The rate of convergence of the GMRES method

Report 90-77

C. Vuik

TU Delft

Faculteit der Technische Wiskunde en Informatica
Faculty of Technical Mathematics and Informatics

Technische Universiteit Delft
Delft University of Technology

ISSN 0922-5641

Copies of these reports may be obtained from the bureau of the Faculty of Technical Mathematics and Informatics, Julianalaan 132, 2628 BL Delft, phone +3115784568.
A selection of these reports is available in PostScript form at the Faculty's anonymous ftp-site. They are located in the directory /pub/publications/tech-reports at ftp.twi.tudelft.nl

The superlinear convergence behaviour of GMRES

H. A. Van der Vorst * and C. Vuik †

April 7, 1994

Abstract

GMRES is a rather popular iterative method for the solution of nonsingular nonsymmetric linear systems. It is well-known that GMRES often has a so-called superlinear convergence behaviour, i.e., the rate of convergence seems to improve as the iteration proceeds. For the conjugate gradients method this phenomenon has been related to a (modest) degree of convergence of the Ritz values. It has been observed in experiments that for GMRES too, changes in the convergence behaviour seem to be related to the convergence of Ritz values. In this paper we prove that as soon as eigenvalues of the original operator are sufficiently well approximated by Ritz values, GMRES from then on converges at least as fast as for a related system in which these eigenvalues (and their eigenvector components) are missing.

Introduction

In this paper we study the superlinear convergence behaviour of GMRES, observed when solving relevant nonsymmetric nonsingular linear systems. That is, our aim is to understand when and why the method converges faster than in a previous phase of the iteration process. This should not be confused with the fact that one can derive bounds for the residual that describe convergence with almost any desired rate of convergence. These bounds are usually obtained by replacing the actual Krylov iteration polynomial (which produces the actual minimum norm residuals) by polynomials that are the

*Mathematical Institute, University of Utrecht, Budapestlaan 6, Utrecht, the Netherlands

†Department of Technical Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, Delft, the Netherlands

product of factors that vanish at selected eigenvalues of the operator and a Chebyshev polynomial that takes small values over the remaining part of the interval. This is a well-known technique (see, e.g., [3], [1], and [10]), which sometimes leads to useful bounds for the residuals, but which fails to predict the gradual changes that one observes in actual convergence histories.

Our main motivation for this theoretical study are the results of the experiments reported in [6] (see also [12]), which suggest a relation between the convergence of the GMRES iterates and the convergence of the Ritz values to the eigenvalues of the matrix. Such a relation for the conjugate gradients method has been derived in [11]. The main problem in obtaining a similar relation for GMRES is that the Ritz values are not the zeros of the iteration polynomial for GMRES, as is the case for conjugate gradients.
Another difficulty is that a nonsymmetric matrix cannot always be transformed to diagonal form by similarity transformations. One might argue that in finite precision arithmetic defectiveness of a matrix does not play a role, because when a matrix has Jordan blocks of dimension larger than one in exact arithmetic, then in the presence of rounding errors the multiplicity of the eigenvalues will be one in practice. However, the eigenvector matrix will be very ill-conditioned in that case. In our experiments we could not see the difference in convergence behaviour for operators with single eigenvalues, but ill-conditioned eigenvectors for some almost multiple eigenvalues, and operators with those almost multiple eigenvalues replaced by Jordan blocks of appropriate size. Therefore, we believe that it is worthwhile to include the Jordan block case in the analysis, even if we wish to explain the phenomena that one observes in rounding error arithmetic.

Our approach is based upon the relation with GMRES and the Full Orthogonalization Method (FOM), for which the iteration polynomial has the Ritz values as its zeros. In our analysis we assume exact arithmetic. Nevertheless, it appears that our analysis may help to understand phenomena that are observed in actual floating point computation.

In Section 1 we briefly describe GMRES and FOM, and we summarize some relevant properties of these methods.
In Section 2 we present relations between the convergence of GMRES and FOM, and the convergence of the Ritz values.
Finally, in Section 3, we describe some numerical experiments, which illustrate our theoretical results.

1 Definitions and Preliminaries

In this section we recall some fundamental properties of the FOM and the GMRES method [10], [2], which are iterative methods for solving linear systems with a non-symmetric matrix.

Consider the linear system $Ax = b$ with $x, b \in \mathbb{R}^n$ and with a non-singular $A \in \mathbb{R}^{n\times n}$. The Krylov subspace $K^k(A; r_0)$ is defined by $K^k = span\{r_0, Ar_0, ..., A^{k-1}r_0\}$. In both methods, FOM and GMRES, Arnoldi's method is used for the construction of an orthonormal basis $\{v_1, ..., v_k\}$ for $K^k(A; r_0)$. The modified Gram-Schmidt version of Arnoldi's method can be described as follows ([10]:p.857 or [4]: p.155):

1. Start: Choose x_0 and compute $r_0 = b - Ax_0$ and $v_1 = r_0/\|r_0\|_2$,
2. Iterate: For $j = 1, ..., k$ do:
   $v_{j+1} := Av_j$
   for $i = 1, ..., j$ do:
   $h_{ij} := v_{j+1}^T v_i,\ v_{j+1} := v_{j+1} - h_{ij}v_i,$
   $h_{j+1,j} := \|v_{j+1}\|_2,\ v_{j+1} = v_{j+1}/h_{j+1,j}.$

(the non-defined $h_{i,j}$ are assumed to be zero).

With the $n \times k$ matrix $V_k = [v_1, \ldots, v_k]$ we have that $H_k = V_k^T A V_k$ is an upper $k \times k$ Hessenberg matrix whose entries are the scalars h_{ij}.

In the FOM method we construct an approximate solution x_k^F of the form $x_k^F = x_0 + z_k^F$ where z_k^F is an element of $K^k(A; r_0)$ with the following property

$$r_k^F = b - Ax_k^F \perp K^k(A; r_0). \tag{1}$$

Note that if A is symmetric then FOM is equivalent to CG (compare [11]: relation (2.3)). If H_k is nonsingular then it is easy to show that $z_k^F = V_k y_k^F$, where y_k satisfies $H_k y_k^F = \|r_0\|_2 e_1$ and e_1 is the first unit vector in $\mathbb{R}^k$. However, when H_k is singular, it can be proved that a solution x_k^F does not exist (see [2]: Section 3).

We describe the solution method followed in GMRES in little more detail since some of the iteration coefficients are needed in Section 3. In GMRES the approximate solution $x_k^G = x_0 + z_k^G$ with $z_k^G \in K^k(A; r_0)$ is such that

$$\|r_k^G\|_2 = \|b - Ax_k^G\|_2 = \min_{z\in K^k(A;r_0)} \|r_0 - Az\|_2. \tag{2}$$

As a consequence of (2) we have that r_k^G is orthogonal to $AK^k(A;r_0)$, or $r_k^G \perp K^k(A;Ar_0)$. If A is symmetric then GMRES is equivalent to MINRES [7].

Defining the matrix $\bar{H}_k \in I\!R^{k+1\times k}$ as

$$\bar{H}_k = \begin{pmatrix} H_k \\ 0...0 \ h_{k+1,k} \end{pmatrix}$$

it follows that $AV_k = V_{k+1}\bar{H}_k$. Using this equation it is shown in ([10]: Section 3.1) that $x_k^G = x_0 + V_k y_k^G$, where y_k^G solves the following least squares problem

$$\|\beta e_1 - \bar{H}_k y_k^G\|_2 = \min_{y\in I\!R^k} \|\beta e_1 - \bar{H}_k y\|_2, \tag{3}$$

with $\beta = \|r_0\|_2$ and e_1 is the first unit vector in $I\!R^{k+1}$. In contrast with FOM, the approximations x_k^G obtained with GMRES exist for all k ([10]: Section 3.4, or [2]: Section 3).

To solve the least squares problem (3) $\bar{H}_k$ is factorized as $Q_k^T R_k$ by Givens rotations, where $Q_k \in I\!R^{(k+1)\times(k+1)}, Q_k^T Q_k = I_{k+1}$, and $R_k \in I\!R^{(k+1)\times k}$ is an upper triangular matrix.
The matrix Q_k is taken as $Q_k = F_1 \cdots F_k$, where the matrix $F_j \in I\!R^{(k+1)\times(k+1)}$ is the following Givens rotation

$$F_j = \begin{pmatrix} 1 & & & & & & \text{Ø} \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c_j & -s_j & & \\ & & & s_j & c_j & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ \text{Ø} & & & & & & 1 \end{pmatrix},$$

which eliminates the element in the $(j+1,j)$ position of $F_{j-1}F_{j-2}\cdots F_1\bar{H}_k$.

The product $F_{k-1}...F_1\bar{H}_k = \begin{pmatrix} * & \dots & * & * \\ & \ddots & \vdots & \vdots \\ & & * & * \\ & \text{Ø} & 0 & \rho_k \\ & & 0 & h_{k+1,k} \end{pmatrix}$,

(where an asterisk stands for a non-zero element) implies that c_k and s_k should be chosen as follows:

$$c_k = \rho_k / \sqrt{\rho_k^2 + h_{k+1,k}^2} \quad \text{and} \quad s_k = -h_{k+1,k} / \sqrt{\rho_k^2 + h_{k+1,k}^2}\ . \tag{4}$$

Using this factorization the least squares problem (3) is equivalent to

$$\|\beta e_1 - \bar{H}_k y_k^G\|_2 = \min_{y \in \mathbb{R}^k} \|Q_k \beta e_1 - R_k y\|_2. \tag{5}$$

Since the last row of R_k is zero, y_k^G is the solution of the linear system with the leading $k \times k$ submatrix of R_k as matrix and the first k components of $Q_k \beta e_1$ as right-hand side.

We cite the following important results for FOM and GMRES.

Lemma 1 *([10]: p. 862, Proposition 1)*
The GMRES residual satisfies the following equation

$$\frac{\|r_k^G\|_2}{\|r_{k-1}^G\|_2} = |s_k| \ . \tag{6}$$

Lemma 2 *([2]: theorem 5.1)*
If $c_k \neq 0$ then the FOM and the GMRES residuals satisfy the following equation

$$\| \ r_k^G \ \|_2 = | \ c_k \ | \ \| \ r_k^F \ \|_2 \tag{7}$$

Note: For a discussion on the implications of these properties for FOM and GMRES see [2].

The Ritz values and the Ritz vectors are defined as follows (compare [11]: Section 2.3 or [5]: p. 274):

Definition 1 *The Ritz values $\theta_1^{(k)}, ..., \theta_k^{(k)}$ are the eigenvalues of the matrix $H_k = V_k^T A V_k$ (note that $V_k^T V_k = I_k$). If $y_i^{(k)} \in \mathbb{R}^k$ is a normalized eigenvector of H_k corresponding to $\theta_i^{(k)}$ then the vector $z_i^k = V_k y_i^{(k)}$ is called a Ritz vector.*

Definition 2 *For all the iterative methods to be discussed, it follows that the residual r_k at the k-th iteration step is a member of $K^{k+1}(A; r_0)$, and hence it can be written as a k-th degree polynomial in A, acting on r_0. In connection with this we will speak about the polynomials for method M as, the "M polynomial".*

The following result is an immediate consequence of Theorem 5.1 in [8].

Lemma 3 *If $c_k \neq 0$, where c_k is defined in (4) then*

1. *The FOM polynomial satisfies $p_k^F(H_k) = 0$*
2. *$\theta_i^{(k)} \neq 0, i = 1, \cdots, k$*
3. $$p_k^F(t) = \frac{(\theta_1^{(k)} - t) \cdots (\theta_k^{(k)} - t)}{(\theta_1^{(k)}) \cdots (\theta_k^{(k)})}. \tag{8}$$

2 The convergence of FOM and GMRES

In this section we shall prove some relations between the convergence of the Ritz values and the convergence of the FOM and the GMRES method. The proofs of our theorems follow closely the lines set forth in a proof given by ([11]: p. 547) where the matrix A is symmetric and positive definite. In that proof two important properties of the CG method are exploited: an optimality property and the property that the Ritz values are the roots of the "CG polynomial". However, when A is non-symmetric neither the FOM method nor the GMRES method has both properties. This has led us to base our proofs on the optimality property of GMRES (see (2)) together with the fact that the Ritz values are the roots of the FOM polynomial (see Lemma 3).

For each matrix A there exists a nonsingular $n \times n$ matrix X which reduces the matrix A to its Jordan normal form, i.e.

$$X^{-1}AX \equiv J \equiv \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix},$$

where each of the $n_j \times n_j$ submatrices J_j has the form

$$J_j = \begin{pmatrix} \lambda_j & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{pmatrix}.$$

We assume that the J_j have been ordered so that if $\lambda_s = \lambda_t$ for $s > t$ then $\lambda_i = \lambda_t$ for all $t \leq i \leq s$, i.e., blocks with equal λ have been grouped together. Furthermore, we assume that if $\lambda_t = \lambda_{t+1}$ then $n_t \geq n_{t+1}$, i.e., blocks with equal λ have been ordered in descending size. Note that these orderings can be realized by simple symmetric permutations to X.
Occasionally we will write a vector with respect to the basis represented by the columns of X, and this representation will be partitioned correspondingly to the partition of J, in particular $r_0 = X\gamma$, with $\gamma = (\gamma_1, \ldots, \gamma_m)^T$.

Definition 3 *If t is such that $\lambda_1 = \lambda_t$ and $\lambda_{t+1} \neq \lambda_t$, then the blockdiagonal matrix $D^{(1)}$, with $n_j \times n_j$ blocks $D_j^{(1)}$ along its diagonal, corresponding to the blocksizes of J, is defined by*

$$D_j^{(1)} = \begin{cases} 0, & \text{if } j \leq t \\ \prod_{i=1}^{n_1} (\frac{\theta_i^{(k)}}{\lambda_1}(\theta_i^{(k)} I - J_j)^{-1}(\lambda_1 I - J_j)) & \text{if } j > t \end{cases}$$

(we assume that $n_1 \leq k$).
Furthermore we define a blockdiagonal matrix $D^{(2)}$ with a blockstructure corresponding to J as

$$D_j^{(2)} = q_\ell(J_j), \text{ for } j = 1, \ldots, m, \text{ and } q_\ell \in \Pi_\ell^1,$$

where Π_k^1 denotes the class of polynomials of degree at most k and constant term 1.

It is readily verified that $D^{(1)}$ and $D^{(2)}$ commute.
Note that $D^{(2)}\gamma = X^{-1}q_\ell(A)r_0$, so that it describes the iteration effect for the polynomial q_ℓ with respect to the basis defined by X. Multiplication of this vector by $D^{(1)}$ has the effect that the part corresponding to the first Jordan block is removed and the parts corresponding to the other blocks are multiplied by block matrices that converge to identity blocks as the first n_1 Ritz values converge to λ_1 (i.e., roots $\theta_i^{(k)}$ converging to λ_1 are replaced by λ_1, see the corresponding polynomial defined in (9)).

Theorem 1 relates the residuals in a certain phase of GMRES to the residuals of a comparison process in which the influence of λ_1 has been removed. This comparison process is defined as follows.

Definition 4 *Let r_k^G (or r_k^F) denote the residual in the k-th step of GMRES (or FOM) applied to $Ax = b$, with starting residual $r_0 = X\gamma$ i.e., $r_k^G = p_k^G(A)r_0$ (or $r_k^F = p_k^F(A)r_0$).*
For k such that H_k is nonsingular we define a comparison GMRES (or FOM) process which starts with $r_0' = p_k^F(A)X\gamma'$, where $\gamma_j' = 0$ for $j \leq t$ and $\gamma_j' = \gamma_j$ for $j > t$. I.e., the comparison processes start with the k-th residual of FOM, after having deleted all the components corresponding to λ_1. The residual at the ℓ-th step of the comparison GMRES (or FOM) process is denoted by $r_\ell^{G'}$ (or $r_\ell^{F'}$), and the corresponding GMRES polynomial by $q_\ell^{G'}$.

Theorem 1 *Let k be such that H_k is nonsingular and let $n_1 \leq k$, then for all $\ell > 0$ we have that*

$$\|r_{k+\ell}^G\|_2 \leq \|XD^{(1)}X^{-1}\|_2 \|r_\ell^{G'}\|_2.$$

Proof: For the comparison process we have the relation

$$r_\ell^{G'} = q_\ell^{G'}(A)r_0'.$$

We can write $r_\ell^{G'}$ with respect to X as

$$r_\ell^{G'} = Xq_\ell^{G'}(J)p_k^F(J)\gamma'.$$

Now we define the auxiliary polynomial $h_k \in \Pi_k^1$ as:

$$h_k(t) = \prod_{i=1}^{n_1} \frac{\theta_i^{(k)}}{\lambda_1} \frac{\lambda_1 - t}{\theta_i^{(k)} - t} p_k^F(t). \tag{9}$$

(note that $p_k^F(t)$ is defined since H_k is nonsingular).

It follows from the optimality property of GMRES that

$$\begin{aligned} \|r_{k+\ell}^G\|_2 &\leq \|q_\ell^{G'}(A) h_k(A) r_0\|_2 \\ &= \|X D^{(1)} X^{-1} X D^{(2)} p_k^F(J) \gamma'\|_2 \\ &\leq \|X D^{(1)} X^{-1}\|_2 \|r_\ell^{G'}\|_2, \end{aligned} \tag{10}$$

which completes the proof. □

With Lemma 2 we have immediately:

Corollary 1 *If k and ℓ are such that H_k, $H_{k+\ell}$ and H'_ℓ are nonsingular, then*

$$\|r_{k+\ell}^F\|_2 \leq \|X D^{(1)} X^{-1}\|_2 \frac{|c'_\ell|}{|c_{k+\ell}|} \|r_\ell^{F'}\|_2.$$

We obtain the following bound for $\|X D^{(1)} X^{-1}\|_2$:

$$\|X D^{(1)} X^{-1}\|_2 \leq \kappa_2(X) \|D^{(1)}\|_2$$

$$\leq \kappa_2(X) \{\prod_{i=1}^{n_1} \frac{|\theta_i^{(k)}|}{|\lambda_1|}\} \max_{\lambda_j \neq \lambda_1} (\prod_{i=1}^{n_1} \{|\frac{\lambda_j - \lambda_1}{\lambda_j - \theta_i^{(k)}}| + \sum_{p=2}^{n_j} |\frac{\lambda_1 - \theta_i^{(k)}}{(\lambda_j - \theta_i^{(k)})^p}|\}). \tag{11}$$

In the derivation of this bound we have used that $(\alpha I - J_j)^{-1}$ and $(\beta I - J_j)$ commute, and, furthermore, that $\|(\alpha I - J_j)^{-1}(\beta I - J_j)\|_2 \leq \|(\alpha I - J_j)^{-1}(\beta I - J_j)\|_\infty$.
In order to prove this inequality we use the norm inequality $\| \ \|_2 \leq \sqrt{\| \ \|_1 \| \ \|_\infty}$ ([5]: Corollary 2.3.2). By writing out the matrix expression in the norm one may verify that $\|(\alpha I - J_j)^{-1}(\beta I - J_j)\|_1 = \|(\alpha I - J_j)^{-1}(\beta I - J_j)\|_\infty$.

When A is diagonalizable, i.e., $n_i = 1$, for all i, then this rather unpleasant expression reduces to

$$\|X D^{(1)} X^{-1}\|_2 \leq \kappa_2(X) \cdot \frac{|\theta_1^{(k)}|}{|\lambda_1|} \cdot \max_{\lambda_j \neq \lambda_1} |\frac{\lambda_j - \lambda_1}{\lambda_j - \theta_1^{(k)}}| \equiv F_k \kappa_2(X). \tag{12}$$

This expression is quite similar as the expression derived in [11] for the CG process. In [11] the factor $\kappa_2(X)$ vanishes, since for symmetric A we have

that X is orthogonal.
When there is a Jordan block with λ_1 of dimension $n_1 \neq 1$ then we have to wait for a phase in the GMRES process in which there are n_1 Ritz values close to λ_1. From then on the factor $\|XD^{(1)}X^{-1}\|_2$ is bounded by $\kappa_2(X)$ times a modest constant, and we may expect that the method will further converge about as fast as for a problem in with λ_1 is absent. This is in quite good agreement with an experiment in ([6]: p.22 and Fig.23).

Theorem 1 says that from the k-th step on, the continued GMRES process behaves, except for a certain factor, as a GMRES process that has been started with the FOM result after k steps, in which the components corresponding to λ_1 have been deleted. In order to get this factor small enough, the value of k should be chosen so large that $\theta_1^{(k)}$ is relatively close to λ_1. The weak point in this comparison is that we relate the convergence behaviour of the continued GMRES process to the result of a FOM process, whereas it would be more natural to use the result of the first k GMRES iterations. In view of the relation between GMRES and FOM the theorem is then only of practical value in situations where $|c_k|$ is not too far from 1, since $\|r_k^F\|_2 = \frac{1}{|c_k|}\|r_k^G\|_2$. In that case we may compare the continued GMRES process from the k-th step on with another GMRES process that has been started with a slightly modified GMRES residual at the k-th step. The assumption about $|c_k|$ does not put a too severe restriction on the applicability of the theorem, since it holds as soon as there is a noticeable reduction in the norms of the residuals in GMRES (cf. Lemma 1).

Theorem 1 may also be used in combination with convergence estimates for the GMRES method, and then it is a powerful tool to analyse the actual convergence behaviour of GMRES. Such convergence estimates for the GMRES process are given in ([10]: Section 3.4). Note that, with Lemma 2 and Corollary 1 this procedure can also be followed for the FOM process.
However, straight-forward application of Theorem 1 with, for instance, ([10]: p. 866, Proposition 4) leads to a bound for the continued GMRES process which contains the factor $\kappa_2(X)^2$ instead of the $\kappa_2(X)$. The following theorem does not have this disadvantage and it also relates the residuals of the continued GMRES process with the residuals of a related GMRES process. For simplicity this theorem has been formulated for the situation that all $n_j = 1$, i.e., that A is diagonalizable. The extension to the Jordan form case is rather straight-forward.

Theorem 2 *Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable. Let k be such that H_k is*

nonsingular and let $\varepsilon^{(\ell)}$ *be defined as follows*

$$\varepsilon^{(\ell)} = \min_{q \in \Pi_\ell^1} \max_{\lambda_i \neq \lambda_1} |q(\lambda_i)|$$

(compare [10]: p. 866, Proposition 4) then

$$\|r_{k+\ell}^G\|_2 \leq \frac{F_k}{|c_k|} \kappa_2(X) \varepsilon^{(\ell)} \|r_k^G\|_2,$$

where F_k *has been defined in equation (12).*

Proof: We use h_k as in (9) and we define another auxiliary polynomial $q_\ell \in \Pi_\ell^1$ as follows: q_ℓ is the polynomial for which

$$\varepsilon^{(\ell)} = \max_{\lambda_i \neq \lambda_1} |q_\ell(\lambda_i)|.$$

For existence and uniqueness of this polynomial we refer to ([9]: p. 115). Since $q_l h_k \in \Pi_{k+\ell}^1$ we obtain

$$\|r_{k+\ell}^G\|_2 \leq \|q_\ell(A) h_k(A) r_0\|_2 = \|X D^{(1)} D^{(2)} X^{-1} X \begin{pmatrix} p_k^F(\lambda_1) & & \emptyset \\ & \ddots & \\ \emptyset & & p_k^F(\lambda_n) \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} \|_2$$

$$\leq \kappa_2(X) F_k \varepsilon^{(\ell)} \|r_k^F\|_2 = \kappa_2(X) F_k \varepsilon^{(\ell)} \frac{1}{|c_k|} \|r_k^G\|_2,$$

which proves the theorem. □

Theorem 2 can be generalized as follows (compare [11]: Theorem 5.1):

Theorem 3 *Let* $A \in \mathbb{R}^{n \times n}$ *be diagonalizable and let* k *be such that* H_k *is nonsingular. Let* Λ_r *denote a set of* r *different eigenvalues of* A*, and* $\Lambda_r^c \equiv \{\lambda_i | \lambda_i \neq \lambda_j \text{ for all } \lambda_j \in \Lambda_r\}$*.* $\varepsilon^{(\ell,r)}$ *is defined as* $\varepsilon^{(\ell,r)} = \min_{q \in \Pi_\ell^1} \max_{\lambda_i \in \Lambda_r^c} |q(\lambda_i)|$ *and*

$$F_{k,r} = \max_{\lambda_j \in \Lambda_r^c} \prod_{\lambda_i \in \Lambda_r} \frac{|\theta_i^{(k)}|}{|\lambda_i|} \frac{|\lambda_j - \lambda_i|}{|\lambda_j - \theta_i^{(k)}|}.$$

Then

$$\|r_{k+\ell}^G\|_2 \leq \frac{F_{k,r}}{|c_k|} \kappa_2(X) \varepsilon^{(\ell,r)} \|r_k^G\|_2.$$

Proof: The proof follows the lines set forth by the proof of Theorem 2. The only differences are in the choice of h_k and q_ℓ. Here they are chosen as

$$h_k(t) = \prod_{\lambda_i \in \Lambda_r} \frac{\theta_i^{(k)}}{\lambda_i} \frac{(\lambda_i - t)}{(\theta_i^{(k)} - t)} p_k^F(t) \quad \text{and } q_\ell \in \Pi_\ell^1 \text{ is such that } \varepsilon^{(\ell,r)} = \max_{\lambda_j \in \Lambda_r^c} |q_\ell(\lambda_j)|.$$

□

By Lemma 2 we obtain the following result for FOM :

Corollary 2 *Let k and ℓ be such that H_k and $H_{k+\ell}$ are nonsingular, then*

$$\|r_{k+\ell}^F\|_2 \leq \frac{F_{k,r}}{|c_{k+\ell}|}\kappa_2(X)\varepsilon^{(\ell,r)}\|r_k^F\|_2.$$

3 Numerical experiments

We have chosen rather simple examples in order to have all the information available to understand the convergence behaviour. These examples also make it easy for the reader to repeat the experiments. First we give an example for which GMRES is superlinear convergent. Then we present an example in which the matrix has a Jordan block of size 3. Though even small rounding errors perturb the matrix to a diagonalizable system, we will see that the actual situation is quite different from the multiple eigenvalue situation and that the size of the Jordan block has a strong influence on the convergence behaviour. Finally, we discuss a well-known example where the Ritz values do not converge. In this example the GMRES method is only linear convergent, which corresponds with the results given in Section 2.
Our problems have been taken from ([6]: p.16, 17). The matrix is of the form $A = SBS^{-1}$ with $A, S, B \in \mathbb{R}^{100\times 100}$. We have selected S to be equal to

$$S = \begin{bmatrix} 1 & \beta & & & \\ & 1 & \beta & \emptyset & \\ & & \ddots & \ddots & \\ & \emptyset & & \ddots & \beta \\ & & & & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & & & & \\ & 1+\alpha & & \emptyset & \\ & & 3 & & \\ & \emptyset & & \ddots & \\ & & & & 100 \end{bmatrix}.$$

The system $Ax = b$ is solved for right-hand sides, such that $x = (1, \ldots, 1)^T$. The iterative methods start with $x_0 = (0, \ldots, 0)^T$.

In our first example we consider the convergence behaviour for the linear system with $\alpha = 0$ and $\beta = 0.9$ in view of Theorem 1. The eigenvalues are numbered as $\lambda_1 = 1$, $\lambda_2 = 1$ and $\lambda_i = i$, i = 3,...,100, and the Ritz values, which are real, are numbered such that $\theta_i^{(k)} < \theta_{i+1}^{(k)}$.
The actual reduction factors $\|r_{k+1}^G\|_2/\|r_k^G\|_2$ are shown in Figure 1. ¿From $k = 16$ on, GMRES appears to be superlinear convergent. For this problem the residuals can be bounded by an upperbound that has a decay rate $(\sqrt{\lambda_{max}/\lambda_{min}} - 1)/(\sqrt{\lambda_{max}/\lambda_{min}} + 1)$ (see [10]: Theorem 5 (with $\nu = 0$), and use the fact that all eigenvalues are on the real axis). Of course, this gives only a rough indication for the actual reduction factors, but it shows us that we may expect faster convergence for problems for which the eigenvalue ratio

$\lambda_{max}/\lambda_{min}$ is smaller. Obviously, since the eigenvalues, except for the first two, are equidistantly distributed, the decay rate changes more if we delete a small eigenvalue instead of a large eigenvalue. For this reason we restrict our attention to the lower part of the spectrum. Note that the eigenvalue $\lambda_2 = 1$ does not play a role in this example.

The following table shows the convergence of the smallest Ritz value. It appears that $F_{k,1}$ with $\Lambda_1 = \{\lambda_1\}$, has a moderate value from $k = 16$,

k	13	14	15	16	17	18	19	20
$\theta_1^{(k)}$	3.26	2.77	2.35	1.99	1.71	1.49	1.34	1.23
$F_{k,1}$	24	24	7	4	2.6	2.0	1.6	1.4

Table 1: The convergence of the first Ritz value for $\alpha = 0$

and, from the discussion to Theorem 1, this implies that the convergence behaviour is comparable with a process for a system in which the residual has no component in the direction of the eigenvector corresponding to $\lambda_1 = 1$. After some iterations the second Ritz value converges to λ_3 (see Table 2,

k	20	21	22	23	24	25	26	27	28	29	30
$\theta_2^{(k)}$	4.19	3.98	3.8	3.66	3.54	3.45	3.38	3.3	3.25	3.2	3.17
$F_{k,2}$	9	85	8	4	2.8	2.2	1.9	1.6	1.5	1.4	1.3

Table 2: The convergence of the second Ritz value for $\alpha = 0$

where Λ_2 has been taken as $\Lambda_2 = \{\lambda_1, \lambda_3\}$). After $k = 23$ the process converges as if the eigenvalues λ_1 (as well as $\lambda_2 = \lambda_1$) and λ_3 are absent. This is quite in line with the superlinear convergence behaviour of GMRES as observed from Figure 1. The decay rates for the upperbounds, corresponding to the comparison processes for the respective phases of the GMRES iteration process, are $(\sqrt{\lambda_{100}/\lambda_1}-1)/(\sqrt{\lambda_{100}/\lambda_1}+1) = .82$, $(\sqrt{\lambda_{100}/\lambda_3}-1)/(\sqrt{\lambda_{100}/\lambda_3}+1) = .7$, and $(\sqrt{\lambda_{100}/\lambda_4}-1)/(\sqrt{\lambda_{100}/\lambda_4}+1) = .67$, respectively. We see that the relative reduction in the decay rates is relatively large after the elimination of λ_1, and then it becomes smaller and smaller. We see a similar trend reflected in the reduction factors, displayed in Figure 1: their decrease is, indeed, much larger just beyond $k = 16$ than it is near $k = 23$. Note also that the predicted decay rates are in quite good agreement with the observed decay rates.

We consider in little more detail the situation for GMRES applied with $\alpha = 0.1$ and $\beta = 0.9$, also displayed in Figure 1. In this example we take $\lambda_1 = 1, \lambda_2 = 1.1$ and $\lambda_i = i$ for $i = 3, \cdots, 100$. Until $k = 24$ the convergence behaviour is virtually the same as for GMRES applied to the system with $\alpha = 0$ and $\beta = 0.9$. From $k = 24$ to $k = 30$ we observe only linear convergence. We note that $\theta_1^{(24)} = 1.107$ and $\theta_2^{(25)} = 1.091$, thus from $k = 25$ the smallest Ritz value is in $[\lambda_1, \lambda_2]$.
¿From $k = 31$ until $k = 36$ the reduction factor increases. The second Ritz value is 3.096 for $k = 30$ and 2.951 for $k = 31$, so in that phase GMRES "discoveres" a second eigenvalue less than $\lambda_3 = 3$. From $k = 37$ the reduction factor decreases again, which is in agreement with the results given in Table

k	30	31	32	33	34	35	36	37	38	39	40
$\theta_2^{(k)}$	3.03	2.95	2.87	2.77	2.63	2.46	2.24	1.98	1.74	1.54	1.39
$F_{k,2}$	227	112	40	22	13	8	5.4	3.6	2.5	1.9	1.5

Table 3: The convergence of the second Ritz value for $\alpha = 0.1$

3 (where $\Lambda_2 = \{\lambda_1, \lambda_2\}$).

In Figure 2 we show the GMRES residuals for the choice $\alpha = 0$ and $\alpha = 0.1$. It appears that the close eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1.1$ for $\alpha = 0.1$ has a decelerating effect. However the number of steps that GMRES for $\alpha = 0.1$ lags behind GMRES for $\alpha = 0$ is rather small (compare [11]: Section 6.7). It is not clear whether this phenomenon also occurs for problems with a more realistic spectrum.

In our second example we consider a situation with a multiple eigenvalue. In exact arithmetic the multiplicity of an eigenvalue plays no role, as long as the matrix A is diagonalizable. To illustrate this we consider the case where $\lambda_1 = \lambda_2 = \lambda_3 = 1$, and $\lambda_j = j$ for $j > 3$, and $\beta = 0.9$. In Figure 3 we see the error reduction plot for this case. The first, by now familiar bulge arises at the 15-th iteration, after which the first Ritz value gets below 4, i.e., enters its final interval, and thereafter we recognize the expected faster convergence.
The picture changes completely if A is not diagonalizable, i.e, when the matrix is defective. In Figure 4 we see the error reduction history for a system in which the first three eigenvalues (all equal to 1) occur in a Jordanblock of size 3.
At the first bulge, we have that $\theta_1^{(i)}$ comes close to 1, at the second bulge

$\theta_2^{(i)}$ arrives near 1 and at the third bulge a third Ritzvalue comes close to 1. Only after these three Ritzvalues have arrived close enough to $\lambda = 1$ we see convergence at a reduced rate. This is quite in line with Theorem 1 which states that faster convergence, corresponding to a process in which only λ_j for $j > 3$ are present, may be expected only after three Ritz values have been converged sufficiently close enough to $\lambda = 1$.
Because of rounding errors one might have expected the matrix to be diagonalizable, but in that case the transformation matrix X is very ill-conditioned so that the system behaves like having a Jordan block, i.e., all three eigenvalues have to be discovered, and the corresponding subspace has to be removed, before faster convergence can really take place.

In our final experiment we take the matrix B as follows [2], [6]:

$$B = \begin{bmatrix} \alpha & 0 & \cdots & 0 & 1 \\ 1 & \ddots & \emptyset & & 0 \\ & \ddots & & & \vdots \\ & \emptyset & & \ddots & 0 \\ & & & 1 & \alpha \end{bmatrix}.$$

We choose $\alpha = 1.2, \beta = 0$, the right-hand side $b = (1, 0, \cdots, 0)^T$ and starting vector $x_0 = 0$.
It is easily seen that the Hessenberg matrix H_k, obtained by Arnoldi's process, is equal to the $k \times k$ upper part of B. So for $k \leq 99$ the Ritz value $\theta_1^{(k)} = 1.2$ does not move to one of the eigenvalues $\lambda_k = \alpha + exp(2k\pi i/100), k = 0, \cdots, 99$ of A. This is in agreement, of course, with our numerical results.
In Figure 5 it can be seen that FOM has a linear convergence behaviour, which means that the reduction factors are constant. Applying GMRES the reduction factor changes only in the first iterates. Experiments with other values of β and b show more or less the same convergence behaviour.

4 Conclusions

We have analyzed in some detail the often observed superlinear convergence behaviour of GMRES. Our results show that a reduction in the speed of convergence, in a certain phase of the iteration process, can be expected as soon as certain eigenvalues of the given matrix are sufficiently well approximated by Ritz values. However, we have no theory why these eigenvalues should be approximated in that phase of the process. Our findings are reasonably well illustrated by numerical experiments and we believe that they have the

following important implication. In practice full GMRES is seldomly used, instead one restarts GMRES after m iteration steps. It is then to be expected that if m is not large enough in order to allow critical eigenspaces to be approximated by Ritz values, then slow convergence may be expected. This is quite in line with observations reported in [12] and [6].

Figure 1: GMRES, I: $\alpha = 0$, II: $\alpha = 0.1$

Figure 2: GMRES, I: $\alpha = 0$, II: $\alpha = 0.1$

Figure 3. GMRES for a spectrum with multiple eigenvalues

Figure 4. GMRES for a spectrum with a Jordanblock of size 3

Figure 5: $\alpha = 1.2$. I: FOM, II: GMRES

References

[1] O. Axelsson, *Solution of linear systems of equations: iterative methods*, In: V. A. Barker, editor, *Sparse Matrix Techniques*, Proc. of Copenhagen 1976, Springer Verlag, Berlin, 1977.

[2] P. N. Brown, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., **12**, 1991, p58-78

[3] P. Concus, G. H. Golub and D. P. O'Leary, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, In: J. R. Bunch and D. J. Rose, editors, *Sparse Matrix Computations*, Academic Press, New York, 1976.

[4] J. J. Dongarra, I. S. Duff, D. C. Sorenson and H. A. Van der Vorst, *Solving linear systems on vector and shared memory computers*, SIAM, Philadelphia, 1991

[5] G. H. Golub and C. F. Van Loan, *Matrix computations*, second edition, The Johns Hopkins University Press, Baltimore, 1989

[6] H. Huang and H. A. Van der Vorst, *Some observations on the convergence behavior of GMRES*, Delft University of Technology, Report 89-09, 1989

[7] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Num. Anal.,**12**, 1975, p617-629

[8] Y. Saad, *Projection methods for solving large sparse eigenvalue problems*, in: B. Kågström and A. Ruhe (eds.), *Matrix Pencils*, Proceedings, Pite Havsbad, 1982, Lecture Notes in Mathematics, **973**, Springer Verlag, Berlin, 1983

[9] Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp.,**37**, 1981, p105-127

[10] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving non symmetric linear systems*, SIAM J. Sci. Stat. Comput., **7**, 1986, p856-869

[11] A. Van der Sluis and H. A. Van der Vorst, *The rate of convergence of conjugate gradients*, Numer. Math., **48** 543-560, 1986

[12] H. A. Van der Vorst, *the convergence behaviour of conjugate gradient type methods*, in: R. Beauwens and P. De Groen, editors, *Iterative methods in linear algebra*, Proc. of the IMACS Int. Symp., Brussels, April 2-4, 1991, North-Holland, Amsterdam, pp. 67-76, 1992.