

# On the Occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods\*

Valeria Simoncini<sup>†</sup>  
Daniel B. Szyld<sup>‡</sup>

**Abstract.** Krylov subspace methods often exhibit superlinear convergence. We present a general analytic model which describes this superlinear convergence, when it occurs. We take an invariant subspace approach, so that our results apply also to inexact methods, and to nondiagonalizable matrices. Thus, we provide a unified treatment of the superlinear convergence of GMRES, conjugate gradients, block versions of these, and inexact subspace methods. Numerical experiments illustrate the bounds obtained.

**Key words.** linear systems, iterative methods, Krylov subspace methods, convergence, superlinear convergence

**AMS subject classifications.** 65F10, 46C05

**DOI.** 10.1137/S0036144503424439

**I. Introduction.** We study certain aspects of the convergence of Krylov subspace methods. These methods are widely used for the solution of  $n \times n$  linear systems of equations of the form

$$(1.1) \quad Ax = f,$$

with  $A$  nonsingular. Let  $r_0 = f - Ax_0$  be the initial residual and let

$$(1.2) \quad \mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, A^2r_0, \dots, A^{m-1}r_0\}$$

be the Krylov subspace of dimension  $m$  defined by  $A$  and  $r_0$ . The shorthand notation  $\mathcal{K}_m$  is used when the dependence on  $A$  and on the specific vector  $r_0$  is clear from the context. Note that  $\mathcal{K}_m \subseteq \mathcal{K}_{m+1}$ .

Krylov subspace methods are iterative methods in which at the  $m$ th step an approximation to the solution of (1.1),  $x_m$ , is found in  $x_0 + \mathcal{K}_m$ . This approximation

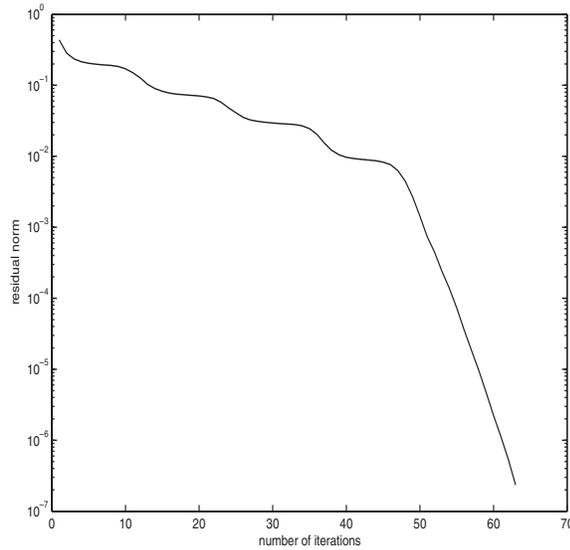
---

\*Received by the editors March 13, 2003; accepted for publication (in revised form) May 8, 2004; published electronically April 29, 2005. Work on this paper was supported in part by the National Science Foundation under grant DMS-0207525. In addition, part of this paper was prepared while the second author visited the Università di Bologna, under the auspices of the Istituto Nazionale di Alta Matematica–GNCS, Italy.

<http://www.siam.org/journals/sirev/47-2/42443.html>

<sup>†</sup>Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy; IMATI-CNR, Pavia, Italy; CIRSA, Ravenna, Italy (valeria@dm.unibo.it).

<sup>‡</sup>Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, PA 19122-6094 (szyld@math.temple.edu).



**Fig. 1.1** *Sample convergence history of GMRES.*

is found by requiring  $x_m$  to be the minimizer of some functional. Different methods depend on the choice of this functional, on the characteristics of the matrix, and on some implementation details. For example, the conjugate gradient method is a Krylov subspace method for the case when  $A$  is symmetric positive definite, and the functional is the  $A$ -norm of the error, i.e., the norm induced by the inner product  $\langle x, y \rangle = x^T A y$ . In the process of iteratively constructing a basis of  $\mathcal{K}_m$ , the method can be implemented so that at each iteration only one matrix-vector multiplication with  $A$  is required, in the form

$$(1.3) \quad z = Av.$$

We discuss further particulars in subsequent sections, but we refer the reader to, e.g., [15], [30] for detailed descriptions of these methods.

Most convergence analyses of Krylov subspace methods indicate a linear convergence rate. Nevertheless, on many occasions, these methods display faster convergence, and this has been called superlinear convergence (in spite of the fact that in exact arithmetic several Krylov subspace methods have final termination). By (nontraditional) “superlinear convergence” we mean that the reduction of the residual norms is linear, but that the linear rate often increases, i.e., it accelerates as the iterations proceed. See Figure 1.1 for a simple example of this behavior. For an example where a Krylov method displays linear and not superlinear convergence, see [39, Fig. 5]. We mention here that a formal analysis of superlinear convergence in the conventional sense has been carried out in the general framework of bounded linear operators in complex separable Hilbert spaces by several authors (see, e.g., [7], [19] [23], [25], [42], and references therein), while the role of the right-hand side has been studied in some special cases [6].

The described phenomenon of superlinear convergence of Krylov subspace methods has been widely observed, and some models explaining this behavior have been

proposed; see [2], [38], [39], and also [9], [10], [12]. The analysis proposed in most of the cited papers relies on the polynomial representation of the approximate solution in  $\mathcal{K}_m$ ; see section 2. In particular, it is argued that once some of the roots of the polynomial defining the Krylov subspace method approximate certain eigenvalues of  $A$ , from then on the process behaves in the same manner as a new process with the same Krylov subspace method, where the initial residual has been stripped of all eigenvector components corresponding to these eigenvalues, and the rate changes. We review some of these results in the next section.

In this paper we propose a different analytic model of superlinear convergence for Krylov subspace methods where we do not use the minimizing polynomials (or their roots). While the polynomial framework provides a good model in the normal case, different techniques seem to be required to reproduce more faithfully the influence of nonnormality on the behavior of Krylov subspace solvers. Thus, we develop a general analytic model of superlinear convergence based in part on the proximity of some invariant subspace to the subspace where the approximation is found. We are motivated by the desire to explain the observed superlinear convergence of inexact Krylov subspace methods. In these methods, the matrix-vector multiplication (1.3) is inexact, i.e., of the form  $(A+E)v$  for some error matrix  $E$  which may change from step to step, and in fact its norm is allowed to grow as the process progresses; see [35] and the references therein. Since the matrix changes from one iteration to the next, the underlying subspaces cannot be written in the form (1.2) with the same matrix at every step, and therefore one cannot write the approximation  $x_m$  in terms of a polynomial on  $A$ .

Our general model is thus applied to the generalized minimal residual (GMRES) [32] and conjugate gradient (CG) methods, and to block Krylov subspace methods, as well as to inexact Krylov subspace methods. We believe that this is the first time that the superlinear convergence of block methods has been addressed. The analysis presented here provides, then, a unified theory of superlinear convergence of Krylov subspace methods, when it occurs.

In this new analytic model, we bound the residual norm after  $m+j$  steps by quantities including the residual norm after  $j$  steps of the same Krylov subspace method, where, however, the initial residual does not have components on a certain invariant subspace. The philosophy of this interpretation is similar to that of the models already proposed, e.g., in [38], [39]; see further section 2 and Remark 4.2.

In the application to inexact Krylov subspace methods, the model developed here provides us with an additional possible explanation of the phenomenon that the norm of  $E$  needs to be small in the initial iterations while it can be relaxed and allowed to grow in later steps: the subspace generated by the inexact Arnoldi method needs to approximate an invariant subspace of the appropriate linear operator; see section 6. Once this is accomplished, the rate of convergence changes, and larger errors are allowed.

We mention that in [4], [5], invariant subspaces have been used in the convergence analysis of Krylov subspace methods in the context of eigenvalue problems. We also point out that using invariant subspaces and not the spectrum of  $A$  directly allows us to treat diagonalizable and nondiagonalizable matrices in the same manner. In addition, our treatment is consistent with the observation, e.g., in [16], that the convergence of GMRES is not determined by the eigenvalues alone. In fact, given any monotonically decreasing convergence curve, it may correspond to any eigenvalue distribution; see further Remark 4.3.

In the next section we first review some of the bounds obtained with analytic models which use polynomial theory for CG and GMRES. Then, we preview the

kinds of bounds we obtain for GMRES with the new analytic model. This serves as an example of our theory developed in section 3, where we prove a very general result on approximation in subspaces in a Hilbert space. We specifically show it for  $\mathbb{R}^n$  and subspaces defined by their bases, but the result applies to any Hilbert space, and in particular to vector spaces over  $\mathbb{C}$ . We remark that in fact in the whole section there is no mention of a matrix  $A$  or of a Krylov subspace.

In section 4 we apply the bounds developed in section 3 to study the superlinear convergence of the GMRES method, and similarly we study the superlinear convergence of the CG method in section 5. Block methods are also analyzed in section 5. Finally, in section 6 we apply the general results of section 3 to inexact Krylov subspace methods. Numerical experiments illustrate the bounds obtained in several cases.

**2. Some Analytic Models of Superlinear Convergence.** We first review the analytic models of superlinear convergence with one result on CG from [38]. As we mentioned, the CG method is used when  $A$  is symmetric positive definite. Let  $x_\star$  be the exact solution of (1.1). After  $m$  iterations of the CG method, the approximate solution  $x_m$  is chosen so as to minimize the  $A$ -norm of the error on  $x_0 + \mathcal{K}_m(A, r_0)$ , i.e.,

$$(2.1) \quad \|x_\star - x_m\|_A = \min_{z \in \mathcal{K}_m} \|x_\star - (x_0 + z)\|_A = \min_p \|x_\star - x_0 - p(A)r_0\|_A,$$

where  $p$  is a polynomial of degree at most  $m - 1$ , and the last equality follows from the fact that  $z \in \mathcal{K}_m(A, r_0)$  must be of the form  $z = p(A)r_0$  for some polynomial  $p$  of degree not greater than  $m - 1$ . The polynomial which attains the minimization in (2.1) is called the CG polynomial. Let  $\lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$  be the eigenvalues of  $A$ . The usual estimate for the linear convergence of CG is  $\|x_m - x_\star\|_A \leq 2\nu^m \|x_0 - x_\star\|_A$ , where  $\nu = (\sqrt{\kappa(A)} - 1)/(\sqrt{\kappa(A)} + 1)$ , with  $\kappa(A) = \|A\| \|A^{-1}\|$  the condition number of  $A$ , and these norms are 2-norms, so that in this case  $\kappa(A) = \lambda_n/\lambda_1$ .

We review the simplest result in [38], taking into account only one eigenvalue. Let  $\theta_1^{(m)} \leq \dots \leq \theta_m^{(m)}$  be the Ritz values, i.e., the roots of the CG polynomial. Let  $\bar{x}_0$  be such that  $x_\star - \bar{x}_0$  is the projection of  $x_\star - x_m$  onto the subspace spanned by all eigenvectors of  $A$  not corresponding to  $\lambda_1$ . Theorem 3.1 of [38] states that if  $\bar{x}_j$  is the  $j$ th approximation of CG starting with  $\bar{x}_0$ , then, for any  $j$ , it holds that

$$(2.2) \quad \|x_\star - x_{m+j}\|_A \leq \alpha_m \|x_\star - \bar{x}_j\|_A,$$

where

$$(2.3) \quad \alpha_m = \left| \frac{\theta_1^{(m)}}{\lambda_1} \right| \max_{\lambda_k \neq \lambda_1} \left| \frac{\lambda_k - \lambda_1}{\lambda_k - \theta_1^{(m)}} \right|.$$

This result implies that if at the  $m$ th iteration the smallest root  $\theta_1^{(m)}$  of the CG polynomial approximates the smallest eigenvalue of  $A$  well, then from this point CG behaves as a CG process with starting vector  $\bar{x}_0$ . For this process the convergence rate is now governed by  $\lambda_n/\lambda_2 < \lambda_n/\lambda_1$ . Similar reasoning applies when considering eigenvalues on the other end of the spectrum. In fact, the bound (2.2) with (2.3) is valid when  $\lambda_1$  is any simple eigenvalue. Bounds of the form (2.2) can also be provided for cases involving more than one eigenvalue. In these cases, the formula for  $\alpha_m$  in (2.3) is more complicated [38].

A similar analytic model of superlinear convergence was presented for GMRES in [39]. The GMRES method is applied to nonsymmetric matrices. It finds at each step the minimum residual solution with respect to the 2-norm over the Krylov subspace; i.e.,  $x_m = x_0 + z_m \in x_0 + \mathcal{K}_m(A, r_0)$  is such that  $r_m = r_0 - Az_m$  satisfies

$$(2.4) \quad \|r_m\| = \min_{d \in \mathcal{AK}_m(A, r_0)} \|r_0 - d\| = \min_{q \in \mathbb{P}_m} \|q(A)r_0\| \leq \|\bar{q}(A)r_0\|,$$

where  $\mathbb{P}_m$  is the set of all polynomials of degree at most  $m$  such that  $q(0) = 1$  and  $\bar{q}$  is any polynomial in that set. The polynomial which attains the minimization in (2.4) is called the GMRES residual polynomial, and its roots  $\theta_i^{(m)}$ ,  $i = 1, \dots, m$ , are called harmonic Ritz values. Assume that  $A$  is diagonalizable, and let  $\lambda_i$ ,  $i = 1, \dots, n$ , be the eigenvalues of  $A$ , with  $\lambda_1$  simple. Let  $\mathbf{Q}$  be the matrix of eigenvectors of  $A$ . If  $\bar{r}_0$  is the projection of  $r_0$  onto the subspace spanned by all eigenvectors of  $A$  not corresponding to  $\lambda_1$ , and if we denote by  $\bar{r}_j$  the residual after  $j$  steps of GMRES starting with  $\bar{r}_0$ , then Theorem 2.3 and Corollary 2.4 in [39] state that

$$(2.5) \quad \|r_{m+j}\| \leq \kappa(\mathbf{Q})\alpha_m \|\bar{r}_j\|,$$

where  $\kappa(\mathbf{Q})$  is the condition number of the eigenvector matrix, and  $\alpha_m$  is given by (2.3). We also point out that similar, although more complicated, bounds can be obtained for  $A$  having eigenvalues with a nontrivial Jordan block [39]. The techniques used in [39] to obtain these results are based on particular choices of the polynomial  $\bar{q}$  in (2.4).

A similar technique was used in [9] when considering the special case where the spectrum of  $A$  has the form  $\sigma(A) = \sigma_c \cup \sigma_o$ , such that, e.g.,  $\sigma_c \subset \{z : |z - 1| < \rho\}$  and  $\sigma_o = \{\lambda_1, \dots, \lambda_s\} \subset \{z : |z - 1| > \rho\}$ , i.e., the case where the spectrum consists of a cluster of eigenvalues and a set of  $s$  outliers. It is shown that in that case

$$\|r_{s+k}\| \leq \beta \rho^k \|r_0\|, \text{ where } \beta = \rho \delta^s \max_{|z-1|=\rho} \|(zI - P_o A)^{-1}\|,$$

$$\delta = \max_{|z-1|=\rho} \max_{1 \leq j \leq s} \frac{|\lambda_j - z|}{|\lambda_j|},$$

and  $P_o$  is the spectral projector onto the invariant subspace corresponding to the eigenvalues in  $\sigma_o$ . Cases with several clusters of eigenvalues were also considered in the same paper. A similar analysis can be found in [12].

We now illustrate our new analytic model of superlinear convergence for GMRES (given in detail in section 4). Here and in the rest of the paper we denote by  $\mathcal{R}(B)$  the range of the matrix  $B$ , and by  $P_B$ , a projection onto  $\mathcal{R}(B)$ . When this projection is orthogonal, we denote it by  $\Pi_B$ . Let  $P_Q$  be a spectral projector onto a simple invariant subspace  $\mathcal{R}(Q)$  of  $A$  of dimension  $k$ , and  $\Pi_Y$  the orthogonal projector onto a  $k$ -dimensional subspace  $\mathcal{R}(Y)$  of  $\mathcal{AK}_m(A, r_0)$ . We show that

$$(2.6) \quad \|r_{m+j}\| \leq \min_{d \in \mathcal{AK}_j(A, r_m)} \{ \|(I - P_Q)(r_m - d)\| + \gamma \|P_Q(r_m - d)\| \},$$

where  $\gamma = \|(I - \Pi_Y)P_Q\|$ . We also relate the upper bound above to the ‘‘most desired’’ problem

$$(2.7) \quad \min_{d \in \mathcal{AK}_j(A, \bar{r}_0)} \|\bar{r}_0 - d\|, \quad \bar{r}_0 = (I - P_Q)r_m,$$

which corresponds to the  $j$ th step of a GMRES process started with  $r_m$  where all components in the invariant subspace  $\mathcal{R}(Q)$  have been eliminated.

The key difference of our new model with respect to previous ones is that the second addend in the upper bound (2.6) replaces the amplification factor  $\kappa(\mathbf{Q})\alpha_m$  in (2.5). Similarly to  $\alpha_m$ , this represents a “noise” term due to the fact that the invariant subspace has not been exactly eliminated. In practice, however, the product  $\gamma\|P_Q(r_m - d)\|$  does not have to be tiny for the first addend to prevail in the bound. Nonetheless, our numerical experience indicates that the desired problem in (2.7) would not by itself represent a correct upper bound for the residual  $\|r_{m+j}\|$ . We discuss in detail the role of  $\gamma$  and  $\|P_Q(r_m - d)\|$  in the next section.

**3. General Theorem.** We use the vector and matrix norms induced by the underlying inner product  $\langle x, y \rangle$  throughout, and when we say “orthogonal” we mean “orthogonal with respect to the underlying inner product.” Similarly, angles, singular values, transposes (the transpose of  $B$  is denoted  $B^*$ ), and pseudoinverses are also with respect to this inner product. In later sections we will apply the results here to different situations and will specify to which inner products those situations apply.

We begin with a very general theorem on minimum residuals over certain subspaces. This result was inspired in part by a related result on augmented methods proved in [31]. We will make repeated use of the fact that given  $X \in \mathbb{R}^{n \times k}$  and  $Y \in \mathbb{R}^{n \times \ell}$ , for any  $r_0 \in \mathbb{R}^n$  it holds that

$$(3.1) \quad \min_{z \in \mathcal{R}([X, Y])} \|r_0 - z\| \leq \min_{x \in \mathcal{R}(X)} \min_{y \in \mathcal{R}(Y)} \|r_0 - x - y\|.$$

**THEOREM 3.1.** *Let  $W_m \in \mathbb{R}^{n \times m}$ ,  $X_j \in \mathbb{R}^{n \times j}$ ,  $Y \in \mathbb{R}^{n \times \ell}$ , and  $Q \in \mathbb{R}^{n \times k}$  be four matrices of full column rank and such that  $\mathcal{R}(Y) \subset \mathcal{R}(W_m)$ . Let  $P_Q$  be any projection onto  $\mathcal{R}(Q)$  and  $\Pi_Y$  be an orthogonal projection onto  $\mathcal{R}(Y)$ . Let  $r_0 \in \mathbb{R}^n$ ,  $d_\star = \arg \min_{d \in \mathcal{R}(W_m)} \|r_0 - d\|$ , and  $r_m = r_0 - d_\star$ . Then*

$$(3.2) \quad \min_{d \in \mathcal{R}([W_m, X_j])} \|r_0 - d\| \leq \min_{d \in \mathcal{R}(X_j)} \{ \|(I - P_Q)(r_m - d)\| + \gamma \|P_Q(r_m - d)\| \},$$

where  $\gamma = \|(I - \Pi_Y)P_Q\|$ .

*Proof.* We have

$$\begin{aligned} \min_{d \in \mathcal{R}([W_m, X_j])} \|r_0 - d\| &= \min_{d \in \mathcal{R}([W_m, X_j, Y])} \|r_0 - d\| \\ &\stackrel{(a)}{\leq} \min_{d_2 \in \mathcal{R}([X_j, Y])} \min_{d_1 \in \mathcal{R}(W_m)} \|r_0 - d_1 - d_2\| \\ &\leq \min_{d_2 \in \mathcal{R}([X_j, Y])} \|r_m - d_2\| \stackrel{(a)}{\leq} \min_{d_3 \in \mathcal{R}(X_j)} \min_{d_4 \in \mathcal{R}(Y)} \|r_m - d_3 - d_4\| \\ &\leq \min_{d_3 \in \mathcal{R}(X_j)} \min_{d_4 \in \mathcal{R}(Y)} \{ \|(I - P_Q)(r_m - d_3)\| + \|P_Q(r_m - d_3) - d_4\| \} \\ &= \min_{d_3 \in \mathcal{R}(X_j)} \{ \|(I - P_Q)(r_m - d_3)\| + \min_{d_4 \in \mathcal{R}(Y)} \|P_Q(r_m - d_3) - d_4\| \} \\ &= \min_{d_3 \in \mathcal{R}(X_j)} \{ \|(I - P_Q)(r_m - d_3)\| + \|(I - \Pi_Y)P_Q(r_m - d_3)\| \} \\ &\stackrel{(b)}{\leq} \min_{d_3 \in \mathcal{R}(X_j)} \{ \|(I - P_Q)(r_m - d_3)\| + \gamma \|P_Q(r_m - d_3)\| \}. \end{aligned}$$

The inequalities marked (a) follow from (3.1), while (b) follows from  $P_Q^2 = P_Q$ .  $\square$

Clearly,  $\gamma = \gamma(Y, Q)$ . We next study how the quantity  $\gamma = \|(I - \Pi_Y)P_Q\|$  depends on the proximity of the two subspaces  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$  when  $\ell = k$ , which is the case we use in our applications of Theorem 3.1; see sections 4–6.

PROPOSITION 3.2. *Let  $Y, Q \in \mathbb{R}^{n \times k}$  be of full column rank. Let  $\Pi_Y, \Pi_Q$  be orthogonal projections onto  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$ , respectively, and let  $P_Q$  be any oblique projection onto  $\mathcal{R}(Q)$ . Then*

$$(3.3) \quad \gamma = \|(I - \Pi_Y)P_Q\| \leq \|(I - \Pi_Y)\Pi_Q\| \|P_Q\| = \|\Pi_Y - \Pi_Q\| \|P_Q\|.$$

*Proof.* Since  $P_Q = \Pi_Q P_Q$  we have that  $\gamma = \|(I - \Pi_Y)\Pi_Q P_Q\|$  and the first inequality in (3.3) follows. Since  $\dim \mathcal{R}(Y) = \dim \mathcal{R}(Q)$ , it follows, e.g., from [18, Theorem I.6.34], that

$$\|(I - \Pi_Y)\Pi_Q\| = \|(I - \Pi_Q)\Pi_Y\| = \|\Pi_Y - \Pi_Q\|.$$

This completes the proof.  $\square$

We note that Proposition 3.2 holds for any norm induced by an inner product. The orthogonality of the projections is with respect to this inner product.

The quantity  $\|\Pi_Y - \Pi_Q\|$  is called the *gap*<sup>1</sup> between two subspaces and is equal to the sine of the maximum canonical angle between the subspaces; see, e.g., [1, sect. 34], [13, sect. 2.5], [21, sect. 2], [36, sect. II.4.1]. The gap is thus a measure of how close the two subspaces are. It follows then from (3.3) that  $\gamma$  is bounded by the product of two quantities: the norm of the projection  $P_Q$ , and the gap between the two subspaces  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$ . We can say that if the gap satisfies  $\|\Pi_Y - \Pi_Q\| \leq \varepsilon$  for some  $\varepsilon > 0$ , then

$$(3.4) \quad \gamma \leq \varepsilon \|P_Q\|.$$

We briefly discuss the case  $\ell \neq k$ . In this case, one has

$$\|\Pi_Y - \Pi_Q\| = \max\{\|(I - \Pi_Y)\Pi_Q\|, \|(I - \Pi_Q)\Pi_Y\|\}.$$

If  $\ell < k$  then  $\|\Pi_Y - \Pi_Q\| = 1$ . If  $\ell > k$ , the *one-sided gap*  $\|(I - \Pi_Y)\Pi_Q\|$  can be interpreted as the opening between  $\mathcal{R}(Q)$  and the “closest” subspace of  $\mathcal{R}(Y)$  of dimension  $k$ . This one-sided gap was used in [4], [5].

All results in this section pertain to subspaces of the form  $\mathcal{R}(Q)$ ,  $\mathcal{R}(Y)$ ,  $\mathcal{R}(X_j)$ , etc., and do not depend on the particular bases chosen to represent them. For simplicity, and without loss of generality, in the rest of the section we assume that the columns of  $X_j$  are orthonormal. This assumption is used explicitly in the proof of Proposition 3.4.

To have a better understanding of the estimate (3.2), we consider a problem of dimension  $2n$  and the following inner product (with its induced vector norm). Let  $x_i, y_i \in \mathbb{R}^n, i = 1, 2$ ; then, if  $x^T = [x_1^T, x_2^T]$  and  $y^T = [y_1^T, y_2^T]$ ,  $\langle x, y \rangle_\star = \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle$ . We use the inequality  $|\alpha| + |\beta| \leq \sqrt{2}\sqrt{\alpha^2 + \beta^2}$  and obtain the following upper bound:

$$(3.5) \quad \begin{aligned} & \min_{d \in \mathcal{R}(X_j)} \{ \|(I - P_Q)(r_m - d)\| + \gamma \|P_Q(r_m - d)\| \} \\ & \leq \sqrt{2} \min_{d \in \mathcal{R}(X_j)} (\|(I - P_Q)(r_m - d)\|^2 + \|\gamma P_Q(r_m - d)\|^2)^{\frac{1}{2}} \\ & = \sqrt{2} \min_{d \in \mathcal{R}(X_j)} \left\| \begin{bmatrix} I - P_Q \\ \gamma P_Q \end{bmatrix} (r_m - d) \right\|_\star. \end{aligned}$$

<sup>1</sup>Also called *distance*, *opening*, or *aperture* between two subspaces.

The advantage of this new bound is that the solution to the problem (3.5) is readily computable. Moreover, we show below that if  $\gamma$  is small (cf. (3.4)), the minimum in (3.5) is close to the minimum of the problem

$$(3.6) \quad \min_{d \in \mathcal{R}[(I-P_Q)X_j]} \|(I-P_Q)r_m - d\|.$$

To see this, consider (3.5) and let  $d = X_j y$  for  $y \in \mathbb{R}^j$ . Define

$$\begin{aligned} \begin{bmatrix} (I-P_Q)X_j \\ \gamma P_Q X_j \end{bmatrix} &= \begin{bmatrix} (I-P_Q)X_j \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma P_Q X_j \end{bmatrix} \equiv B + \Delta B, \\ \begin{bmatrix} (I-P_Q)r_m \\ \gamma P_Q r_m \end{bmatrix} &= \begin{bmatrix} (I-P_Q)r_m \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma P_Q r_m \end{bmatrix} \equiv b + \Delta b. \end{aligned}$$

With this notation, the problem in (3.5) can be written as

$$(3.7) \quad \min_{y \in \mathbb{R}^j} \|(b + \Delta b) - (B + \Delta B)y\|_*,$$

which is a perturbation of the problem

$$\min_{y \in \mathbb{R}^j} \|b - By\|_* = \min_{y \in \mathbb{R}^j} \|(I-P_Q)r_m - (I-P_Q)X_j y\|,$$

which is precisely (3.6).

*Remark 3.3.* Under the assumption that  $\mathcal{R}(Q) \cap \mathcal{R}(X_j) = \{0\}$ , i.e., that  $(I-P_Q)X_j$  has full rank, the least squares problem (3.7) is well posed, since in this case Lemmas A.1 and A.2 in the appendix imply that  $\text{rank } B = \text{rank } B + \Delta B = j$  [8]. We note that implicit in this assumption is the hypothesis that  $j \leq n - k$ . Similarly, implicit in the results used in the rest of the section is the hypothesis that  $2j \leq n$ ; see, e.g., [36, sect. I.5.2]. These conditions are not at all restrictive in practice; in general  $n$  is much larger than  $j$  or  $k$ .

We proceed now to analyze how close the residual of the problem (3.6) is to that of (3.5). Let  $\sigma_{\min}(M)$  denote the minimum singular value of  $M$ . The proof of the following result is given in the appendix.

**PROPOSITION 3.4.** *With the current notation and assumptions we have that*

$$\min_{y \in \mathbb{R}^j} \|(b + \Delta b) - (B + \Delta B)y\|_* \leq \min_{d \in \mathcal{R}[(I-P_Q)X_j]} \|(I-P_Q)r_m - d\| + \|\Delta r\|_*,$$

where

$$\|\Delta r\|_* \leq \gamma \min_{d \in \mathcal{R}[(I-P_Q)X_j]} \|(I-P_Q)r_m - d\| + \frac{\gamma}{\tau_j} \|(I - \Pi_{X_j})r_m\|,$$

$\Pi_{X_j}$  is the orthogonal projector onto  $\mathcal{R}(X_j)$ , and

$$(3.8) \quad \tau_j = \sigma_{\min}(Z^* X_j),$$

with  $Z$  a matrix having orthonormal columns and spanning  $\mathcal{R}[(I-P_Q)^*(I-P_Q)X_j]$ .

*Remark 3.5.* We note that if  $\|\Pi_Y - \Pi_Q\| < 1$ , then  $\tau_j$  defined in (3.8) is bounded from below by  $1 - \|\Pi_Y - \Pi_Q\|$  [28]. Thus, if the subspaces  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$  are close to each other, this quantity is close to 1.

In summary, Proposition 3.4 says that if the subspaces  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$  are close enough, as measured by the gap, then the minima of problems (3.5) and (3.6) are close

to each other. In the experiments in sections 4–6 we display the residual of the two minimization problems, and their small difference can be appreciated. In addition, by collecting our bounds the difference between the original approximation problem and the “desired” bound (3.6) can also be appreciated. Numerical values for each term are given for Example 4.4 in the next section.

**COROLLARY 3.6.** *Let the hypotheses of Theorem 3.1 and Proposition 3.4 hold. Then*

$$\min_{d \in \mathcal{R}([W_m, X_j])} \|r_0 - d\| \leq \sqrt{2}(1 + \gamma) \min_{d \in \mathcal{R}((I - P_Q)X_j)} \|(I - P_Q)r_m - d\| + \frac{\sqrt{2}\gamma}{\tau_j} \|(I - \Pi_{X_j})r_m\|.$$

**4. Superlinear Convergence of GMRES.** Now consider the GMRES method applied to (1.1). We assume here and in the following that  $x_0 = 0$  and thus that  $r_0 = f$ . Moreover, throughout this section we use the 2-norm.

Let  $W_m, X_j$  be two matrices such that  $\mathcal{R}(W_m) = AK_m(A, r_0)$  and let  $\mathcal{R}(X_j) = AK_j(A, r_m)$ , where  $r_m = r_0 - d_*$  is the GMRES solution after  $m$  steps. In this case, we have that  $\mathcal{R}([W_m, X_j]) = AK_{m+j}(A, r_0)$ . Let  $Q$  be a matrix whose columns are a basis of a  $k$ -dimensional simple invariant subspace of  $A$ ; this implies that there is a complementary subspace in  $\mathbb{R}^n$  which is also invariant [36, sect. V.1].

Let  $\mathbf{Q} = [Q, Q_o]$  be an eigenvector matrix of  $A$ , and let  $\mathbf{T} = (\mathbf{Q}^{-1})^* = [T, T_o]$ . Let  $P_Q = QT^*$  be the spectral projector onto  $\mathcal{R}(Q)$ . Then  $I - P_Q$  is also a spectral projector, and we have the identities

$$(I - P_Q)\mathcal{K}_j(A, r_m) = \mathcal{K}_j(A, (I - P_Q)r_m)$$

and  $P_Q\mathcal{K}_j(A, r_m) = \mathcal{K}_j(A, P_Qr_m)$ . This follows from the fact that  $A$  commutes with its spectral projectors.

We further assume that the GMRES process on the linear system  $Ax = (I - P_Q)r_m$  does not encounter early termination, i.e., that the problem (3.6) has a solution. Let the columns of  $U_j$  be a basis of  $\mathcal{K}_j(A, (I - P_Q)r_m)$ . Then this assumption implies that  $AU_j$  is not rank deficient, or equivalently that  $(I - P_Q)X_j$  is full rank; cf. Remark 3.3. With this notation and assumptions, the results of section 3 imply the following theorem.

**THEOREM 4.1.** *Let  $Y$  be a matrix whose columns are a basis of a  $k$ -dimensional subspace of  $AK_m(A, r_0)$ . Let  $\bar{r}_j$  be the residual after  $j$  steps of a GMRES method applied to  $(I - P_Q)r_m$ , i.e.,*

$$(4.1) \quad \|\bar{r}_j\| = \min_{d \in AK_j(A, (I - P_Q)r_m)} \|(I - P_Q)r_m - d\|.$$

Let  $\Pi_Y$  be the orthogonal projection onto  $\mathcal{R}(Y)$ , and let  $\gamma = \|(I - \Pi_Y)P_Q\|$ . Then

$$(4.2) \quad \begin{aligned} \|r_{m+j}\| &= \min_{d \in AK_{m+j}(A, r_0)} \|r_0 - d\| \\ &\leq \min_{d \in AK_j(A, r_m)} \{ \|(I - P_Q)(r_m - d)\| + \gamma \|P_Q(r_m - d)\| \} \end{aligned}$$

$$(4.3) \quad \leq \sqrt{2} \min_{d \in AK_j(A, r_m)} \left\| \begin{bmatrix} I - P_Q \\ \gamma P_Q \end{bmatrix} (r_m - d) \right\|$$

$$(4.4) \quad \leq \sqrt{2}(1 + \gamma)\|\bar{r}_j\| + \frac{\sqrt{2}\gamma}{\tau_j} \|(I - \Pi_{X_j})r_m\|,$$

where  $\tau_j = \sigma_{\min}(Z^*X_j)$ ,  $X_j$  is a matrix whose orthonormal columns span  $AK_j(A, r_m)$ , and  $Z$  is a matrix whose orthonormal columns span  $(I - P_Q)^*AK_j(A, (I - P_Q)r_m)$ .

Observe that

$$\|(I - \Pi_{X_j})r_m\| = \|r_m - \Pi_{X_j}r_m\| = \min_{d \in AK_j(A, r_m)} \|r_m - d\|$$

is the norm of the residual after  $j$  steps of GMRES with initial residual  $r_m$ . This observation implies that the second term in (4.4) decreases as  $j$  increases. In our numerical experiments later in the paper, we report values for the bounds (4.3) and (4.4) for several examples.

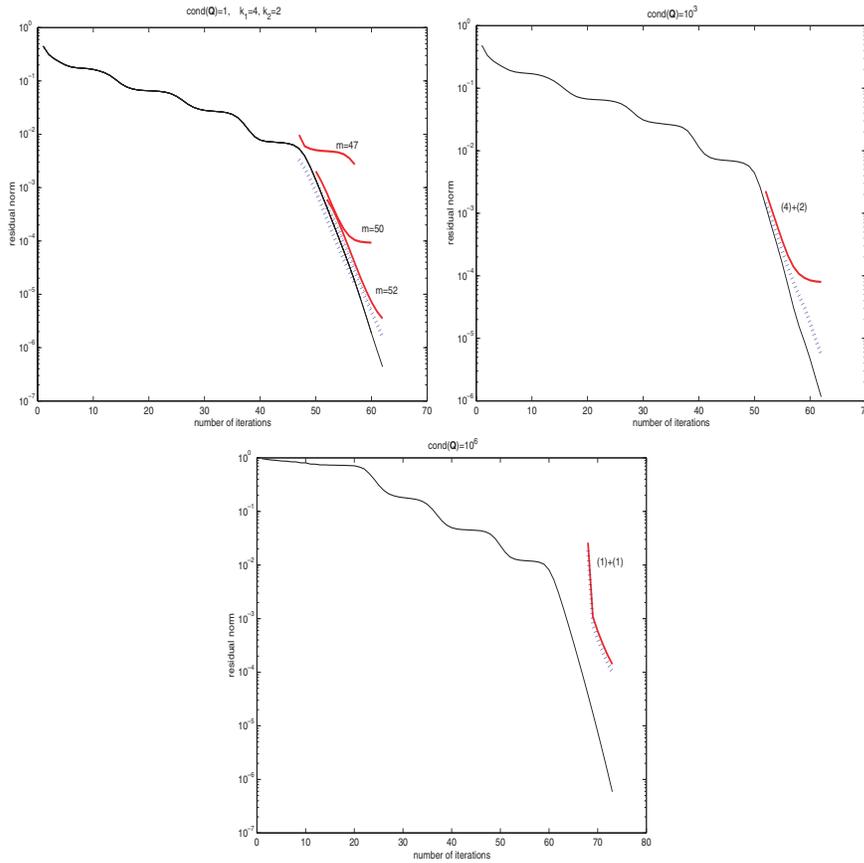
We remark that Theorem 4.1 is valid for *any* linearly independent set of  $k$  columns of  $Y$  taken from  $AK_m(A, r_0)$ . In particular, this holds for  $\mathcal{R}(Y)$  being the closest to the invariant subspace  $\mathcal{R}(Q)$ , i.e., the one making  $\|(I - \Pi_Y)P_Q\|$  smallest. Note also that the only place where the choice of  $Y$  plays a role is in the definition of  $\gamma$ . Several candidates for  $Y$  can be considered, with columns of the form  $y_i = Az_i$ ,  $z_i \in \mathcal{K}_m(A, r_0)$ . In particular,  $z_i$  can be taken to be a Ritz or harmonic Ritz vector (Galerkin or Petrov–Galerkin approximation) in the Hermitian case, while refined Ritz vectors (constrained least squares approximation) could also be considered in the non-Hermitian case; see [3] for definitions and examples. In all cases, by means of Proposition 3.2, the magnitude of  $\gamma$  is directly related to the gap between  $\mathcal{R}(Y)$  and  $\mathcal{R}(Q)$ .

*Remark 4.2.* The interpretation of Theorem 4.1 is that when the subspace  $AK_m(A, r_0)$  has captured  $k$  linearly independent vectors close to vectors in an invariant subspace  $\mathcal{R}(Q)$ , the GMRES method behaves almost like another GMRES process applied to an initial vector  $(I - P_Q)r_m$  with no components in the invariant subspace  $\mathcal{R}(Q)$ ; see (4.1). The difference between the estimate (4.1) and the bound proved in (4.3) is not greater than the second addend in (4.4). As long as this quantity is small, e.g., when the gap is small enough (cf. (3.4)), Theorem 4.1 presents a good model for the superlinear behavior of GMRES as follows: once the subspace  $\mathcal{K}_m(A, r_0)$  captures  $k$  linearly independent vectors close to some  $k$ -dimensional simple invariant subspace  $\mathcal{R}(Q)$ , then there is a change in the convergence rate to that of problem (4.1).

*Remark 4.3.* In [16] it is shown that the convergence of GMRES is not determined by the eigenvalues alone; the initial residual  $r_0$  also plays a role. In (4.4) information on the initial residual is maintained through  $\bar{r}_j$  and  $r_m$ . The spectral information from the matrix  $A$  is present through the spectral projector  $P_Q$  and also by the choice of the subspace  $\mathcal{R}(Y) \subset AK_m$ .

We would like to also remark that our bound depends on  $\|P_Q\|$ . Indeed, for  $j = 0$ , we obtain  $\|r_m\| \leq \|(I - P_Q)r_m\| + \gamma\|P_Qr_m\|$ . When  $P_Q$  is not an orthogonal projector,  $\|(I - P_Q)r_m\|$  and  $\|P_Qr_m\|$  may be larger than  $\|r_m\|$ . The quantity on the right-hand side of the bound for  $j = 0$  provides our starting estimate for the residual norm  $\|r_{m+j}\|$ ,  $j = 1, 2, \dots$ , which may be far from tight when  $\|P_Q\|$  is large; we analyze such a situation in our first set of numerical experiments, Example 4.4 below. We point out that in contrast to some of the previous models (cf. (2.5)), our bound does not depend on  $\kappa(\mathbf{Q}) = \|\mathbf{Q}\| \|\mathbf{Q}^{-1}\|$ , the condition number of the whole eigenbasis  $\mathbf{Q}$  of  $A$ . As before, and without loss of generality, we assume  $\mathbf{Q} = [Q, Q_o]$  and write  $\mathbf{T} = (\mathbf{Q}^{-1})^* = [T, T_o]$ . We recall that  $\|P_Q\| \leq \kappa(\mathbf{Q}) = \|\mathbf{Q}\| \|\mathbf{Q}^{-1}\|$ . This can be seen by writing  $\|P_Q\| = \|QT^*\| \leq \|T^*\| \leq \|\mathbf{Q}^{-1}\| = \kappa(\mathbf{Q})$  and noting that  $\|\mathbf{Q}\| = 1$ . Thus our bound is only influenced by a local sensitivity of the basis and not by the ill-conditioning of the whole basis  $\mathbf{Q}$ .

*Example 4.4.* We consider a family of  $100 \times 100$  matrices  $A^{(i)} = \mathbf{Q}^{(i)}\Lambda(\mathbf{Q}^{(i)})^{-1}$ , where  $\mathbf{Q}^{(i)}$  is determined so as to have condition number  $\kappa_i$  taking values  $1, 10^3, 10^6$ .



**Fig. 4.1** Example 4.4. Convergence history of GMRES,  $\kappa(\mathbf{Q}) = 1, 10^3, 10^6$ .

For each  $\kappa_i$ , the matrix  $\mathbf{Q}^{(i)}$  is defined as  $\mathbf{Q}^{(i)} = D^{(i)}U^*$ , where  $D^{(i)}$  is a diagonal matrix of positive real values uniformly distributed between 1 and  $\kappa_i$ . The matrix  $U$  is the orthogonal matrix in the QR factorization of the lower triangular portion of

$$\begin{bmatrix} 1 & n+1 & 2n+1 & \cdots & \vdots \\ 2 & n+2 & 2n+2 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & 2n & 3n & \cdots & n^2 \end{bmatrix}.$$

By construction, we thus have  $\kappa(\mathbf{Q}^{(i)}) = \kappa_i$ . Matrix  $\Lambda$  is the diagonal matrix of the (fixed for the whole family) eigenvalues of  $A^{(i)}$ , which were chosen to be 0.1, 0.2, 0.3, 0.4, 5, 6, ..., 100. The right-hand side  $f$  is the normalized vector of all ones.

The GMRES convergence curves in the analyzed cases are given in the three plots of Figure 4.1 (solid line). The curves of our bound (4.3) (solid line, in red) are also reported starting at specific values of  $m$  and different values of  $j$ . The dotted line is the “optimal” estimate (4.1). The same convention is used in all figures in the paper.

**Table 4.1** Example 4.4.  $\kappa(\mathbf{Q}) = 1$ ,  $k_1 = 4$ ,  $k_2 = 2$ ,  $m = 52$ .

$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
7.6155e-02	1.00	4.2272e-04	2.7102e-05	4.2185e-04
$j$	$\ r_{m+j}\ $	Term 1	Term 2	Optimal
1	2.2891e-04	2.8463e-04	2.0553e-06	2.8463e-04
2	1.2311e-04	1.6835e-04	2.0373e-06	1.6835e-04
3	6.5794e-05	9.3375e-05	2.0189e-06	9.3375e-05
4	3.4534e-05	5.0366e-05	2.0020e-06	5.0366e-05
5	1.7492e-05	2.6965e-05	1.9864e-06	2.6965e-05

**Table 4.2** Example 4.4.  $\kappa(\mathbf{Q}) = 10^6$ ,  $k_1 = 1$ ,  $k_2 = 1$ ,  $m = 68$ .

$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
2.6041e-03	3.325015e+04	3.8053e-05	1.8419e-02	1.8419e-02
$j$	$\ r_{m+j}\ $	Term 1	Term 2	Optimal
1	1.7299e-05	7.6656e-04	3.2841e-07	7.6656e-04
2	7.6487e-06	4.1386e-04	1.3463e-07	4.1386e-04
3	3.3319e-06	2.4337e-04	1.0216e-07	2.4337e-04
4	1.4182e-06	1.5063e-04	6.5251e-08	1.5063e-04
5	5.8866e-07	9.9942e-05	4.7947e-08	9.9942e-05

We used as the relevant invariant subspace  $\mathcal{R}(Q)$  the one which corresponds to the  $k_1$  smallest and  $k_2$  largest eigenvalues, i.e.,  $k = k_1 + k_2$  (in the figures, the corresponding bound is labeled  $(k_1) + (k_2)$ ). We used this convention in all examples where only real eigenvalues are considered. In the cases of  $\kappa(\mathbf{Q}) = 1$  and  $\kappa(\mathbf{Q}) = 10^3$ , we used  $k_1 = 4$ ,  $k_2 = 2$ , and thus  $k = 6$ . For  $\kappa(\mathbf{Q}) = 10^6$  we used  $k_1 = k_2 = 1$ . Ritz vectors were used to determine the eigenspace approximations.

We call the reader's attention to the figure corresponding to  $\kappa(\mathbf{Q}) = 1$ , where it can be observed that the "optimal" estimate (4.1) is in fact below the curve for the residual ( $m = 47$ ). This indicates that the second term in (4.2) cannot be removed. We also note that for this value of  $m$ , the curve for the bound is only able to capture the change of slope of the convergence curve after a significant delay. These two observations are related to the fact that the subspace  $\mathcal{R}(Y)$  is not close enough to an invariant subspace of  $A$ .

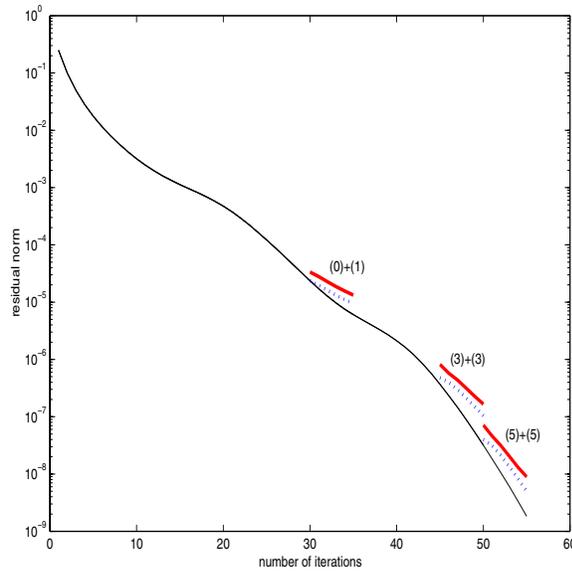
We especially note that for the case  $\kappa(\mathbf{Q}) = 10^3$ , we have  $\|P_Q\| \approx 34$ . This is why in spite of the condition number being two orders of magnitude larger, our bound (4.3) approximates the convergence curve so well.

To make the description of the experiments more informative, we report in Tables 4.1–4.2 some quantities monitored during the iteration, for  $\kappa(\mathbf{Q}) = 1$ ,  $10^6$ . Recall that  $\|P_Q\| = \|I - P_Q\|$  [17], [22]. The columns labeled "Term 1" and "Term 2" contain the values of the norms of the first and second component of (4.3), respectively. The columns labeled "Optimal" have the values of (4.1). We note that the first component of (4.3), i.e., the quantity labeled "Term 1" in the tables, always dominates and is very close to the value labeled "Optimal," i.e., the estimate (4.1).

In Table 4.3 we explore the sharpness of the bound in (4.4) and how much each of the two terms (labeled (4.4)-I and (4.4)-II in the table) contribute to this bound. We report the values of these addends as well as the value of (4.3) for  $\kappa(\mathbf{Q}) = 1$  ( $m = 52$ ) and for  $\kappa(\mathbf{Q}) = 10^6$  ( $m = 68$ ). Note that in the latter case, the first term of (4.4) is

**Table 4.3** Example 4.4. Evaluation of the sharpness of the first and second addends of the bound in (4.4).

$j$	$\kappa(\mathbf{Q}) = 1$			$\kappa(\mathbf{Q}) = 10^6$		
	Problem (4.3)	Bound (4.4)-I	Bound (4.4)-II	Problem (4.3)	Bound (4.4)-I	Bound (4.4)-II
1	4.02e-04	4.33e-04	3.07e-05	1.08e-03	1.08e-03	1.97e-04
2	2.38e-04	2.56e-04	1.83e-05	5.85e-04	5.86e-04	1.05e-04
3	1.32e-04	1.42e-04	1.04e-05	3.44e-04	3.45e-04	6.24e-05
4	7.12e-05	7.66e-05	6.11e-06	2.13e-04	2.13e-04	3.85e-05
5	3.82e-05	4.10e-05	4.04e-06	1.41e-04	1.41e-04	2.56e-05
6	2.08e-05	2.22e-05	3.20e-06	9.72e-05	9.75e-05	1.76e-05
7	1.18e-05	1.24e-05	2.90e-06	6.76e-05	6.78e-05	1.23e-05
8	7.21e-06	7.17e-06	2.79e-06	4.69e-05	4.70e-05	8.58e-06
9	4.75e-06	4.19e-06	2.74e-06	3.29e-05	3.30e-05	6.04e-06
10	3.51e-06	2.41e-06	2.70e-06	2.44e-05	2.45e-05	4.45e-06



**Fig. 4.2** Example 4.5. Convergence history of GMRES.

always larger than the value of (4.3). In the first case, this is only true for  $j \leq 7$ . For larger values of  $j$ , the second term in (4.4) is essential.

*Example 4.5.* This example is from [39, sect. 3]. Let  $A = SBS^{-1}$ , and let  $B \in \mathbb{R}^{n \times n}$  be a diagonal matrix with entries  $1, 1.2, 3, \dots, 100$ , while  $S \in \mathbb{R}^{n \times n}$  is the bidiagonal matrix with ones on the main diagonal and  $\beta = 0.9$  on the upper diagonal. The right-hand side is  $f = Ae$ , normalized to have unit norm, where  $e$  is the vector of all ones.

In Figure 4.2 we plot the convergence history of GMRES and the curves of our bound (4.3), corresponding to three different stages of the procedure, for  $m = 30, 45, 50$ . Curves for (4.1) are also shown (dotted line). The number of iterations to evaluate the bound was  $j = 1, \dots, 5$  in all cases. Tables 4.4–4.5 show some details of these runs for  $m = 30$  and  $m = 45$ .

**Table 4.4** Example 4.5.  $m = 30$ ,  $k_1 = 0$ ,  $k_2 = 1$ .

$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
4.6157e-03	1.345362e+00	2.3673e-05	1.3321e-06	2.3709e-05
$j$	$\ r_{m+j}\ $	Term 1	Term 2	Optimal
1	1.7201e-05	1.9738e-05	6.0151e-09	2.3709e-05
2	1.2738e-05	1.6108e-05	5.9924e-09	1.9738e-05
3	9.6771e-06	1.3232e-05	5.7235e-09	1.6108e-05
4	7.5631e-06	1.1061e-05	4.9958e-09	1.3232e-05
5	6.0603e-06	9.3903e-06	4.0432e-09	1.1061e-05

**Table 4.5** Example 4.5.  $m = 45$ ,  $k_1 = 3$ ,  $k_2 = 3$ .

$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
2.8908e-01	2.294157e+00	3.6788e-07	3.5084e-07	4.7727e-07
$j$	$\ r_{m+j}\ $	Term 1	Term 2	Optimal
1	2.3419e-07	3.9671e-07	9.5605e-08	4.7727e-07
2	1.4665e-07	2.9969e-07	8.0615e-08	3.9671e-07
3	9.0057e-08	2.1461e-07	6.6688e-08	2.9966e-07
4	5.4007e-08	1.5020e-07	5.8012e-08	2.1456e-07
5	3.1763e-08	1.0432e-07	5.3443e-08	1.5016e-07

We comment here that in contrast to what is seen in [39], we choose eigenvectors corresponding to eigenvalues on both ends of the spectrum. In particular, it can be observed that the first approximating curve, for  $m = 30$ , captures the behavior of the GMRES approximation with information from the right end of the spectrum.

Our next example illustrates the applicability of our results to nondiagonalizable matrices.

*Example 4.6.* Consider  $A = SBS^{-1}$  with  $S$  as in Example 4.5, while  $B$  is the block diagonal matrix

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & D \end{bmatrix}, \quad D = \text{diag}(4, 5, \dots, 100).$$

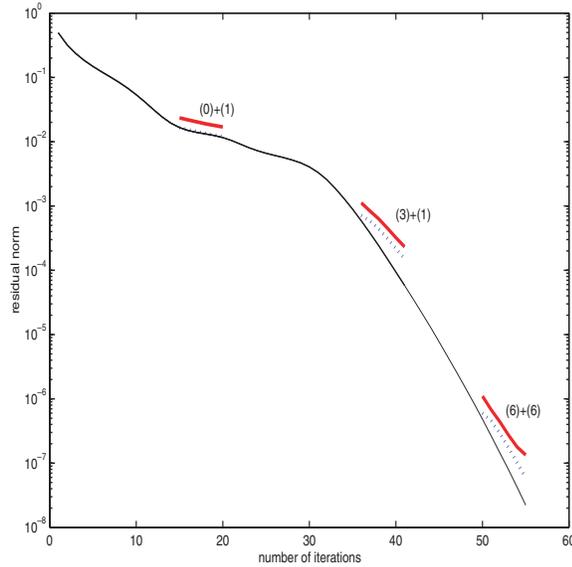
The right-hand side is the vector  $e$  normalized. The matrix  $A$  has the multiple eigenvalue  $\lambda = 1$  with one Jordan block of size 3. A basis for the associated invariant subspace  $Q$  was computed by explicitly generating the chain of principal vectors [36].

In Figure 4.3 we report the convergence history of GMRES and our bounds for  $m = 15, 36, 50$  and  $j = 1, \dots, 5$ . We present some details in Table 4.6.

We finally consider a problem with a complex spectrum, in which superlinear convergence can be observed.

*Example 4.7.* The linear system is constructed so that the experiment can be fully replicated. We consider the preconditioned system  $MP^{-1}x = f$ , where  $f$  is the normalized vector of all ones, while

$$M = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}, \quad P = \begin{bmatrix} D & B \\ 0 & -B^T D^{-1} B \end{bmatrix},$$



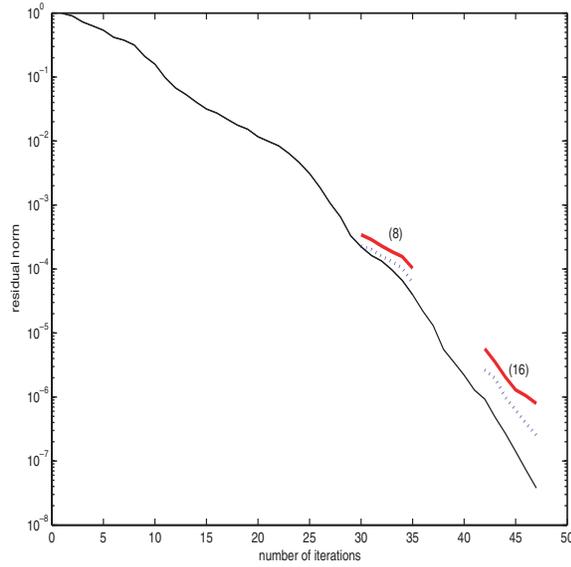
**Fig. 4.3** Example 4.6. Convergence history of GMRES.

**Table 4.6** Example 4.6. Relevant quantities after  $m$  GMRES iterations.  $Q$  is given by the eigenvectors corresponding to the  $k_1$  smallest and  $k_2$  largest eigenvalues.

$m$	$k_1$	$k_2$	$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
15	0	1	1.6983e-01	1.345362	1.6551e-02	3.5886e-04	1.6549e-02
36	3	1	1.7726e-01	2.294157	5.8776e-04	4.4794e-04	7.1121e-04
50	6	6	4.4461e-01	2.294157	4.8169e-07	4.0699e-07	5.9786e-07

where  $A \in \mathbb{R}^{n_1 \times n_1}$ ,  $n_1 = 100$ , is the scaled finite difference discretization of the Laplace operator in  $[0, 1]^2$  (see, e.g., [30]),  $D$  is the diagonal of  $A$ , and  $B \in \mathbb{R}^{n_1 \times 25}$  is bidiagonal with ones on the diagonal and minus ones on the lower diagonal. The overall system dimension is thus  $n = 125$ . This block structure is typical in the solution of saddle-point problems, and preconditioners of type  $P$  have been used in the literature; see, e.g., [20], [33].

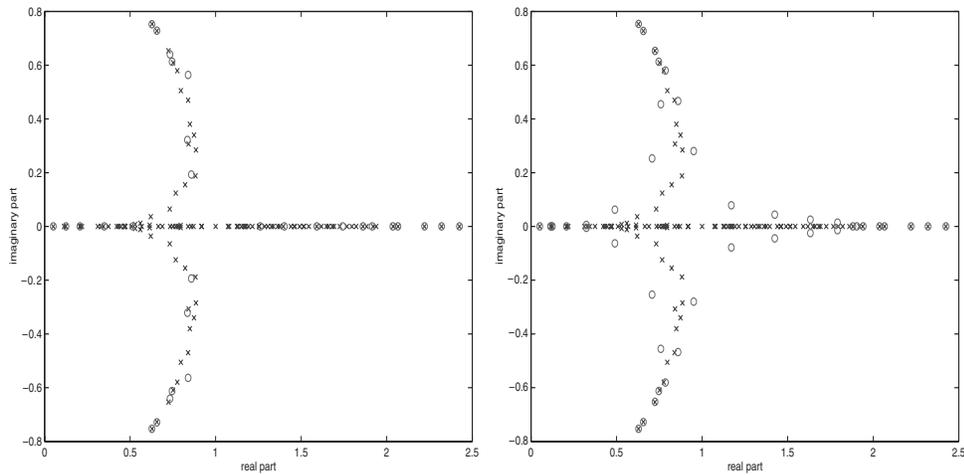
Figure 4.4 reports the GMRES convergence history and our bounds starting at  $m = 30$  and  $m = 42$ , for  $j = 1, \dots, 5$ . Table 4.7 provides some details. In Figure 4.5 we report the eigenvalues of the matrix  $MP^{-1}$  (“ $\times$ ”) and their approximations by means of harmonic Ritz values (“ $\circ$ ”) for  $m = 30$  (left) and for  $m = 42$  (right). We mention in passing that this figure illustrates the fact that all complex eigenvalues (with nonzero imaginary part) for these types of preconditioned problems lie in the circle centered at one with radius smaller than one [33]. It is worth noting that the exterior eigenvalues are well approximated already for  $m = 30$ , as clearly confirmed by the figure. For  $m = 30$ , the invariant subspace of dimension  $k = 8$  corresponds to the smallest real eigenvalue, the four most extreme complex ones, and the first three largest real eigenvalues. For  $m = 42$  we have  $k = 16$  with the invariant subspace chosen in a similar manner.



**Fig. 4.4** Example 4.7. Complex spectrum. Convergence history of GMRES.

**Table 4.7** Example 4.7. Complex spectrum. Relevant quantities after  $m$  GMRES iterations.  $Q$  is given by the eigenvectors corresponding to the  $k$  most exterior eigenvalues (see text).

$m$	$k$	$\gamma$	$\ P_Q\ $	$\ r_m\ $	$\ P_Q r_m\ $	$\ (I - P_Q)r_m\ $
30	8	1.7106e-01	2.359412	2.2290e-04	4.1922e-05	2.2646e-04
42	16	2.5793e-01	6.628733	9.2831e-07	2.4653e-06	2.6379e-06



**Fig. 4.5** Example 4.7. Spectrum of  $MP^{-1}$  ( $\times$ ) and harmonic Ritz values ( $o$ ) for  $m = 30$  (left) and  $m = 42$  (right).

We have assumed in Theorem 4.1 that  $\ell = k$ . We now consider the case  $\ell < k$ , which is important when an eigenvalue  $\lambda_1$  of  $A$  has algebraic multiplicity  $\mu_1 > 1$  equal to its geometric multiplicity. Thus, in this case, Proposition 3.2 does not hold, and  $\gamma$  may not be small. Nevertheless Theorem 3.1 and our bound (4.2) hold. In addition, as we show below,  $\|P_Q r_m\|$  is small, implying that the second term in (4.2) is small, thus providing us with the same analytic model of superlinear convergence for GMRES. In this case we quantify the closeness of  $\mathcal{R}(Y) \subset AK_m(A, r_0)$  and  $\mathcal{R}(Q)$  by the fact that there is one harmonic Ritz value  $\theta_1$  close to  $\lambda_1$ .

**PROPOSITION 4.8.** *Let  $Q$  be an invariant subspace of  $A$  corresponding to  $\lambda_1$  of multiplicity  $\mu_1 > 1$ . Assume that  $A$  is diagonalizable. Let  $\varphi_m$  be the residual polynomial of GMRES after  $m$  steps, i.e.,  $r_m = \varphi_m(A)r_0$ . Then  $\|P_Q r_m\| = |\varphi_m(\lambda_1)| \|P_Q r_0\|$ .*

*Proof.* We write  $r_0 = P_Q r_0 + (I - P_Q)r_0$ . Thus, we have  $P_Q r_m = \varphi_m(\lambda_1)P_Q r_0$ .  $\square$

The harmonic Ritz values  $\theta_i$  are the roots of the residual polynomial  $\varphi_m$ , and thus if  $\theta_1 \approx \lambda_1$ , Proposition 4.8 indicates that  $\|P_Q r_m\| \approx 0$ . The more general case where the Jordan blocks are of order larger than one can be shown in a similar manner; we sketch this in the appendix. In this situation, as is well known, for a sufficiently general starting vector, Krylov subspaces can at best determine approximations to a basis for an invariant subspace associated with the largest Jordan block of a given multiple eigenvalue [29], [41].

We remark that due to the well-known relation between the norm of the FOM residuals and those of GMRES (see, e.g., [15], [30]), then our model of superlinear convergence of GMRES applies to FOM as well. Also, the minimal residual method (MINRES) is an implementation of GMRES for the case of  $A$  symmetric (see, e.g., [15], [30]), and therefore our analysis is likewise applicable to this method.

**5. Superlinear Convergence of the CG and Block Krylov Subspace Methods.**

The CG method for the solution of linear systems of the form  $Ax = f$  with  $A$  symmetric positive definite minimizes at each step the  $A$ -norm of the error, as in (2.1). Equivalently, since  $\|x_* - x\|_A = \|f - Ax\|_{A^{-1}}$  (see, e.g., [40]),  $x_m$  minimizes the  $A^{-1}$ -norm of the residual; i.e., the residual  $r_m = f - Ax_m$  satisfies

$$\|r_m\|_{A^{-1}} = \min_{d \in AK_m(A, r_0)} \|r_0 - d\|_{A^{-1}}.$$

We can thus apply our general Theorem 3.1 and Propositions 3.2 and 3.4 to this situation, by using the  $A^{-1}$ -inner product and choosing two matrices  $W_m, X_j$  such that  $\mathcal{R}(W_m) = AK_m(A, r_0)$  and  $\mathcal{R}(X_j) = AK_j(A, r_m)$ , as was done in the previous section. We thus obtain the superlinear convergence of the CG method, and one can write a theorem analogous to Theorem 4.1. This result is then consistent with that in [38], where the proof uses polynomials.

Block methods are used, e.g., when one needs to solve linear systems of the form  $Ax = f$  with multiple right-hand sides, say,  $f^{(1)}, \dots, f^{(p)}$ ; see, e.g., [27], [34]. Given the starting approximate solutions  $x_0^{(1)}, \dots, x_0^{(p)}$  and associated residual matrix  $R_0 = [r_0^{(1)}, \dots, r_0^{(p)}]$ , with  $r_0^{(i)} = f^{(i)} - Ax_0^{(i)}$ , block methods aim at computing approximate solutions in the subspace  $\mathbf{K}_m = \mathbf{K}_m(A, R_0) = \text{span}\{R_0, AR_0, \dots, A^{m-1}R_0\}$ . Note that  $\mathbf{K}_m$  has dimension at most  $m \cdot p$ , and therefore for each system  $Ax^{(i)} = f^{(i)}$  (with solution  $x_*^{(i)}$ ) a better approximation is expected with the block method, compared with approximating  $x^{(i)}$  in the subspace  $K_m(A, r_0^{(i)})$ , of dimension at most  $m$ . We note that an advantage of these block methods is that the subspaces generated by

them can approximate complete invariant subspaces of multiple eigenvalues (with several Jordan blocks).

In the case of block CG, for  $A$  symmetric positive definite, at the  $m$ th step one finds the solution to the minimization problem

$$\min_{x^{(i)} \in \mathbf{K}_m, i=1, \dots, p} \sum_{i=1}^p \|x_\star^{(i)} - x^{(i)}\|_A^2;$$

see, e.g., [15, sect. 7.4]. This is equivalent to finding the solution to

$$(5.1) \quad \min_{d^{(i)} \in \mathbf{AK}_m, i=1, \dots, p} \sum_{i=1}^p \|r_0^{(i)} - d^{(i)}\|_{A^{-1}}^2.$$

This problem can be analyzed in the context of the results of section 3 if we consider as our space  $\mathbb{R}^{n \times p}$  equipped with the inner product

$$(5.2) \quad \langle V, W \rangle = \sum_{i=1}^p \langle v^{(i)}, w^{(i)} \rangle_{A^{-1}},$$

where  $V = [v^{(1)}, \dots, v^{(p)}]$  and  $W = [w^{(1)}, \dots, w^{(p)}]$ . Thus, if  $D = [d^{(1)}, \dots, d^{(p)}]$ , the problem (5.1) can be rewritten as

$$\min_{\mathcal{R}(D) \subset \mathbf{AK}_m} \|R_0 - D\|^2,$$

where the norm is the one induced by the inner product (5.2). We can readily apply Theorem 3.1 and the other results on superlinear convergence by choosing  $W_m$  with  $\mathcal{R}(W_m) = \mathbf{AK}_m$  and  $X_j$  with  $\mathcal{R}(X_j) = \mathbf{AK}_j(A, R_m)$ , where  $R_m = [r_m^{(1)}, \dots, r_m^{(p)}]$ , in a manner analogous to the last two sections.

We demonstrate in a similar manner the superlinear convergence of block GMRES. In this case, one solves the problem

$$\min_{d^{(i)} \in \mathbf{AK}_m, i=1, \dots, p} \sum_{i=1}^p \|r_0^{(i)} - d^{(i)}\|_2^2 = \min_{\mathcal{R}(D) \subset \mathbf{AK}_m} \|R_0 - D\|^2,$$

where the norm in the last expression is the Frobenius norm and is induced by the following inner product in  $\mathbb{R}^{n \times p}$ :  $\langle V, W \rangle = \sum_{i=1}^p \langle v^{(i)}, w^{(i)} \rangle$ , the latter being the Euclidean inner product in  $\mathbb{R}^n$ ; see, e.g., [30, sect. 6.12].

**6. Superlinear Convergence of Inexact Krylov Subspace Methods.** Inexact Krylov subspace methods are those where the matrix-vector multiplication (1.3) is not performed exactly. Instead one has  $(A + E_k)v$ , where  $E_k$  is the error matrix at the  $k$ th step, which may be different at every step, but its norm can be monitored. The (exact) Arnoldi method produces at the  $m$ th step a matrix  $V_m$  such that its orthonormal columns span  $\mathcal{K}_m(A, r_0)$  [15], [30] and satisfy the Arnoldi relation

$$AV_m = V_{m+1}H_m,$$

where  $H_m$  is upper Hessenberg. In inexact Krylov subspace methods the Arnoldi relation does not hold, but instead one has for each  $m$

$$(6.1) \quad (A + \mathcal{E}_m)V_m = V_{m+1}H_m, \quad \mathcal{E}_m = \sum_{k=1}^m E_k v_k v_k^T,$$

where  $V_m = [v_1, \dots, v_m]$  has orthonormal columns. The subspace where the approximation lies is

$$(6.2) \quad \mathcal{R}(V_m) = \mathcal{K}_m(A + \mathcal{E}_m, r_0);$$

see [35] for a full description and further references. Note that in (6.2) we have a different Krylov subspace at each step.

The property that makes these methods attractive is that  $\|E_k\|$  can be allowed to grow once  $\|r_k\|$  starts to diminish, and, in fact,  $\|E_k\|$  can be very large after certain steps and still achieve convergence to the solution of  $Ax = f$ . We refer the reader to [35], where criteria are given to successively relax the size of  $\|E_k\|$  while maintaining the convergence of the method. In the same reference applications can be found where allowing the error to grow decreases the cost of the matrix-vector product.

We now consider, for the application of Theorem 3.1 and Proposition 3.4, matrices  $W_{m+j} \in \mathbb{R}^{n \times (m+j)}$ ,  $X_j \in \mathbb{R}^{n \times j}$ ,  $Y \in \mathbb{R}^{n \times k}$ , and  $Q \in \mathbb{R}^{n \times k}$  with the following properties:

$$\begin{aligned} \mathcal{R}(W_{m+j}) &= (A + \mathcal{E}_{m+j})\mathcal{K}_{m+j}(A + \mathcal{E}_{m+j}, r_0) = \mathcal{K}_{m+j}(A + \mathcal{E}_{m+j}, (A + \mathcal{E}_{m+j})r_0), \\ \mathcal{R}(X_j) &= (A + \mathcal{E}_{m+j})\mathcal{K}_j(A + \mathcal{E}_{m+j}, r_m), \\ \mathcal{R}(Y) &\subset (A + \mathcal{E}_m)\mathcal{K}_m(A + \mathcal{E}_m, r_0), \end{aligned}$$

and  $Q$  such that its columns span a simple invariant subspace of  $A + \mathcal{E}_m$ .

LEMMA 6.1. *With this notation, we have that  $(A + \mathcal{E}_m)\mathcal{K}_m(A + \mathcal{E}_m, r_0) \subset \mathcal{R}(W_{m+j})$ .*

*Proof.* The inclusion follows from (6.1), i.e., that

$$\mathcal{E}_{m+j} = \mathcal{E}_m + \sum_{k=m+1}^{m+j} E_k v_k v_k^T,$$

implying that  $\mathcal{E}_{m+j}V_m = \mathcal{E}_mV_m$  and that a basis of  $\mathcal{R}(W_{m+j})$  is  $(A + \mathcal{E}_{m+j})V_{m+j} = (A + \mathcal{E}_{m+j})[V_m, \hat{V}_j]$  for an appropriate matrix  $\hat{V}_j$  with orthonormal columns.  $\square$

As a consequence we have that  $\mathcal{R}(Y) \subset \mathcal{R}(W_{m+j})$ , and since

$$r_m \in \mathcal{K}_{m+1}(A + \mathcal{E}_{m+1}, r_0),$$

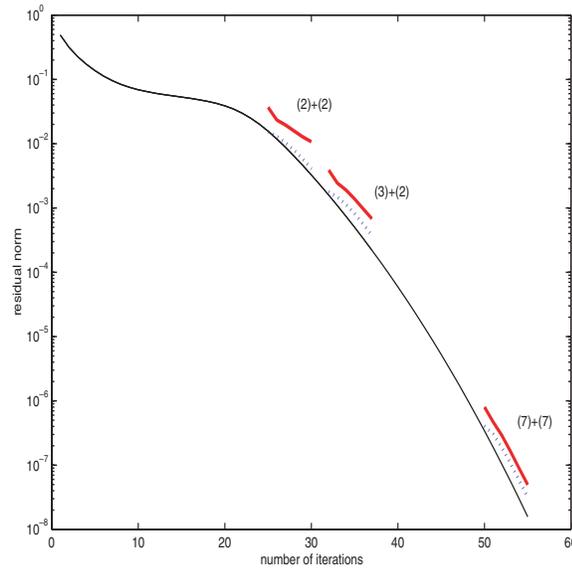
using a similar argument, we also have that  $\mathcal{R}(X_j) \subset \mathcal{R}(W_{m+j})$ . We can now interpret Theorem 3.1 and Proposition 3.4 to say that

$$(6.3) \quad \begin{aligned} \|r_{m+j}\| &= \min_{d \in (A + \mathcal{E}_{m+j})\mathcal{K}_{m+j}(A + \mathcal{E}_{m+j}, r_0)} \|r_0 - d\| \\ &\leq \min_{d \in (A + \mathcal{E}_{m+j})\mathcal{K}_j(A + \mathcal{E}_{m+j}, r_m)} \{ \|(I - P_Q)(r_m - d)\| + \gamma \|P_Q(r_m - d)\| \} \end{aligned}$$

$$(6.4) \quad \leq \sqrt{2} \min_{d \in (A + \mathcal{E}_{m+j})\mathcal{K}_j(A + \mathcal{E}_{m+j}, r_m)} \left\| \begin{bmatrix} I - P_Q \\ \gamma P_Q \end{bmatrix} (r_m - d) \right\|,$$

where  $\gamma = \|(I - \Pi_Y)P_Q\|$ . Proposition 3.4 and Corollary 3.6 are also valid in this context.

We observe that our analysis in the inexact case involves the spectral projector associated with an invariant subspace of  $A + \mathcal{E}_m$  and not of  $A$ . In particular, the projections of  $r_m$  onto the invariant subspaces of  $A + \mathcal{E}_m$  may differ substantially from those in the exact case. The consequences of this fact are twofold.



**Fig. 6.1** Example 6.2. Convergence history of GMRES.

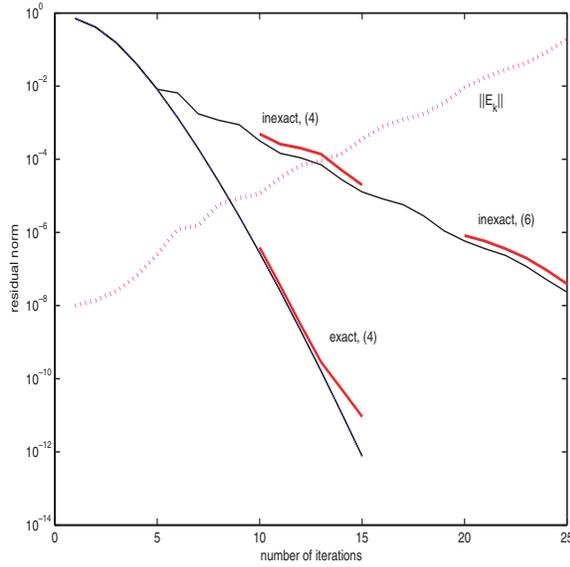
First, a different number  $m$  of iterations may be required to obtain a sufficiently good approximation to  $\mathcal{R}(Q)$  and thus a small  $\gamma$ . Such a difference depends on how sensitive the corresponding invariant subspace of  $A$  is to perturbations, and thus the problem may be harmless at an early stage of the process, when the perturbations are indeed small.

Second, even when the space spanned by the columns of  $Q$  is not too sensitive to perturbations and thus is (almost) deflated as in the exact case, subsequent iterations are affected by the composition of  $(I - P_Q)r_m$ . We provide experimental evidence of this phenomenon in Example 6.3.

In practice, one can experience convergence delay (as compared to the exact case) either because the relevant invariant subspaces of  $A$  are very sensitive to perturbations or because the computed residual has more significant eigencomponents than the residual in the exact case. The worst-case scenario is thus lack of superlinear convergence in the inexact case, though maintaining a linear rate of convergence.

If we use the 2-norm in (6.3)–(6.4), we have an inexact GMRES method. We first illustrate our bounds with two examples for the inexact GMRES method and later discuss inexact conjugate gradients.

*Example 6.2.* Consider the bidiagonal matrix  $A$  with ones on the upper diagonal and  $0.1, 2, \dots, 100$  on the diagonal. The right-hand side is the normalized vector of all ones. We run the inexact GMRES method using as the perturbation a random matrix so that  $\|E_i\| = \varepsilon/\|r_i\|$ , where  $r_i$  is the computed residual at iteration  $i$ , and  $\varepsilon = 10^{-8}$ . In Figure 6.1 we show the plot of the convergence curve as well as those of our bounds corresponding to three different stages of the procedure, for  $m = 25, 32, 50$ . The number of iterations to evaluate the bound was  $j = 1, \dots, 5$  in all cases. We point out that this plot is identical to that obtained with exact GMRES. This fact illustrates the potential for savings for the inexact methods. The fact that our bounds apply to the inexact method can also be appreciated. We mention that the



**Fig. 6.2** Example 6.3. Convergence history of exact and inexact GMRES.

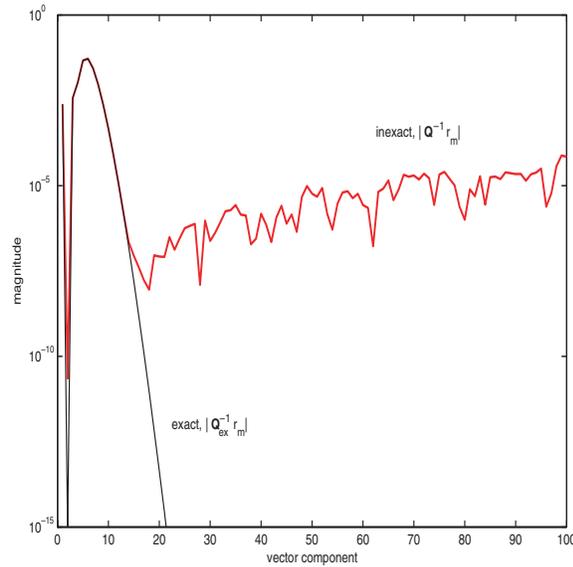
quantities in the inexact case are very close to those of the exact case, even though when  $m = 50$ ,  $\|E_m\| \approx 9 \cdot 10^{-3}$ , and when  $m = 55$ ,  $\|E_m\| \approx 3 \cdot 10^{-1}$ .

The residual norms reported in this section are the computed residuals, i.e.,  $r_0 - V_{m+1}H_m y_m$ , where  $x_m = V_m y_m$ . The difference between these and the true residuals  $r_0 - Ax_m$  is not detected by the eye in the plot for this example. In [35] and [37] bounds for this difference are provided.

The following example is taken from [37], where it was used to show experimentally that the inexact method can indeed delay convergence. We use this example to illustrate that even in these circumstances our results, and in particular our bounds of the form (6.4), hold as well.

*Example 6.3.* Let  $A$  be a bidiagonal matrix with diagonal  $[1, 2, \dots, 100]$  and lower diagonal of all ones. The right-hand side is  $f = e_1$ , the first vector of the standard basis, which has very large projection onto the eigenvector corresponding to the smallest eigenvalue.

The convergence behaviors of both the exact and inexact solvers are depicted in Figure 6.2 for  $m = 10$  (exact case) and  $m = 10, 20$  (inexact case),  $j = 1, \dots, 5$ . In the inexact case, the perturbation is a random matrix so that  $\|E_i\| = \varepsilon/\|r_i\|$ , where  $r_i$  is the computed residual at iteration  $i$  and  $\varepsilon = 10^{-8}$ ; the values of these norms are shown with a dotted line in magenta. Note that after  $m = 10$  iterations, we observe  $\|E_m\| \approx 8 \cdot 10^{-6}$ , while for  $m = 20$ ,  $\|E_m\| \approx 4 \cdot 10^{-3}$ . The smallest  $k$  eigenvalues were considered to build  $Q$ . In the exact case,  $k = 4$ , which yields  $\gamma = 7.36 \cdot 10^{-4}$  after  $m = 10$  iterations, while in the inexact case,  $k = 4$  for  $m = 10$  ( $\gamma = 0.14$ ) and  $k = 6$  for  $m = 20$  ( $\gamma = 4.68 \cdot 10^{-2}$ ); see Figure 6.2. Only the bound (6.4) is shown in the plot. The experiment fully confirms our theoretical analysis, showing complete agreement between the inexact GMRES curve and our bound. The very modest projection norm,  $\|P_Q\| \approx 1.6$  for all considered cases, is a key ingredient to assess the sharpness of the bound.



**Fig. 6.3** Example 6.3. Projected residual components,  $m = 4$ .

We linger over this example to emphasize that the quantities we use in our analysis also provide a better understanding of the delay in convergence in the inexact case. We first observe in Figure 6.2 that the exact and inexact curves split at  $m = 5$ . We therefore discuss the situation after  $m = 4$  iterations. Let  $Q_{ex}$  and  $Q$  be, respectively, the exact and inexact eigenvector matrices corresponding to the two smallest eigenvalues (i.e.,  $k = k_1 = 2$ ); then we have  $\|P_{Q_{ex}} - P_Q\| \approx 5 \cdot 10^{-9}$ . However, the spectral projection of the residual onto the eigenvector space provides a quite different picture in the exact and inexact cases. Let  $\mathbf{Q}_{ex}$  and  $\mathbf{Q}$  denote the eigenvector matrices of  $A$  and of  $A + \mathcal{E}_m$ , respectively. The columns of  $\mathbf{Q}_{ex}$  and  $\mathbf{Q}$  are ordered according to the absolute value of the associated eigenvalues, in increasing order. In Figure 6.3 we display the absolute values of the components of  $\mathbf{Q}_{ex}^{-1} r_m$  (in black) and  $\mathbf{Q}^{-1} r_m$  (in red), where  $r_m$  is the exact GMRES residual or the inexact GMRES computed residual, respectively. Clearly, the plot shows that in the inexact case, the residual is polluted by several eigenvector components associated with the largest eigenvalues. On the other hand, in the exact case, the residual at iteration  $m = 4$  has very few nonzero eigencomponents, all corresponding to the small eigenvalues (this is because of the special right-hand side  $f$ ). Our theory enables us to interpret this situation as follows: after  $m$  iterations, our bound predicts that inexact GMRES should almost behave as an inexact GMRES iteration on  $(I - P_Q)r_m$ . This example shows that in the inexact process,  $(I - P_Q)r_m$  may have larger components than in the exact process. Therefore, although  $k$  components are indeed deflated, inexactness causes more components to grow compared with the exact case, causing delay in convergence. This experiment also confirms that our analytic model takes into account spectral information contained in  $r_m$ , as discussed in Remark 4.3.

We conclude by discussing the inexact CG method, which is used when  $A$  is symmetric and positive definite; see [14], [26]. Our bounds (6.3)–(6.4) apply to this case when the  $A^{-1}$ -norm is used. We mention that in our setting, the three-term recurrence usually associated with implementations of CG is lost.

**7. Concluding Remarks.** We have presented a general analytic model which describes the superlinear convergence of exact and inexact Krylov subspace methods in a unified manner. Our model is based on the proximity of certain invariant subspaces  $\mathcal{R}(Q)$  to spaces generated by the Krylov subspace methods and is consistent with other models of superlinear convergence which apply to specific (exact) methods.

Deflation and augmentation strategies using the columns of  $Q$  that correspond to eigenvalues closest to zero have been proposed, e.g., in [11], [24], [31]. Our analysis confirms the effectiveness of such approaches, and in addition it suggests that other spectral regions could be considered in the augmented procedure. The examples we present here support this idea.

**Appendix.** In this appendix we present some auxiliary results.

*Proof of Proposition 3.4.* Let  $y_b = B^\dagger b$  be the minimizing element of (3.6), where  $B^\dagger$  is the pseudoinverse of  $B$ . Then

$$\begin{aligned} \min_{y \in \mathbb{R}^j} \|(b + \Delta b) - (B + \Delta B)y\|_* &\leq \|b - By_b\|_* + \|\Delta b - \Delta B y_b\|_* \\ &= \min_{d \in \mathcal{R}[(I - P_Q)X_j]} \|(I - P_Q)r_m - d\| + \|\Delta r\|_*, \end{aligned}$$

where  $\Delta r = \Delta b - \Delta B y_b$ , and thus  $\|\Delta r\|_* = \gamma \|P_Q(r_m - X_j y_b)\|$ . Therefore

$$\begin{aligned} \|\Delta r\|_* &\leq \gamma \|(I - P_Q)(r_m - X_j y_b)\| + \gamma \|r_m - X_j y_b\|, \text{ i.e.,} \\ \|\Delta r\|_* &\leq \gamma \min_{d \in \mathcal{R}[(I - P_Q)X_j]} \|(I - P_Q)r_m - d\| + \gamma \|r_m - X_j y_b\|. \end{aligned}$$

We use the explicit representation of  $y_b$  and obtain

$$(A.1) \quad \|r_m - X_j y_b\| = \|(I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q))r_m\|.$$

We observe now that  $(I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q))\Pi_{X_j} = 0$ . Indeed, since  $\Pi_{X_j} = X_j(X_j^*X_j)^{-1}X_j^*$ , then

$$(A.2) \quad \begin{aligned} &(I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q))X_j \\ &= (I - X_j[X_j^*(I - P_Q)^*(I - P_Q)X_j]^{-1}X_j^*(I - P_Q)^*(I - P_Q))X_j = 0. \end{aligned}$$

From (A.1) we obtain

$$\begin{aligned} \|r_m - X_j y_b\| &= \|(I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q))(I - \Pi_{X_j})r_m\| \\ &\leq \|(I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q))\| \cdot \|(I - \Pi_{X_j})r_m\|. \end{aligned}$$

Since  $X_j[(I - P_Q)X_j]^\dagger(I - P_Q)$  is a projection, we have

$$\|I - X_j[(I - P_Q)X_j]^\dagger(I - P_Q)\| = \|X_j[(I - P_Q)X_j]^\dagger(I - P_Q)\|;$$

see, e.g., [17], [22]. To complete the proof all that remains to be shown is that  $\tau_j^{-1} = \|X_j[(I - P_Q)X_j]^\dagger(I - P_Q)\|$ . Let  $\hat{Z} = (I - P_Q)^*(I - P_Q)X_j$ . Writing explicitly the pseudoinverse as in (A.2), we obtain  $X_j[(I - P_Q)X_j]^\dagger(I - P_Q) = X_j(\hat{Z}^*X_j)^{-1}\hat{Z}^*$ . As  $\hat{Z} = ZR$  is the reduced QR decomposition of  $\hat{Z}$ ,  $X_j(\hat{Z}^*X_j)^{-1}\hat{Z}^* = X_j(Z^*X_j)^{-1}Z^*$  and the result follows.  $\square$

**LEMMA A.1.** Let  $X \in \mathbb{R}^{n \times j}$  be of full column rank (with  $j \leq n$ ) and  $P_Q$  be any  $n \times n$  matrix. Let  $\gamma \in \mathbb{R}$ ,  $\gamma \neq 0$ . Then,

$$M = \begin{bmatrix} (I - P_Q)X \\ \gamma P_Q X \end{bmatrix}$$

is also of full column rank  $j$ .

*Proof.* We show that the columns of  $M$  are linearly independent. Let  $My = 0$ ; then  $(I - P_Q)Xy = 0$  and  $\gamma P_Q Xy = 0$ , which implies that  $P_Q Xy = 0$ . Then  $Xy = (I - P_Q)Xy + P_Q Xy = 0$ , implying that  $y = 0$ .  $\square$

LEMMA A.2. *Let  $Q \in \mathbb{R}^{n \times k}$  and  $X \in \mathbb{R}^{n \times j}$  be of full column rank, with  $j \leq n - k$ . Then  $(I - P_Q)X$  also has full column rank  $j$  if and only if  $\mathcal{R}(Q) \cap \mathcal{R}(X) = \{0\}$ .*

*Proof.* We show that the columns of  $(I - P_Q)X$  are linearly independent if and only if the two subspaces only intersect at the zero vector. Let  $X = [w_1, \dots, w_j]$ ; then

$$\begin{aligned} & (I - P_Q)Xy \neq 0 \quad \text{for all nonzero } y \in \mathbb{R}^j \\ \Leftrightarrow & Xy \notin \mathcal{N}(I - P_Q) = \mathcal{R}(Q) \quad \text{for all nonzero } y \in \mathbb{R}^j \\ \Leftrightarrow & w_i \notin \mathcal{R}(Q), \quad i = 1, \dots, j, \\ \Leftrightarrow & \mathcal{R}(X) \cap \mathcal{R}(Q) = \{0\}. \quad \square \end{aligned}$$

In the following we show the counterpart to Proposition 4.8 when the Jordan blocks corresponding to  $\lambda_1$  are of order larger than 1.

Let  $A = \mathbf{Q}J\mathbf{Q}^{-1}$  be the Jordan canonical form, and without loss of generality assume that  $\|\mathbf{Q}\| = 1$ . Let  $J_1, \dots, J_j$  be the Jordan blocks of  $A$  of size  $\ell_1, \dots, \ell_j$  associated with  $\lambda_1$  and denote by  $Q = [Q_1, \dots, Q_j]$  the associated portion of  $\mathbf{Q}$  with analogous partitioning. Similarly, let  $\mathbf{Q}^{-1}r_0$  be partitioned conformally, and call the corresponding vectors  $\hat{r}_0^{(1)}, \dots, \hat{r}_0^{(j)}$ . Without loss of generality, we can assume that  $\ell_1, \dots, \ell_j$  are in ascending order. Then we have

$$P_Q r_m = P_Q \varphi_m(A)r_0 = Q_1 \varphi_m(J_1) \hat{r}_0^{(1)} + \dots + Q_j \varphi_m(J_j) \hat{r}_0^{(j)}.$$

We observe that for any  $i = 2, \dots, j$ ,  $\varphi_m(J_{i-1})$  is a submatrix of  $\varphi_m(J_i)$ , and therefore  $\|\varphi_m(J_{i-1})\| \leq \|\varphi_m(J_i)\|$ . Hence, using the fact that  $\|Q_i\| \leq 1$ ,

$$\|P_Q r_m\| \leq \|\varphi_m(J_j)\| \left( \|\hat{r}_0^{(1)}\| + \dots + \|\hat{r}_0^{(j)}\| \right).$$

To complete the result all we need to show is that  $\|\varphi_m(J_j)\| \approx 0$ . Let  $w$  be a generalized eigenvector of  $\lambda_1$  of maximal grade. Assume that  $r_0$  is general enough, i.e., that  $P_Q r_0$  has some nonzero component in the direction of  $w$ . Our assertion then follows from the fact that for each copy of  $\lambda_1$  there is a harmonic Ritz value  $\theta_i \approx \lambda_1$ , so that  $\varphi_m(\lambda_1) \approx 0, \varphi'_m(\lambda_1) \approx 0, \dots, \varphi_m^{(j-1)}(\lambda_1) \approx 0$ .

**Acknowledgments.** We benefitted from comments and suggestions from several colleagues, including Michele Benzi, Eric de Sturler, Mark Embree, Jörg Liesen, and Michael Parks, and from two anonymous referees.

#### REFERENCES

- [1] N. I. AKHIEZER AND I. M. GLAZMAN, *Theory of Linear Operators in Hilbert Space*, F. Ungar Publishing, New York, 1961, 1963. Reprinted by Dover, New York, 1993.
- [2] O. AXELSSON AND G. LINDSOK, *On the rate of convergence of the preconditioned conjugate gradient method*, Numer. Math., 48 (1986), pp. 499–523.
- [3] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE, AND H. VAN DER VORST, EDs., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.
- [4] C. BEATTIE, M. EMBREE, AND J. ROSSI, *Convergence of restarted Krylov subspaces to invariant subspaces*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 1074–1109.
- [5] C. A. BEATTIE, M. EMBREE, AND D. C. SORENSEN, *Convergence of polynomial restart Krylov methods for eigenvalue computation*, SIAM Rev., 47 (2005), to appear.
- [6] B. BECKERMANN AND A. B. J. KUIJLAARS, *Superlinear CG convergence for special right-hand sides*, Electron. Trans. Numer. Anal., 14 (2002), pp. 1–19.

- [7] B. BECKERMANN AND A. B. J. KUIJLAARS, *Superlinear convergence of conjugate gradients*, SIAM J. Numer. Anal., 39 (2001), pp. 300–329.
- [8] A. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
- [9] S. L. CAMPBELL, I. C. IPSEN, C. T. KELLEY, AND C. D. MEYER, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–675.
- [10] Z.-H. CAO, *A note on the convergence behavior of GMRES*, Appl. Numer. Math., 25 (1997), pp. 13–20.
- [11] E. DE STURLER, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Anal., 36 (1999), pp. 864–889.
- [12] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Performance and analysis of saddle point preconditioners for the discrete steady-state Navier–Stokes equations*, Numer. Math., 90 (2002), pp. 665–688.
- [13] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
- [14] G. H. GOLUB AND Q. YE, *Inexact preconditioned conjugate gradient method with inner-outer iteration*, SIAM J. Sci. Comput., 21 (1999), pp. 1305–1320.
- [15] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
- [16] A. GREENBAUM, V. PTÁK, AND Z. STRAKOŠ, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.
- [17] T. KATO, *Estimation of iterated matrices, with application to the von Neumann condition*, Numer. Math., 2 (1960), pp. 22–29.
- [18] T. KATO, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, New York, 1980.
- [19] C. T. KELLEY AND Z. Q. XUE, *GMRES and integral operators*, SIAM J. Sci. Comput., 17 (1996), pp. 217–226.
- [20] A. KLAWONN, *Block-triangular preconditioners for saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 172–184.
- [21] M. G. KREIN AND M. A. KRASNOSELSKII, *Fundamental theorems concerning the extension of Hermitian operators and some of their applications to the theory of orthogonal polynomials and the moment problem*, Uspekhi Mat. Nauk., 2 (1947), pp. 60–106 (in Russian).
- [22] V. E. LANCIE, *Some properties of idempotent operators*, Teoret. i Prikl. Mat., 1 (1958), pp. 16–22 (in Russian).
- [23] I. MORET, *A note on the superlinear convergence of GMRES*, SIAM J. Numer. Anal., 34 (1997), pp. 513–516.
- [24] R. B. MORGAN, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.
- [25] O. NEVANLINNA, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, Boston, Berlin, 1993.
- [26] Y. NOTAY, *Flexible conjugate gradients*, SIAM J. Sci. Comput., 22 (2000), pp. 1444–1460.
- [27] D. P. O’LEARY, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.
- [28] M. L. PARKS, *The Iterative Solution of a Sequence of Linear Systems Arising from Nonlinear Finite Element Analysis*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2005.
- [29] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, Halstead Press, New York, 1992.
- [30] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
- [31] Y. SAAD, *Analysis of augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 435–449.
- [32] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [33] V. SIMONCINI, *Block triangular preconditioners for symmetric saddle-point problems*, Appl. Numer. Math., 49 (2004), pp. 63–80.
- [34] V. SIMONCINI AND E. GALLOPOULOS, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.
- [35] V. SIMONCINI AND D. B. SZYLD, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.
- [36] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, San Diego, London, 1990.
- [37] J. VAN DEN ESHOF AND G. L. G. SLEIJPEN, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153.
- [38] A. VAN DER SLUIS AND H. A. VAN DER VORST, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–586.

- [39] H. A. VAN DER VORST AND C. VUIK, *The superlinear convergence behaviour of GMRES*, J. Comput. Appl. Math., 48 (1993), pp. 327–341.
- [40] R. WEISS, *Parameter-Free Iterative Linear Solvers*, Akademie-Verlag, Berlin, 1996.
- [41] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, UK, 1965.
- [42] R. WINTHER, *Some superlinear convergence results for the conjugate gradient method*, SIAM J. Numer. Anal., 17 (1980), pp. 14–17.