Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.2010.33.4

Vol. **3**, No. 3, pp. 338-351
August 2010

An Investigation of Restarted GMRES Method by Using Flexible Starting Vectors†

Qiang Niu[1,*] and Lin-Zhang Lu[2,3]

[1] *BNU-HKBU United International College, Zhuhai 519085, China.*
[2] *School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, China.*
[3] *School of Mathematical Sciences, Xiamen University, Xiamen 361005, China.*

Received 30 September 2009; Accepted (in revised version) 27 January 2010

Available online 2 July 2010

Abstract. We discuss a variant of restarted GMRES method that allows changes of the restarting vector at each cycle of iterations. The merit of the variant is that previously generated information can be utilized to select a new starting vector, such that the occurrence of stagnation be mitigated or the convergence be accelerated. The more appealing utilization of the new method is in conjunction with a harmonic Ritz vector as the starting vector, which is discussed in detail. Numerical experiments are carried out to demonstrate that the proposed procedure can effectively mitigate the occurrence of stagnation due to the presence of small eigenvalues in modulus.

AMS subject classifications: 65F10, 65F15

Key words: Linear systems of equations, Arnoldi process, GMRES, harmonic Ritz vector.

1. Introduction

In this paper, we consider the solution of large sparse linear systems of equations

$$Ax = b, \tag{1.1}$$

where A is an $n \times n$ nonsingular matrix, b is an n dimensional vector. Krylov subspace methods are particularly appealing for this kind of problems and they are widely investigated, see, e.g., [10, 16, 17, 21, 24]. We refer to [22] for a recent survey.

For nonsymmetric matrix A, the GMRES method proposed by Y. Saad and M. H. Schultz [17] is one of the most popular choices. The GMRES method is an iterative method in nature, it generates a sequence of approximations converging to the exact solution to (1.1),

†This paper is presented at the International Conference on Preconditioning Techniques for Scientific and Industrial Applications, August 24-26, 2009, Hong Kong.

*Corresponding author. *Email addresses:* kangniu@gmail.com (Q. Niu), lzlu@xmu.edu.cn (L.-Z. Lu)

and the approximations have the property of minimizing the 2-norm of the residual vector over the Krylov subspace involved. However, the optimality of the GMRES comes at a cost. As the computational costs and storage requirements are prohibitive, so the full GMRES is limited for practical applications, the restarted version of the algorithm is proposed in [17]. The idea is that the GMRES method is restarted after fixed number of iterations, say m ($\ll n$ usually), the resulting restarted version algorithm is denoted by GMRES(m). However, since GMRES(m) only keeps the current approximate solution as the new initial guess for the next cycle (the next m iterations), restarting would lose most information obtained from the previous cycle of the iteration and the convergence may slow down and even stagnation occurs. Stagnation means that there is little decrease in the residual norm at the end of a restart cycle, which is often encountered in the GMRES(m), especially in the case that m is small though there are exceptions [8]. In this paper, the occurrence of the stagnation is studied in restarted GMRES method, the result reveals that stagnation of restarted GMRES method occurs if and only if the first row of $\tilde{H}_m$, the projection matrix of the coefficient matrix A in Krylov subspace, is zero.

Techniques for reducing the negative effect of restarting have been investigated in [13, 14, 19]. These methods aim to recovering the superlinear convergence behavior of full GMRES (see, e.g., [21]) by retaining some eigenvector information generated in the former cycle of iterations. M. Eiermann et al. [7] provided an overview of the augmentation strategies and some comparisons are done with some preconditioning approaches [4, 9]. Generalizations of augmenting procedures that aim to retaining information other than approximate subspace are investigated in [2, 5]. Some other equivalent formulations and variants of restarted GMRES method have been proposed in [1, 18, 20, 26]. All these methods take the residual vector at the end of each restart cycle as the starting vector at the new cycle of iterations, continuing the process until convergence. In this paper, after analysis of convergence behavior including stagnation of restarted GMRES algorithm, we present a variant of GMRES(m) that allows change of the restarting vector at every cycle of the iterative process. The flexibility of choosing the starting vector of the new method provides us a frame work of using inner-outer iterations, in which other iterative methods can be used to get the next starting vector. A simple strategy of taking the harmonic vector associated with the smallest harmonic Ritz value as the starting vector is discussed in details. Numerical experiments are done to compare the variant of GMRES(m) combining with this strategy with the original GMRES(m) and show the former superiority.

In Section 2, we give a brief review of restarted GMRES and an analysis of convergence behavior including stagnation of GMRES(m). In Section 3, we present a variant of restarted GMRES and a simple strategy of choosing the restarting vector. Some numerical results are reported in Section 4. The last section gives a brief concluding remarks.

Throughout the paper, by f^* we denote the conjugate transpose of vector f, by $\mathscr{R}^m$ the m-dimensional complex space and by e_j the jth column of the unit matrix whose order is determined from the context. The Euclidean inner product $\langle x, y\rangle = y^*x$ is used and norm $\|\cdot\|$ denotes both the Euclidean vector norm and the subordinate spectral matrix norm.

2. Restarted GMRES and occurrence of stagnation

In the following discussions, we refer to the group of m iterations between successive restarts as a cycle. The restart number is denoted with a subscript: r_i is the residual after i cycles or $m \times i$ iterations. At the end of each restart cycle i, GMRES(m) finds an approximate solution of the form:

$$x_{i+1} = x_i + V_m y_{i+1}, \tag{2.1}$$

where $V_m = (v_1, v_2, \cdots, v_m)$, whose columns form an orthonormal basis for the Krylov subspace of dimension m defined by

$$\mathscr{K}_m(A, r_i) = \text{span}\{r_i, Ar_i, \cdots, A^{m-1} r_i\}. \tag{2.2}$$

The m dimensional vector y_{i+1} is determined so that the norm of the residual $r_{i+1} = b - Ax_{i+1}$ is minimal over $\mathscr{K}_m(A, r_i)$. The orthonormal basis $v_1, v_2, \cdots, v_m$ for the Krylov subspace $\mathscr{K}_m(A, r_i)$ is obtained via the well known Arnoldi process, which satisfies the relationship

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^* = V_{m+1} \tilde{H}_m, \tag{2.3}$$

where H_m is an $m \times m$ upper Hessenberg matrix,

$$\tilde{H}_m = \begin{pmatrix} H_m \\ h_{m+1,m} e_m^* \end{pmatrix}$$

is an $(m+1) \times m$ upper Hessenberg matrix.

Let $\beta = \|r_i\|$ and $v_1 = r_i/\beta$, then the residual r_{i+1} associated with the approximate solution x_{i+1} computed from $(i+1)$-th cycle satisfies

$$\begin{aligned} r_{i+1} &= b - Ax_{i+1} \\ &= r_i - AV_m y_{i+1} \\ &= \beta v_1 - V_{m+1} \tilde{H}_m y_{i+1} \\ &= V_{m+1}(\beta e_1 - \tilde{H}_m y_{i+1}). \end{aligned} \tag{2.4}$$

Suppose $h_{m+1,m} \neq 0$, then the residual norm $\|r_{i+1}\| = \|\beta e_1 - \tilde{H}_m y_{i+1}\|$ is minimal when y_{i+1} solves the following least square problem

$$\min_{y \in \mathscr{R}^m} \|\beta e_1 - \tilde{H}_m y\|. \tag{2.5}$$

It was known in this case that (2.5) is equivalent to

$$r_{i+1} \perp A\mathscr{K}_m(A, r_i). \tag{2.6}$$

So the approximate solution x_{i+1} of the form (2.1) can be determined uniquely. The details on the practical implementation of GMRES algorithm can be found in [17].

From (2.1) it is easy to see that the residual vectors satisfies the relation

$$r_{i+1} = r_i - AV_m y_{i+1}.$$

Taking the inner product of r_{i+1} with itself, then from (2.6) we have

$$\begin{aligned}\langle r_{i+1}, r_{i+1}\rangle &= \langle r_i, r_i\rangle - \langle AV_m y_{i+1}, r_i\rangle - \langle r_i - AV_m y_{i+1}, AV_m y_{i+1}\rangle \\ &= \langle r_i, r_i\rangle - \langle AV_m y_{i+1}, r_i\rangle - \langle r_{i+1}, AV_m y_{i+1}\rangle \\ &= \langle r_i, r_i\rangle - \langle AV_m y_{i+1}, r_i\rangle. \end{aligned} \tag{2.7}$$

As y_{i+1} solves the least square problem (2.5), so we have

$$y_{i+1} = ||r_i||(\tilde{H}_m^*\tilde{H}_m)^{-1}\tilde{H}_m^* e_1. \tag{2.8}$$

It follows from (2.7) and (2.8) that

$$\begin{aligned}||r_{i+1}||^2 &= ||r_i||^2 - ||r_i||^2 e_1^*\tilde{H}_m(\tilde{H}_m^*\tilde{H}_m)^{-1}\tilde{H}_m^* e_1 \\ &= ||r_i||^2\big(1 - e_1^*\tilde{H}_m(\tilde{H}_m^*\tilde{H}_m)^{-1}\tilde{H}_m^* e_1\big) \\ &= ||r_i||^2(1 - g^*\big(\tilde{H}_m^*\tilde{H}_m\big)^{-1} g\big), \end{aligned} \tag{2.9}$$

where

$$g = \tilde{H}_m^* e_1 = H_m^* e_1. \tag{2.10}$$

Suppose the m-step Arnoldi process does not terminate, i.e., $h_{j+1,j} \neq 0$, $j = 1, 2, \cdots, m$, then we have rank$(\tilde{H}_m) = m$. It follows that $\tilde{H}_m^*\tilde{H}_m$ is symmetric positive definite (SPD), and hence $(\tilde{H}_m^*\tilde{H}_m)^{-1}$ is also SPD. From (2.9) it is easy to see that $||r_{i+1}||^2 = ||r_i||^2$ if and only if $g = 0$ if and only if the first row of $\tilde{H}_m$ or H_m is zero. Thus we have the following theorem.

Theorem 2.1. *Suppose* $\dim\big(\mathscr{K}_{m+1}(A, r_i)\big) = m+1$, *i.e., the m-step Arnoldi process does not terminate at the i-th cycle, then GMRES*(m) *stagnates if and only if* $g = \tilde{H}_m^* e_1 = 0$, *where* $\tilde{H}_m$ *is the* $(m+1)\times m$ *upper Hessenberg matrix generated by the m-step Arnoldi process using* $v_1 = r_i/||r_i||$ *as the starting vector.*

At the m-th iteration of the i-th cycle, GMRES(m) seeks the approximate solution of the form $x_{i+1} = x_i + z$ with

$$z \in \mathscr{K}_m(A, r_i).$$

Then $r_{i+1} = b - Ax_{i+1}$ satisfies the minimum property that

$$||r_{i+1}|| = \min_{z\in\mathscr{K}_m(A,r_i)} ||r_i - Az|| = ||r_i|| \sin\angle\big(r_i, A\mathscr{K}_m(A, r_i)\big). \tag{2.11}$$

So the method stagnates equivalently means $\sin\angle\big(r_i, A\mathscr{K}_m(A, r_i)\big) = 1$, i.e., $r_i \perp A\mathscr{K}_m(A, r_i)$. This has been discussed in [3, 12, 27].

3. A variant of restarted GMRES method

Algorithm 3.1 or usual GMRES(m) takes the residual vector at the end of each restart cycle as the starting vector for the next cycle of iterations. This means that at each restart cycle, GMRES(m) loses all information except an approximation solution generated in the previous cycles. From the analysis given in the previous section, we have seen that this might lead to the convergence failure of GMRES(m) if stagnation or nearly stagnation happens. Moreover, this also means that GMRES(m) is sensitive to the starting vector [8]. To overcome the occurrence of the stagnation, a number of strategies have been proposed in the literature [11, 20, 27].

Algorithm 3.1: A variant of restarted GMRES algorithm

1. Start:
   Choose a stopping criterion ϵ and an initial guess x; set the size of the Krylov subspace to be m and compute $r = b - Ax$.
2. Choose a starting vector $\tilde{r}$; compute $u_1 = \tilde{r}/||\tilde{r}||$. Iteration and form the orthonormal basis matrix U_m of $\mathscr{K}_m(A, \tilde{r})$:
   for $j = 1 : m$ do:
   $u = Au_j$;
   for $i = 1 : j$ do:
   $f_{i,j} = (u, u_i)$;
   $u = u - f_{i,j}u_i$;
   end
   $f_{j+1,j} = ||u||$;
   $u_{j+1} = u/f_{j+1.j}$;
   end
3. Form the approximate solution:
   Find $\tilde{z} \in \mathscr{K}_m(A, \tilde{r})$ such that $||r - A\tilde{z}|| = \min_{z \in \mathscr{K}_m(A,\tilde{r})} ||r - Az||$;
   Compute $x = x + \tilde{z}$.
4. Restart or stop:
   Compute the residual vector $r = b - Ax$ and relative residual norm $||r||/||b||$. If $||r||/||b|| \leq \epsilon$ then stop, else go to 2.

In this paper, we also intend to tune the starting vector flexibly to avoid occurrence of stagnation or to accelerate the convergence. In principle, any vector with suitable dimension can be utilized as the starting vector. However, detailed analysis show that the traditional choice of the new starting vector generally involves a hybrid procedure [11] or inexactly solving a new linear system equivalent to the original one, like GMRESR [25] with embedded inner outer iteration. In the following text, we will discuss a procedure specialized for coefficient matrices with relatively small eigenvalues in magnitude, by utilizing the eign-information from previous cycle to choose the starting vector at the new cycle. The idea leads to a new variant of restarted GMRES method. We should address

that the idea of utilizing eigen-information, particularly adding approximate eigenvectors at restarting has been proposed discussed by R. B. Morgan [13, 14]. We emphasize particularly on the selection of the starting vector of GMRES(m) in this paper, the augmenting techniques appeared [2, 13] may also be applicable in the new algorithm settings which needs further investigation.

Remark 3.1. The main differences between GMRES(m) and Algorithm 3.1 are in steps 2-4. If we fix $\tilde{r} = r$ at step 2 in Algorithm 3.1, then the two algorithms are equivalent. However, if stagnates or nearly stagnates happen in GMRES(m), as will be shown by several numerical examples in the next section; for Algorithm 3.1, we can adjust the starting vector $\tilde{r}$ such that Algorithm 3.1 may succeed.

Subsequently, we first show how to realize step 4 and then a practical choice of $\tilde{r}$ in step 2. Suppose that x_i is an initial guess ($i = 0$) or an approximation ($i \geq 1$) got from i-th cycle of Algorithm 3.1. In $(i+1)$-th cycle, instead of using $r_i = b - Ax_i$ as the starting vector, we can select another vector by utilizing information generated in the i-th cycle or simply choose a random n dimensional vector $\tilde{r}$ to construct a Krylov subspace $\tilde{\mathscr{K}} \equiv \mathscr{K}_m(A, \tilde{r})$. Letting $u_1 = \tilde{r}/||\tilde{r}||$, then the application of m-step Arnoldi process to matrix A yields the relationship

$$AU_m = U_m F_m + f_{m+1,m} u_{m+1} e_m^* = U_{m+1} \tilde{F}_m, \tag{3.1}$$

where $U_m = [u_1, \cdots, u_m]$, whose columns form an orthonormal basis of $\tilde{\mathscr{K}}$, F_m is an $m \times m$ upper Hessenberg matrix and

$$\tilde{F}_m = \begin{pmatrix} F_m \\ f_{m+1,m} e_m^* \end{pmatrix}.$$

In the $(i+1)$-th cycle, the new method seeks the next approximate solution of the form $x_{i+1} = x_i + \tilde{z} = x_i + U_m \tilde{q}$, where $\tilde{z} \in \tilde{\mathscr{K}}$ and $\tilde{q} \in \mathscr{R}^m$ such that

$$||r_i - A\tilde{z}|| = \min_{z \in \tilde{\mathscr{K}}} ||r_i - Az||, \quad \text{or} \quad ||r_i - AU_m\tilde{q}|| = \min_{q \in \mathscr{R}^m} ||r_i - AU_m q||. \tag{3.2}$$

However, if the starting vector $\tilde{r}$ of the new Krylov subspace does not collinear with r_i, then from (2.4) we can see that the minimization problem (3.2) can not be reduced to a least square problem of the form (2.5). So a new procedure should be given for solving the minimization problem (3.2).

It is well known that $\tilde{F}_m$ can be transformed into an upper triangular matrix by the computation of a QR decomposition

$$Q_m \tilde{F}_m = \tilde{R}_m = \begin{pmatrix} R_{i+1} \\ 0 \end{pmatrix}. \tag{3.3}$$

where $R_{i+1} \in \mathscr{R}^{m \times m}$ is an upper triangular matrix, $Q_m \in \mathscr{R}^{(m+1)\times(m+1)}$ is unitary. Moreover, since $\tilde{F}_m$ is upper Hessenberg, Q_m can be constructed by a product of m Givens rotations

$$Q_m = G_m \begin{pmatrix} G_{m-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} G_{m-2} & 0 \\ 0 & I_2 \end{pmatrix} \cdots \begin{pmatrix} G_1 & 0 \\ 0 & I_{m-1} \end{pmatrix}. \tag{3.4}$$

where, for $j = 1 : m$,

$$G_j = \begin{pmatrix} I_{j-1} & & \\ & c_j & -s_j \\ & s_j & c_j \end{pmatrix}, \quad s_j^2 + c_j^2 = 1.$$

From (3.1) and (3.3), we have

$$AU_m = U_{m+1}Q_m^*\tilde{R}_m = \tilde{U}_{m+1}\tilde{R}_m = \tilde{U}_m R_{i+1}, \tag{3.5}$$

where $\tilde{U}_{m+1} = U_{m+1}Q_m^T = [\tilde{U}_m, \tilde{u}_{m+1}]$, and the columns of $\tilde{U}_m$ form an orthonormal basis for the shifted Krylov subspace $A\tilde{\mathscr{K}}$. Based on (3.2) and (3.5), then the norm of the residual vector r_{i+1} satisfies the following relationship

$$||r_{i+1}|| = \min_{q \in \mathscr{R}^m} ||r_i - \tilde{U}_m R_{i+1} q||. \tag{3.6}$$

Suppose $\tilde{q} = arg\{\min_{q\in\mathscr{R}^m} ||r_i - \tilde{U}_m R_{i+1}q||\}$. Then we have

$$r_i - \tilde{U}_m R_{i+1}\tilde{q} \perp \tilde{U}_m,$$

which is equivalent to

$$R_{i+1}\tilde{q} = \tilde{U}_m^* r_i. \tag{3.7}$$

So $\tilde{q}$ can be obtained by solving (3.7), then the new approximate solution $x_{i+1} = x_i + U_m\tilde{q}$ can be formed. That is, step 4 of Algorithm 3.1 can be completed. Moreover, we know from the above discussion that costs of step 4 of both algorithms are almost the same. So the cost of Algorithm 3.1 is almost the same as original GMRES algorithm except possibly cost of forming $\tilde{r}$.

The residual vector r_{i+1} of the new method at cycle $i+1$ satisfies

$$||r_{i+1}|| = ||r_i|| \sin\angle(r_i, A\tilde{\mathscr{K}}). \tag{3.8}$$

That is, the residual norm of the new method correlates with the angle between the residual vector r_i and the shifted Krylov subspace $A\tilde{\mathscr{K}}$. If $r_i \in A\tilde{\mathscr{K}}$, then we will obtain the exact solution in the affine subspace $x_i + \tilde{\mathscr{K}}$. This suggests a reasonable way of finding the starting vector $\tilde{r}$, that is, trying to get an approximate solution of the linear system $Ar = r_i$ by using another iteration method at step 2 in Algorithm 3.1. Similar strategy in augmented GMRES method has also been analyzed by Y. Saad in [19]. On the other hand, based on the same analysis for occurrence of stagnation as given in the previous section, it is easy to see that the new algorithm stagnates if and only if $r_i \perp A\tilde{\mathscr{K}}$. So a careful selected starting vector $\tilde{r}$ should be able to mitigate occurrence of stagnation in the new algorithm.

To estimate $||r_{i+1}||$, we first recall the definition of gap between two subspaces [23].

Definition 3.1. *For any pair of subspaces $(\mathscr{H}_1, \mathscr{H}_2)$ of $\mathscr{R}^n$, the gap $\Theta(\mathscr{H}_1, \mathscr{H}_2)$ between $\mathscr{H}_1$ and $\mathscr{H}_2$ is defined by*

$$\Theta(\mathscr{H}_1, \mathscr{H}_2) = \max\left\{ \sup_{x\in\mathscr{H}_1, ||x||=1} ||(I - P_2)x||, \sup_{x\in\mathscr{H}_2, ||x||=1} ||(I - P_1)x|| \right\}, \tag{3.9}$$

where P_i is the orthogonal projector onto subspace $\mathscr{H}_i$, $i = 1, 2$.

Thus $\Theta(\mathscr{H}_1, \mathscr{H}_2)$ represents the sine of the largest possible angle between vectors in $\mathscr{H}_1$ and their projection in $\mathscr{H}_2$. It is proved that

$$\Theta(\mathscr{H}_1, \mathscr{H}_2) = \max\Big\{||(I - P_2)P_1||, ||(I - P_1)P_2||\Big\}.$$

Theorem 3.1. *Suppose that Φ is a subspace of $\tilde{\mathscr{K}}$ such that $\Theta(A\Phi, \Phi) \le \varepsilon$. Then the residual vector r_{i+1} at $(i+1)$-th cycle obtained by the new variant of restarted GMRES method using $\tilde{r}$ as the starting vector satisfies*

$$||r_{i+1}|| \le ||(I - P_\Phi)(I - P_{A\tilde{\mathscr{K}}})r_i|| + \varepsilon||r_i||,$$

where P_Φ is the orthogonal projector onto subspace Φ.

Proof. From (3.6), (3.7) we have $r_{i+1} = r_i - \tilde{U}_m\tilde{U}_m^T r_i = (I - P_{A\tilde{\mathscr{K}}})r_i$, so

$$P_\Phi r_{i+1} = P_\Phi(I - P_{A\tilde{\mathscr{K}}})r_i.$$

where $P_{A\tilde{\mathscr{K}}}$ is the orthogonal projector on to subspace $A\tilde{\mathscr{K}}$. As

$$||P_\Phi(I - P_{A\tilde{\mathscr{K}}})|| \le ||P_\Phi(I - P_{A\Phi})||,$$

based on the definition of gap and assumption of the theorem, we have

$$||P_\Phi r_{i+1}|| \le ||P_\Phi(I - P_{A\Phi})||||r_i|| \le \Theta(A\Phi, \Phi)||r_i|| \le \varepsilon||r_i||.$$

Decompose r_{i+1} as

$$r_{i+1} = (I - P_\Phi)r_{i+1} + P_\Phi r_{i+1},$$

then the result holds by taking norm on both sides of the above equation. □

Theorem 3.1 shows that if the subspace $\tilde{\mathscr{K}}$ includes an exactly invariant subspace Φ, then the new method will get the approximate solution with the components of the starting vector $\tilde{r}$ in the direction of Φ has been removed completely. Similar results have been given for the GMRES method in [19,21] when analyzing the superlinear convergence behavior of the augmented restarted GMRES method. Thus it is seen that the new method will benefit from fast convergence of eigenspaces during the iteration. This suggests that we can restart the process by favoring the convergence of eigenspaces, meanwhile, the reduction of the GMRES residual norm would also be accelerated according to Theorem 3.1. So the restarting strategies for solving large eigenvalue problems may be adopted to restart the GMRES process. The flexibility of the starting vector proposed in the new method provides us a framework to accomplish the idea.

In the rest of this section, we just present a simple strategy to choose the starting vector $\tilde{r}$ in Algorithm 3.1, which will be demonstrated to be effective by numerical experiments in the next section. We believe that there must be other better strategies.

Suppose that in Algorithm 3.1, the residual r_0 is used as the starting vector at the first iteration (of the first cycle) and step 3 of i-th cycle has been completed, we get the

relationship (3.1) and Krylov subspace $\tilde{\mathcal{K}}$. Let $(\hat{\lambda}_j, \hat{\varphi}_j)$ be the harmonic Ritz pair obtained at the i-th cycle, then $(\hat{\lambda}_j, \hat{\varphi}_j)$ satisfies the relationship

$$\begin{cases} \hat{\varphi}_j \in \tilde{\mathcal{K}}, \\ (A - \hat{\lambda}_j I)\hat{\varphi}_j \perp A\tilde{\mathcal{K}}. \end{cases} \tag{3.10}$$

From (3.1), (3.5) and (3.10) we know that $(\hat{\lambda}_j, \hat{\varphi}_j)$ are the solution of the following generalized eigenvalue problem.

$$\begin{cases} \hat{\varphi}_j = U_m g_j, \\ R_{i+1} g_j = \hat{\lambda}_j Q_m(1:m, 1:m) g_j, \end{cases} \tag{3.11}$$

where U_m is defined in (3.1), Q_m is defined by (3.3) and (3.4).

Define the harmonic residual vector by

$$\hat{r}_j = A\hat{\varphi}_j - \hat{\lambda}_j \hat{\varphi}_j.$$

Then Theorem 5.5 of [14] can be stated as follows.

Theorem 3.2. *In GMRES(m), the $(i+1)$-th GMRES residual vector r_{i+1} is collinear with the residual vector associated with the residual vectors associated with harmonic ritz pairs computed within the i-th cycle of iterations.*

The result is helpful to us in determining the starting vectors. By this theorem, we know that if the smallest harmonic Ritz value $\hat{\lambda}_1$ is very small in magnitude, then $A\hat{\varphi}_1 \approx \hat{r}_1$, that is, $A\hat{\varphi}_1$ will lie in the direction close to r_i. This property indicates that

$$\sin\angle(r_i, A\hat{\varphi}_1) \approx \sin\angle(r_i, \hat{r}_1)$$

is very small. Furthermore, since $\hat{\varphi}_1 \in \tilde{\mathcal{K}}$ (see (3.10)), we have

$$\sin\angle(r_i, A\tilde{\mathcal{K}}) \leq \sin\angle(r_i, A\hat{\varphi}_1).$$

So it follows from (3.8) that fast convergence of the new variant of restarted GMRES method can be expected if we use $\hat{\varphi}_1$ to construct the new Krylov subspace $\tilde{\mathcal{K}}$. So for the new method or Algorithm 3.1, a simple strategy of choosing the starting vector is to set $\tilde{r} = \hat{\varphi}_1$ [§] from the beginning of the second cycle of the iteration, whereas the first cycle of the iteration is the same as the original GMRES(m) using $r_0 = b - Ax_0$ as the starting vector. In numerical experiments of the next section, the new method combining with this strategy of choice is denoted by NGMRES(m), which is demonstrated to be very effective for problems with well separated small eigenvalues. However, it should be emphasized that there are other choices like using a combination of the harmonic Ritz vectors as the starting vector or using an inner-outer iteration to determine the starting vector in the new method.

[§]In case that A is real matrix but $\hat{\varphi}_1$ is a complex vector, we use $real(\hat{\varphi}_1) + imag(\hat{\varphi}_1)$ to avoid the complex operation in the process. However, no complex $\hat{\varphi}_1$ appear in all numerical examples of the next section.

4. Numerical experiments

In this section, some numerical experiments are performed to demonstrate the analysis and theory presented in the previous sections. The matrices with small separated eigenvalues (see Table 3) are selected and restart parameter m is chosen to be relatively small so that there is possibly occurrence of nearly stagnation in GMRES(m) (see Examples 4.2, 4.3 and 4.4 below).

We report some numerical experiments on NGMRES(m), which is tested and compared with GMRES(m). All numerical tests have been carried out on an INTEL PENTIUM IV 2.4GHz with main memory 256MB and the machine precision $eps = 2.22 \times 10^{-16}$ using MATLAB 6.5 on a Window XP-based system. In all tables of the paper, m is the dimension of the Krylov subspace, $iter$ denotes the number of restarts, cpu is the CPU timings in seconds, res is the relative residual norm. In the following experiments, the stopping criterion is set to be $res = 10^{-7}$ and the initial approximate vector is taken to be the vector with random entries uniformly distributed on $[0, 1]$ in each examples.

Example 4.1. Two matrices sherman1 and sherman4 from [6] are tested in this example. The right hand sides b are also taken from the collection. Both matrices are real nonsymmetric generated in oil reservoir simulation. The dimensions of sherman1 is 1000, and sherman4 is 1104. Tested results on sherman1 are listed in Table 1, from which we can see that for subspaces with different dimensions, NGMRES(m) is faster than GMRES(m), both in terms of restarts and of the elapsed time. The convergence history for matrix sherman4 are shown in Fig. 1. GMRES(15) and GMRES(20) take 0.48s and 0.46s to reach $res = 10^{-7}$, while NGMRES(15) and NGMRES(20) need only 0.29s and 0.25s, respectively. In the example, no stagnation or nearly stagnation phenomena occurs.

Table 1: Tested results on sherman1.

m	methods	$iter$	cpu	res
15	GMRES	337	4.5	9.6E-8
	NGMRES	143	1.03	9.8E-8
20	GMRES	194	2.7	9.5E-8
	NGMRES	80	0.93	9.7E-8
25	GMRES	123	2.2	9.3E-8
	NGMRES	53	1.02	8.8E-8

Example 4.2. Real symmetric positive definite matrix nos3 from [6] is tested in this example. It is of order 960, arising from structure engineering. Fig. 2 shows convergence curves of both algorithms. We can see from Fig. 2 that GMRES(10) nearly stagnates from the first few cycles of the iteration while NGMRES(10) converges more quickly. The convergence of GMRES(20) is steady but slowly, its residual norm reaches 5.1×10^{-6} after 500 restarts. The NGMRES(20) exhibits faster convergence behavior, it needs only 76 restarts in 1.6s to reduce the relative residual norm below 10^{-7}. So the variant of restarted GMRES is far better than the original one for this problem.

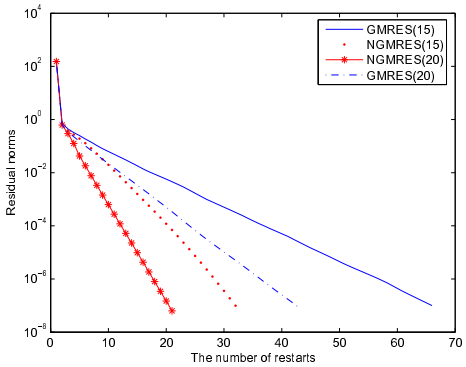

Figure 1: The convergence curves for sherman4.

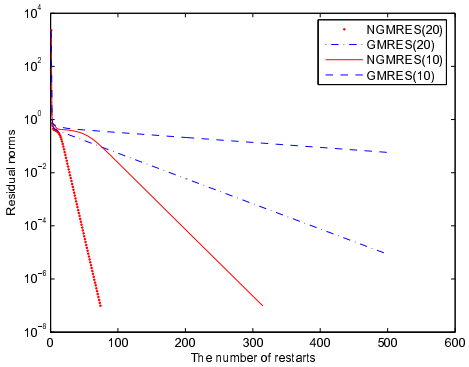

Figure 2: The convergence curves for nos3.

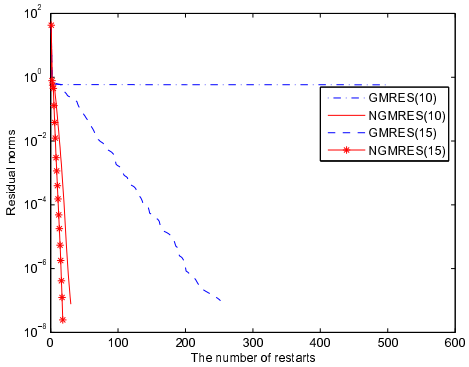

Figure 3: The convergence curves for cdde1.

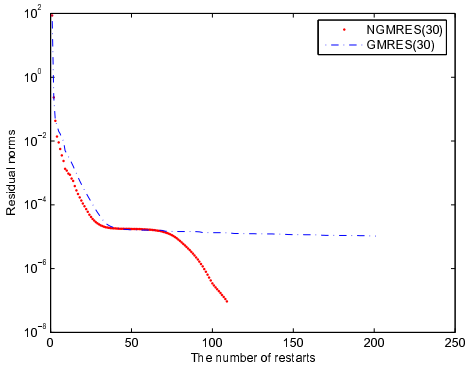

Figure 4: The convergence curves for cavity10.

Example 4.3. We test real nonsymmetric matrix cdde1 from [15] in this example, it is generated in computational fluid dynamics with size 961. Fig. 3 and Table 2 show the comparison of the convergence behavior of the two methods with different subspace sizes. From Fig. 3 we can see that GMRES(10) nearly stagnate after several times of restarting, while NGMRES(10) exhibits consistently fast convergence behavior. When enlarging the dimension of the subspace involved, Table 2 shows that convergency of both methods

Table 2: Tested results on cdde1.

m	methods	$iter$	cpu	res
15	GMRES	200	1.62	8.5E-8
	NGMRES	15	0.15	6.8E-8
20	GMRES	84	0.91	7.7E-8
	NGMRES	12	0.16	1.2E-8
25	GMRES	46	0.86	3.8E-8
	NGMRES	8	0.17	1.9E-8
30	GMRES	27	0.76	3.2E-8
	NGMRES	7	0.24	1.4E-8

improve accordingly, but NGMRES(m) behaves much better in terms both the timing and the number of the iteration.

Example 4.4. In this example, we test real nonsymmetric matrix cavity10 from [6]. This matrix is of size 2597 arising from finite element modeling. The right hand side b used is the normalized vector accompanies with cavity10 in matrix market. Fig. 4 depicts the convergence history of both algorithms with different dimensions of the Krylov subspaces, from which we can see that GMRES(30) stagnate at the level of 10^{-5}, whereas NGMRES(30) effectively mitigate the occurrence of stagnation, and exhibit much better convergence behavior.

Table 3: Matrices tested in the examples.

Example	*matrix*	*size*	*SM*	*LM*
4.1	sherman1	1000	-3.0 × 10^{-4}, -1.0 × 10^{-3}, -1.1 × 10^{-3}	-5.04
4.1	sherman4	1104	0.03, 0.085, 0.28	66.5
4.2	nos3	960	0.0183, 0.249, 0.267	689.9
4.3	cdde1	961	-0.0052, 0.024, 0.024	7.95
4.4	cavity10	2597	4.4 × 10^{-6},4.3 × 10^{-4}, 4.3 × 10^{-4}	13.05

LM: the largest eigenvalue in magnitude (estimated)
SM: the three smallest eigenvalues in magnitude (estimated)

In Table 3, we list some information on the matrices used in the test. From this table, we can see that all the matrices tested in the above examples have one common property: the smallest eigenvalues are well separated and the ratio of the second smallest eigenvalue in magnitude to the smallest one is large. For such problems, our numerical experiments show that NGMRES(m) always outperforms GMRES(m).

5. Closing remarks

After an analysis of the occurrence of stagnation of GMRES algorithm, we discuss the implementation of restarted GMRES from a different view of point. It is shown that the starting vector at the each restart cycle of GMRES(m) can be chosen flexibly for mitigating occurrence of stagnation or for accelerating the convergence. A simple strategy of taking the harmonic vector associated with the smallest harmonic Ritz value in magnitude as the starting vector is discussed. The new method combining with the strategy is demonstrated to be effective by a few numerical examples with small well-separated eigenvalues. However, other strategies like inner-outer iteration methods may also be adopted in the framework.

Acknowledgments The authors would like to thank Prof. Michael Ng for his valuable comments on the earlier version of the paper.

References

[1] E. H. Ayachour, *A fast implementation for GMRES method,* J. Comput. Appl. Math., 159 (2003), pp.269-283.

[2] A. Baker, L. Jessup, and T. Manteuffel, *A technique for accelerating the convergence of restarted GMRES*, SIAM J. Matrix Anal Appl., 26 (2005), pp.962-984,.

[3] P. N. Brown. *A theoretical comparison of the Arnoldi and the GMRES algorithm*, SIAM J. Sci. Stat. Comput., 12 (1999), pp.58-78.

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

[5] E. De Sturler, *Truncation strategy for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp.864-889.

[6] I. S. Duff, R. G. Grimes, and J. G. Lewis, *Sparse matrix test problems*, ACM Trans. Math. Soft., 15 (1989), pp.1–14.

[7] M. Eiermann, O. G. Ernst, and O. Schneider, *Analysis of Acceleration Stategies for Restarted Minimal Residual Methods,* J. Comput. Appl. Math., 123 (2000), pp.261-292.

[8] M. Embree *The tortoise and the hare restarted GMRES*, SIAM Review., 45 (2003), pp.259-266.

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

[10] Q. W. Freund and N. M. Nachnical, *QMR: a quasi-minimal residual residual algorithm for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp.315-339.

[11] M. A. Gomes-Ruggiero, V. L. Rocha Lopes and J. V. Toledo-Benavides, *A new choice for the forcing term and a global convergent inexact Newton method*, Tech. Rep. 52-03, 2003, Departamento de Matema'tica Aplicada, Universidade Estadual de Campinas, Brasil.

[12] A. Greenbaum, *Iterative methods for solving linear systems*, SIAM, Philadelphia, 1997.

[13] R. B. Morgan, *A restarted method augmented with eigenvectors* , SIAM J. Matrix Anal Appl., 16 (1995), pp.1154-1171.

[14] R. B. Morgan, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp.1112–1135.

[15] National Institute of Standards and Technology, Mathematical and Computatinal Sciences Division, *Matrix Market*. Available at $http://math.nist.gov/MatrixMarket/$.

[16] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations,* SIAM J. Numer. Anal., 12 (1975), pp.617-629.

[17] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856-869.

[18] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp.461-469.

[19] Y. Saad, *Analysis of augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp.435-449.

[20] V. Simoncini, *A new variant of restarted GMRES*, Numer. Linear Algebra Appl., 6 (1999), pp.61-77.

[21] V. Simoncini and D. B. Szyld. *On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods*, SIAM Review, 47 (2005), pp.247-272.

[22] V. Simoncini and D. B. Szyld. *Recent computational developments in Krylov subspace methods for linear systems*, preprint., 2006, pp.1-55.

[23] G. W. Stewart and J. Sun,*Matrix perturbation theory*, Academic Press, London, 1990.

[24] H. A. Van der vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems,* SIAM J. Sci. Stat. Comput., 13 (1992), pp.631-644.

[25] H. A. van der Vorst and C. Vuik. *GMRESR: a family of nested GMRES methods*, Numer. Linear. Algebra. Appl., 1 (1994), 369-386.
[26] H. F. Walker and L. Zhou, *A simpler GMRES,* Numer. Linear Algebra Appl., 1 (1994), 571-581.
[27] I. Zavorin, D. P. *O'*Leary and H. Elman, *Complete stagnation of gmres,* Linear. Algebra Appl., 367 (2003), 165-183.