NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2005; **12**:373–390
Published online 9 August 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.421

Convergence conditions for a restarted GMRES method augmented with eigenspaces

Jan Zítko*,†

Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University of Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic

SUMMARY

We consider the GMRES(m,k) method for the solution of linear systems $Ax=b$, i.e. the restarted GMRES with restart m where to the standard Krylov subspace of dimension m the other subspace of dimension k is added, resulting in an augmented Krylov subspace. This additional subspace approximates usually an A-invariant subspace. The eigenspaces associated with the eigenvalues closest to zero are commonly used, as those are thought to hinder convergence the most. The behaviour of residual bounds is described for various situations which can arise during the GMRES(m,k) process. The obtained estimates for the norm of the residual vector suggest sufficient conditions for convergence of GMRES(m,k) and illustrate that these augmentation techniques can remove stagnation of GMRES(m) in many cases. All estimates are independent of the choice of an initial approximation. Conclusions and remarks assessing numerically the quality of proposed bounds conclude the paper. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: restarted GMRES; augmented Krylov subspace; convergence; non-Hermitian systems

1. INTRODUCTION

In this paper we consider the GMRES method for solving the non-Hermitian and non-singular system

$$Ax=b, \quad A\in\mathbb{C}^{n\times n}, \quad x,b\in\mathbb{C}^n \tag{1}$$

In order to limit both computation and memory requirements, a restarted version is often used, in which the Krylov subspace is restricted to be of fixed dimension m and the process is restarted using the last iterate x_m as a new initial approximation for the next restart.

*Correspondence to: Jan Zítko, Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University of Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic.
†E-mail: zitko@karlin.mff.cuni.cz

Contract/grant sponsor: Grant Agency of the Czech Republic; contract/grant number: 201/02/0684
Contract/grant sponsor: Grant Agency of the Czech Republic; contract/grant number: MSM 113200007

Received 27 November 2001
Revised 22 March 2004

But restarting slows down the convergence. It is known that a trouble may be caused by the presence of the eigenvalues nearest to zero. Hence the technique for improving the convergence is based on the idea to remove the smallest eigenvalues from the spectrum of A. The research for reducing the negative effects of a restart developes in two ways. The first way [1–3] consists of constructing a preconditioner for GMRES on the base of information gathered from the Arnoldi process during one restart. The preconditioner is updated after each restart and removes the influence of small eigenvalues. The other way attempts to remove the components of the initial residual vector associated with the eigenvalues closest to zero. The corresponding eigenvectors and principal vectors, or their approximations, are added directly to the Krylov subspace. This idea proposed by Morgan [4] will be the topic of our further investigations. Let us remark (see References [5, 6]) that the vector $x \neq 0$ is a principal vector of the matrix A of degree v belonging to λ if v is the smallest positive integer fulfilling the equality $(A - \lambda I)^v x = 0$. An eigenvector is therefore a principal vector of degree 1. It is an often heuristic assumption in practice that the matrix A is diagonalizable and that approximations to the eigenvectors corresponding to the smallest eigenvalues in magnitude are added to the Krylov subspace. According to Reference [4] the motivation for this is that if good approximations are added, the corresponding eigenvalues are effectively eliminated from the spectrum. Both the above-mentioned approaches are efficient when there is a cluster of a few eigenvalues that have the strongest influence on the rate of convergence.

Let us remark that there exist another techniques for improving the GMRES(m) method which can be applied with success in many cases. Let us mention two ones here.

The procedure proposed by Simoncini is based on the construction of the residual polynomial which is smaller than the GMRES(m) one in a small neighbourhood of zero. A more exact formulation is in Reference [7, pp. 67, 68]. This procedure is used in the case of stagnation or slow convergence of GMRES(m).

In the paper [8], the following procedure is applied before GMRES starts: by using some regular splitting of the matrix A, the original system (1) is rewritten in the form $(I - T)x = c$ and p successive approximations $x_{l+1} = Tx_l + c$ for $l = 1, 2, \ldots, p-1$ are performed at the beginning of the calculation. The last iteration x_p is taken for the initial guess for GMRES(m) applied to the rewritten system $(I - T)x = c$. This process, called pre-iterations, is often applied if a stagnation occurs. The technique gives very good results in many cases. We must be only very careful and restrict the number of pre-iterations if the matrix T is divergent. The same idea called pre-processing was developed in Reference [9].

In this paper, the behaviour of residual bounds for various situations which can arise during the GMRES(m,k) process is described. These estimates suggest sufficient conditions for the convergence of GMRES(m,k) which are also formulated.

The paper is organized as follows. In Section 2, we briefly mention the practical implementation of GMRES(m,k). The new results are concentrated in Section 3 where the estimates of norm of the residual vector are presented in dependence on spaces which are added to the Krylov subspace. Sufficient convergence conditions resulting from these theorems are independent of an initial approximation and illustrate that the augmentation technique can remove stagnation or slow convergence in many cases. The quality of proposed bounds is assessed on some numerical examples in the last section.

In the whole paper the following notational conventions will be used. The set of all complex $n \times k$ matrices is a complex vector space denoted by $\mathbb{C}^{n \times k}$. The superscript H will be used for the conjugate transpose. If $u_i \in \mathbb{C}^n$, $i = 1, \ldots, r$, then the symbol $(u_1, \ldots, u_r)$ denotes a rectangular $n \times r$ matrix with columns $u_1, \ldots, u_r$. For $k \in \{1, 2, \ldots\}$ and $i \in \{1, 2, \ldots, k\}$, the vector e_i^k is the ith column of the identity matrix $I_k \in \mathbb{C}^{k \times k}$ and $e^k = \sum_{i=1}^k e_i^k$. The symbol S_n denotes the unit sphere in $\mathbb{C}^n$, i.e. $S_n = \{z \in \mathbb{C}^n \mid z^H z = 1\}$. If $W \in \mathbb{C}^{n \times k}$ then the symbol span$\{W\}$ denotes the k-dimensional vector space generated by the columns of the matrix W. In the whole paper we employ the Euclidean vector and matrix norm and exact arithmetic is assumed in the whole paper.

2. PRACTICAL IMPLEMENTATION OF GMRES(m, k)

Now we briefly describe the GMRES(m, k) process. Let x_0 be an initial approximation, $r_0 = b - Ax_0 \neq 0$ the associated residual and $v_1 = r_0 / \|r_0\|$. Put $s = m + k$. Let $y_1, y_2, \ldots, y_k$ be such vectors that the space span$\{v_1, Av_1, \ldots, A^m v_1, Ay_1, Ay_2, \ldots, Ay_k\}$ has dimension $s + 1$. Define for $j = 1, 2, \ldots, s$ vectors $v_{j+1} = P_j^{\perp} Au_j / \|P_j^{\perp} Au_j\|$ where

1. for $j \leqslant m$, we substitute $u_j = v_j$ and $P_j^{\perp}$ denotes the orthogonal projection onto the orthogonal complement of the Krylov subspace $\mathcal{K}_j(A, v_1) = \text{span}\{v_1, Av_1, \ldots, A^{j-1} v_1\}$;
2. for $j = m + 1, \ldots, s$, we put $u_j = y_{j-m}$ and $P_j^{\perp}$ is the orthogonal projection onto the orthogonal complement of the space span$\{v_1, v_2, \ldots, v_j\}$.

This definition includes (see Reference [10]) all the known processes (Arnoldi, Arnoldi-Modified Gram–Schmidt, Householder–Arnoldi) for the construction of an orthogonal basis in Krylov subspaces. Morgan [4] chooses the vectors $y_1, y_2, \ldots, y_k$ to be approximations of the k eigenvectors belonging to the smallest in magnitude k eigenvalues. Let us define the matrices

$$W_s = (v_1, v_2, \ldots, v_m, y_1, \ldots, y_k) \tag{2}$$

$$\tilde{W}_{s+1} = (v_1, v_2, \ldots, v_m, v_{m+1}, \ldots, v_{s+1}) \tag{3}$$

Then

$$AW_s = \tilde{W}_{s+1} \tilde{H}_{s+1,s} \tag{4}$$

where $\tilde{H}_{s+1,s}$ is the known $(s+1) \times s$ upper Hessenberg matrix with elements $h_{i,j} = v_i^H A v_j$ for $i < j+2$, $j \leqslant m$, and $h_{i,j} = v_i^H A y_{j-m}$ for $i < j+2$, $m < j \leqslant s$. In the sequel we will suppose that the matrix W_s has the maximal rank for all restarts. Now we describe one run of GMRES(m, k). As usual we put $\beta = \|r_0\|$. For the construction of a new approximation x_s we need m vectors of the Krylov subspace and k vectors $y_1, \ldots, y_k$, i.e. in the whole s vectors, and therefore this approximation is denoted by x_s. This vector is written in the form $x_s = x_0 + w_s$, where $w_s \in \text{span}\{W_s\}$ such that

$$\|r_0 - Aw_s\| = \min_{w \in \text{span}\{W_s\}} \|r_0 - Aw\| \tag{5}$$

Every vector $w \in \text{span}\{W_s\}$ can be written in the form $w = W_s u$, where $u \in \mathbb{C}^s$. Analogously as in the papers [11–13]

$$\begin{aligned}\|r_0 - Aw\| &= \|\beta v_1 - AW_s u\| = \|\beta v_1 - \tilde{W}_{s+1} \tilde{H}_{s+1,s} u\| \\ &= \|\tilde{W}_{s+1}(\beta e_1^{s+1} - \tilde{H}_{s+1,s} u)\| = \|\beta e_1^{s+1} - \tilde{H}_{s+1,s} u\|\end{aligned}$$

The matrix $\tilde{H}_{s+1,s}$ is factored into $Q_{s+1}R_{s+1,s}$, where $Q_{s+1} \in \mathbb{C}^{(s+1)\times(s+1)}$ is a unitary and $R_{s+1,s} \in \mathbb{C}^{(s+1)\times s}$ an upper triangular matrix having zero in the last row. The factorization is carried out in detail in Reference [12]. In the same way as for the standard GMRES we have $w_s = W_s u_s$, where

$$u_s = \arg\min_{u\in\mathbb{C}^s} \|\beta e_1^{s+1} - \tilde{H}_{s+1,s}u\| = \arg\min_{u\in\mathbb{C}^s} \|\beta Q_{s+1}^{\mathrm{T}} e_1^{s+1} - R_{s+1,s}u\|$$

and the finding of u_s leads to the solution of the system with the upper triangular matrix $R_{s,s}$, where $R_{s,s}$ is formed from the first s rows of $R_{s+1,s}$.

3. THEORETICAL ANALYSIS OF CONVERGENCE

Let us recall the definition of the gap between two subspaces. If P_i for $i = 1,2$ is the orthogonal projector of $\mathbb{C}^n$ onto the space $H_i \subset \mathbb{C}^n$ and $P^{(i)} = I - P_i$ then the gap $\Theta(H_1,H_2)$ between H_1 and H_2 is defined (see Reference [14]) by the formula

$$\Theta(H_1,H_2) = \max\left\{ \sup_{x\in H_1,\|x\|=1} \|P^{(2)}x\|, \sup_{x\in H_2,\|x\|=1} \|P^{(1)}x\| \right\}$$

Evidently $0 \leqslant \Theta(H_1,H_2) \leqslant 1$ and $\Theta(H_1,H_2) = 0 \Leftrightarrow H_1 = H_2$. It is proved (see Reference [14]) that the equalities

$$\Theta(H_1,H_2) = \|P_1 - P_2\| = \max\{\|P^{(2)}P_1\|, \|P^{(1)}P_2\|\}$$

hold. Let $Y_k = (y_1, y_2, \ldots, y_k)$ be a matrix from $\mathbb{C}^{n\times k}$ and define the spaces $\mathscr{Y} = \operatorname{span}\{Y_k\}$ and $\mathscr{K} = \mathscr{K}_m(A, r_0) + \mathscr{Y}$. Let $\mathscr{Z}$ be an invariant subspace and $P_{\mathscr{Z}}$ any projector onto $\mathscr{Z}$. In paper [15], it is proved that if $\Theta(\mathscr{Z}, A\mathscr{Y}) = \varepsilon$ then for the residual r_s after one restarted run the inequality

$$\|r_s\| \leqslant \min_{p\in\mathscr{P}_m^*} \{\|p(A)(I - P_{\mathscr{Z}})r_0\| + \varepsilon\|p(A)P_{\mathscr{Z}}r_0\|\} \tag{6}$$

holds, where $\mathscr{P}_m^*$ is the space of polynomials p of degree $\leqslant m$ satisfying the constraint $p(0) = 1$. If ε is small then the estimate on the right-hand side is given by the first term which represents the norm of the residual after one restarted run of GMRES(m) starting with an initial residual $(I - P_{\mathscr{Z}})r_0$. The space $\operatorname{span}\{z_1, z_2, \ldots, z_k\}$, where $z_1, z_2, \ldots, z_k$ are eigenvectors and principal vectors corresponding to the smallest eigenvalues of A, is often considered for the space $\mathscr{Z}$. Formula (6) is in detail discussed in Reference [15], because different bounds may be obtained depending on which type of projector $P_{\mathscr{Z}}$ is used. We compare our estimate and (6) later.

In what follows let the vectors $z_1, z_2, \ldots, z_k$ denote a basis of an invariant space $\mathscr{Z}$.

Now we analyse the course of GMRES(m,k) algorithm. At the beginning of the GMRES(m,k) process we usually put $y_j = A^{m+j-1}r_0$ for $j = 1, 2, \ldots, k$. Hence the estimate for the quotient $\|r_s\|^2/\|r_0\|^2$ in the first restarted run of the GMRES(m,k) is equivalent with the estimate for GMRES(s). Let us briefly mention the construction. We assume that $\dim(\mathscr{K}_s(A, r_0)) = s$. The approximation $x_s \in x_0 + \mathscr{K}_s(A, r_0)$ is such that the Min-Res condition is fulfilled.

If $v := r_0/\|r_0\|$,

$$D_\alpha := \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_s), \quad \alpha_i \in \mathbb{C}, \quad \prod_{j=1}^{s} \alpha_j \neq 0 \tag{7}$$

$$V := (Av, A^2v, \ldots, A^s v) \quad \text{and} \quad U := VD_\alpha \tag{8}$$

then $r_s = b - Ax_s = r_0 - Pr_0$ where $P = U(U^H U)^{-1}U^H$ is the orthogonal projector for the space span$\{U\}$. It is

$$\|r_s\|^2 = \|r_0\|^2 - r_0^H P r_0 \tag{9}$$

and for the quotient $\|r_s\|^2/\|r_0\|^2$ we successively obtain the estimates

$$\|r_s\|^2/\|r_0\|^2 = 1 - v^H U(U^H U)^{-1}U^H v \leqslant 1 - \|U^H v\|^2 \lambda_{\min}((U^H U)^{-1}) \tag{10}$$

$$= 1 - \|U^H v\|^2/\lambda_{\max}(U^H U) \leqslant 1 - \|U^H v\|^2/\mathrm{tr}(U^H U) \tag{11}$$

Let us denote

$$D_\beta = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_s) := |D_\alpha|^2 \quad \text{and} \tag{12}$$

$$h(p, w, D_\beta) := \left(\sum_{j=1}^{p} \beta_j |w^H A^j w|^2 \right) \Bigg/ \sum_{j=1}^{p} \beta_j \|A^j w\|^2 \tag{13}$$

In the last formula $w \in \mathbb{C}^n$ and $p \leqslant s$ is a positive integer. We have from (11) the estimates

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - h(s, v, D_\beta) \leqslant 1 - \min_{w \in S_n} h(s, w, D_\beta) \tag{14}$$

Remark 3.1
If the matrix A and the initial approximation are real then we can apparently consider the above (and of course the following) estimates only in real vector spaces. Hence instead of $w \in S_n$ we can write only $w \in S_n \cap \mathbb{R}^n$.

In the next considerations we will use the symbol v for the quotient $r_0/\|r_0\|$, where r_0 is the initial residual vector for an arbitrary restarted run. On the basis of information obtained during the previous restart, the vectors $y_1, y_2, \ldots, y_k$ are determined and the new approximation x_s is constructed among vectors from the set $x_0 + \mathscr{K}_m(A, r_0) + \mathscr{Y}$ where $\mathscr{Y} = \mathrm{span}\{y_1, y_2, \ldots, y_k\}$. Here x_0 and r_0 is the resulting approximation and residual from the previous restart, respectively.

Remark 3.2
The vectors y_i are often calculated as approximations for the eigenvectors corresponding to the small in magnitude eigenvalues of the matrix A using Ritz–Galerkin projection (see References [4, 11]).

We can see that the choice of y_i may result in three main possibilities:
the space $\mathscr{Y}$ is an arbitrary k-dimensional space in $\mathbb{C}^n$;
the space $\mathscr{Y}$ is an invariant subspace or is close to some invariant subspace.

We will investigate these three possibilities independent of r_0, i.e. uniformly for all restarted runs.

Let $y_1, y_2, \ldots, y_k$ be random vectors choosen only under the condition that the space span $\{v, Av, \ldots, A^m v, Ay_1, \ldots, Ay_k\}$ has dimension $s+1$. Analogous to relation (13) we define the function

$$g(m, w, D_\beta, Y_k) := \left(\sum_{j=1}^{m} \beta_j |w^H A^j w|^2 + \sum_{j=m+1}^{s} \beta_j |w^H A y_{j-m}|^2 \right) \Bigg/ \left(\sum_{j=1}^{m} \beta_j \|A^j w\|^2 + \sum_{j=m+1}^{s} \beta_j \|A y_{j-m}\|^2 \right) \tag{15}$$

where $w \in S_n$, D_β is defined by (12) and Y_k is the matrix with columns $y_1, y_2, \ldots, y_k$. The behaviour of the quotient $\|r_s\|^2/\|r_0\|^2$ is described by the following theorem.

Theorem 3.1

Let the vectors $y_1, y_2, \ldots, y_k$ be added to the Krylov subspace $\mathscr{K}_m(A, v)$ such that the space span$\{v, Av, \ldots, A^m v, Ay_1, \ldots, Ay_k\}$ has dimension $s+1$.

Then the estimate

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j > 0} \min_{w \in S_n} g(m, w, D_\beta, Y_k) \stackrel{\text{def}}{=} q_1 \tag{16}$$

holds.

Proof

If we change the matrix V in (8) so that we put

$$V := (Av, A^2 v, \ldots, A^m v, Ay_1, Ay_2, \ldots, Ay_k) \tag{17}$$

then according to (9)–(11) we obtain

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \min_{w \in S_n} g(m, v, D_\beta, Y_k) \tag{18}$$

Since β_j are arbitrary positive numbers, the estimate remains true considering $\sup_{\beta_j > 0}$. This gives (16). □

Remark 3.3

Any concrete choice of the matrix D_β in (18) gives an upper bound for $\|r_s\|^2/\|r_0\|^2$ which is possible to evaluate. This is discussed in greater detail in Section 4.

Remark 3.4

Substituting $\|A^j\|$ and $\|A\|$ instead of $\|A^j w\|$ and $\|Ay_{j-m}\|$ in formula (15), we obtain further estimates for $\|r_s\|^2/\|r_0\|^2$. In this case we assume that $\|y_i\| = 1 \ \forall i \in \{1, 2, \ldots, k\}$.

Remark 3.5

Estimate (16) remains true if we write $\min_{w \in S_n} \sup_{\beta_j > 0}$ instead of $\sup_{\beta_j > 0} \min_{w \in S_n}$.

Corollary 3.2
Let $m,k,s \in \{1,2,\dots,n-1\}$, $s=m+k$. The vectors $y_1,y_2,\dots,y_k$ are added to the Krylov subspace $\mathscr{K}_m(A,v)$ in all restarts and let all augmented Krylov subspaces have dimension $s+1$. Let the system of equations

$$w^H A^j w = 0 \quad \text{for } j=1,\dots,m \tag{19}$$

$$w^H A y_j = 0 \quad \text{for } j=1,\dots,k \tag{20}$$

has only the solution $w=0$ in $\mathbb{C}^n$.

Then GMRES(m,k) is convergent, i.e. the iterations converge to the unique solution of (1) for arbitrary initial approximation if the number of restarts tends to infinity.

Proof
Let us consider one particular choice: $\tilde{\beta}_1,\tilde{\beta}_2,\dots,\tilde{\beta}_s$, i.e. $D_{\tilde{\beta}}=\mathrm{diag}(\tilde{\beta}_1,\tilde{\beta}_2,\dots,\tilde{\beta}_s)$ is a given matrix. From (19) and (20) it follows that a number $\gamma>0$ exists such that

$$g(m,w,D_{\tilde{\beta}},Y_k)>\gamma$$

for all w lying in the compact set S_n. Hence

$$\min_{w\in S_n} g(m,w,D_{\tilde{\beta}},Y_k)>\gamma \Rightarrow \sup_{\beta_j>0}\min_{w\in S_n} g(m,w,D_\beta,Y_k)>\gamma$$

The last inequality guarantees convergence according to (16). □

Corollary 3.2 with only equalities (19) give a sufficient condition for convergence of GMRES(m). If this condition does not hold then an element $u\in S_n$ exists such that

$$u^H A^j u = 0 \quad \forall j \tag{21}$$

But the second condition (20) offers to find $Y_k=(y_1,y_2,\dots,y_k)$ such that

$$u^H A Y_k \neq 0 \tag{22}$$

If we are successful in completing (22) for all u keeping (21) then GMRES(m,k) converges even though GMRES(m) stagnates or converges slowly.

In the next part of this section we assume that an invariant subspace $\mathscr{Z}$ is added to $\mathscr{K}_m(r_0,A)$ in all restarted runs. Let the vectors $z_1,z_2,\dots,z_k$ form a basis of $\mathscr{Z}$. In order to obtain the estimate in the first restarted run we put

$$V:=(Av,A^2v,\dots,A^m v,z_1,z_2,\dots,z_k),\quad U:=VD_\alpha \tag{23}$$

where D_α is defined in (7) and $v=r_0/\|r_0\|$. If $\beta_j:=|\alpha_j|^2$ for $j\in\{1,2,\dots,m\}$, $\beta_j:=|\alpha_j|^2\|z_{j-m}\|^2$ for $j\in\{m+1,m+2,\dots,s\}$ and

$$g_1(m,v,D_\beta,Z_k)$$
$$:= \left(\sum_{j=1}^{m}\beta_j|v^H A^j v|^2+\sum_{j=m+1}^{s}\beta_j\cos^2\angle(v,z_{j-m})\right)\Bigg/\left(\sum_{j=1}^{m}\beta_j\|A^j w\|^2+\sum_{j=m+1}^{s}\beta_j\right)\ \forall w\in\mathbb{C}^n \tag{24}$$

where $D_\beta = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_s)$ and Z_k is the matrix with columns $z_1, z_2, \ldots, z_k$ then analogous to (9)–(11) we obtain

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j > 0} g_1(m, v, D_\beta, Z_k) \stackrel{\text{def}}{=} q_2 \tag{25}$$

After the implementation of this restart, in which the invariant subspace $\mathscr{X}$ for the first time was added to the Krylov subspace, we have

$$r_s \perp z_i \quad \text{for } i \in \{1, 2, \ldots, k\} \tag{26}$$

If we put $r_0 := r_s$ then $r_0 \perp z_i$ for $i \in \{1, 2, \ldots, k\}$ and these orthogonality conditions are valid in all next restarted runs and change the estimate for $\|r_s\|^2/\|r_0\|^2$ in view of the fact that last k components of the vector $V^H v$ equal zero. Now we estimate the quotient $\|r_s\|^2/\|r_0\|^2$ under these orthogonality conditions. For this purpose we prove the following lemma. Let us still remark that a Hermitian matrix M is said to be positive definite in case $x \neq 0 \Rightarrow x^H A x > 0$. For non-Hermitian matrices the concept is not defined.

Lemma 3.3
Let $M \in \mathbb{C}^{s \times s}$ be a positive definite matrix, $W \in \mathbb{C}^{s \times k}$, where $k \leqslant s$, and $W^H W = I_k$. Then

$$\lambda_{\min}(W^H M^{-1} W) \geqslant 1/\lambda_{\max}(W^H M W) \tag{27}$$

Proof
From the Cauchy–Schwarz inequality we have

$$(x^H x)^2 = (x^H M^{-\frac{1}{2}} M^{\frac{1}{2}} x)^2 \leqslant \|M^{-\frac{1}{2}} x\|^2 \|M^{\frac{1}{2}} x\|^2 = (x^H M^{-1} x)(x^H M x)$$

and hence

$$\frac{(x^H M^{-1} x)}{x^H x} \frac{(x^H M x)}{x^H x} \geqslant 1 \quad \forall x \in \mathbb{C}^n \tag{28}$$

For simplicity we write $\lambda_{\min}$ and $\lambda_{\max}$ instead of $\lambda_{\min}(W^H M^{-1} W)$, and $\lambda_{\max}(W^H M W)$, respectively. Let $x = Wy$ where $(\lambda_{\min}, y)$ is the eigenpair of $W^H M^{-1} W$. Upon substitution into (28) we obtain

$$\underbrace{\frac{(y^H W^H M^{-1} W y)}{y^H y}}_{\lambda_{\min}} \frac{(y^H W^H M W y)}{y^H y} \geqslant 1 \tag{29}$$

However

$$\lambda_{\max} = \max_{z \neq 0} \frac{(z^H W^H M W z)}{z^H z} \geqslant \frac{(y^H W^H M W y)}{y^H y}$$

and therefore $\lambda_{\min}\ \lambda_{\max} \geqslant 1$, which proves the lemma. $\square$

We present an elementary example which shows that inequality (27) can be sharp.

Let us consider the matrices

$$M = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

We have

$$W^H M W = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad W^H M^{-1} W = \frac{1}{4}\begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

$$1/\lambda_{\max}(W^H M W) = \frac{1}{3} = 0.3333$$

$$\lambda_{\min}(W^H M^{-1} W) = (7 - \sqrt{17})/8 = 0.3596$$

In this case we obtain $\lambda_{\min}(W^H M^{-1} W) > 1/\lambda_{\max}(W^H M W)$.

Using this lemma, we easily estimate $\|r_s\|/\|r_0\|$ for the second and further restarts in the case that the same invariant subspace $\mathscr{Z}$ is added to the actual Krylov subspace. Let us define the space $\mathscr{L} = \mathscr{Z}^{\perp}$. Hence $\mathbb{C}^n = \mathscr{L} \oplus \mathscr{Z}$, $\dim(\mathscr{L}) = n - k$ and the residual vectors r_s lie in $\mathscr{L}$ after each above-mentioned restart. We put $r_0 := r_s$ and $v := r_0/\|r_0\|$ after every restarted run and hence the relation $v \perp \mathscr{Z}$ always holds. Let E_s denote all unit vectors from $\mathbb{C}^s$ having the last k components zero and write $S_{n-k} = S_n \cap \mathscr{L}$.

Theorem 3.4

Let the invariant subspace $\mathscr{Z} = \operatorname{span}\{z_1, z_2, \ldots, z_k\}$ be added to the corresponding Krylov subspace for all restarted runs and let all augmented Krylov subspaces $\mathscr{K}_{m+1}(A, r_0) + \mathscr{Z}$ have dimension $s+1$. Then for all the restarts, except for the first one, we have the estimate

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j > 0} \min_{w \in S_{n-k}} h(m, w, D_\beta) \stackrel{\text{def}}{=} q_3 \tag{30}$$

where $h(m, w, D_\beta)$ was defined by formula (13).

Proof

We use the process described by (8)–(11) substituting in (8) the matrix

$$V := (Av, A^2 v, \ldots, A^m v, z_1, z_2, \ldots, z_k) \tag{31}$$

instead of V. Putting $d_s := U^H v$ where $U = V D_\alpha$ we have

$$\begin{aligned} \|r_s\|^2/\|r_0\|^2 &\leqslant 1 - \|d_s\|^2 \left(\frac{d_s}{\|d_s\|}\right)^H (U^H U)^{-1} \left(\frac{d_s}{\|d_s\|}\right) \\ &\leqslant 1 - \left(\sum_{j=1}^{m} |\alpha_j|^2 |v^H A^j v|^2\right) \left(\min_{y \in E_s} (y^H (U^H U)^{-1} y)\right) \end{aligned} \tag{32}$$

In the last estimate we used the fact that the last k components of d_s equal zero because $v \perp z_i$ for $i \in \{1, 2, \ldots, k\}$. If D_m denotes the rectangular $s \times m$ matrix of the form $D_m = (e_1^s, e_2^s, \ldots, e_m^s)$

and if S_m denotes the unit sphere in $\mathbb{C}^m$ then we can rewrite, putting $\beta_j = |\alpha_j|^2$, the last estimate (32) in the form

$$\begin{aligned}\|r_s\|^2/\|r_0\|^2 &\leqslant 1 - \left(\sum_{j=1}^{m} \beta_j |v^H A^j v|^2\right) \left(\min_{z\in S_m}(z^H(D_m^H(U^H U)^{-1}D_m))z\right) \\ &= 1 - \left(\sum_{j=1}^{m} \beta_j |v^H A^j v|^2\right) \lambda_{\min}(D_m^H(U^H U)^{-1}D_m) \\ &\leqslant 1 - \left(\sum_{j=1}^{m} \beta_j |v^H A^j v|^2\right) \Big/ \lambda_{\max}(D_m^H(U^H U)D_m) \end{aligned} \tag{33}$$

according to Lemma 3.3. But $D_m^H(U^H U)D_m = U_m^H U_m$, where $U_m = (\alpha_1 Av,\ \alpha_2 A^2 v, \ldots, \alpha_m A^m v)$. We continue in our estimate:

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j>0} \min_{w\in S_{n-k}} \left(\sum_{j=1}^{m} \beta_j |w^H A^j w|^2 \Big/ \lambda_{\max}(U_m^H U_m)\right) \tag{34}$$

$$\leqslant 1 - \sup_{\beta_j>0} \min_{w\in S_{n-k}} h(m, w, D_\beta) \tag{35}$$

We have obtained estimate (30). □

Let us sum up all the previous estimates in the following theorem.

Theorem 3.5
Let $m, k, s \in \{1, 2, \ldots, n-1\}$, $s = m + k$ and all augmented Krylov subspaces have dimension $s + 1$. Let r_0 be the initial residual vector and $\hat{r}$ the residual vector after the ith restart of GMRES(m, k) if arbitrary k vectors are added to the Krylov subspace in every restart. Then

$$\|\hat{r}\| \leqslant q_1^{i/2} \|r_0\| \tag{36}$$

where the expression on the right-hand side of (16) is substituted for q_1.

If an invariant subspace $\mathscr{Y}$ is added to the Krylov subspace in all restarts then

$$\|\hat{r}\| \leqslant q_2 q_3^{(i-1)/2} \|r_0\| \tag{37}$$

where q_2 is defined by formula (25) and the expression on the right-hand side of (30) can be substituted for q_3.

We have seen from the Theorem 3.1 that if arbitrary vectors are added to the Krylov subspace in every restart then the estimate is not better than the estimate for GMRES(m) (see Reference [16]). But the improvement of the convergence yields the augmentation of the Krylov subspace with the same invariant subspace in all restarts. Theorem 3.5, moreover, yields sufficient conditions for the convergence of GMRES(m, k). We will present them only for the case when an invariant subspace is added. The proof is analogous to that of Theorem 2.1 in Reference [16] and, therefore, will be omitted.

Corollary 3.6
Let $m, k, s \in \{1, 2, \ldots, n-1\}$, $s = m + k$ and all augmented Krylov subspaces have dimension $s + 1$. Let the same invariant subspace $\mathscr{Z}$ be added to the Krylov subspace in all restarted runs. Write $S_{n-k} = S_n \cap \mathscr{L}$ where $\mathscr{L} = \mathscr{Z}^{\perp}$. Let at least one of these conditions be valid.

(i) There exists $j \in \{1, \ldots, m\}$ such that $\min_{w \in S_{n-k}} |w^H A^j w| > 0$.
(ii) There exists $j \in \{1, \ldots, m\}$ such that for every $w \in S_{n-k}$, for which $w^H A^j w = 0$, an integer $l = l(w) \in \{1, \ldots, m\}$ exists such that $|w^H A^l w| > 0$.
(iii) The system of equations

$$w^H A^j w = 0 \quad \text{for } j = 1, \ldots, m \tag{38}$$

has only the solution $w = 0$ in $\mathscr{L}$.

Then GMRES(m, k) is convergent independently of the choice of an initial guess.

Remark 3.6
Let us compare estimates (30) and (6) for $m = 1$ in the case that the exact invariant subspace $\mathscr{Z}$ is added to the Krylov space and $P_{\mathscr{Z}}$ is the orthogonal projection into $\mathscr{Z}$. In this case $\varepsilon = 0$. Estimate (6) has the form

$$\|r_s\|^2 \leqslant \min_{\alpha \in \mathbb{C}^1} \|(I - \alpha A)(I - P_{\mathscr{Z}})r_0\|^2$$

From this we obtain, for $r_0 \in \mathscr{L}$, the estimate

$$\|r_s\|^2/\|r_0\|^2 = \|r_s\|^2/\|(I - P_{\mathscr{Z}})r_0\|^2 \leqslant \min_{\alpha \in \mathbb{C}^1} (1 - \bar{\alpha} v^H A^H v - \alpha v^H A v + |\alpha|^2 v^H A^H A v)$$

where $v = r_0/\|r_0\| \in S_{n-k}$. We obtain the minimum for $\alpha = v^H A^H v / v^H A^H A v$ and the estimate has the form

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - |v^H A v|^2 / v^H A^H A v \tag{39}$$

Hence

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \min_{w \in S_{n-k}} |w^H A^j w|^2 / (Aw)^H (Aw) \tag{40}$$

But the last estimate is identical with (35) for $m = 1$ and $\beta_1 = 1$.

In the last part of this section we investigate the case that the added vectors lie in a small neighbourhood of the invariant subspace. Let us formulate this idea more precisely.

Let the space $\mathscr{Y} = \operatorname{span}\{y_1, y_2, \ldots, y_n\}$ be added to the Krylov subspace in some restart with the starting residual vector r_0. After one restarted run we obtain the residual r_s and we determine new vectors $y_{1+}, y_{2+}, \ldots, y_{k+}$ which are added to the Krylov subspace in the next restart. Let $\mathscr{Y}_+ = \operatorname{span}\{y_{1+}, y_{2+}, \ldots, y_{k+}\}$. First we investigate the case that $\Theta(A\mathscr{Y}, A\mathscr{Y}_+) \leqslant \varepsilon$ for some small real number ε. Let us remark that the Min-Res condition yields the orthogonality condition $r_s \perp A\mathscr{Y}$.

Lemma 3.7
Let the space $\mathscr{Y}$ be added to the Krylov subspace $\mathscr{K}_m(A, r_0)$ and the residual vector r_s be obtained after one restarted run of GMRES(m, k). Let us write $r_{0+} = r_s$ and let the space $\mathscr{Y}_+$

be added to $\mathscr{K}_m(A, r_{0+})$ in the next restart with the resultant residual vector r_{s+}. We assume that the augmented spaces $\mathscr{K}_{m+1}(A, r_0) + A\mathscr{Y}$ and $\mathscr{K}_{m+1}(A, r_{0+}) + A\mathscr{Y}_+$ have maximal dimension,

$$\|Ay_{i+}\| = 1 \quad \forall i \quad \text{and} \quad \Theta(A\mathscr{Y}, A\mathscr{Y}_+) \leqslant \varepsilon \tag{41}$$

Let $S_n^+ = \{w \in S_n | w \perp A\mathscr{Y}\}$

Then the numbers $\vartheta_i \in [0,1]$, $i = 1, \dots, k$ exist such that

$$\|r_{s+}\|^2/\|r_{0+}\|^2 \leqslant 1 - \sup_{\substack{\beta_j > 0 \\ j=1,\dots,m}} \min_{w \in S_n^+} h(m, w, D_\beta) - \varepsilon^2 \sup_{\substack{\beta_{m+j} > 0 \\ j=1,\dots,k}} \sum_{j=1}^{k} \beta_{m+j}\vartheta_j \Big/ \sum_{j=1}^{k} \beta_{m+j} \tag{42}$$

$$\leqslant 1 - \sup_{\substack{\beta_j > 0 \\ j=1,\dots,m}} \min_{w \in S_n^+} h(m, w, D_\beta) \tag{43}$$

where $h(m, w, D_\beta)$ was defined by formula (13).

Proof

Let us write $v_+ = r_{0+}/\|r_{0+}\|$. As we have mentioned,

$$v_+ \perp A\mathscr{Y} \tag{44}$$

If we define the matrix

$$U := (\alpha_1 A v_+, \alpha_2 A^2 v_+, \dots, \alpha_m A^m v_+, \alpha_{m+1} A y_{1+}, \alpha_{m+2} A y_{2+}, \dots, \alpha_{m+k} A y_{k+})$$

where $\prod_{i=1}^{m+k} \alpha_i \neq 0$, then according to our previous investigations

$$r_{s+} = r_{0+} - U(U^H U)^{-1} U^H r_{0+} \quad \text{and}$$
$$\|r_{s+}\|^2 = \|r_{0+}\|^2 - r_{0+}^H U(U^H U)^{-1} U^H r_{0+}$$

Let us define the matrices

$$U_1 := (\alpha_1 A v_+, \alpha_2 A^2 v_+, \dots, \alpha_m A^m v_+), \quad U_2 := (\alpha_{m+1} A y_{1+}, \alpha_{m+2} A y_{2+}, \dots, \alpha_{m+k} A y_{k+})$$

i.e. $U = (U_1, U_2)$. For the quotient $\|r_{s+}\|/\|r_{0+}\|$ we have the estimate

$$\begin{aligned}\|r_{s+}\|^2/\|r_{0+}\|^2 &= 1 - v_+^H U(U^H U)^{-1} U^H v_+ \\ &= 1 - x^H (U^H U)^{-1} x - y^H (U^H U)^{-1} y - 2\mathscr{R}e\, x^H (U^H U)^{-1} y\end{aligned} \tag{45}$$

where we write

$$x^H = (v_+^H U_1, \underbrace{0, \dots, 0}_{k\text{-times}}), \qquad y^H = (\underbrace{0, \dots, 0}_{m\text{-times}}, v_+^H U_2)$$

According to Lemma 3.3 we have the following estimates for the second and third term in (45):

$$x^H(U^H U)^{-1}x \geqslant \min_{w\in S_n^+}\left(\sum_{j=1}^{m}|\alpha_j|^2|w^H A^j w|^2\Big/\lambda_{\max}(U_1^H U_1)\right) \tag{46}$$

$$y^H(U^H U)^{-1}y \geqslant \|U_2^H v_+\|^2/\lambda_{\max}(U_2^H U_2) \tag{47}$$

In the final estimates we have only the numbers $|\alpha_j|^2$ for $j\in\{1,2,\dots,m+k\}$. Hence α_j can be choosen in such a way that $\mathscr{Re}\, x^H(U^H U)^{-1}y\geqslant 0$. Therefore this term can be crossed out in estimate (45). In accordance with (41) for every $j\in\{1,2,\dots,k\}$, a vector $\hat{y}_j\in\mathscr{Y}$ exists such that the inequality

$$\|Ay_{j+}-A\hat{y}_j\|\leqslant\varepsilon$$

holds. The orthogonal projection of Ay_{j+} into $A\mathscr{Y}$ can be taken for $A\hat{y}_j$. Moreover, due to orthogonality conditions (44) the relations

$$|v_+^H Ay_{j+}| = |v_+^H A\hat{y}_j + v_+^H(Ay_{j+}-A\hat{y}_j)| = |v_+^H(Ay_{j+}-A\hat{y}_j)|\leqslant\varepsilon$$

hold. Hence numbers $\zeta_j\in[0,1]$ exist such that

$$|v_+^H Ay_{j+}| = |v_+^H(Ay_{j+}-A\hat{y}_{j+})| = \zeta_j\varepsilon$$

and we have from (47) that

$$y^H(U^H U)^{-1}y\geqslant \varepsilon^2\left(\sum_{j=1}^{k}\beta_{m+j}\vartheta_j\right)\Big/\sum_{j=1}^{k}\beta_{m+j} \tag{48}$$

where we have put $\beta_{m+j}=|\alpha_{m+j}|^2$ and $\vartheta_j=\zeta_j^2$. If we substitute estimates (46), (48) into (45) we obtain inequality (42) after some short manipulation. Inequality (43) is evident. □

Let $\mathscr{X}$ be an invariant space. We have defined $\mathscr{L}=\mathscr{X}^\perp$. Let $\hat{\mathscr{L}}$ denote the set of all vectors $v\in\mathbb{C}^n$ such that $\sin\angle(v,\mathscr{L})\leqslant\varepsilon$ and write $\hat{S}=\hat{\mathscr{L}}\cap S_n$.

Theorem 3.8

Let the assumptions of Lemma 3.7 be valid. Let relations (41) be substituted by the following ones:

$$\|Ay_{i+}\|=1,\quad \Theta(A\mathscr{Y},\mathscr{X})\leqslant\varepsilon\quad\text{and}\quad\Theta(A\mathscr{Y}_+,\mathscr{X})\leqslant\varepsilon \tag{49}$$

Then the numbers $\vartheta_j\in[0,1]$ exist such that

$$\|r_{s+}\|^2/\|r_{0+}\|^2 \leqslant 1-\sup_{\substack{\beta_j>0\\ j=1,\dots,m}}\min_{w\in\hat{S}_n} h(m,w,D_\beta) - 4\varepsilon^2\sup_{\substack{\beta_{m+j}>0\\ j=1,\dots,k}}\sum_{j=1}^{k}\beta_{m+j}\vartheta_j\Big/\sum_{i=1}^{k}\beta_{m+j} \tag{50}$$

$$\leqslant 1-\sup_{\substack{\beta_j>0\\ j=1,\dots,m}}\min_{w\in\hat{S}_n} h(m,w,D_\beta) \tag{51}$$

where $h(m,w,D_\beta)$ was defined by formula (13).

Proof
Let $P_{A\mathscr{Y}}$, $P_{A\mathscr{Y}_+}$, $P_{\mathscr{X}}$ and $P_{\mathscr{L}}$ denote the orthogonal projections onto $A\mathscr{Y}$, $A\mathscr{Y}_+$, $\mathscr{X}$ and $\mathscr{L}$, respectively. We have

$$\begin{aligned}\Theta(A\mathscr{Y},A\mathscr{Y}_+) &= \|P_{A\mathscr{Y}} - P_{A\mathscr{Y}_+}\| = \|P_{A\mathscr{Y}} - P_{\mathscr{X}} + P_{\mathscr{X}} - P_{A\mathscr{Y}_+}\| \\ &\leqslant \|P_{A\mathscr{Y}} - P_{\mathscr{X}}\| + \|P_{\mathscr{X}} - P_{A\mathscr{Y}_+}\| = \Theta(A\mathscr{Y},\mathscr{X}) + \Theta(A\mathscr{Y}_+,\mathscr{X}) \leqslant 2\varepsilon\end{aligned}$$

Moreover, we have

$$\Theta(\mathscr{L},P_{A\mathscr{Y}}^{\perp}\mathbb{C}^n) = \|P_{\mathscr{X}}^{\perp} - P_{A\mathscr{Y}}^{\perp}\| = \|I - P_{\mathscr{X}} - (I - P_{A\mathscr{Y}})\| = \|P_{A\mathscr{Y}} - P_{\mathscr{X}}\| = \Theta(A\mathscr{Y},\mathscr{X}) \leqslant \varepsilon$$

where $P_{A\mathscr{Y}}^{\perp}$ denotes the orthogonal projection onto the orthogonal complement of the subspace $A\mathscr{Y}$. Because $v_+ = r_{0,+}/\|r_{0,+}\| \perp A\mathscr{Y}$ it is $v_+ \in P_{A\mathscr{Y}}^{\perp}\mathbb{C}^n$ and hence $\sin\angle(v_+,\mathscr{L}) \leqslant \varepsilon$, i.e. $v_+ \in \hat{S}$. The rest of the proof is the same as in the previous lemma. We only have to substitute $2\varepsilon \to \varepsilon$ in formula (42), and we obtain (50). □

We can see that for $\varepsilon \to 0$, estimate (50) converges to estimate (30), i.e. to the estimates for the case that an invariant subspace is added to the Krylov subspace. Therefore if the right-hand side of (30) is less than 1 then a number $\varepsilon_0 > 0$ exists such that the right-hand side of (50) is less than 1 for all $\varepsilon \in [0,\varepsilon_0]$. Hence, if for $l = 1,2,\ldots$ in the lth restarted run the space $\mathscr{Y}_l$, that approximates well some invariant subspace $\mathscr{X}$, is added to the Krylov subspace then the condition $1 - \sup_{\beta_j>0}\min_{w\in S_{n-k}} h(m,w,D_\beta) < 1$ (yielding the convergence of GMRES(m,k) by the addition of $\mathscr{X}$ in all restarts) implies the convergence of GMRES(m,k) by the addition of $\mathscr{Y}_l$ in the lth restart ($l = 1,2,\ldots$).

4. NUMERICAL CONCLUSIONS AND REMARKS

In this section, we will present very shortly some examples assessing numerically the quality of the proposed bound (30), i.e.

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j>0}\min_{w\in S_{n-k}} h(m,w,D_\beta) \stackrel{\text{def}}{=} q_3$$

and (16), i.e.

$$\|r_s\|^2/\|r_0\|^2 \leqslant 1 - \sup_{\beta_j>0}\min_{w\in S_n} g(m,w,D_\beta,Y_k) \stackrel{\text{def}}{=} q_1$$

Example 1
Let us consider the tridiagonal matrix $A \in \mathbb{R}^{10\times 10}$ of the form

$$A = \text{tri}[-0.7, 2.5, -1.1]$$

Let (λ_i, v_i) denote the eigenpairs of A. For information, all eigenvalues are real, $\lambda_1 = 4.184$, $\lambda_{10} = 0.816$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_{10}$. We consider restart $m = 3$. We put $k = 0$ if no vector is

Table I. Behaviour of $1-\min_{w\in S_{n-k}} h(3,w,D_\beta)$.

β_1	β_2	β_3	k	Added vectors	$1-\min_{w\in S_{n-k}} h(3,w,D_\beta)$
1.0	1.0	1.0	0		0.724
10^3	1.0	1.0	0		0.481
10^3	1.0	1.0	1	v_{10}	0.403
10^3	1.0	1.0	2	v_9, v_{10}	0.309

Table II. Behaviour of $1-\min_{w\in S_n} g(3,w,D_\beta,Y_1)$.

β_1	β_2	β_3	k	Added vector	$1-\min_{w\in S_n} g(3,w,D_\beta,Y_1)$
1.0	1.0	1.0	0		0.724
10^3	1.0	1.0	1	v_{10}	0.481
10^3	1.0	1.0	1	e_1	0.482
1.0	10^3	1.0	1	e_1	0.879
1.0	1.0	10^3	1	e_1	0.974

appended to the Krylov subspace. We start with $D_\beta = I$ and then we choose some D_β giving a remarkable decrease.

Estimate (30) gives the following estimate for q_3 (see Table I):

$$q_3 \leqslant 0.403 \text{ for } k=1,\ q_3 \leqslant 0.309 \text{ for } k=2 \text{ and } q_3 \leqslant 0.724 \text{ for GMRES(3)}$$

Estimate (16) gives $q_1 \leqslant 0.481$ being worse than (30) for q_3 (see Table II). Let us remark that e_1 denotes the first column of the identity matrix I_{10}.

We try to explain the behaviour of the number $\min_{w\in S_n} g(3,w,D_\beta,Y_1)$ in dependence on D_β. It is

$$\min_{w\in S_{10}\cap\mathbb{R}^{10}} \frac{|w^{\mathrm{T}}Aw|^2}{\|Aw\|^2} \gg \min_{w\in S_{10}\cap\mathbb{R}^{10}} \frac{|w^{\mathrm{T}}A^j w|^2}{\|A^j w\|^2}$$

for $j=2,3$ and the dominating position of this quotient is set off by choosing $\beta_1 \gg \beta_j$ for $j=2,3$. The following theorem proves that the dominant quotient between $|w^H A^j w|^2/\|A^j w\|^2$ $j\in\{1,2,\dots,m\}$ plays a leading role for the magnitude of q_1.

Theorem 4.1

Let $j_0 \in \{1,2,\dots,m\}$ exists such $\min_{w\in S_n} |w^H A^{j_0} w| > 0$. Then for q_1 the following estimate

$$q_1 \leqslant 1 - \max_{j=1,\dots,m} \left\{ \min_{w\in S_n} \frac{|w^H A^j w|^2}{\|A^j w\|^2} \right\} \tag{52}$$

holds.

Proof
Let $\varepsilon > 0$ be an arbitrary real number. For every matrix Y_k a diagonal matrix $\hat{D}_\beta^{(j)}$ depending on Y_k exists such that

$$g(m, w, \hat{D}_\beta^{(j)}, Y_k) \geqslant \frac{|w^H A^{j_0} w|^2}{\|A^{j_0} w\|^2} \frac{1}{1+\varepsilon} \quad \forall w \in S_n$$

Consequently,

$$\min_{w \in S_n} g(m, w, \hat{D}_\beta^{(j)}, Y_k) \geqslant \min_{w \in S_n} \frac{|w^H A^j w|^2}{\|A^j w\|^2} \frac{1}{1+\varepsilon} \quad \forall j \in \{1, 2, \ldots, m\}$$

and hence

$$\sup_{\beta_j > 0} \min_{w \in S_n} g(m, w, D_\beta\, Y_k) \geqslant \min_{w \in S_n} \frac{|w^H A^j w|^2}{\|A^j w\|^2} \quad \forall j \in \{1, 2, \ldots, m\}$$

The last inequality immediately yields

$$\sup_{\beta_j > 0} \min_{w \in S_n} g(m, w, D_\beta, Y_k) \geqslant \max_{j=1,\ldots,m} \left\{ \min_{w \in S_n} \frac{|w^H A^j w|^2}{\|A^j w\|^2} \right\} \tag{53}$$

Let us remark that

$$\min_{w \in S_n} \frac{|w^H A y_j|^2}{\|A y_j^2\|} = 0$$

Inequality (52) is a consequence of (53) and Theorem 3.1. □

Remark 4.2
An analogous estimate can be derived for q_3. We have

$$q_3 \leqslant 1 - \max_{j=1,\ldots,m} \left\{ \min_{w \in S_{n-k}} \frac{|w^H A^j w|^2}{\|A^j w\|^2} \right\} \tag{54}$$

We can see that the estimates for q_1 given by (16) depend on the added vectors $y_1, y_2, \ldots, y_k$ and if we eliminate this dependence we obtain an estimate which is identical with the estimate for GMRES(m) (see Reference [16]). The uniform estimate in the whole S_n eliminates the dependence on the vectors $y_1, y_2, \ldots, y_k$ in the estimate presented in Theorem 4.1. Therefore, we can expect that if we augment the Krylov subspace with arbitrary vectors the convergence need not be better than restarted GMRES. This is confirmed by numerical experiments [11, 17] and results in a nice paper [18, Section 4] and our example too. (Compare the second rows in Tables I and II.) Comparing the second and third row in Table II only confirms the above-mentioned fact, that the estimates for q_1 need not depend on appended vectors. However both the estimates are very sensitive in relation to β.

Table III. Behaviour of $1-\min_{w\in S_{n-k}} h(3,w,D_\beta)$.

β_1	β_2	β_3	k	Added vectors	$1-\min_{w\in S_{n-k}} h(3,w,D_\beta)$
1.0	1.0	1.0	0		1.0
1.0	1.0	1.0	1	v_{10}	1.0
1.0	1.0	1.0	1	v_9	$1-0.678\times10^{-4}$
1.0	1.0	1.0	2	v_{10}, v_9	$1-0.417\times10^{-3}$
10^3	1.0	1.0	3	$v_{10}, v_9, \mathscr{Re}\, v_8, Im\, v_8$	0.6516
10^3	1.0	1.0	4	$v_{10}, v_9, \mathscr{Re}\, v_8, Im\, v_8, v_6$	0.4371

Table IV. Convergence of GMRES$(3,k)$ for k considered in Table III.

Method	Added vectors	Number of restarts	$\log_{10}\|r_k\|/\|r_0\|$
GMRES$(3,0)$		600	8.463×10^{-6}
GMRES$(3,1)$	v_{10}	600	1.147×10^{-6}
GMRES$(3,1)$	v_9	600	3.952×10^{-6}
GMRES$(3,2)$	v_{10}, v_9	395	9.853×10^{-11}
GMRES$(3,3)$	$v_{10}, v_9, \mathscr{Re}\, v_8, Im\, v_8$	20	3.029×10^{-11}
GMRES$(3,4)$	$v_{10}, v_9, \mathscr{Re}\, v_8, Im\, v_8, v_6$	14	8.429×10^{-11}

Example 2
Let

$$A_1=\begin{pmatrix} 0.4163 & 0.3176 & 0 & 0 & 0\\ 0.0001 & 0.4132 & 0.8175 & 0 & 0\\ 0.6321 & 0.3157 & 0.4823 & 0.6614 & 0\\ 0.5157 & 0.8321 & 0.5642 & 0.6541 & 0.4321\\ 0.5563 & 0.4431 & 0.2567 & 0.8325 & 0.8475 \end{pmatrix}$$

let the matrix A_2 have the numbers 4, 3, 2, 1, 0.01 on the main diagonal and 0.1 on the super diagonal and put $A=\mathrm{diag}(A_1,A_2)$, i.e. $A\in\mathbb{R}^{10\times10}$. For the matrix A_1 (see Reference [19]). We restrict only on the behaviour of q_3. The estimates are present only for β_i giving approximately the smallest last column.

According to Table III, we cannot say whether GMRES(3) or GMRES$(3,1)$ are convergent. The convergence can be slow due to small last eigenvalue. However GMRES$(3,k)$ is convergent for $k=2,3,4$ according to Theorem 3.4.

Calculating iterations of GMRES$(3,k)$ for $k=0,1,2,3,4$, (GMRES$(3,0)\equiv$GMRES(3)), we have considered the first iteration for $\log_{10}\|r_k\|/\|r_0\|$ to be less than 10^{-10}. Here r_0 is an initial residual vector. The results are shown in Table IV. Maximum number of restarts equals 600.

Although estimates for q_3 do not describe the true progress of $\log_{10}\|r_k\|/\|r_0\|$, the number of restarts descends in accordance with the decrease of estimate for q_3 in convergent cases.

All calculations were performed on a PENTIUM PC using double precision arithmetic.

ACKNOWLEDGEMENTS

The author gratefully thanks both the anonymous referees for very valuable remarks leading to improving this paper.

This paper was supported by the Grant Agency of the Czech Republic under Grant No. 201/02/0684 and under Grant MSM 113200007.

REFERENCES

1. Baglama J, Calvetti D, Golub GH, Reichel L. Adaptively preconditioned GMRES algorithms. *SIAM Journal on Scientific Computing* 1998; **20**(1):243–269.
2. Burrage K, Erhel J. On the performance of various adaptive preconditioned GMRES strategies. *Numerical Linear Algebra with Applications* 1998; **5**:101–121.
3. Erhel J, Burrage K, Pohl B. Restarted GMRES preconditioned by deflation. *Journal of Computational and Applied Mathematics* 1996; **69**:303–318.
4. Morgan RB. A restarted GMRES method augmented with eigenvectors. *SIAM Journal on Matrix Analysis and Applications* 1995; **16**(4):1154–1171.
5. Golub GH, Van Loan CHF. *Matrix Computation* (2nd edn). The John Hopkins University Press: Baltimore, London, 1990.
6. Householder AS. *The Theory of Matrices in Numerical Analysis.* Blaisdell: New York, 1964.
7. Simoncini V. A new variant of restarted GMRES. *Numerical Linear Algebra with Applications* 1999; **6**: 61–77.
8. Zítko J. Using successive approximations for improving the convergence of GMRES method. *Applications of Mathematics* 1998; **43**:321–350.
9. Brandts JH. Explanation of a phenomenon witnessed in pre-processed GMRES. *Proceedings of the 3rd European Conference Numerical Mathematics and Advanced Applications*. World Scientific: Singapore, 2000; 440–447.
10. Walker HF, Zhou Lu. A simpler GMRES. *Numerical Linear Algebra with Applications* 1994; **1**(6):571–588.
11. Chapman A, Saad Y. Deflated and augment Krylov subspace techniques. *Numerical Linear Algebra with Applications* 1997; **4**:43–66.
12. Saad Y, Schultz MH. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
13. Saad Y. *Iterative Methods for Sparse Linear Systems.* PWS Publishing Company: Boston, MA, 1996.
14. Krasnoselskii MA, Vainiko GM, Zabreiko PP, Rutickii JaB, Otecenko VJa. *Approximate Solution of Operator Equations.* Nauka: Moscow, 1969.
15. Saad Y. Analysis of augmented Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications* 1997; **18**(2):435–449.
16. Zítko J. Generalization of convergence conditions for a restarted GMRES. *Numerical Linear Algebra with Applications* 2000; **7**:117–131.
17. Tajovský M. Augmentation of Krylov subspace by using the GMRES method. *Diploma work*, Faculty of Mathematics and Physics, Charles University of Prague, Czech Republic, 2001.
18. Eiermann M, Ernst OG, Schneider O. Analysis of acceleration strategies for restarted minimal residual methods. *Journal of Computational and Applied Mathematics* 2000; **123**:261–292.
19. Gregory RT, Karney DL. *The Collection of Matrices for Testing Computational Algorithms.* Wiley: New York, 1969.
20. Saad Y. Further analysis of minimum residual iterations. *Numerical Linear Algebra with Applications* 2000; **7**:67–93.
21. Sorensen DC. Implicit application of polynomial filters in k-step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**(1):357–385.
22. Saad Y. *Numerical Methods for Large Eigenvalue Problems.* Halstead Press: NY, 1992.
23. Weiss R. *Parameter-Free Iterative Linear Solvers.* Academie Verlag: Berlin, 1996.
24. Zítko J. Improving the convergence of iterative methods. *Applications of Mathematics* (*Aplikace Matematiky*) 1983; **28**:215–229.
25. Eiermann M, Ernst OG. Geometric aspects in the theory of Krylov subspace methods. *Acta Numerica* 2000; 1–61.
26. Morgan RB. GMRES with deflated restarting. *SIAM Journal on Scientific Computing* 2002; **24**(1):20–37.

