cases where an l-dimensional Krylov subspace is augmented with k eigenvectors or error approximations.

Algorithm 3: The LGMRES Algorithm

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, $\{\boldsymbol{z}_i\}$ (the set of k error approximations from previous restart cycles), the initial guess $\boldsymbol{x}_0$ and l (the Krylov subspace dimension)

Result: The approximate solution $\boldsymbol{x}_l$ to the system of equations

1 *Compute* $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$ *and* $\boldsymbol{q}_1 = \frac{\boldsymbol{r}_0}{\|\boldsymbol{r}_0\|}$

/* Arnoldi's Method: */

2 **for** $j = 1, 2, ..., l$ **do**

3 $\quad \boldsymbol{w}_j = \boldsymbol{A}\boldsymbol{q}_j$

4 $\quad$ *Use an orthogonalization process (ex. Modified Gram − Schmidt) to get* $\boldsymbol{q}_{j+1}$ *and entries in* $\tilde{\boldsymbol{H}}$ *from* $\boldsymbol{w}_j$

5 **end**

/* Augment the subspace using the prior error approximations */

6 **for** $j = 1, 2, ..., k$ **do**

7 $\quad \boldsymbol{w}_j = \boldsymbol{A}\boldsymbol{z}_j$

8 $\quad$ *Use an orthogonalization process to get* $\boldsymbol{q}_j$ *and entries in* $\tilde{\boldsymbol{H}}$ *from* $\boldsymbol{w}_j$

9 **end**

/* Solve the least squares problem using the augmented subspace for the approximate solution */

10 $\boldsymbol{y}_l = \arg\min_{\boldsymbol{y}} \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_1 - \tilde{\boldsymbol{H}} \cdot \boldsymbol{y} \right\|_2$

11 $\boldsymbol{z}_l = \boldsymbol{W} \cdot \boldsymbol{y}_l$ // The error approximation for this restart iteration

12 $\boldsymbol{x}_l = \boldsymbol{x}_0 + \boldsymbol{z}_l$

/* Return solution or restart if necessary */

13 *If* $\boldsymbol{r}_l$ *is small, return* $\boldsymbol{x}_l$. *Else, set* $\boldsymbol{x}_0 = \boldsymbol{x}_m$ *and GoTo* 1

3.1 GMRES-E

We begin by looking at a simple numerical case to verify the theoretical properties of the GMRES-E method. Consider a bi-diagonal matrix $\boldsymbol{A} \in \mathbb{R}^{1000\times1000}$ whose main diagonal entries are $1, 2, ..., 1000$ and whose super-diagonal entries are all 0.1. Furthermore, the right hand side matrix is taken to have all entries 1.0 and the initial guess $\boldsymbol{x}_0$ is the zero vector.

In Figure (1) we display the results obtained using the standard restarted GMRES algorithm and the GMRES-E algorithm. The GMRES(25) and GMRES-E(21,4) were used in the study (therefore, the subspaces used to find the approximate solution for both methods was the same with a dimension of 25). It can be seen that for this case, the GMRES-E method does indeed accelerate convergence as its residual drops almost twice as fast. Shown also in Figure (1) is the convergence in the approximate eigenvalues and eigenvectors generated from the Rayleigh-Ritz procedure. It can be seen that in the first few restart iterations (about the first 5), when the approximate eigenvectors are still quite off, no real acceleration in the convergence rate is observed. However, after this point when the eigenvector estimates become more accurate and their relative error decreases by at least one order of magnitude, the faster convergence is observed.

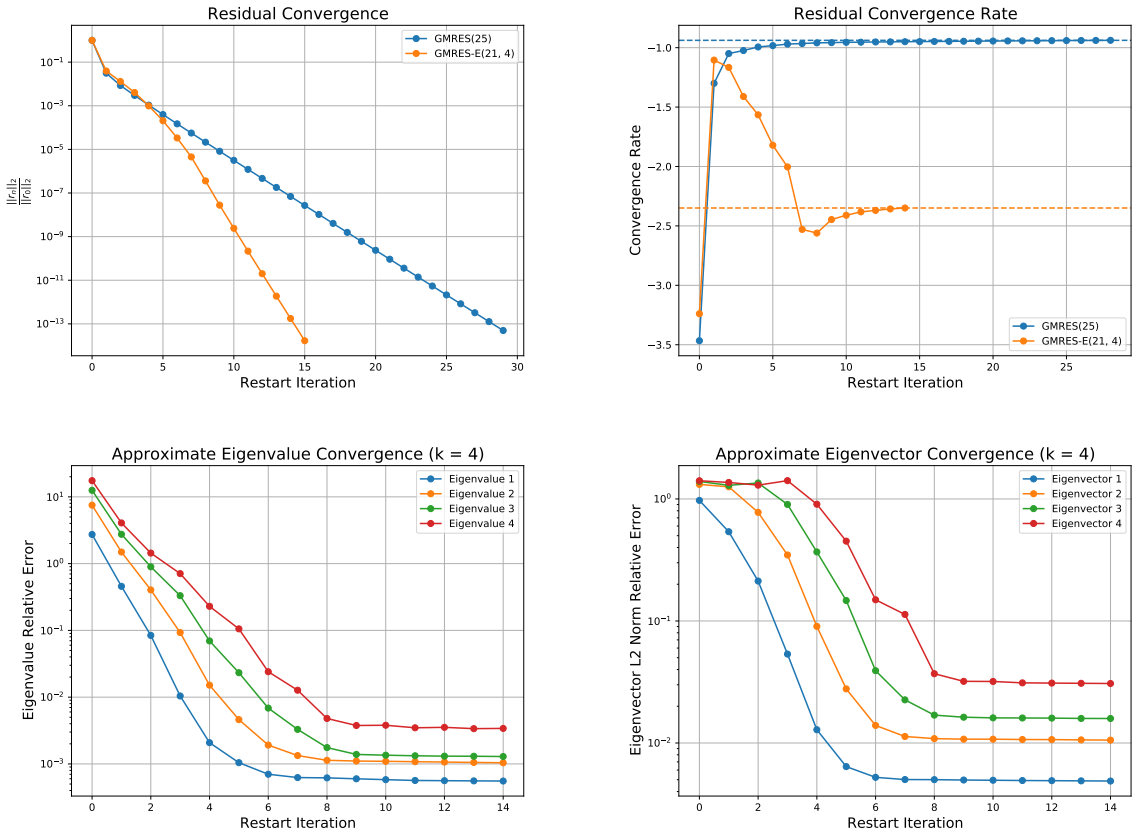

Figure 1: The residual convergence for the GMRES and GMRES-E method for the sample case (top left). The spectral convergence rate (i.e. the slope on the semilog plot) of the residual (top right). The convergence in the eigenvalue (bottom left) and eigenvector (bottom right) estimates.

3.2 LGMRES

We consider in this section a simple case to verify the theoretical properties of the LGMRES algorithm. The system of equations used was from a problem labeled "add20" (a circuit simulation problem) from the University of Florida Sparse Matrix Collection. The right hand side vector was taken to be a vector of ones and the initial guess was a vector of zeros. Figure (2) shows the convergence properties of the standard restarted GMRES and LGMRES algorithms. To perform a fair comparison, the same subspace size was used to search for the optimal solution as the GMRES(30) and LGMRES(26,4) solvers were considered. It can be seen that LGMRES is indeed able to accelerate convergence as it uses approximately half as many restart iterations. Moreover, the slow convergence in GMRES indeed accompanies a small skip angle $\angle(\boldsymbol{r}_{i+1}, \boldsymbol{r}_{i-1})$ observed in the residuals. It can be seen that LGMRES, through augmenting the subspace with information of prior Krylov subpsaces through the error approximations, is able to significantly increase the skip angles of the residuals and allow for the theoretically predicted faster convergence.

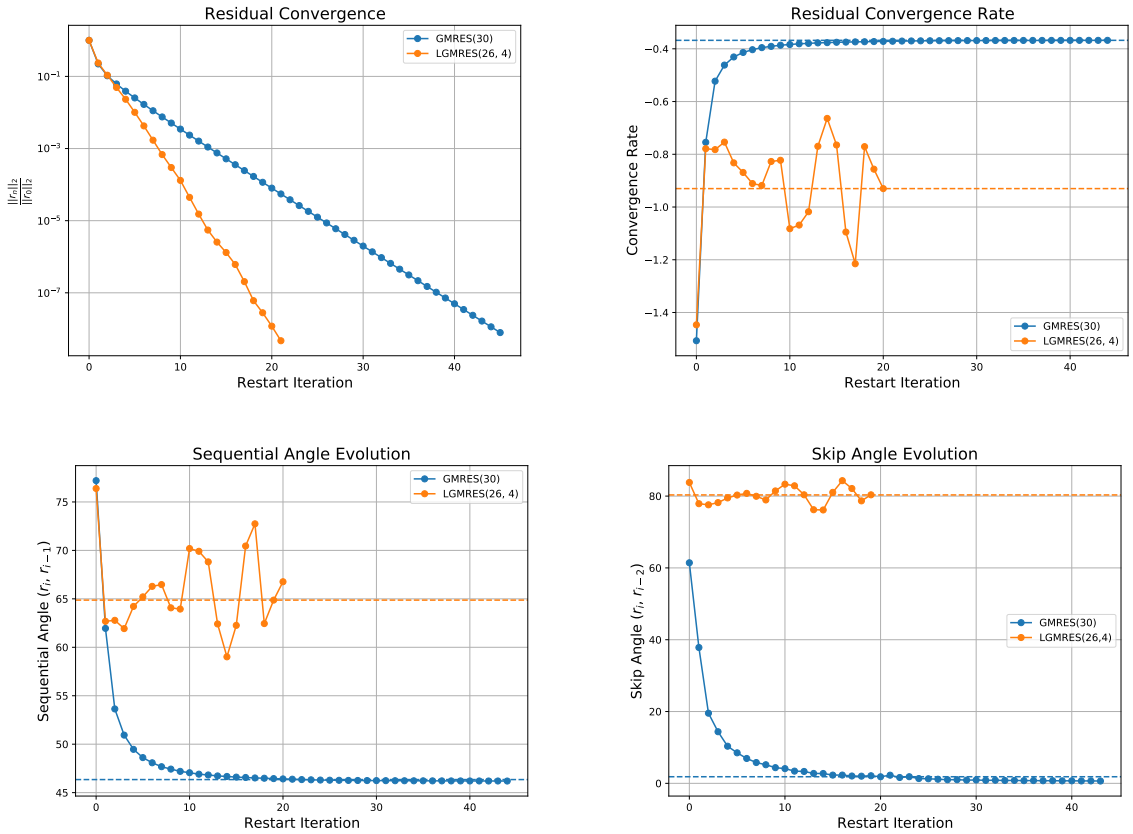

Figure 2: The residual convergence for the GMRES and LGMRES method for the sample case (top left) and its spectral convergence rate (top right). The evolution of the sequential angle of the residuals for the GMRES and LGMRES solve (botom left). Shown in the dashed line is the median sequential angle. Similarly, the evolution of the skip angle of the residuals (bottom right).

3.3 Experimental Results

In this section, we study the performance of the standard restarted GMRES, GMRES-E and LGMRES algorithm together on several test cases taken from the University of Florida Sparse Matrix Collection. For all cases, the GMRES(30) solver was used as a baseline to compare to the GMRES-E(26, 4) and LGMRES(26, 4) solvers.

For each case, Table (1) summarizes key theoretical properties that dictate how well the GMRES-E and LGMRES method perform as accelerators. Shown first are the median sequential angles, $\angle(\boldsymbol{r}_i, \boldsymbol{r}_{i-1})$, and skip angles, $\angle(\boldsymbol{r}_{i+1}, \boldsymbol{r}_{i-1})$, obtained from the GMRES(30) runs. Additionally, to predict the performance of the GMRES-E method, two metrics are listed for each case. The "normality metric" is a measure of how normal the matrix of interest is. It is computed by first forming the matrix $\boldsymbol{D} = \boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{A}\boldsymbol{A}^T$. The normality metric was then defined as:

$$N = \frac{\|\boldsymbol{D}\|_F^2}{\text{size}(\boldsymbol{D})} \tag{16}$$

The metric is therefore just a simple measure of the size of the average entry in the matrix

squared. Therefore, a small value indicates a matrix that is close to normal, while a high value indicates the opposite. As discussed in Section 2.2, we expect GMRES-E to perform best for those matrices that are close to normal. The next metric is denoted as the κ-ratio and is a measure to see if a few eigenvalues are indeed impeding convergence. Let the eigenvalues, sorted in increasing absolute value, be $\lambda_1, \lambda_2, ..., \lambda_n$. Motivated by the discussion in Section 2.2, we let the parameter κ_0 be defined as $\kappa_0 = \frac{|\lambda_n|}{|\lambda_1|}$. Furthermore, if k eigenvectors plan to be augmented to the subspace in the LGMRES method, we let $\kappa_k = \frac{|\lambda_n|}{|\lambda_{k+1}|}$. The κ-ratio is then given as:

$$\kappa\text{-ratio} = \frac{\kappa_k}{\kappa_0} = \frac{|\lambda_1|}{|\lambda_{k+1}|} \tag{17}$$

If there are a few eigenvalues (we consider the k-smallest) which are much smaller relative to all others, the κ-ratio will be small, and this indicates that being able to remove the influence of these eigenvalues by approximating their corresponding eigenvectors in the GMRES-E method should in theory accelerate convergence by a large amount. On the contrary, if the value is large and close to 1, it is an indication that there may not be a handful of small eigenvalues affecting convergence and so the GMRES-E algorithm may not perform as well.

Figure (3) shows the acceleration of each solver relative to GMRES(30). All cases were run until the relative residual, $\frac{\|r\|}{\|r\|_0}$, was less than 10^{-11}. In Figure (3), the ratio of the number of restart cycles needed for GMRES(30) to solve the system to the number of restarts needed by each GMRES variant is shown. For GMRES-E, we see that indeed the cases that are most "normal" and have the smaller κ-ratio, such as raefsky1 and cavity05, result in the largest speed-up. Additionally, LGMRES predictably provides a larger speed-up to cases with a smaller skip angle. Residual convergence plots are shown as well for certain cases in Figure (4).

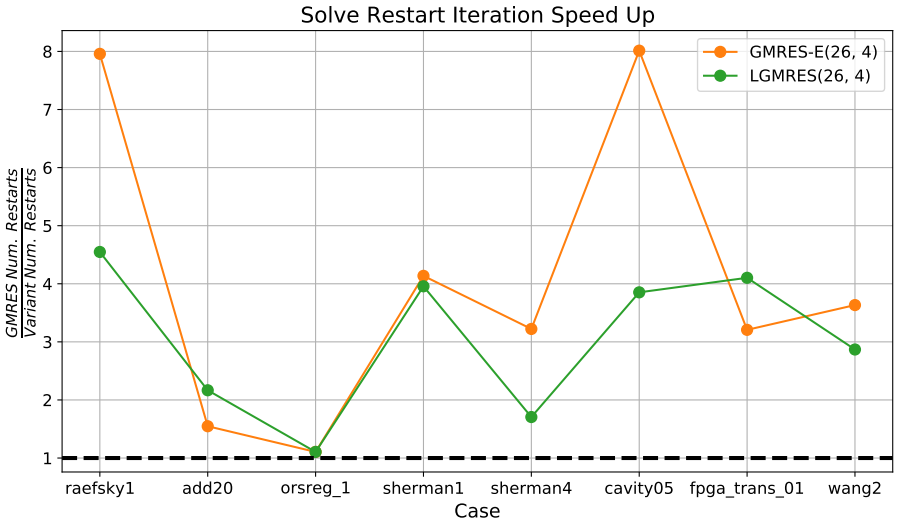

Figure 3: The factor speed-up in the number restart cycles needed to converge to a residual $\frac{\|r\|}{\|r\|_0} \leq 10^{-11}$ for the LGMRES(26,4) and GMRES-E(26,4) methods as compared to the GMRES(30) solver.

Table 1: Several properties for the cases from the sparse matrix library considered. Included are the median sequential (seq.) and skip angles from the GMRES(30) solves. Furthermore, the normality metric (16) and κ-ratio (for the case of $k = 4$ eigenvectors being approximated) is shown as well.

Case	Median Seq. Angle (°)	Median Skip Angle (°)	Normality Metric	κ-Ratio
raefsky1	26.33	8.29	5.834e-03	1.923e-02
add20	46.24	0.93	2.534e-07	9.813e-01
orsreg_1	67.81	32.52	4.003e+06	8.837e-01
sherman1	28.35	0.17	0.000e+00	1.684e-01
sherman4	64.77	10.42	6.305e-01	7.120e-02
cavity05	8.68	3.36	2.086e-01	2.115e-02
fpga_trans_01	23.33	0.18	2.374e-05	9.968e-01
wang2	34.75	7.10	2.065e-03	9.788e-02

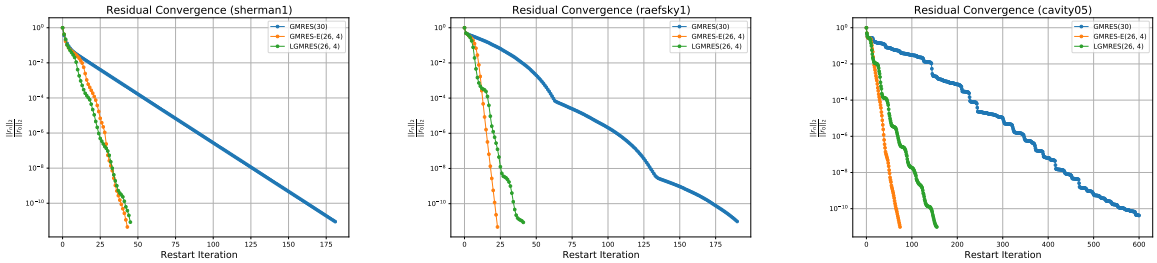

Figure 4: The residual convergence as a function of restart cycles for various cases from the sparse matrix library.

4 Conclusion

In this paper, two popular techniques for accelerating the convergence of the restarted GMRES algorithm have been studied: the GMRES-E and LGMRES algorithms. Both these approaches look to augment the standard Krylov subspace to fix slow convergence issues. In the GMRES-E method, approximate eigenvectors are appended to the Krylov subspace, while in the LGMRES algorithm, approximations for the error in the solution are used. Numerical tests, moreover, do indeed confirm that a significant reduction in the number of restart cycles needed to converge can be observed (if certain properties are satisfied) for both methods of interest.

References

[1] Y. Saad, *Iterative methods for sparse linear systems*, vol. 82. siam, 2003.

[2] R. B. Morgan, "A restarted gmres method augmented with eigenvectors," *SIAM Journal on Matrix Analysis and Applications*, vol. 16, no. 4, pp. 1154–1171, 1995.

[3] A. H. Baker, E. R. Jessup, and T. Manteuffel, "A technique for accelerating the convergence of restarted gmres," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 4, pp. 962–984, 2005.

[4] C. E. Schaerer, "On adaptive strategy for overcome stagnation in lgmres (m, l)," 2017.