Computable convergence bounds for GMRES

Jörg Liesen
Fakultät für Mathematik
Universität Bielefeld
33501 Bielefeld
Germany

May 27, 1998

Abstract

The main purpose of this paper is the derivation of computable bounds on the residual norms of (full) GMRES. The new bounds depend on the initial guess and thus are conceptually different from standard 'worst-case' bounds. The analysis is valid for nonsingular linear systems and for any singular linear system, provided a certain condition on the initial residual is satisfied. It is shown that approximations to all factors in the new bounds can be obtained early in the GMRES run. The approximations serve to predict the convergence behavior of GMRES in later phases of the iteration. Numerical examples demonstrate that the new bounds are capable to describe the actual convergence behavior of GMRES for the given linear system and initial guess.

Key words. linear systems, convergence analysis, GMRES method, Krylov subspace methods, iterative methods

AMS Subject Classifications. 65F10, 65F15, 65F50, 65N12, 65N15

1 Introduction

The GMRES algorithm by Saad and Schultz [22] is one of the most popular iterative methods for solving a system of linear equations,

$$A\hat{x} = b,\ A \in \mathbf{C}^{N\times N},\ b \in \mathbf{C}^N. \tag{1.1}$$

The convergence analysis of GMRES has been an attractive field of research in recent years. Among others, Saad and Schultz [22], Nachtigal, Reddy and Trefethen [19], Starke [24] and Eiermann [7] derived upper bounds on the residual norms. However, none of these bounds leads to a complete characterization of the algorithm's convergence behavior. Some bounds are only applicable to a restricted class of nonsingular linear systems, others include factors that are not computable. All of these bounds are independent of the initial guess. As demonstrated in examples by Faber et al. [10] and Toh [27] this might lead to an over-estimation even of the algorithm's worst-case behavior. It is generally agreed, that the convergence analysis of GMRES needs further work (see e.g. [25, p. 187]).

In response, this paper tries to shed more light on the situation. Based on new characterizations of GMRES-equivalent matrices, we derive new convergence results for GMRES, which differ from those mentioned above in three main ways: First, our bounds on the algorithm's residual norms depend on the initial guess. Second, they give computable reduction factors for the residual norms for any nonsingular linear system and are applicable even to the singular case, if a certain condition is satisfied by the initial residual. Third, approximations to all factors in our bounds can be obtained during the GMRES run. These approximations lead to a heuristic which often allows us to predict after only a few steps, how GMRES will perform in later stages of the iteration. The paper is organized as follows: In Section 2 we introduce the GMRES algorithm and its mathematical properties. We then briefly review several convergence results in Section 3, before deriving new characterizations of GMRES-equivalent matrices in Section 4. Section 5 includes our new bounds for GMRES residual norms. In Section 6 we explain how to obtain approximations to the factors in our bounds from information generated during the GMRES run. We also devise heuristics for the behavior of these approximations and for the prediction of the performance of the GMRES algorithm. Section 7 includes numerical examples, and Section 8 contains our concluding remarks.

We point out that the linear system (1.1) might be singular, if not assumed otherwise. Throughout this paper, we assume that all calculations are performed in exact arithmetic. We let π_n denote the set of polynomials of degree at most n with value one at the origin. Furthermore, $\|\cdot\|$ denotes the Euclidean vector and corresponding matrix norm. By $\Lambda(M)$ we denote the spectrum and by M^H the Hermitian transpose of the matrix $M \in \mathbf{C}^{N\times N}$, respectively. If M is nonsingular, we define $\kappa(M) := \|M\|\|M^{-1}\|$, and $\kappa(M) := \infty$, otherwise. For $n = 1, 2, ...,$ we define $K_n(M, v) := \text{span}\{v, Mv, ..., M^{n-1}v\}$, the nth Krylov space generated by the matrix M and the vector $v \in \mathbf{C}^N$. For technical reasons we also define $K_0(A, v) := \{0\}$.

2 The mathematical properties of GMRES

In this section we introduce the GMRES algorithm and state several of its properties. We are not concerned with implementational details.

Definition 2.1 *Suppose that x_0 is an initial guess for the solution of a linear system (1.1) and define the initial residual $r_0 := b - Ax_0$. For $n = 1, 2, ...,$ let*

$$x_n \in x_0 + K_n(A, r_0) \tag{2.1}$$

be a vector, such that

$$r_n := b - Ax_n \in r_0 + AK_n(A, r_0) \tag{2.2}$$

satisfies

$$\|r_n\| = \min_{r \in r_0 + AK_n(A, r_0)} \|r\|, \tag{2.3}$$

then the vectors x_n and r_n are called nth GMRES iterate and residual, respectively.

As a comparison with [22, Formulas (4) and (8)] shows, our definition of the GMRES iterates and residuals is mathematically equivalent to their original derivation by Saad and Schultz. The next well known result (see e.g. [5, p. 8]) justifies Definition 2.1.

Lemma 2.2 *Suppose that $A \in \mathbf{C}^{N\times N}$ and $r_0 \in \mathbf{C}^N$. Then for every $n = 1, 2, ...,$ there exists a unique vector $r_n \in r_0 + AK_n(A, r_0)$ that satisfies (2.3).*

Thus, for $n = 1, 2, ...,$ the GMRES residuals are uniquely defined. In case of a nonsingular linear system, uniqueness of the residuals results in uniqueness of the iterates. Definition 2.1 leads to the following algorithm, which does not require uniqueness of the iterates.

Algorithm 2.3 (GMRES)

Given: A linear system (1.1) and an initial guess x_0.
Initialize: $r_0 := b - Ax_0$, $n = 0$.
While $r_n \neq 0$
$n = n + 1$
Construct $x_n \in x_0 + K_n(A, r_0)$ *so that*

$$\|r_n\| = \|b - Ax_n\| = \min_{r \in r_0 + AK_n(A,r_0)} \|r\|.$$

End While

If $r_n = 0$ for some $n \geq 0$, or, equivalently, x_n is a solution of (1.1), we say that GMRES terminates at step n. Before analyzing the termination properties of the algorithm, we point out an equivalent way to define the GMRES residuals.

Lemma 2.4 *Let* $S \subset \mathbf{C}^N$ *be a subspace and suppose that* $y_0 \in \mathbf{C}^N$ *and* $y_n \in y_0 + S$. *Then* $\|y_n\| = \min_{y \in y_0 + S} \|y\|$ *if and only if* $y_n \perp S$.

From this classic result (see e.g. [5, p. 9]) follows that our definition of the GMRES residuals via (2.2) and (2.3) is equivalent to defining them by (2.2) combined with

$$r_n \perp AK_n(A, r_0). \tag{2.4}$$

The following lemma is a direct consequence of (2.3).

Lemma 2.5 *The GMRES algorithm applied to a linear system (1.1) and initial guess* x_0 *terminates at step* $d \geq 0$ *if and only if* d *is the smallest index for which* $r_0 \in AK_d(A, r_0)$.

To further analyze this situation, we state an important property of the Krylov spaces, which is easy to prove.

Lemma 2.6 *If* $A \in \mathbf{C}^{N \times N}$ *and* $r_0 \in \mathbf{C}^N$, *then*

$$\dim AK_N(A, r_0) = d \ \ for \ \ some \ \ d, \ \ 0 \leq d \leq N,$$

is equivalent to

$$\begin{aligned} \dim AK_j(A, r_0) &= j \ \ for \ \ 0 \leq j \leq d \ \ and \\ \dim AK_{d+j}(A, r_0) &= d \ \ for \ \ all \ \ j \geq 1. \end{aligned}$$

Since the spaces $AK_n(A, r_0)$ coincide for all $n \geq \dim AK_N(A, r_0)$, the following is an immediate consequence.

Corollary 2.7 *Suppose that a linear system (1.1) and an initial guess* x_0 *are given, and define* $d := \dim AK_N(A, r_0)$. *Then for the GMRES residuals follows* $r_n = r_d$ *for all* $n \geq d$.

Thus, either GMRES terminates at some step $n \leq d := \dim AK_N(A, r_0) \leq N$, which requires $r_0 \in AK_N(A, r_0) = AK_d(A, r_0)$, or the algorithm fails to terminate, but stagnates after step d. The next theorem can be shown using Lemma 2.6.

Theorem 2.8 (cf. [22, Proposition 2]) *Suppose that the matrix* A *in (1.1) is non-singular. Then for any initial guess* x_0, *GMRES applied to (1.1) and* x_0 *terminates at step* $d \geq 0$ *if and only if* $\dim AK_N(A, r_0) = d$.

Obviously, the convergence analysis of GMRES is concerned with the derivation of upper bounds on the residual norms. Note that (2.3) implies that the residual norms are monotonically decreasing. Furthermore, (2.3) is equivalent to

$$\|r_n\| = \min_{p_n \in \pi_n} \|p_n(A) r_0\|. \tag{2.5}$$

If A is nonsingular and $r_n \neq 0$, then from the uniqueness of the residuals follows that the iteration polynomials in (2.5) are uniquely defined as well. Several bounds on the residual norms are derived from (2.5).

While GMRES must fail to terminate for inconsistent linear systems, for consistent singular systems the question of termination depends on the initial guess. As noted above, the residuals do not change after step $d := \dim AK_N(A, r_0)$. It is therefore not useful to apply the standard mathematical concept of convergence for $n \to \infty$ to the convergence analysis of GMRES. To derive a reasonable bound on the residual norms, one must make the assumption that GMRES will terminate, which is equivalent to assuming $r_0 \in AK_d(A, r_0) = AK_N(A, r_0)$. In the derivation of our new bounds we will therefore always make this assumption. For further analysis of the application of GMRES to singular systems, in particular the case $r_0 \notin AK_d(A, r_0)$, we refer to [4].

3 A review of GMRES convergence results

In this section we briefly review the most important results on convergence behavior of GMRES. Further analysis of some of these results can be found, for example, in [12, Section 3.2]. We suppose that a linear system (1.1) and an initial guess x_0 are given.

If A is diagonalizable, $X^{-1}AX = D$, then for the GMRES residuals follows from (2.5) that

$$\frac{\|r_n\|}{\|r_0\|} \leq \kappa(X) \min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(A)} |p_n(\lambda)|. \tag{3.1}$$

This bound was the first convergence result for GMRES [22, Proposition 4]. If A is normal, then $\kappa(X) = 1$ and (3.1) is sharp in the sense that for each n, there exists an initial guess for which equality holds [13, 17]. This suggests that if A is normal, the convergence of GMRES is mainly determined by $\Lambda(A)$ and can be analyzed by studying $\mu_n(\Omega)$, where

$$\mu_n(\Omega) := \inf_{p_n \in \pi_n} \max_{z \in \Omega} |p_n(z)|, \tag{3.2}$$

and $\Omega \subset \mathbf{C}$ is compact. If $\Lambda(A) \subseteq \Omega$, then $\mu_n(\Omega)$ is an upper bound for the min-max expression in (3.1). It is a fundamental result of approximation theory, that if $0 \notin \Omega$, then for every $n = 1, 2, \ldots$, there exists a unique polynomial $p_n \in \pi_n$ for which $\mu_n(\Omega)$ is attained (see e.g. [23, Chapter 5]). In these cases, we can replace the 'inf' in (3.2) by 'min'. On the other hand, if $0 \in \Omega$, we have $\mu_n(\Omega) = 1$ for all n. Note that the eigenvalues of a non-Hermitian matrix A can be anywhere in the complex plane. But $\mu_n(\Omega)$ is only known for a small class of sets Ω, including ellipses and one-parameter curves. It is therefore often not possible to find a compact set Ω, for which $\mu_n(\Omega)$ is known, simultaneously satisfying $\Lambda(A) \subseteq \Omega$ and $0 \notin \Omega$. But these requirements have to be met, for (3.2) to yield a reasonable reduction factor for the bound (3.1). Thus, even if A is normal, (3.1) often gives no useful information about the rate of reduction of the residual norms. For diagonalizable but nonnormal A, $\kappa(X)$ might be very large, and the right hand side of (3.1) might be a large overestimate of the actual residual norm. It is furthermore not clear, how the conditioning of the *eigenvectors* of A should influence the convergence behavior of GMRES applied to a *linear system* with A.

The sharpness results from [13, 17] can lead to confusion. They show that the convergence behavior of GMRES is better understood if A is normal, provided that the solution to the corresponding approximation problem (3.2) is known. They show neither that the algorithm converges faster for normal matrices, nor that the degree of normality of a matrix has a direct influence on the convergence behavior. In fact, Greenbaum and Strakoš [15, Section 3.1] show that any nonincreasing curve of residual norms,

$$\|r_0\| \geq \|r_1\| \geq \ldots \geq \|r_{N-1}\| > \|r_N\| = 0,$$

can be produced by GMRES applied to a normal, even unitary $N \times N$ matrix.

To analyze cases of large $\kappa(X)$, Trefethen [28] introduces the $\epsilon-$pseudospectrum of a matrix A. For every $\epsilon > 0$, this is defined as

$$\Lambda_\epsilon(A) := \{z \in \mathbf{C} \ : \ \|(zI - A)^{-1}\| \geq \epsilon^{-1}\}.$$

Let L_ϵ denote the arc length of the boundary of $\Lambda_\epsilon(A)$. Using the integral representation of the operator $p_n(A)$ (see e.g. [5, p. 201]), and choosing the boundary of $\Lambda_\epsilon(A)$ as the path of integration, from (2.5) follows [19, Theorem 4],

$$\frac{\|r_n\|}{\|r_0\|} \leq \frac{L_\epsilon}{2\pi\epsilon} \inf_{p_n \in \pi_n} \max_{\lambda \in \Lambda_\epsilon(A)} |p_n(\lambda)|. \tag{3.3}$$

While (3.1) assumes A to be diagonalizable, (3.3) holds for all linear systems. Clearly, both (3.1) and (3.3) yield no reduction factor if A is singular. Generally, it is very costly to compute $\Lambda_\epsilon(A)$, in particular when the cost of this computation is compared to the cost of a single GMRES run. Just like (3.1), the bound (3.3) in many cases can not describe the rate of convergence of GMRES, because $\mu_n(\Lambda_\epsilon(A))$ is not known. We also note that it is easy to construct examples in which the right hand side of (3.3) considerably overestimates the left hand side (see e.g. [15, Section 1]).

Another bound on GMRES residual norms is based on the field of values

$$F(A) := \{x^H A x \ : \ \|x\| = 1\}.$$

Defining $\nu(A) := \min_{z \in F(A)} |z|$, Starke [24, Theorem 3.2] shows that

$$\frac{\|r_n\|}{\|r_0\|} \leq (1 - \nu(A)\nu(A^{-1}))^{n/2}. \tag{3.4}$$

This bound gives a reduction factor for the residual norms, provided that $0 \notin F(A)$ and $0 \notin F(A^{-1})$, which excludes indefinite matrices. It is possible to obtain approximations to the right hand side of (3.4) during the GMRES run by solving small eigenproblems. Thus (3.4) is interesting from a practical point of view. However, while $F(A)$ in many cases provides useful information about the matrix A, the concept has its limitations in this context. For example, suppose that A is Hermitian positive definite and denote $\kappa := \kappa(A)$. Then $\nu(A)\nu(A^{-1}) = \kappa^{-1}$. But if $\kappa > 1$, then

$$\left(1 - \kappa^{-1}\right)^{1/2} \ > \ \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$

Hence, for each $\kappa > 1$, there exists an index n_0, such that for all $n \geq n_0$, the bound (3.4) is weaker than the classic bound (see e.g. [12, p. 51])

$$\frac{\|r_n\|}{\|r_0\|} \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^n, \tag{3.5}$$

which is derived from (3.1) using the Chebyshev polynomials of the first kind.

There exists a second bound based on the field of values. Defining the asymptotic convergence factor associated with (3.2) by

$$\gamma(\Omega) := \lim_{n\to\infty} (\mu_n(\Omega)))^{1/n}, \tag{3.6}$$

Eiermann [7] shows that

$$\frac{\|r_n\|}{\|r_0\|} \leq 2c_n[\gamma(F(A))]^n, \tag{3.7}$$

where the c_n are some positive constants satisfying $c_n < 2/(1-[\gamma(F(A))]^n)$. As is the case for (3.4), the bound (3.7) is only applicable to definite matrices. In most cases it is expensive to compute $\gamma(F(A))$. Also note that if $\gamma(F(A)) \approx 1$, then the constants c_n might be very large. With regard to this, we point out that in all examples of a recent study by Ernst [9], computed asymptotic convergence factors $\gamma(F(A))$ range between 0.995 and 0.9999, although GMRES shows very different types of convergence behavior.

Note that neither of the bounds presented above depends on x_0 and b, or, equivalently, on r_0. One problem with such bounds surfaces, when we consider unitary similarity transformations of A. Since $\Lambda(A)$, $\Lambda_\epsilon(A)$ and $F(A)$ are invariant under such transformations, all of the above bounds are the same for the linear systems (1.1), and (1.1) with A replaced by U^HAU for any unitary matrix U. It is obvious that the solutions to these two linear systems are generally different. But also the behavior of GMRES applied to these systems with the same initial guess might be completely different, as we will show next.

Lemma 3.1 *Suppose that $A \in \mathbf{C}^{N\times N}$ has a nonzero eigenvalue and that $b \in \mathbf{C}^N$. Then there exists a unitary matrix $U \in \mathbf{C}^{N\times N}$, such that GMRES applied to $U^HAUx = b$ with $x_0 = 0$ terminates at step one.*

Proof. Suppose that u_1 is an eigenvector of A that corresponds to a nonzero eigenvalue λ. We complete u_1 by unit vectors $u_2, ..., u_N$ to an orthonormal basis of $\mathbf{C}^N$ and denote $U_1 := [u_1, u_2, ..., u_N]$. Let $e_1 := [1, 0, ..., 0]^T \in \mathbf{C}^N$. Suppose further that $U_2 \in \mathbf{C}^{N\times N}$ is a unitary matrix for which $U_2b = \|b\|e_1$. We define $U := U_1U_2$, then the application of GMRES to $U^HAUx = b$ with $x_0 = 0$ yields

$$\|r_1\| = \min_{p_1\in\pi_1} \|p_1(U^HAU)b\| = \|b\| \min_{\alpha\in\mathbf{C}} \|e_1 - \alpha\lambda e_1\| = 0.$$

□

Lemma 3.1 shows that the the bounds (3.1), (3.3), (3.4) and (3.7) can lead to a misprediction of the actual behavior of GMRES.

We remark that not considering the dependence of GMRES on the initial residual can also lead to a misprediction of the worst-case behavior. Since GMRES minimizes $\|p_n(A)r_0\|$ for a given r_0, worst-case GMRES is the maximization of $\min_{p_n\in\pi_n} \|p_n(A)r_0\|$ over all r_0 with norm one. On the other hand, (3.1), (3.3) and (3.7) are derived from the 'ideal GMRES' [16] approximation problem

$$\min_{p_n\in\pi_n} \|p_n(A)\|.$$

It is clear that

$$\max_{\|r_0\|=1} \min_{p_n\in\pi_n} \|p_n(A)r_0\| \leq \min_{p_n\in\pi_n} \|p_n(A)\|, \tag{3.8}$$

which means that 'ideal GMRES' forms an upper envelope for all possible actual GMRES convergence curves. It was recently shown that there exist examples for which this envelope is never attained. Faber et al. [10] and Toh [27] constructed matrices for which for a certain n, the left hand side of (3.8) is strictly less than the right hand side. For linear systems with these matrices, bounds based on 'ideal GMRES' will overestimate the worst-case behavior. In the example of [27], the ratio of left and right hand side can be made arbitrarily small, which means that the over-estimation is arbitrarily large. However, the practical implications of such results are not yet clear.

4 GMRES-equivalent matrices

The main result of this section is the following: For a given matrix $A \in \mathbf{C}^{N\times N}$ and vector $r_0 \in \mathbf{C}^N$, for which $r_0 \in AK_N(A, r_0)$, we characterize the matrices $B \in \mathbf{C}^{N\times N}$ that satisfy $AK_n(A, r_0) = BK_n(B, r_0)$, for $n = 1, 2, ..., \dim AK_N(A, r_0)$.

Since A might be singular and we allow $\dim AK_N(A, r_0) < N$, we obtain generalizations of results of Greenbaum and Strakoš [15, Section 2] (also cf. [1, 14]). In the following, by $r_n^{M,r}$ we will denote the nth residual, when GMRES is applied to a linear system with the matrix M and the initial residual is r.

General assumptions. A linear system (1.1) and initial guess x_0 are given such that $2 \leq \dim AK_N(A, r_0) =: d \leq N$ and $r_0 \in AK_d(A, r_0)$. The latter assumption is always satisfied when A is nonsingular (cf. Theorem 2.8). With assuming $d \geq 2$ we exclude the trivial cases of GMRES termination at the steps zero and one. Furthermore, let $w_1, w_2, ..., w_d$ be orthonormal vectors that satisfy

$$\mathrm{span}\{w_1, w_2, ..., w_n\} = AK_n(A, r_0), \quad \text{for} \ \ n = 1, 2, ..., d. \tag{4.1}$$

Denoting $W_d := [w_1, w_2, ..., w_d] \in \mathbf{C}^{N\times d}$, we have

$$r_0 = W_d h, \ \ \text{where} \ \ h := [\eta_1, ..., \eta_d]^T, \ \ \text{and} \ \ \eta_n := w_n^H r_0, \ \ \text{for} \ \ n = 1, 2, ..., d. \tag{4.2}$$

It is easy to see that

$$AW_d = W_d H, \tag{4.3}$$

where $H \in \mathbf{C}^{d\times d}$ is an unreduced upper Hessenberg matrix. If H was not unreduced, then for some n, $1 \leq n \leq d-1$, Aw_n would be a linear combination of $w_1, ..., w_n$. This contradicts the dimension assumption on $AK_N(A, r_0)$.

Lemma 4.1 *The matrix H in (4.3) is nonsingular. Furthermore, if we define*

$$\hat{H} := \begin{bmatrix} 0 & \dots & 0 & 1/\eta_d \\ 1 & & & -\eta_1/\eta_d \\ & \ddots & & \vdots \\ & & 1 & -\eta_{d-1}/\eta_d \end{bmatrix}, \tag{4.4}$$

then $H = \tilde{R}\hat{H}$, where $\tilde{R} \in \mathbf{C}^{d\times d}$ is some nonsingular and upper triangular matrix.

Proof. From $r_0 = W_d h$ and the definition of $\hat{H}$ follows that

$$A[r_0, W_{d-1}] = AW_d \begin{bmatrix} \eta_1 & 1 & & \\ \eta_2 & & \ddots & \\ \vdots & & & 1 \\ \eta_d & 0 & \dots & 0 \end{bmatrix} = AW_d\hat{H}^{-1}. \tag{4.5}$$

From (4.1) follows that $w_1 = e^{i\alpha} Ar_0/\|Ar_0\|$ for some $\alpha \in \mathbf{R}$. Without loss of generality we assume $\alpha = 0$. We let H_{d-1} denote the first $d-1$ columns of the matrix H to get

$$A[r_0, W_{d-1}] = [Ar_0, AW_{d-1}] = W_d[\|Ar_0\|e_1, H_{d-1}].$$

Together with $W_d^H AW_d = H$ and (4.5), this yields

$$H = [\|Ar_0\|e_1, H_{d-1}]\hat{H}.$$

Since H is unreduced, $[\|Ar_0\|e_1, H_{d-1}]$ is nonsingular and upper triangular. Together with nonsingularity of $\hat{H}$ also follows that H is nonsingular. □

Definition 4.2 *Suppose that $A \in \mathbf{C}^{N\times N}$ and $r_0 \in \mathbf{C}^N$ with $d := \dim AK_N(A, r_0)$. Then $B \in \mathbf{C}^{N\times N}$ is called GMRES-equivalent to A with respect to r_0, if*

$$\dim BK_N(B, r_0) = d \;\; and \;\; AK_n(A, r_0) = BK_n(B, r_0), \;\; for \;\; n = 1, 2, ..., d.$$

Clearly, if B is GMRES-equivalent to A with respect to r_0, then

$$r_n^{A,r_0} = r_n^{B,r_0} \;\; \text{for} \;\; n = 1, 2, ..., d.$$

Our main result of this section generalizes [15, Theorems 2.1 and 2.2].

Theorem 4.3 *Under our general assumptions, for a matrix $B \in \mathbf{C}^{N\times N}$ the following three statements are equivalent.*

(a) B is GMRES-equivalent to A with respect to r_0.

(b) B satisfies

$$BW_d = W_d\tilde{R}\hat{H},$$

where $\hat{H}$ is defined in (4.4) and $\tilde{R} \in \mathbf{C}^{d\times d}$ is a nonsingular upper triangular matrix.

(c) B satisfies

$$BW_d = W_d\tilde{R}H,$$

where H is defined in (4.3) and $\tilde{R} \in \mathbf{C}^{d\times d}$ is a nonsingular upper triangular matrix.

Proof. The equivalence of (b) and (c) is a direct consequence of Lemma 4.1. We thus only have to show that (a) and (b) are equivalent. Suppose that (a) holds. Since $r_0 = W_d h$, we have

$$B[r_0, W_{d-1}] = BW_d\hat{H}^{-1}. \tag{4.6}$$

Furthermore, from (4.1) and the GMRES-equivalence of A and B follows that

$$B[r_0, W_{d-1}] = W_d\tilde{R}, \tag{4.7}$$

for some upper triangular matrix $\tilde{R}$. Now (4.6) and (4.7) yield $BW_d = W_d\tilde{R}\hat{H}$. To show that (b) holds, we have to show that the matrix $\tilde{R} = (r_{ij})$ in (4.7) must be nonsingular. Since $\dim BK_N(B, r_0) = d$, we have $B^n r_0 \neq 0$ for $n = 1, 2, ..., d$. In particular, $Br_0 \neq 0$ and hence $r_{11} \neq 0$. Moreover,

$$w_n = \sum_{k=1}^{n} \alpha_k B^k r_0, \;\; \text{for} \;\; n = 1, 2, ..., d,$$

and some complex numbers $\alpha_1, ..., \alpha_n$, not all zero. Therefore,

$$Bw_n = \sum_{k=1}^{n} \alpha_k B^{k+1} r_0 \neq 0, \text{ for } n = 1, 2, ..., d-1.$$

Now suppose that $r_{nn} = 0$ for some n, $2 \leq n \leq d$. Then the nth column of (4.7) is

$$Bw_{n-1} = \sum_{k=1}^{n-1} r_{kn} w_n \neq 0.$$

But from this follows $\dim BK_n(B, r_0) \leq n-1$, which contradicts our assumption on the dimension of $BK_N(B, r_0)$. Thus, $r_{nn} \neq 0$ for $n = 1, 2, ..., d$, i.e. $\tilde{R}$ is nonsingular.

Conversely, if (b) holds, then

$$B^n r_0 = B^n W_d h = r_{11} W_d (\tilde{R}\hat{H})^{n-1} e_1.$$

Since $\tilde{R}$ is nonsingular,

$$AK_n(A, r_0) = \text{span}\{w_1, w_2, ..., w_n\} = BK_n(B, r_0) \text{ for } n = 1, 2, ..., d.$$

We conclude that (a) holds, since $B^{d+1} r_0 \in BK_d(B, r_0)$, so that $\dim BK_N(B, r_0) = d$. □

Note that by varying $\tilde{R}$ in Theorem 4.3, we obtain a whole class of matrices which are GMRES-equivalent to the given matrix A with respect to the initial residual $r_0 = W_d h$. We remark that if $d < N$, then the matrix B in Theorem 4.3 must be singular, regardless of A being singular or not. We obtain a corollary that will be useful in the following section.

Corollary 4.4 *Suppose that, in the notation of Theorem 4.3,*

$$B = W_d \tilde{R} H W_d^H \text{ and } \hat{B} = W_d \tilde{R} \hat{H} W_d^H,$$

where $\tilde{R}$ is any nonsingular upper triangular matrix. Then B and $\hat{B}$ are both GMRES-equivalent to A with respect to r_0, and in particular,

$$\|r_n^{B,r_0}\| = \|r_n^{\hat{B},r_0}\| = \|r_n^{A,r_0}\|, \text{ for } n = 1, 2, ..., d.$$

The following corollary to Lemma 4.1 relates the RQ-decompositions of the matrices $\hat{H}$ and H.

Corollary 4.5 *$\hat{H} = \hat{R}\hat{Q}$ is an RQ-decomposition of $\hat{H}$ if and only if $H = R\hat{Q}$ is an RQ-decomposition of H.*

It is interesting to note that the Q-factors of the RQ-decompositions of $\hat{H}$ and H can be chosen identical. This is not true for the Q-factors of the corresponding QR-decompositions. If $\hat{H} = \hat{Q}\hat{R}$, then $\hat{Q}$ is a unitary shift matrix, while generally if $H = QR$, then Q might be any unitary upper Hessenberg matrix.

5 New bounds on GMRES residual norms

The results of this section are derived under the general assumptions stated at the beginning of Section 4. Using Theorem 4.3, we can relate the convergence of GMRES for the given linear system to the convergence for some other system, which is easier to analyze. Our starting point for the derivation of the new bounds therefore is the construction of suitable matrices that are GMRES-equivalent to A with respect to the given r_0.

As stated in Corollary 4.4, for any nonsingular upper triangular matrix $\tilde{R} \in \mathbf{C}^{d\times d}$, the matrices

$$W_d \tilde{R} H W_d^H \quad \text{and} \quad W_d \tilde{R} \hat{H} W_d^H \tag{5.1}$$

are both GMRES-equivalent to A with respect to r_0. We consider the following decompositions of H and $\hat{H}$,

$$H = QR, \quad \text{and} \quad \hat{H} = \hat{R}\hat{Q}. \tag{5.2}$$

The matrices Q and $\hat{Q}$ are unitary and the matrices R and $\hat{R}$ are nonsingular and upper triangular. All four are of size $d \times d$.

Remark 5.1 Two other similar decompositions are possible, $H = R_1 Q_1$ and $\hat{H} = Q_2 R_2$. In the remainder of this paper we will be particularly interested in the spectra of the Q-factors of these decompositions. As noted in Section 4, Q_2 is simply a unitary shift matrix, which is uninteresting for our purposes. We will therefore not consider the QR-decomposition of $\hat{H}$. Furthermore, Corollary 4.5 shows that the results derived for $\Lambda(\hat{Q})$ also hold for $\Lambda(Q_1)$. Hence, we do not need to consider the RQ-decomposition of H explicitly.

For later use we note that $W_d^H A W_d = QR$ yields

$$\kappa(R) \leq \kappa(A), \tag{5.3}$$

with equality if $d = N$. Often $d = N$ is a reasonable assumption, in particular when the 'generic' case of an initial residual without special properties is to be analyzed.

Choosing $\tilde{R} = R^{-1}$ and $\tilde{R} = \hat{R}^{-1}$ in (5.1), and using (5.2), we define the matrices

$$B := W_d R^{-1} Q R W_d^H \quad \text{and} \quad \hat{B} := W_d \hat{Q} W_d^H, \tag{5.4}$$

which are both GMRES-equivalent to A with respect to r_0.

Remark 5.2 Greenbaum and Strakoš [15, Section 3.1] consider the RQ-decomposition of the matrix H. They point out that if $d = N$, then the matrix $\hat{B} = W_N \hat{Q} W_N^H$ is unitary. Thus, whenever A and r_0 are such that $d = N$, there exists a unitary matrix which is GMRES-equivalent to A with respect to r_0.

Theorem 5.3 *Suppose that B and $\hat{B}$ are defined as in (5.4). Then*

$$\frac{\|r_n\|}{\|r_0\|} \leq \kappa(R) \min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(Q)} |p_n(\lambda)|, \quad \textit{for} \;\; n = 1, 2, ..., d, \tag{5.5}$$

and

$$\frac{\|r_n\|}{\|r_0\|} \leq \min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(\hat{Q})} |p_n(\lambda)|, \quad \textit{for} \;\; n = 1, 2, ..., d. \tag{5.6}$$

Proof. Corollary 4.4 yields for $n = 1, 2, ..., d$,

$$\begin{aligned} \|r_n\| \equiv \|r_n^{A,r_0}\| &= \|r_n^{B,r_0}\| = \min_{p_n \in \pi_n} \|W_d R^{-1} p_n(Q) R W_d^H r_0\| \\ &\leq \|R^{-1}\| \, \|R\| \min_{p_n \in \pi_n} \|p_n(Q)\| \, \|r_0\|, \end{aligned}$$

from which (5.5) follows. The proof of (5.6) is similar. □

To study the approximation problem

$$\min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(U)} |p_n(\lambda)|, \quad U \;\; \text{unitary}, \tag{5.7}$$

we need the following definition.

Definition 5.4 *Suppose that the spectrum of the $d \times d$ unitary matrix U is given by $\Lambda(U) := \{e^{i\beta_j} \; : \; 0 \le \beta_1 \le \beta_2 \le \cdots \le \beta_d < 2\pi\}$ and let $\beta_{d+1} := 2\pi + \beta_1$. We define the d gaps in $\Lambda(U)$ by*

$$g_j := \beta_{j+1} - \beta_j, \quad \text{for} \;\; j = 1, 2, ..., d. \tag{5.8}$$

For $j = 1, 2, ..., d$, the complex number $e^{i(\beta_{j+1}+\beta_j)/2}$ is called the center of the gap g_j. $\phi := \max\{g_1, g_2, ..., g_d\}$ denotes the largest gap in $\Lambda(U)$.

Lemma 5.5 *For every unitary matrix U and every $\alpha \in \mathbf{R}$,*

$$\min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(U)} |p_n(\lambda)| = \min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(e^{i\alpha} U)} |p_n(\lambda)|. \tag{5.9}$$

We omit the easy proof. As Lemma 5.5 shows, in the study of (5.7) we can assume without loss of generality, that 1 is the center of the largest gap ϕ in $\Lambda(U)$. More explicitly, we define

$$\Omega_\phi := \left\{ e^{i\alpha} \; : \; \frac{\phi}{2} \le \alpha \le 2\pi - \frac{\phi}{2} \right\}, \tag{5.10}$$

to conclude from Lemma 5.5,

$$\min_{p_n \in \pi_n} \max_{\lambda \in \Lambda(U)} |p_n(\lambda)| \le \min_{p_n \in \pi_n} \max_{z \in \Omega_\phi} |p_n(z)|. \tag{5.11}$$

Suppose that $0 < \phi < 2\pi$, or, equivalently, $\Lambda(U)$ contains at least two distinct points. Then for $n = 1, 2, ...$, we can introduce the polynomials

$$\begin{aligned} p_n(z) &:= \prod_{k=0}^{n-1} \left(1 - z \frac{(1 - \gamma e^{i\psi_k})}{e^{i\psi_k} (\gamma - e^{i\psi_k})} \right), \text{ where} && (5.12) \\ \psi_k &:= \frac{2\pi k}{n}, \;\; \text{for} \;\; k = 0, 1, ..., n-1, && (5.13) \\ \gamma &:= \left(\cos \frac{\phi}{4} \right)^{-1}. && (5.14) \end{aligned}$$

By assumption, $1 < \gamma < \infty$ and hence the polynomials (5.12) are well defined for every $n \ge 1$. Furthermore, we have $p_n \in \pi_n$.

Remark 5.6 The polynomials p_n in (5.12) are constructed using a conformal mapping technique from [18]. It is shown in [18], that $\Psi(z) := \frac{z(\gamma - z)}{1 - \gamma z}$ maps the exterior of the unit circle conformally onto the exterior of Ω_ϕ. Thus, the zeros of p_n are Fejér points (see e.g. [23, Chapter 1]) associated with Ω_ϕ. However, we do not explicitly use these properties in this paper.

Lemma 5.7 *The polynomials introduced in (5.12) satisfy*

$$\max_{z \in \Omega_\phi} |p_n(z)| \le \frac{4}{\gamma^n - 1}, \quad \text{for} \quad n = 1, 2, ... \tag{5.15}$$

Proof. We will start with noting some elementary formulas for later use:

$$\begin{aligned} e^{i\beta} + e^{-i\beta} &= 2\cos\beta, \;\; \forall \, \beta \in \mathbf{R}, && (5.16) \\ \prod_{k=0}^{n-1} \left(z - e^{i\psi_k} \right) &= z^n - 1, \;\; \forall \, z \in \mathbf{C}, && (5.17) \\ \frac{1}{2c} \left| c - e^{i\beta} \right|^2 &= \left(\frac{1 + c^2}{2c} - \cos\beta \right), \;\; \forall \, \beta \in \mathbf{R}, \; c \in \mathbf{C} - \{0\}. && (5.18) \end{aligned}$$

We now turn our attention to the polynomials (5.12) for some fixed but arbitrary $z = e^{i\alpha} \in \Omega_\phi$. A simple manipulation yields

$$\begin{aligned} p_n(e^{i\alpha}) &= \prod_{k=0}^{n-1} \left(1 - e^{i\alpha} \frac{(1-\gamma e^{i\psi_k})}{e^{i\psi_k}(\gamma - e^{i\psi_k})}\right) \\ &= \prod_{k=0}^{n-1} \frac{e^{i\frac{\alpha}{2}}\left(\gamma\left(e^{i\frac{\alpha}{2}} + e^{-i\frac{\alpha}{2}}\right) - \left(e^{i(\psi_k - \frac{\alpha}{2})} + e^{-i(\psi_k - \frac{\alpha}{2})}\right)\right)}{\gamma - e^{i\psi_k}} \\ &= \frac{\left(2e^{i\frac{\alpha}{2}}\right)^n}{\gamma^n - 1} \prod_{k=0}^{n-1} \left(\gamma \cos\frac{\alpha}{2} - \cos\left(\psi_k - \frac{\alpha}{2}\right)\right), \end{aligned} \tag{5.19}$$

where, in the last equation, we have used (5.16) and (5.17). Let us define the complex number c by

$$c := \gamma \cos\frac{\alpha}{2} + i\left(1 - \left(\gamma\cos\frac{\alpha}{2}\right)^2\right)^{1/2}.$$

Since $\gamma = (\cos\frac{\phi}{4})^{-1}$, and $\frac{\phi}{2} \le \alpha \le 2\pi - \frac{\phi}{2}$, we have $-1 \le \gamma\cos\frac{\alpha}{2} \le 1$, so that

$$1 - \left(\gamma\cos\frac{\alpha}{2}\right)^2 \ge 0,$$

and thus,

$$|c| = \left(\gamma\cos\frac{\alpha}{2}\right)^2 + \left(1 - \left(\gamma\cos\frac{\alpha}{2}\right)^2\right) = 1.$$

Furthermore,

$$\gamma\cos\frac{\alpha}{2} = \frac{1+c^2}{2c}.$$

With this definition of c, we apply (5.17) and (5.18) to the product in (5.19) to get

$$\begin{aligned} p_n(e^{i\alpha}) &= \frac{\left(2e^{i\frac{\alpha}{2}}\right)^n}{\gamma^n - 1} \frac{1}{(2c)^n} \prod_{k=0}^{n-1} \left|c - e^{i(\psi_k - \frac{\alpha}{2})}\right|^2 \\ &= \frac{1}{\gamma^n - 1} \left(\frac{e^{i\frac{\alpha}{2}}}{c}\right)^n \left|\left(ce^{i\frac{\alpha}{2}}\right)^n - 1\right|^2. \end{aligned} \tag{5.20}$$

Since $|c| = 1$, we have

$$\left|\left(ce^{i\frac{\alpha}{2}}\right)^n - 1\right|^2 \le 4.$$

Taking absolute values in (5.20) yields

$$\left|p_n(e^{i\alpha})\right| \le \frac{4}{\gamma^n - 1}, \tag{5.21}$$

which completes the proof. □

As Lemma 5.7 shows, the polynomials (5.12) converge to zero on Ω_ϕ, with the speed of convergence being directly related to the gap ϕ: The larger ϕ, the faster the convergence and vice versa. Since the polynomials (5.12) are constructed using Fejér points, they converge to zero on Ω_ϕ with the largest geometric degree of convergence (cf. [31, Chapter 4.7]). Because of the moderate constant 4, Lemma 5.7 is useful not only asymptotically, but also for small n.

We now combine (5.3), (5.11), Theorem 5.3 and Lemma 5.7 for our main convergence result. Note that from our general assumption $d \ge 2$ follows that $\Lambda(Q)$ and $\Lambda(\hat{Q})$ both contain at least two distinct points, so that Lemma 5.7 is applicable.

Theorem 5.8 *Suppose that a linear system (1.1) and an initial guess x_0 are given, such that $2 \le d := \dim AK_N(A, r_0) \le N$ and $r_0 \in AK_d(A, r_0)$. Define the matrices Q, R and $\hat{Q}$ as in (5.2), and let ϕ and $\hat{\phi}$ denote the largest gaps in $\Lambda(Q)$ and $\Lambda(\hat{Q})$, respectively. Define $\gamma := (\cos \frac{\phi}{4})^{-1}$ and $\hat{\gamma} := (\cos \frac{\hat{\phi}}{4})^{-1}$, then the GMRES residuals for (1.1) and x_0 satisfy*

$$\frac{\|r_n\|}{\|r_0\|} \le 4 \frac{\kappa(R)}{\gamma^n - 1}, \quad for \;\; n = 1, 2, ..., d, \tag{5.22}$$

and

$$\frac{\|r_n\|}{\|r_0\|} \le \frac{4}{\hat{\gamma}^n - 1}, \quad for \;\; n = 1, 2, ..., d. \tag{5.23}$$

Furthermore, $\kappa(R) \le \kappa(A)$, with equality if $d = N$.

To analyze Theorem 5.8 and its implications, we first point out that γ^{-1} and $\hat{\gamma}^{-1}$, the rates of reduction of the bounds (5.22) and (5.23) depend on the initial guess, and that both rates lie strictly between zero and one. We have not found an algebraic relation between the two rates. As we will show by numerical examples in Section 7, γ^{-1} and $\hat{\gamma}^{-1}$ might be of different sizes and have the potential to describe the rate of convergence of GMRES for the given initial residual.

At first sight, (5.23) seems to be more useful than (5.22), since it does not include the constant $\kappa(R)$ (or $\kappa(A)$). However, it is possible to implement GMRES in a way that approximations to (5.22) are easily computable in each step of the algorithm. Moreover, these approximations lead to a heuristic which often allows us to early predict the behavior of GMRES in later stages of the iteration. The situation is different for (5.23). The computation of approximations to this bound during the GMRES run is sometimes not possible and always more expensive. Furthermore, a heuristic on the behavior of these approximations is not so clear. We also point out that there are cases in which (5.22) better describes the actual behavior of GMRES, simply because the constant $\kappa(R)$ (or $\kappa(A)$) is involved. We give an example in Section 7.

From our point of view, the appearance of the constant $\kappa(R)$ (or $\kappa(A)$) in an upper bound on GMRES residual norms is not a nuisance, but rather shows the algorithm's dependence on the conditioning of the coefficient matrix. Ill-conditioning of A often leads to slow convergence of GMRES. Since well-conditioning does not imply fast GMRES convergence, it is generally difficult to strengthen this statement. The worst possible GMRES behavior, no progress until step N, can even occur if A is perfectly conditioned. Just consider an $N \times N$ unitary matrix, having the first N roots of unity as its eigenvalues (cf. [19, Example C]).

We finally remark that to the best of our knowledge, (5.22) is the first convergence result that shows the dependence of GMRES on the conditioning of the coefficient matrix. In most of the earlier results, the emphasis was laid on the conditioning of associated eigenproblem, which sometimes is unrelated to the convergence behavior.

6 Approximating our bounds to predict the behavior of GMRES

In this section we explain how to obtain approximations to the upper bounds (5.22) and (5.23) from information generated during the GMRES run. We point out that if close approximations to the right hand sides of (5.22) and (5.23) are available in early stages of the GMRES iteration, then these potentially *predict* the behavior of the algorithm in later stages.

6.1 WZ-GMRES

To obtain the approximations, we consider an implementation of GMRES first proposed by Walker and Zhou [30]. In the following we will refer to this implementation as WZ-GMRES. For a description of WZ-GMRES we need to introduce the Arnoldi [2] process.

Suppose that a matrix $M \in \mathbf{C}^{N\times N}$ and a unit vector $w_1 \in \mathbf{C}^N$ satisfy dim $K_N(M, w_1) = d \leq N$. Then the Arnoldi process applied to M and w_1 is the construction of an orthonormal basis, $w_1, w_2, ..., w_d$, of $K_d(M, w_1)$ by means of any orthogonalization method.

Remark 6.1 In many references (see e.g. [11]), the orthogonalization performed by the Arnoldi process is implemented using the Classical or Modified Gram-Schmidt procedure. As pointed out by Walker [29], it is also possible to use Householder transformations for this purpose. For a collection of different implementations of the Arnoldi process, we refer to [6] (also cf. [20]). Our development is independent of the particular implementation, so that any (numerically stable) one will serve our purpose.

In practice, d is not known, and the basis vectors are constructed sequentially. After $n < d$ steps of the Arnoldi process, we have

$$MW_n = W_{n+1}H_n^e, \;\; W_n := [w_1, ..., w_n], \tag{6.1}$$

where

$$H_n^e = \begin{bmatrix} h_{11} & \dots & h_{1n} \\ h_{21} & \ddots & \vdots \\ & \ddots & h_{nn} \\ & & h_{n+1,n} \end{bmatrix} \tag{6.2}$$

is an $(n+1)\times n$ unreduced upper Hessenberg matrix. Clearly, in exact arithmetic, the process has to terminate at step d with the final matrix relation

$$MW_d = W_dH_d, \;\; W_d \in \mathbf{C}^{N\times d}, \;\; H_d \in \mathbf{C}^{d\times d}. \tag{6.3}$$

The classical implementation of GMRES [22] is based on the Arnoldi process applied to the matrix A and the vector $r_0/\|r_0\|$. However, following [30], we will use

$$w_1 := \frac{Ar_0}{\|Ar_0\|} \tag{6.4}$$

as starting the vector for the Arnoldi process. Recall that for our given linear system and initial guess we have assumed dim $AK_N(A, r_0) = d$. Then, since

$$K_n(A, w_1) = AK_n(A, r_0), \;\; \text{for} \;\; n = 1, 2, ..., d,$$

the application of the Arnoldi process to A and w_1 yields orthonormal vectors $w_1, w_2, ..., w_d$, that satisfy (4.1). Using these vectors, (2.1) can be rewritten as

$$x_n = x_0 + [r_0/\|r_0\|, W_{n-1}]y_n \tag{6.5}$$

for some $y_n \in \mathbf{C}^n$, $n = 1, 2, ..., d$. To implement GMRES we use the orthogonality relation (2.4) in the computation of y_n. Together with (6.1) and (6.5), this yields

$$0 = W_n^H r_n = W_n^H r_0 - W_n^H A[r_0/\|r_0\|, W_{n-1}]y_n,$$

which is equivalent to

$$[(\|Ar_0\|/\|r_0\|)e_1, H_{n-1}^e]y_n = W_n^H r_0. \tag{6.6}$$

Thus, y_n is the solution of an upper triangular system. Since H_{n-1}^e is unreduced, y_n is well-defined. It is important to note that r_n can be obtained without explicitly computing x_n [21],

$$r_n = r_{n-1} - (w_n^H r_0) w_n. \tag{6.7}$$

The upper triangular solve for y_n therefore only has to be performed if the residual norm computed from (6.7) is small enough. Clearly, in exact arithmetic, the termination of the Arnoldi process at step d is equivalent to the termination of WZ-GMRES at step d. WZ-GMRES is a complete and economical implementation of the GMRES algorithm. We remark that from (6.5) and (6.6) it is easy to see, that WZ-GMRES is (mathematically) independent of the orthogonalization method used for the Arnoldi process. For more details about WZ-GMRES we refer to [30, 21].

6.2 Computing approximations to (5.22)

If (6.4) holds, then the vectors $w_1, w_2, ..., w_d$ produced by the Arnoldi process satisfy (4.1). Consequently, the matrix H_d in (6.3) can be identified with the matrix H in (5.2). We can therefore compute the numbers ϕ and $\kappa(R)$ used in (5.22) from the QR-decomposition of H_d. Thus, the right hand side of (5.22) is computable from information generated during the run of WZ-GMRES. Moreover, because of the sequential nature of the Arnoldi process, the matrix H_n^e is the leading $(n+1)\times n$ submatrix of H_d. For $n = 1, 2, ..., d-1$, we can therefore decompose

$$H_n^e = Q_n \begin{bmatrix} R_n \\ 0 \end{bmatrix}, \quad Q_n \in \mathbf{C}^{(n+1)\times(n+1)}, \quad R_n \in \mathbf{C}^{n\times n}, \tag{6.8}$$

and use ϕ_n, the largest gap in $\Lambda(Q_n)$, as approximation to ϕ and $\kappa(R_n)$ as approximation to $\kappa(R)$. Notice that because H_n^e is unreduced, R_n is nonsingular and (6.8) is well defined.

The decomposition (6.8) is particularly easy to compute. It is well known, that because of the special structure of H_n^e, the matrix Q_n can be derived as a product of n Givens transformations, $Q_n = G_1 \cdot .... \cdot G_n$, where

$$G_j = \begin{bmatrix} I_{j-1} & & & \\ & c_j & -\overline{s}_j & \\ & s_j & c_j & \\ & & & I_{n-j} \end{bmatrix} \in \mathbf{C}^{(n+1)\times(n+1)}, \quad 1 \le j \le n. \tag{6.9}$$

Suppose that for some n, $2 \le n \le d-1$,

$$Q_{n-1}^H H_{n-1}^e = \begin{bmatrix} R_{n-1} \\ 0 \end{bmatrix}, \quad Q_{n-1}^H = G_{n-1}^H \cdot ... \cdot G_1^H. \tag{6.10}$$

We define

$$h^n := [h_{1n}, ..., h_{nn}]^T, [\rho_1, ..., \rho_n]^T := Q_{n-1}^H h^n, \quad \text{and} \quad \nu_n := [\rho_1, ..., \rho_{n-1}]^T.$$

Then

$$\begin{aligned} Q_n^H H_n^e &= G_n^H \begin{bmatrix} Q_{n-1}^H & \\ & 1 \end{bmatrix} \begin{bmatrix} H_{n-1}^e & h^n \\ & h_{n+1,n} \end{bmatrix} \\ &= G_n^H \begin{bmatrix} R_{n-1} & \nu_n \\ & \rho_n \\ & h_{n+1,n} \end{bmatrix}. \end{aligned} \tag{6.11}$$

If $\rho_n = |\rho_n| e^{i\delta_n}$, we define $\omega_n := (|\rho_n|^2 + |h_{n+1,n}|^2)^{1/2}$ and choose

$$c_n = \frac{|\rho_n|}{\omega_n}, \tag{6.12}$$

$$s_n = \frac{e^{-i\delta_n} h_{n+1,n}}{\omega_n}, \tag{6.13}$$

to give (6.11) the required upper triangular form. Hence, to compute Q_n and R_n from Q_{n-1} and R_{n-1}, respectively, only one matrix-vector product and a few flops to compute c_n and s_n are required.

To compute ϕ_n we have to solve an $n \times n$ unitary eigenproblem. For the efficient computation of $\kappa(R_n)$ we can use the method of incremental condition estimation [3]. With this method, the computation of $\kappa(R_n)$ costs only $O(n)$ operations, provided $\kappa(R_{n-1})$ is known.

6.3 Computing approximations to (5.23)

Suppose that we apply WZ-GMRES to (1.1) and x_0 and use the Arnoldi vectors $w_1, w_2, ..., w_d$ to set up the matrix $\hat{H}$ as defined in 4.4. Note that $\hat{H}$ can be obtained at no additional cost, if the updated residuals are computed using (6.7). The bound (5.23) depends only on the largest gap $\hat{\phi}$ in $\Lambda(\hat{Q})$, where $\hat{Q}$ is defined by the RQ-decomposition $\hat{H} = \hat{R}\hat{Q}$ (cf. (5.2)).

From (6.7) follows that

$$|\eta_n| = |w_n^H r_0| = \|r_{n-1} - r_n\|, \quad \text{for} \quad n = 1, 2, ..., d.$$

Hence, if GMRES does not stagnate in step n, $2 \le n \le d-1$, then $\eta_n \neq 0$, and the decomposition

$$\hat{H}_n := \begin{bmatrix} 0 & \dots & 0 & 1/\eta_n \\ 1 & & & -\eta_1/\eta_n \\ & \ddots & & \vdots \\ & & 1 & -\eta_{n-1}/\eta_n \end{bmatrix} = \hat{R}_n \hat{Q}_n \quad \in \mathbf{C}^{n \times n} \tag{6.14}$$

is defined. In this case, $\hat{\phi}_n$, the largest gap in $\Lambda(\hat{Q}_n)$, can be used as an approximation to $\hat{\phi}$. However, we have no easy way to update the decomposition (6.14). $\hat{R}_n$ and $\hat{Q}_n$ have to be recomputed from scratch in each step. We have shown above that this is drastically different for the decomposition (6.8).

Approximations $\hat{\phi}_n$ can also be computed from the RQ-decomposition of the matrix $H_n := W_n^H A W_n$ (cf. (6.3)). In exact arithmetic, stagnation of WZ-GMRES in step n leads to singularity of H_n. Hence $\hat{\phi}_n$ is not well defined in this case. But on the contrary to (6.14), an approximation $\hat{\phi}_n$ might still be computable. This is particularly true in finite precision arithmetic.

6.4 Behavior of the approximations and prediction of the performance of GMRES

Q_n in (6.8) is obtained from Q_{n-1} by a low rank modification. In particular, if $q_1, ..., q_n$ denote the columns of Q_{n-1}, then

$$Q_n = \begin{bmatrix} q_1 & \dots & q_{n-1} & c_n q_n & -\overline{s}_n q_n \\ 0 & \dots & 0 & s_n & c_n \end{bmatrix}.$$

By (6.12), $c_n \neq -1$, so that we can readily apply results of Elsner and He [8, Section 5]. We partition

$$Q_n = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, \quad U_{11} = [q_1, ..., q_{n-1}, c_n q_n] \in \mathbf{C}^{n\times n}, \quad U_{22} = c_n,$$

to get, in the notation of [8, Theorem 5.3],

$$U_n := U_{11} - U_{12}(I + U_{22})^{-1} U_{21} = Q_{n-1}.$$

From [8, Theorem 5.3] follows that if

$$\Lambda(Q_{n-1}) = \Lambda(U_n) = \{e^{i\beta_j} \ : \ 0 \leq \beta_1 \leq ... \leq \beta_n < 2\pi\},$$
$$\Lambda(Q_n) = \{e^{i\tau_j} \ : \ 0 \leq \tau_1 \leq ... \leq \tau_{n+1} < 2\pi\},$$

then

$$\tau_j \leq \beta_j \leq \tau_{j+1}, \quad \text{for} \ \ j = 1, 2, ..., n.$$

Because of this interlacing property, we can expect a 'smooth' behavior of the eigenvalues of the matrices Q_n and, in particular, of their largest gaps. This is not necessarily the case for the matrices $\hat{Q}_n$ in (6.14), since we know of no easy algebraic relation between them. We remark that often for small n, ϕ_n and $\hat{\phi}_n$ are very close to ϕ and $\hat{\phi}$, respectively. We found no clear explanation for these observations.

The matrix R_n in (6.8) is the leading principal submatrix R in (5.2). Thus, we have $\kappa(R_{n-1}) \leq \kappa(R_n) \leq \kappa(R)$ for $n = 2, ..., d-1$. Monitoring the numbers $\kappa(R_n)$ is a useful check on the numerical accuracy of the WZ-GMRES iterates. As (6.6) shows, the conditioning of R_{n-1} determines the accuracy of the vector y_n. If the R_n become very ill conditioned, it might be useful to terminate WZ-GMRES early. In the derivation of our bounds we have assumed $r_0 \in AK_d(A, r_0)$ for some $d \leq N$. From this assumption follows that $\kappa(R_n) \leq \kappa(R) < \infty$. But for ill conditioned (or singular) A, $\kappa(R)$ and thus $\kappa(R_n)$ might still be very large.

We point out that the bounds (5.22) and (5.23) are independent of each other and might be of different use. On the one hand, if GMRES stagnates early in the iteration, then $\hat{\phi}$ must be small, since (5.23) involves no large constant. Thus, $\hat{\phi}$ and, if defined, the $\hat{\phi}_n$ cannot predict if GMRES will converge faster in later phases of the iteration. On the other hand, early stagnation of GMRES does not imply that ϕ is small, since (5.22) involves the potentially large constant $\kappa(R)$ (or $\kappa(A)$). This has interesting practical consequences. Suppose that GMRES stagnates early in the iteration and that the ϕ_n remain constant during this phase of stagnation. Then they are possibly close approximations to ϕ and if they are large, they predict that the algorithm will converge well later in the iteration. We are led to the following heuristic, which will be confirmed by numerical experiments in Section 7.

Heuristic 6.2 *Suppose that for small n, the ϕ_n stagnate on a (relatively) high level. Then they are often close approximations to ϕ and their size predicts that GMRES will converge well in later phases of the iteration. On the other hand, a fast decrease of the ϕ_n to zero predicts that GMRES will converge slowly over the whole course of the iteration.*

7 Numerical experiments

We consider three example matrices, BWM200, PDE225 and FS1836. All three were downloaded from the `MatrixMarket` web site at URL

`http : //math.nist.gov/MatrixMarket.`

BWM200 and PDE225 are from the NEP matrix collection. BWM200 represents a problem in chemical engineering, while PDE225 is a partial differential equation test problem. FS1836, which comes from the Harwell Boeing sparse matrix collection, represents a problem in chemical kinetics. All three matrices are real and nonsymmetric.

We applied a Householder version [29] of WZ-GMRES to linear systems with these matrices. We computed the approximations ϕ_n as described in Section 6.2 and the $\hat{\phi}_n$ from the RQ-decomposition of $H_n = W_n^H A W_n$ (cf. Section 6.3). All computations were performed using MATLAB 5.0 [26]. To check the numerical accuracy of our computations, we compared the norms of computed and 'updated' residuals (cf. (6.7)).

Before analyzing our results we point out that the bounds (5.22) and (5.23), in particular the approximations ϕ_n and $\hat{\phi}_n$, are sensitive to the scaling of the diagonal entries of the R-factors in the QR- and RQ-decompositions. The results of this section were derived with the following choices:

- The QR-decompositions (6.8) were computed using Givens transformations with nonnegative cosines, i.e. $c_n \geq 0$ for all n (cf. (6.12)).
- The RQ-decompositions $H_n = \hat{R}_n \hat{Q}_n$ were computed so that the $\hat{R}_n$ had positive diagonal entries.

Of all scalings we tested, the above choices led to the largest gaps ϕ and $\hat{\phi}$ and thus yielded the smallest upper bounds (5.22) and (5.23).

Example 7.1 (BWM200)

We used $b = [1, 1, ..., 1]^T$ and $x_0 = 0$. We terminated WZ-GMRES at step d when the updated scaled residual norms satisfied

$$\|r_d\|/\|r_0\| < 10^{-15}.$$

Since the residual norm had almost dropped to machine precision, $\epsilon \approx 10^{-16}$, we are justified to identify d with $\dim AK_N(A, r_0)$. Hence, we consider the last computed approximations, ϕ_d, $\hat{\phi}_d$ and $\kappa(R_d)$ as equal to the respective quantities used in Theorem 5.8.

Figure 1 shows the computed (solid) and updated (dashed) scaled residual norms produced by WZ-GMRES. These two curves are the same, except for the last few steps, when the algorithm has reached its final level of finite precision accuracy. We also plot the upper bounds (5.22) (pluses) and (5.23) (dashdot). We stress that the actual right hand sides of (5.22) and (5.23) are plotted, and not just the 'rates of convergence' γ^{-1} and $\hat{\gamma}^{-1}$. Figures 2 and 3 show the approximations ϕ_n (solid), $\hat{\phi}_n$ (dashed) and $\kappa(R_n)$, respectively, during the run of the algorithm. The final quantities are listed in the following table.

	N	d	ϕ	$\hat{\phi}$	$\kappa(R)$	$\kappa(A)$
BWM200	200	99	2.9629	2.9625	3.4	3.4

For this example, (5.22) and (5.23) describe the actual convergence curve very well. Due to the constant $\kappa(R)$, (5.22) is slightly weaker than (5.23). In Figure 2 we see that the ϕ_n converge 'smoothly' to their final level, while the $\hat{\phi}_n$ show an erratic behavior. Only after step 85 the two approximations are about the same. With respect to Heuristic 6.2, we remark that the smooth behavior of the ϕ_n and their early stagnation at a high level predicts that GMRES will not stagnate in later phases of the iteration. This is indeed the case. Figure 3 shows that $\kappa(R_n) \approx \kappa(R) = \kappa(A)$ after only a few steps, so that good approximations to (5.22) are obtainable early in the iteration.

Example 7.2 (PDE225)

We used $b = A[1, 1, ..., 1]^T$, $x_0 = 0$ and the same termination criterion as for BWM200. Figures 4, 5 and 6 have the same notation as Figures 1, 2 and 3, respectively.

	N	d	ϕ	$\hat{\phi}$	$\kappa(R)$	$\kappa(A)$
PDE225	225	97	1.9675	1.4250	35.1	39.1

GMRES converges slowly for the first 30 steps, so that ϕ is potentially larger than $\hat{\phi}$. This heuristic is confirmed by the actual values. Also, in accordance with Heuristic 6.2, the increase in the speed of convergence of GMRES is predicted by the large size of the ϕ_n early in the iteration. However, since our bounds decrease with the constant rates γ^{-1} and $\hat{\gamma}^{-1}$, they are generally not capable to approximate closely a curve with a dramatically changing slope. Hence, the bounds in this example cannot be as tight as for BWM200. Finally, as Figures 5 and 6 show, good approximations to both our bounds are available after 40 steps.

Example 7.3 (FS1836)

We used two different right hand sides,

$$b^{(1)} = [1, 1, ..., 1]^T, \;\; b^{(2)} = A[1, 1, ..., 1]^T,$$

and $x_0 = 0$ in both cases. FS1836 is ill conditioned, $\kappa(A) \approx 10^{11}$, and the performance of GMRES for a linear system with this 183×183 matrix is very sensitive to the initial residual.

Figure 7 shows the WZ-GMRES convergence curves for both right hand sides. For $b^{(1)}$, fast convergence sets in after 25 steps of total stagnation. There is no stagnation for $b^{(2)}$. Due to ill-conditioning of A, in both cases a finite precision accuracy of little more than $\epsilon * \kappa(A)$ is achievable by WZ-GMRES. We terminated WZ-GMRES shortly after we noticed a considerable difference between the computed and updated residuals. The quantities ϕ_n, $\hat{\phi}_n$ and $\kappa(R_n)$ at the termination step n are listed in the table below. After termination of WZ-GMRES we continued the Arnoldi process to compute ϕ and $\hat{\phi}$. For both right hand sides these are close to the last computed approximations. Figure 8 shows the ϕ_n and $\hat{\phi}_n$ during the iteration. Figure 9 shows the condition numbers $\kappa(R_n)$, which for both right hand sides quickly approach $\kappa(A)$.

	n	ϕ_n	ϕ	$\hat{\phi}_n$	$\hat{\phi}$	$\kappa(R_n)$
FS1836 ($b^{(1)}$)	50	2.2044	2.1931	0.2497	0.2497	1.1e+11
FS1836 ($b^{(2)}$)	25	1.3674	1.2548	0.8892	0.8909	3.3e+09

Our bounds, in particular (5.22), overestimate the actual convergence curves, and hence we have not included them in Figure 7. Still our theory provides useful information because of the different largest gaps in $\Lambda(Q)$ and $\Lambda(\hat{Q})$. WZ-GMRES stagnates for $b^{(1)}$, so $\hat{\phi}$ must be small. On the other hand, ϕ for $b^{(1)}$ is large and the behavior of the ϕ_n predicts for small n, that the algorithm eventually will stop to stagnate. If interpreted as 'rates of convergence', $\hat{\gamma}^{-1} \approx 0.99$ describes the first phase of the iteration, and $\gamma^{-1} \approx 0.85$ describes the second phase. For $b^{(2)}$, there is no stagnation and ϕ and $\hat{\phi}$ are closer to each other. Note that the speed of decrease of the scaled residual norms for $b^{(2)}$ is smaller than for $b^{(1)}$ after the stagnation. This corresponds to the fact that $\gamma^{-1} \approx 0.95$ and $\hat{\gamma}^{-1} \approx 0.97$ for $b^{(2)}$ are both smaller than γ^{-1} for $b^{(1)}$. We finally point out that no bound based on 'ideal GMRES' (cf. Section 3) can describe these differences in the actual convergence behavior of GMRES.

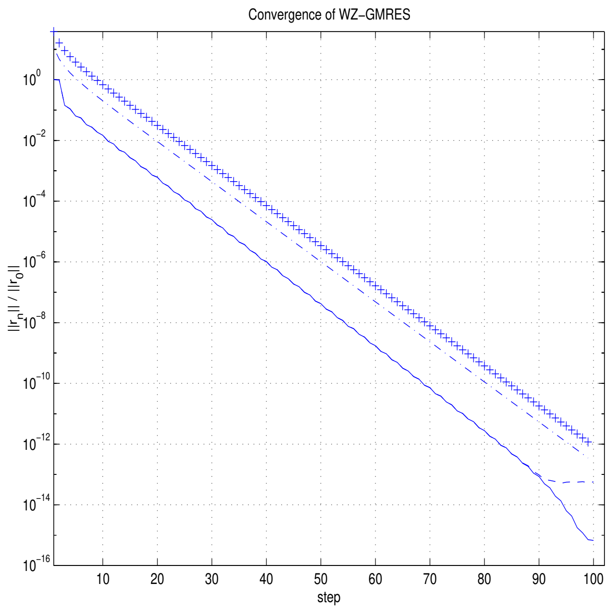

Figure 1: BWM200: Computed (dashed) and updated (solid) scaled residual norms; bounds (5.22) (pluses) and (5.23) (dashdot)

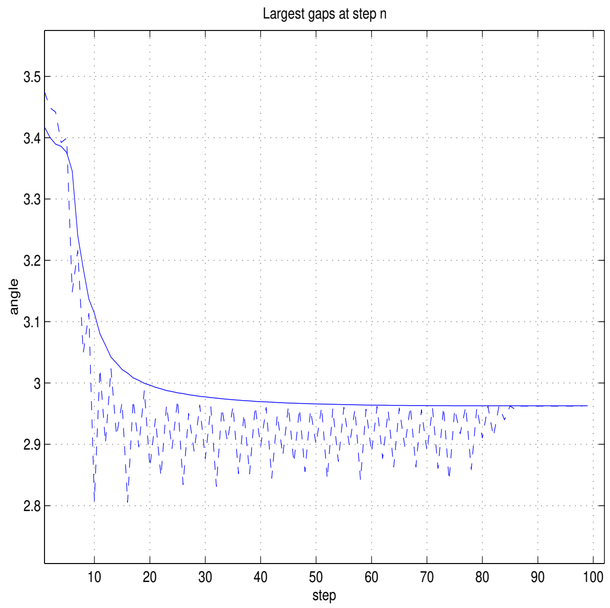

Figure 2: BWM200: ϕ_n (solid) and $\hat{\phi}_n$ (dashed)

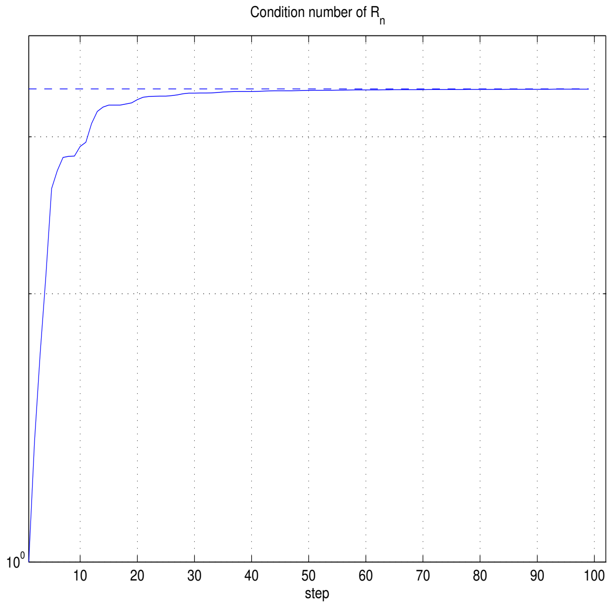

Figure 3: BWM200: $\kappa(R_n)$

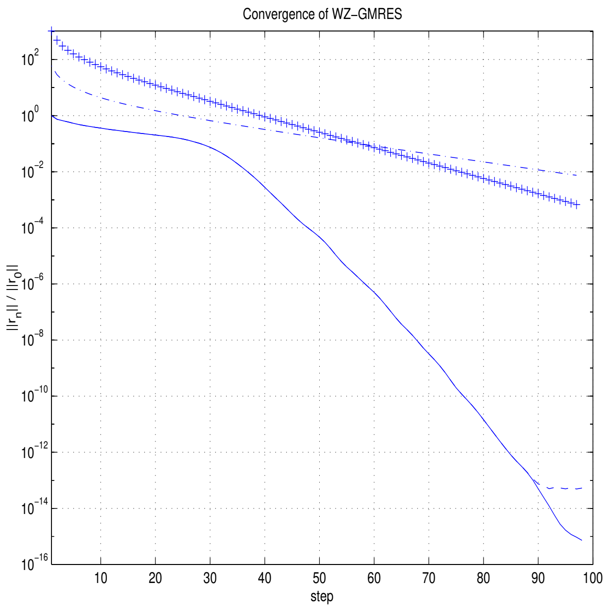

Figure 4: PDE225: Computed (dashed) and updated (solid) scaled residual norms; bounds (5.22) (pluses) and (5.23) (dashdot)

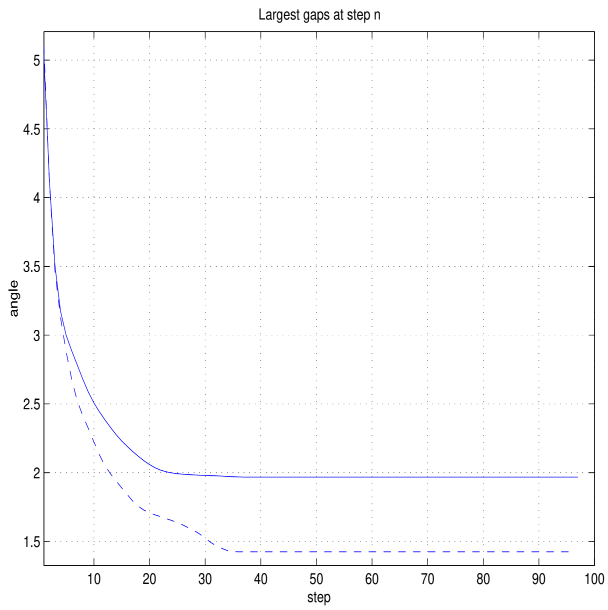

Figure 5: PDE225: ϕ_n (solid) and $\hat{\phi}_n$ (dashed)

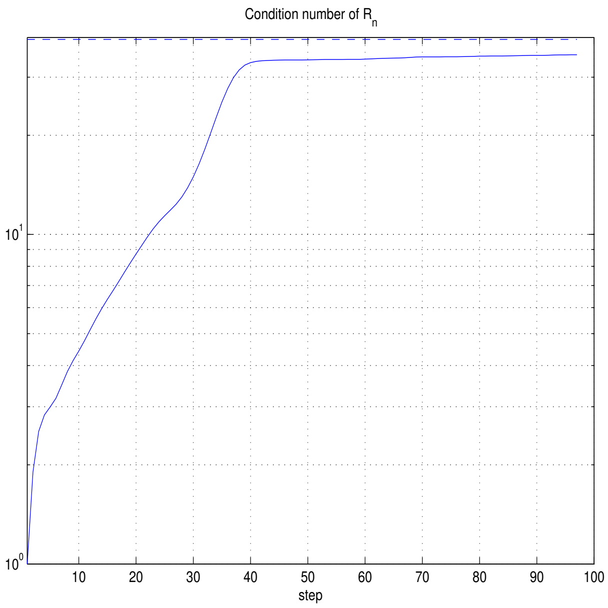

Figure 6: PDE225: $\kappa(R_n)$

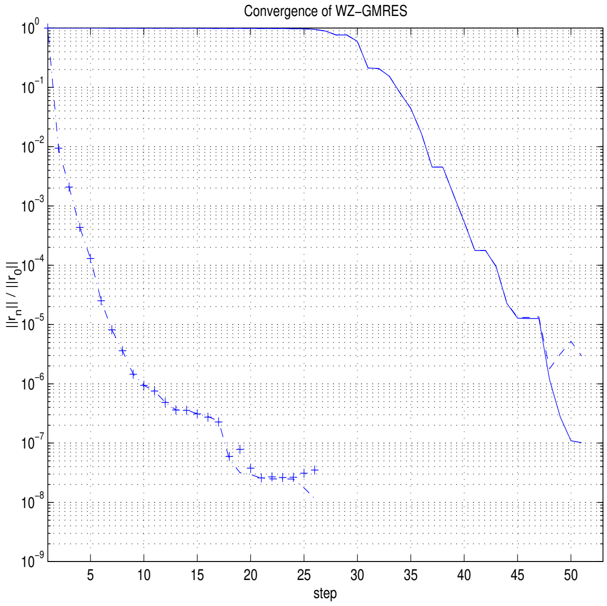

Figure 7: FS1836: Computed (dashed) and updated (solid) scaled residual norms for $b^{(1)}$; computed (pluses) and updated (dashdot) scaled residual norms for $b^{(2)}$

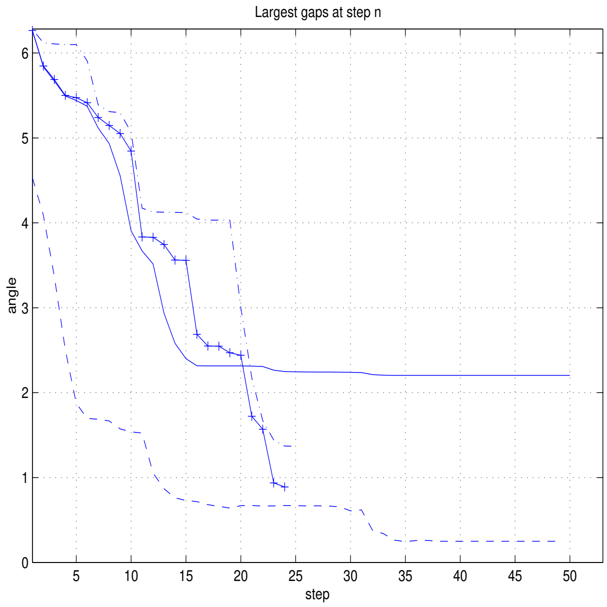

Figure 8: FS1836: ϕ_n (solid) and $\hat{\phi}_n$ (dashed) for $b^{(1)}$; ϕ_n (dashdot) and $\hat{\phi}_n$ (solid-plus) for $b^{(2)}$

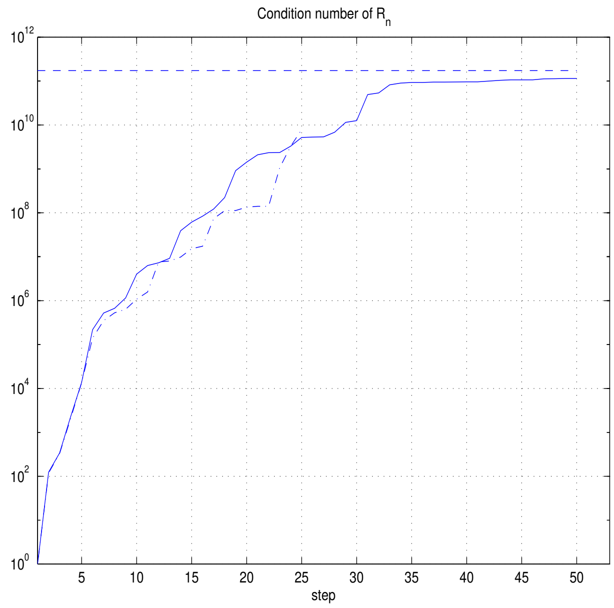

Figure 9: FS1836: $\kappa(R_n)$ for $b^{(1)}$ (solid) and $b^{(2)}$ (dashdot)

8 Conclusions

We have derived new bounds on the residual norms of GMRES, which are conceptually different from the standard 'worst-case' analysis. Our bounds depend on the initial guess and thus are capable to describe the actual convergence behavior of the algorithm. If the bounds are not close to the convergence curve, useful information might still be given by the rates γ^{-1} and $\hat{\gamma}^{-1}$ (cf. Example 7.3). Because our theory not generally excludes singular systems, the bounds are more widely applicable than those reviewed in Section 3. Heuristic 6.2, which was validated by numerical experiments, might prove helpful in large scale applications. The idea of predicting the performance of GMRES during its run might also be useful in the study of other iterative methods.

We have clearly demonstrated the dependence of our bounds on A and r_0 in Example 7.3. Hence, it will be generally impossible to derive reasonable a priori estimates on ϕ and $\hat{\phi}$ based solely on properties of A. Since information about the convergence behavior of GMRES for A and r_0 seems to be 'distilled' into ϕ and $\hat{\phi}$, a priori estimates for these quantities are of great interest and subject of further investigations.

9 Acknowledgments

This paper is part of my Ph.D. work at the Fakultät für Mathematik, Universität Bielefeld, Germany. The work is funded by the Studienstiftung des deutschen Volkes. Thanks to my advisor Ludwig Elsner for his support and guidance, careful reading of the manuscript and several suggestions that improved the paper. I am happy to acknowledge many fruitful discussions with Zdeněk Strakoš, who made very helpful comments and pointed out several references. Part of the work was performed while I was visiting the Institute of Computer Science of the Academy of Sciences of the Czech Republic. Thanks also to Paul Saylor for his careful reading of the manuscript and helpful comments. Finally, thanks to Tino Koch and Miro

Rozložnik for helpful discussions.

References

[1] M. Arioli, V. Pták, and Z. Strakoš. Krylov sequences of maximal length and convergence of GMRES. *To appear in BIT*, 1998.

[2] W. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. of Appl. Math.*, 9:17–29, 1951.

[3] C. Bischof. Incremental condition estimation. *SIAM J. Matrix Anal. Appl.*, 11:312–322, 1990.

[4] P. N. Brown and H. F. Walker. GMRES on (nearly) singular systems. *SIAM J. Matrix Anal. Appl.*, 18:37–51, 1997.

[5] J. B. Conway. *A course in functional analysis.* Springer-Verlag, New York, 2nd edition, 1990.

[6] J. Drkošová, A. Greenbaum, M. Rozložník, and Z. Strakoš. Numerical stability of GMRES. *BIT*, 35:309–330, 1995.

[7] M. Eiermann. Field of values and iterative methods. *In preparation*, 1996.

[8] L. Elsner and C. He. Perturbation and interlace theorems for the unitary eigenvalue problem. *Lin. Alg. Appl.*, 188-189:207–229, 1993.

[9] O. Ernst. Residual-minimizing Krylov subspace methods for stabilized discretizations of convection-diffusion equations. *Submitted to SIAM J. Matrix Anal. Appl.*, 1998.

[10] V. Faber, W. Joubert, E. Knill, and T. Manteuffel. Minimal residual method stronger than polynomial preconditioning. *SIAM J. Matrix Anal. Appl.*, 17:707–729, 1996.

[11] G. H. Golub and C. F. van Loan. *Matrix computations.* The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

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

[13] A. Greenbaum and L. Gurvits. MAX-MIN properties of matrix factor norms. *SIAM J. Sci. Comput.*, 15:348–358, 1994.

[14] A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, 17:465–469, 1996.

[15] A. Greenbaum and Z. Strakoš. Matrices that generate the same Krylov residual spaces. In G. H. Golub, A. Greenbaum, and M. Luskin, editors, *Recent advances in iterative methods*, volume 60 of *The IMA volumes in mathematics and its applications.* Springer-Verlag, New York, 1994.

[16] A. Greenbaum and L. N. Trefethen. GMRES/CR and Arnoldi/Lanczos as matrix approximation problems. *SIAM J. Sci. Comput.*, 15:359–368, 1994.

[17] W. Joubert. A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems. *SIAM J. Sci. Comput.*, 15:427–439, 1994.

[18] T. Koch and J. Liesen. Polynomial iterative methods and the conformal bratwurst map. *In preparation*, 1998.

[19] N. M. Nachtigal, S. Reddy, and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.*, 13:778–795, 1992.

[20] M. Rozložník. *Numerical stability of the GMRES method.* PhD thesis, Institute of Computer Science, Academy of Sciences of the Czech Republic, 1997.

[21] M. Rozložník and Z. Strakoš. Variants of the residual minimizing Krylov space methods. In I. Marek, editor, *Proceedings of the XI-th Summer School on Software and Algorithms of Numerical Mathematics*, 1995.

[22] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.

[23] V. Smirnov and N. Lebedev. *Functions of a complex variable - Constructive theory.* The M.I.T. Press, Cambridge, MA, 1968.

[24] G. Starke. Field-of-values analysis of preconditioned iterative methods for nonsymmetric elliptic problems. *Numer. Math.*, 78:103–117, 1997.

[25] Z. Strakoš. Convergence and numerical behavior of the Krylov space methods. In *Proceedings of the NATO ASI Institute "Algorithms for Large Sparse Linear Algebraic Systems: The State of the Art and Applications in Science and Engineering" (invited lectures)*, pages 175–196, 1998.

[26] The MathWorks. *Using MATLAB (December 1996).* The MathWorks Inc., Natick, 1996.

[27] K.-C. Toh. GMRES vs. ideal GMRES. *SIAM J. Matrix Anal. Appl.*, 18:30–36, 1997.

[28] L. N. Trefethen. Approximation theory and numerical linear algebra. In J. Mason and M. Cox, editors, *Algorithms for approximation II.* Chapman and Hall, London, 1990.

[29] H. F. Walker. Implementation of the GMRES method using Householder transformations. *SIAM J. Sci. Stat. Comput.*, 9:152–163, 1988.

[30] H. F. Walker and L. Zhou. A simpler GMRES. *Numer. Lin. Alg. Appl.*, 1:571–581, 1994.

[31] J. Walsh. *Interpolation and approximation by rational functions in the complex domain.* American Mathematical Society, Providence, RI, 5th edition, 1969.