Convergence analysis of Krylov subspace methods†

Jörg Liesen[*1] and **Petr Tichý**[1]

[1] Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany

Received ..., revised ..., accepted ...
Published online ...

Key words Krylov subspace methods, convergence analysis.
MSC (2000) 15A06, 65F10, 41A10

One of the most powerful tools for solving large and sparse systems of linear algebraic equations is a class of iterative methods called Krylov subspace methods. Their significant advantages like low memory requirements and good approximation properties make them very popular, and they are widely used in applications throughout science and engineering. The use of the Krylov subspaces in iterative methods for linear systems is even counted among the "Top 10" algorithmic ideas of the 20th century. Convergence analysis of these methods is not only of a great theoretical importance but it can also help to answer practically relevant questions about improving the performance of these methods. As we show, the question about the convergence behavior leads to complicated nonlinear problems. Despite intense research efforts, these problems are not well understood in some cases. The goal of this survey is to summarize known convergence results for three well-known Krylov subspace methods (CG, MINRES and GMRES) and to formulate open questions in this area.

1 Introduction

Krylov subspace methods represent one of the most important classes of iterative methods for solving linear algebraic systems. Their main common ingredient are the Krylov subspaces, which are spanned by the initial residual and by vectors formed by repeated multiplication of the initial residual by the system matrix. These subspaces first appeared in a paper by the Russian scientist and navy general Aleksei Nikolaevich Krylov (1863–1945), published in 1931 [44]. Motivated by an application in naval science, Krylov was interested in analyzing oscillations of mechanical systems, and proposed a method for computing the minimal polynomial of a given matrix (see, e.g., [21, Section 42], [25, Chapter VII], or [38, Chapter 6] for detailed accounts of Krylov's method). Independently of Krylov's work, the first Krylov subspace methods for solving linear algebraic systems appeared two decades later with the publication of the conjugate gradient (CG) method for hermitian positive definite matrices by Hestenes and Stiefel [36], and the closely related methods developed by Lanczos [45, 46]. Driven by the need to solve linear systems of vastly increasing dimension and the accompanying rapid development of computational resources, these Krylov subspace methods were

* Corresponding author: e-mail: liesen@math.tu-berlin.de, Phone: +49 30 314 29295, Fax: +49 30 314 79706
† This work was supported by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft.

used in many applications, particularly in the engineering community. In the numerical linear algebra community, the potential of Krylov subspace methods was fully recognized only after an influential paper of Reid appeared in 1971 [58]. Subsequently, numerous additional Krylov subspace methods were developed, with focus on indefinite and nonhermitian matrices. Today, the use of the Krylov subspaces in iterative methods for linear systems is counted among the "Top 10" algorithmic ideas of the 20th century [10]. One of the main reasons for this success is that the Krylov subspaces can be build up using only a function that computes the multiplication of the system matrix and a vector, so that the system matrix itself never has to be formed or stored explicitly. Hence Krylov subspace methods are particularly well suited for application to large and sparse linear systems, which today are commonplace throughout applications in science and engineering.

Mathematically, Krylov subspace methods are based on projection methods. Instead of solving the potentially very large linear system, the idea is to approximate the systems' solution from Krylov subspaces of small dimension. The goal of the convergence analysis of these methods is to *describe the convergence of this process in terms of input data of the given problem*, i.e. in dependence on properties of the system matrix, the right hand side vector and the initial guess. Understanding the convergence of Krylov subspace methods is particularly important to answer the practically relevant questions how to accelerate the convergence (in particular how to precondition the system), and how to choose potential restart parameters.

The goal of this paper is to survey the known theory of convergence of Krylov subspace methods that are based on two basic types of projection methods, namely the Galerkin (orthogonal residual (OR)) method and the minimal residual (MR) method. Both types of methods have been implemented in various commonly used algorithms. An example of the OR Krylov subspace method is the CG method [36] for hermitian positive definite matrices. Implementations of the MR Krylov subspace method are the MINRES method [56] for nonsingular hermitian indefinite matrices and the GMRES method [62] for general nonsingular matrices. The distinction between OR and MR methods made in this paper is not new. In fact it has been extensively used in the past to derive relations between the convergence quantities (e.g. error or residual norms) of different methods, see, e.g., [12, 14, 37]. Here our focus is on giving bounds for the convergence quantities of each method separately.

For normal system matrices A, the (worst-case) convergence behavior of CG, MINRES and GMRES is completely determined by the spectrum of A. The convergence analysis then reduces to analyzing a certain min-max approximation problem on the matrix eigenvalues. In the nonnormal case, however, the convergence behavior of the GMRES method may not be related to the eigenvalues at all. As a consequence, other properties of the input data must be considered to describe the convergence. Despite intense efforts to identify descriptive properties, understanding the convergence of GMRES in the general nonnormal case still remains a largely open problem.

After a brief introduction to the mathematical background of Krylov subspace methods (Section 2), we survey in Section 3 the theory of convergence of these methods. We distinguish between the normal (Section 3.1) and the nonnormal (Section 3.2) case. Section 4 contains concluding remarks. We point out that all convergence results we state in this paper were derived assuming exact arithmetic. A recent survey of the numerical stability of Krylov subspace methods that also discusses effects of finite precision arithmetic on the convergence is given in [65].

2 Krylov subspace methods

In this section we briefly describe the mathematical background of the Krylov subspace methods for solving linear algebraic systems of the form

$$Ax \;=\; b\,, \tag{1}$$

where A is a real or complex nonsingular N by N matrix, and b is a real or complex vector of length N. Suppose that x_0 is an initial guess for the solution x, and define the initial residual $r_0 = b - Ax_0$. As shown originally by Saad [59, 60] (see his book [61] for a summary), Krylov subspace methods can be derived from the following *projection method*: The nth iterate x_n, $n = 1, 2, \ldots$, is of the form

$$x_n \;\in\; x_0 + \mathcal{S}_n\,, \tag{2}$$

where $\mathcal{S}_n$ is some n-dimensional space, called the search space. Because of the n degrees of freedom, n constraints are required to make x_n unique. This is done by choosing an n-dimensional space $\mathcal{C}_n$, called the constraints space, and by requiring that the nth residual is orthogonal to that space, i.e.,

$$r_n \;=\; b - Ax_n \;\in\; r_0 + A\mathcal{S}_n\,, \qquad r_n \;\perp\; \mathcal{C}_n\,. \tag{3}$$

Orthogonality here is meant in the Euclidean inner product. A similar type of projection process appears in many areas of mathematics. As an example, consider the Petrov-Galerkin framework in the context of the finite element method for discretizing partial differential equations, see e.g. [57, Chapter 5]. There the notions of test and trial spaces correspond to search and constraints spaces in (2)–(3).

In this paper we concentrate on the projection method (2)–(3) and two basic relations between $\mathcal{S}_n$ and $\mathcal{C}_n$, that to our mind are among the most important ones:

$$\mathcal{C}_n = \mathcal{S}_n \qquad \text{(Galerkin method)}\,, \tag{4}$$

$$\mathcal{C}_n = A\mathcal{S}_n \qquad \text{(Minimal residual method)}\,. \tag{5}$$

The Galerkin and the minimal residual (MR) method are called a Krylov subspace method when the so-called Krylov subspaces $\mathcal{K}_n(A, r_0)$ are used as search spaces, i.e.

$$\mathcal{S}_n \;=\; \mathcal{K}_n(A, r_0) \;\equiv\; \text{span}\{r_0, Ar_0, \ldots, A^{n-1}r_0\}\,, \quad n = 1, 2, \ldots\,. \tag{6}$$

Using these spaces in the Galerkin method, we construct residuals $r_n = b - Ax_n$ that are orthogonal to all previous residuals $r_{n-1}, \ldots, r_0$. That is why, in the context of Krylov subspaces, the Galerkin method is often called orthogonal residual (OR) method.

There are many possible choices of Krylov subspaces and their variants (e.g. $A\mathcal{K}_n(A, r_0)$, $\mathcal{K}_n(A^H, r_0)$, $A^H\mathcal{K}_n(A^H, r_0)$, etc.) in the projection process (2)–(3). This fact certainly contributes to the overabundant supply of these methods. Also note that for each mathematical description there may be several different implementations that in exact arithmetic satisfy (2)–(3) for given search and constraint spaces, but that may differ in their finite precision behavior. Particularly comprehensive and systematic surveys of existing Krylov subspace methods can be found in [4, 9] and [14].

The Krylov subspaces form a nested sequence that ends with a subspace of maximal dimension $d = \dim \mathcal{K}_N(A, r_0)$, i.e.,

$$\mathcal{K}_1(A, r_0) \subset \cdots \subset \mathcal{K}_d(A, r_0) = \cdots = \mathcal{K}_N(A, r_0)\,.$$

The number of steps of the OR/MR Krylov subspace method is limited by the maximal Krylov subspace dimension d. We say that a projection process *breaks down* in step n if no iterate x_n exists, or if x_n is not unique. Naturally, we are interested in projection methods that ensure existence and uniqueness of their iterates x_n for each step $n \leq d$. Such *well-defined* methods terminate with the exact solution in the step d, which is called the *finite termination property*. If a method is well-defined or not, depends on the properties of the matrix A.

In general, the OR Krylov subspace method yields uniquely defined iterates for each n whenever zero is outside the *field of values* of A, which is defined as

$$\mathcal{F}(A) = \left\{ v^H A v \;:\; \|v\| = 1\,,\; v \in \mathbb{C}^N \right\}\,. \tag{7}$$

However, in this paper we limit our discussion to the OR Krylov subspace method for hermitian positive definite matrices, since only in this case the given system matrix defines a norm in which the errors are minimized (see Section 3.1.1 for details). A particular implementation in this case is the CG method [36].

The MR Krylov subspace method is well defined whenever A is nonsingular. This feature makes this method very popular, since it can be used for general matrices. The most well-known implementations are the MINRES method [56] for hermitian indefinite matrices and the GMRES method [62] for general nonsingular matrices.

Finally, note that the conditions $x_n \in x_0 + \mathcal{K}_n(A, r_0)$ and $r_n \in r_0 + A\mathcal{K}_n(A, r_0)$ imply that the error $x - x_n$ and the residual r_n can be written in the polynomial form

$$x - x_n = p_n(A)(x - x_0), \qquad r_n = p_n(A) r_0, \tag{8}$$

where p_n is a polynomial of degree at most n and with value one at the origin. For a well-defined OR/MR Krylov subspace method, the polynomial p_n is uniquely determined by the constraint conditions (3).

3 Convergence analysis

In exact arithmetic, well-defined Krylov subspace methods terminate in a finite number of steps. Therefore no limit can be formed, and terms like "convergence" or "rate of convergence" loose their classical meaning; see, e.g., [35, Chapter 9.4] for a cautioning in this direction. This situation requires approaches that are substantially different from the analysis of classical fixed point iteration methods such as Gauß-Seidel or SOR. The convergence of the latter methods has typically been described asymptotically, with the "asymptotic convergence factor" of the iteration matrix being the central concept. Surprisingly, this principal difference between the Krylov subspace methods and the classical iteration methods is still not always accepted. For example, the classical convergence *bound* for the CG method that is based on the matrix condition number (see equation (15) below) is sometimes confused with the actual convergence *behavior* of the method. Hence the actual convergence is identified with a bound based on the asymptotic convergence factor of the convex hull of the spectrum, without considering any other properties of the given data. Clearly, this approach can be very misleading in some situations.

A related difficulty in the convergence analysis is the typical requirement of finding an acceptable approximate solution x_n in $n \ll N$ steps. Therefore it is important to understand the convergence from the very beginning, i.e., in the classical terminology, to understand the "transient" behavior. This early stage of convergence, however, can depend significantly on the right hand side b and the initial guess x_0. In general, the non-existing limiting process, the relevance of the transient phase, and the dependence of this phase on b and x_0 make the convergence analysis of Krylov subspace methods a difficult nonlinear problem – although the system to be solved is linear.

We divide our discussion about the convergence of Krylov subspace methods into two parts. In the first part (Section 3.1) we consider normal system matrices A and show that in this case the spectral information is important for analyzing the convergence. The second part (Section 3.2) shows the difficulties with estimating the convergence in the nonnormal case.

3.1 Convergence analysis for normal matrices

Consider a nonsingular and *normal* matrix A, and let

$$A = V \Lambda V^H, \qquad \text{where} \qquad V^H V = I, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)\,,$$

be its eigendecomposition. The orthogonality of the eigenvector basis lead s to a significant simplification in the convergence analysis of Krylov subspace methods: Considering A^n in the form $V \Lambda^n V^H$ and using (8), the errors and residuals of a Krylov subspace method satisfy

$$x - x_n = V p_n(\Lambda) V^H (x - x_0), \qquad r_n = V p_n(\Lambda) V^H r_0\,. \tag{9}$$

Because the projection property usually refers to some sort of optimality, we can expect that Krylov subspace methods for normal matrices solve some weighted polynomial minimization problem on the matrix spectrum. In the following subsections we explain that in the worst-case, the convergence speed of well-known Krylov subspace methods (CG, MINRES, GMRES) is determined by the value

$$\min_{p \in \pi_n} \max_k |p(\lambda_k)|\,, \tag{10}$$

where π_n denotes the set of polynomials of degree at most n and with value one at the origin. Note that the value (10) represents a min-max approximation problem on the discrete set of the matrix eigenvalues. The value (10) depends in a complicated (nonlinear) way on the eigenvalue distribution. Consider, for simplicity, that all eigenvalues are real and distinct. The results in [26, 51] show that there exists a subset of $n+1$ (distinct) eigenvalues $\{\mu_1, \ldots, \mu_{n+1}\} \subseteq \{\lambda_1, \ldots, \lambda_N\}$, such that

$$\min_{p \in \pi_n} \max_k |p(\lambda_k)| \;=\; \left(\sum_{j=1}^{n+1} \prod_{\substack{k=1 \\ k \neq j}}^{n+1} \frac{|\mu_k|}{|\mu_k - \mu_j|} \right)^{-1}\,. \tag{11}$$

If at least one eigenvalue of A is complex, the equality (11) does not hold in general, cf. [51]. Nevertheless, in [51] we formulate a conjecture, supported by numerical experiments and by some theoretical results, that there exist a set of $n+1$ eigenvalues such that the value on the right hand site of (11) is equal to (10) up to a factor between 1 and $4/\pi$.

Of course, except for model problems and special situations, not all eigenvalues of A are known, and hence an analysis based on (11) cannot be applied. In the following we will concentrate on the practically more relevant approach to estimate the value of (10) using only a partial knowledge of the spectrum, in particular only some set that contains all the eigenvalues (a so-called inclusion set). An inclusion set is often known a priori or can be easily estimated. We discuss the resulting convergence bounds for CG (hermitian positive definite A), MINRES (hermitian A) and GMRES (general normal A).

3.1.1 Convergence analysis for CG

Consider a *hermitian positive definite* matrix A. Each such matrix defines a norm (the so-called A-norm),

$$\|u\|_A \;=\; \left(u^H A u\right)^{\frac{1}{2}} , \tag{12}$$

and it is well known (see, e.g., [27]) that the OR Krylov subspace iterates x_n are in this case uniquely defined in each iterative step n and can be computed using the CG method. The CG iterates x_n satisfy

$$\|x - x_n\|_A \;=\; \min_{p \in \pi_n} \|p(A)(x - x_0)\|_A \,. \tag{13}$$

In other words, the CG method constructs an approximation x_n from the affine subspace $x_0 + \mathcal{K}_n(A, r_0)$ with minimal A-norm of the error. It can be shown that the A-norm of the error is strictly monotonically decreasing, i.e., that $\|x - x_n\|_A < \|x - x_{n-1}\|_A$ for $n = 1, \ldots, d$. The A-norm of the error often has a counterpart in the underlying real-world problem. For example, when the linear system comes from finite element approximations of self-adjoint elliptic PDEs, then the A-norm of the error can be interpreted as the discretized measure of energy which is to be minimized; see, e.g., [1, 2].

A simple algebraic manipulation shows that the value (10) represents an upper bound on the relative A-norm of the error,

$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \;\le\; \min_{p \in \pi_n} \max_k |p(\lambda_k)| \,. \tag{14}$$

This convergence bound is sharp, i.e., for each iteration step n there exist a right hand side b or an initial guess x_0 (depending on n and A) such that equality holds in (14), see [26]. In this sense, the bound (14) completely describes the *worst-case behavior* of the CG method. When the whole spectrum of A is known, one can try to determine the value of the right hand side of (14) using the formula (11). However, it is in general an open problem which subset of $n+1$ eigenvalues leads to equality in (11).

Obviously, the bound (14) depends only on the matrix eigenvalues and not on any other properties of A, b, or x_0. If a particular right hand side b is known, it is sometimes possible to incorporate the information about b into the analysis, and thus to obtain a better estimate of the actual convergence behavior.

Estimating the bound (14). Often, the largest and smallest eigenvalue (or at least estimates for them) are known. Then the classical approach is to replace the discrete set of the matrix eigenvalues by an interval containing all eigenvalues and to use Chebyshev polynomials of the first kind to estimate the min-max approximation (14). This results in the following well-known upper bound based on the condition number of A, i.e. the ratio of the largest and the smallest eigenvalue (see, e.g., [27]),

$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n, \qquad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{15}$$

We stress that there is a principal difference between the bounds (14) and (15). The bound (14) represents a min-max approximation problem on the *discrete set* $\lambda_1, \ldots, \lambda_N$, and it describes the convergence behavior in the worst-case sense. On the other hand, the bound (15) represents an estimate of the min-max approximation on the *interval* $[\lambda_{\min}, \lambda_{\max}]$ containing all eigenvalues of A. It therefore bounds the worst-case behavior for all possible eigenvalue distributions in the given interval. In other words, the bounds (14) and (15) describe different approximation problems, and thus their values can differ significantly. Clearly, the bound (15) cannot be identified with the CG convergence, and it represents an overestimate even of the worst-case behavior except for very special eigenvalue distributions in the given interval (see [50] for further discussion of this fact). The bound (15) shows, however, that a small condition number (close to 1) implies fast convergence of the CG method. This justifies the classical goal of "preconditioning", namely to decrease the condition number of the given system matrix. On the other hand, the bound (15) does *not* show that a large condition number implies slow convergence of the CG method.

Example 3.1 Consider two example eigenvalue distributions in the interval $[1/400, 1]$. The first eigenvalue set, given by

$$\lambda_k = k^2/400, \quad k = 1, \ldots, 20, \tag{16}$$

has a cluster close to zero, whereas the second set, given by

$$\lambda_k = \log(k)/\log(20), \quad k = 2, \ldots, 20, \quad \lambda_1 = 1/400, \tag{17}$$

has a cluster close to one. Each hermitian and positive definite matrix having the eigenvalues (16) or (17) has the (moderate) condition number 400. Fig. 1 shows that the worst-case CG convergence behavior differs significantly for the eigenvalue set (16) (solid) and for the eigenvalue set (17) (dashed). Since the bound (15) (dash-dotted) represents an upper bound on the worst-case CG behavior for any eigenvalue distribution in the given interval, it cannot describe the actual CG convergence for a particular eigenvalue set like (17). □

An alternative estimate for the value (10), based on the ratio of arithmetic and geometric averages of the eigenvalues (the so-called K-condition number), was introduced by Kaporin [41]. This and other related estimates can also be found in [5, Chapter 13]. In [6], Axelsson and Kaporin propose convergence estimates for the CG method based on a generalized condition number of A, which also depends on the initial error.

Superlinear convergence of CG. In many applications it has been observed that the A-norm of the error in the CG method converges "superlinearly", which means that speed of

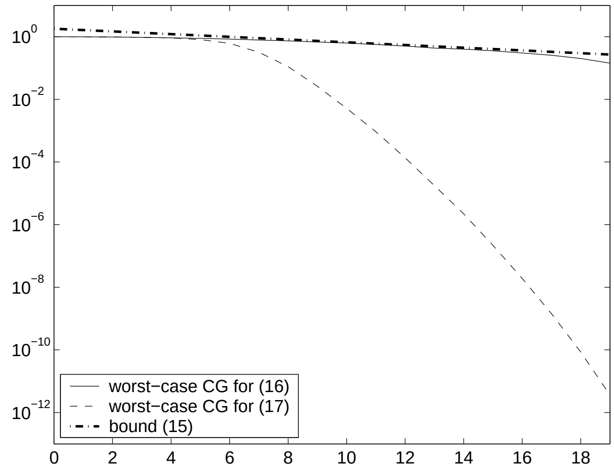

Fig. 1 For a particular eigenvalue distribution (17), the worst-case CG behavior (dashed) can significantly differ from the bound (15) (dash-dotted).

convergence increases during the iteration. Some attempts have been made to explain this behavior using the convergence of Ritz values in the Lanczos process that underlies the CG method. An intuitive explanation of the superlinear behavior, given in the early paper [11], is that when the extremal eigenvalues of A are well approximated by the Ritz values, then the CG method proceeds as if the corresponding eigenvectors were not present . This leads to a smaller "effective" condition number of A, which in turn might explain the faster convergence. This situation is discussed and analyzed, for example, in [52, 71, 69]; see [70, Chapter 5.3] for a recent summary.

The results just mentioned attempt to explain the behavior of the CG method using information that is generated during the run of the method. A different, and certainly not less interesting problem is to identify (a priori) properties of the input data A, b and x_0 that imply superlinear convergence behavior. This problem is considered in an asymptotic setting by Beckermann and Kuijlaars [7, 8]. They show that superlinear CG convergence can be observed when solving a sequence of linear systems with hermitian positive definite matrices whose eigenvalue distributions are far from an equilibrium distribution [7] (see, e.g., [22] for an introduction to these asymptotic concepts). Such favorable eigenvalue distributions occur, for example, when the system matrices come from the standard five-point finite difference discretizations of the two-dimensional Poisson equation. Another situation where superlinear convergence is observed despite an equilibrium distribution of the eigenvalues is when the components of the initial error in the eigenvector basis of the system matrices strongly vary in size [8]. In a finite dimensional setting, analytic examples for this phenomenon in the context of the discretized one-dimensional Poisson equation are given in [50].

Example 3.2 Consider the N by N tridiagonal symmetric and positive definite Toeplitz matrix $A = \mathrm{tridiag}(-1, 2, -1)$ for $N = 120$, that arises by the central finite difference approximation of the one-dimensional Poisson equation. As proved asymptotically by Beckermann and Kuijlaars [8], CG may for this model problem converge superlinearly when the initial error exhibits certain distributions of components in the eigenvector basis of A.

For particular initial errors, the superlinear convergence can in this model problem even be proved in a finite dimensional setting. In particular, consider an initial error whose components in the eigenvector basis of A are given by $\gamma \sin^{-2}(k\pi/(2N+2))$, $k = 1, \ldots, N$, where γ represents a nonzero scaling factor; cf. the solid line in the right part of Fig 2. Apparently, these components strongly vary in size, with larger components corresponding to smaller eigenvalues of A. Using the results of Naiman et al. [54], it can be shown by an elementary computation [50], that the CG errors for this initial error satisfy

$$\frac{\|x - x_n\|_A}{\|x - x_{n-1}\|_A} = \left(\frac{N-n}{N-n+3} \right)^{1/2}, \qquad n = 1, \ldots, N .$$

The right hand side in the above equation is a strictly decreasing function of the iteration step n, which gives an analytic proof for the superlinear CG convergence for A and this initial error. The superlinear CG convergence curve is shown as the solid line in the left part of Fig. 2. For comparison, we use an initial error with eigencomponents that are equally distributed; cf. the dashed line in the right part of Fig 2. As shown by the dashed line in the left part of Fig 2, no superlinear convergence can be observed in this case. □

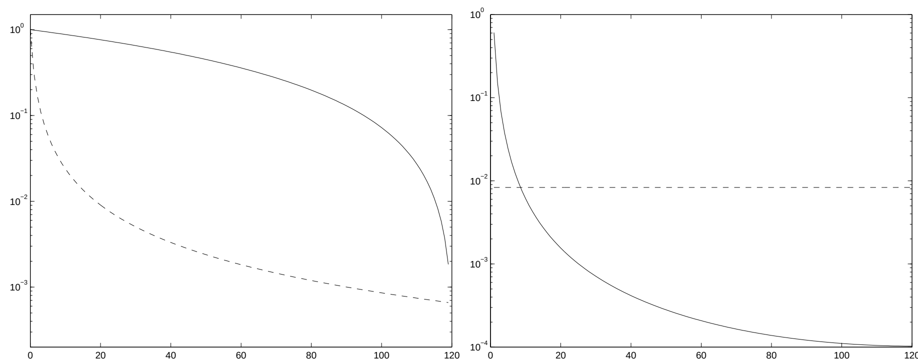

Fig. 2 CG convergence curves (left part) for two distributions of eigencomponents of the initial error (right part).

In summary, the convergence behavior of the CG method is relatively well understood, but some open problems still remain. The right approach for investigating the convergence behavior is to use all information about the eigenvalue distribution we have at our disposal. If a particular right hand side b and initial guess x_0 are given, they should be incorporated in the analysis. An example for such an approach for the model problem of the one-dimensional Poisson equation is given in [50].

3.1.2 Convergence analysis for MINRES and GMRES

In this subsection we consider nonsingular and *normal* matrices A. It is well known (see, e.g., [27]) that the iterates x_n of the MR Krylov subspace method are for any such matrix uniquely defined in each iterative step n, and that the nth residual $r_n = b - Ax_n$ satisfies

$$\|r_n\| = \min_{p\in\pi_n} \|p(A)r_0\|. \tag{18}$$

The residual norms decrease strictly monotonically whenever zero is outside the field of values of A, see [15, 33] for different proofs. However, in general no strict monotonicity is guaranteed. In fact, any (finite) nonincreasing sequence of numbers represents a convergence curve of the MR Krylov subspace residual norms applied to some linear system with a normal system matrix [3, 32, 48]. The normal matrix can even be chosen to have all its eigenvalues on the unit circle.

In the normal case, the relative residual norm of the MR Krylov subspace method can be bounded similarly as in (14),

$$\frac{\|r_n\|}{\|r_0\|} \le \min_{p\in\pi_n} \max_k |p(\lambda_k)| \tag{19}$$

and again, the bound (19) is sharp [31, 40]. In other words, the bound (19) describes the worst-case behavior of the MR Krylov subspace method. If the full spectral information is available, then the approach in [51] (cf. the discussion of formula (11)) can be used for estimating the worst-case convergence behavior. Otherwise, one can try to estimate the worst-case bound (19) similarly as in the hermitian positive definite case, i.e., by replacing the discrete spectrum by a continuous inclusion set. However, as we will see, the estimation of the min-max approximation becomes much more complicated now.

The hermitian indefinite case. When A is hermitian indefinite, the MR Krylov subspace method MINRES can be used. An estimate on the min-max approximation (19) that represents the worst-case MINRES convergence behavior, can be obtained by replacing the discrete set of the eigenvalues by the union of two intervals containing all of them and *excluding the origin*, say $I^- \cup I^+ \equiv [\lambda_{\min}, \lambda_s] \cup [\lambda_{s+1}, \lambda_{\max}]$ with $\lambda_{\min} \le \lambda_s < 0 < \lambda_{s+1} \le \lambda_{\max}$. Note that if zero would be contained in the inclusion set $I^- \cup I^+$, then the optimal min-max polynomial from π_n on this set would be the constant polynomial $p_n(z) = 1$ for all n, and the resulting convergence bounds would be useless.

When both intervals are of the same length, i.e., $\lambda_{\max} - \lambda_{s+1} = \lambda_s - \lambda_{\min}$, the following bound for the min-max value can be found,

$$\min_{p\in\pi_n} \max_k |p(\lambda_k)| \quad \le \quad \min_{p\in\pi_n} \max_{z\in I^-\cup I^+} |p(z)| \tag{20}$$

$$\le \quad 2\left(\frac{\sqrt{|\lambda_{\min}\lambda_{\max}|} - \sqrt{|\lambda_s\lambda_{s+1}|}}{\sqrt{|\lambda_{\min}\lambda_{\max}|} + \sqrt{|\lambda_s\lambda_{s+1}|}}\right)^{[k/2]}, \tag{21}$$

where $[k/2]$ denotes the integer part of $k/2$, see [27, Chapter 3]. For an illustration of this bound suppose that $|\lambda_{\min}| = \lambda_{\max} = 1$ and $|\lambda_s| = \lambda_{s+1}$. Then the condition number of A is $\kappa = \lambda_{s+1}^{-1}$, and the right hand side of (21) reduces to

$$2\left(\frac{\kappa - 1}{\kappa + 1}\right)^{[k/2]}. \tag{22}$$

Apparently, (22) corresponds to the value of right hand side of (15) at step $[k/2]$ for a hermitian positive definite matrix having all its eigenvalues in the interval $[\lambda_{s+1}^2, 1]$, and thus a condition number of λ_{s+1}^{-2}. Hence the convergence bound for an indefinite matrix with condition number κ needs twice as many steps to decrease to the value of the bound for a definite matrix with condition number κ^2. Although neither of the two bounds is sharp, this clearly indicates that solving indefinite problems represents a significant challenge. In the general case when the two intervals I^- and I^+ are not of the same length, the explicit solution of the min-max approximation problem on $I^- \cup I^+$ becomes quite complicated, see, e.g., [22, Chapter 3], and no simple and explicit bound on the min-max value is known. One may of course extend the smaller interval to match the length of the larger one, and still apply (21). But this usually results in a significantly weaker convergence bound, which fails to give relevant information about the actual convergence behavior. Similar as in the case of the CG method we stress that there is a principal difference between the bounds (19) and (21). These bounds describe different approximation problems, and thus their values can differ significantly.

The general normal case. If A is a general normal matrix, the MR Krylov subspace method GMRES can be used. Again, an estimate of the right hand side of (19) can be obtained by replacing the discrete set of the eigenvalues of A by some (compact) inclusion set $\Omega \subset \mathbb{C}$ on which (nearly) optimal polynomials are explicitly known. Usually one works with connected inclusion sets, since polynomial approximation on disconnected sets is not well understood (even in the case of two disjoint intervals; see above). Because of the normalization of the polynomials at zero, the set Ω should not include the origin.

The simplest result is obtained when the spectrum of A is contained in a disk in the complex plane (that excludes the origin), say with radius $r > 0$ and center at $c \in \mathbb{C}$. Then the polynomial $p_n(z) = ((c-z)/c)^n \in \pi_n$ can be used to show that

$$\min_{p \in \pi_n} \max_k |p(\lambda_k)| \leq \left| \frac{r}{c} \right|^n .$$

In particular, a disk of small radius that is far from the origin guarantees fast convergence of the GMRES residual norms.

More refined bounds can be obtained using the convex hull $\mathcal{E}$ of an ellipse instead of a disk. For example, suppose that the spectrum is contained in an ellipse with center at $c \in \mathbb{R}$, focal distance $d > 0$ and major semi axis $a > 0$. If $0 \notin \mathcal{E}$, it can be shown that

$$\min_{p \in \pi_n} \max_k |p(\lambda_k)| \leq \frac{C_n(a/d)}{|C_n(c/d)|} \approx \left(\frac{a + \sqrt{a^2 - d^2}}{c + \sqrt{c^2 - d^2}} \right)^n ,$$

where $C_n(z)$ denotes the nth complex Chebyshev polynomial, see, e.g., [59]. We remark that, as shown by Fischer and Freund [23], the polynomials $C_n(z)/C_n(0)$ are in general not the optimal min-max polynomials from π_n on $\mathcal{E}$. However, these polynomials are asymptotically optimal and hence predict the correct rate of convergence of the min-max approximation problem on $\mathcal{E}$. For more details we refer to [61].

Of course, one would like to find a set Ω in the complex plane that yields the smallest possible upper bound on the right hand side of (19). Both a disk and the convex hull of an ellipse are convex, so one can probably improve the convergence bound by using the smallest convex set containing all the eigenvalues, i.e., the convex hull of the eigenvalues. Since A is assumed normal, this set coincides with the field of values $\mathcal{F}(A)$. Hence the bound (28) studied below in the context of nonnormal matrices can in principle be used in the normal

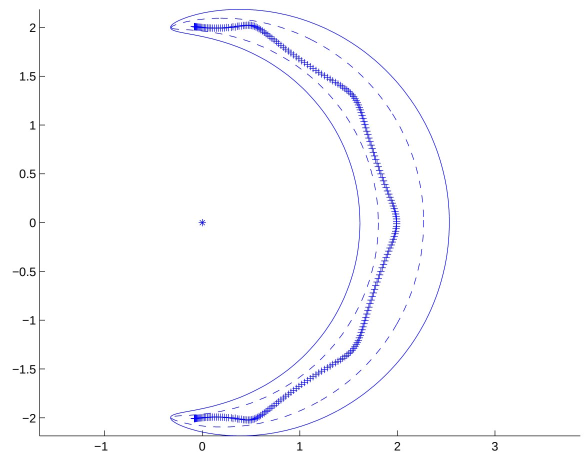

Fig. 3 Tight inclusion of the eigenvalues of the GRCAR matrix by two elements of the class of sets introduced in [43, 47].

case as well. However, all convex inclusion sets Ω are limited in their applicability by the strict requirement that $0 \notin \Omega$. In particular, if zero is inside the convex hull of the eigenvalues of A, then no convex inclusion set for these points can be used. Moreover, if the convex hull is close to the origin, then any bound derived from this set will be poor, regardless of the distance of the eigenvalues to the origin. Another difficulty with using the convex hull of the eigenvalues (or any other inclusion set bounded by a polygon) is that the boundary of this set in not smooth and hence the computation of (nearly) optimal polynomials on these sets such as the Faber polynomials is complicated, see, e.g., [64].

To overcome such difficulties, a parameterized class of non-convex sets with analytic boundaries is constructed in [47] (also see [43]), for which the Faber polynomials are explicitly known. These polynomials give rise to analytic and easily computable bounds for the min-max approximation problem; see [47] for details. Two examples of the inclusion sets are show in Fig. 3. The plus signs in this figure show the eigenvalues of the so-called Grcar matrix of order 250, generated by the MATLAB command `gallery('grcar',250,6)`. Obviously, the convex hull of these eigenvalues contains the origin (indicated by the star). On the other hand, none of the eigenvalues is particularly close to the origin, which should be exploited by the choice of the inclusion set. The boundaries of the two example inclusion sets are shown by the dashed and the solid curves.

3.2 Convergence analysis for nonnormal matrices (GMRES)

In this section we consider the case of a general nonsingular and *nonnormal* matrix A. In this general case, an MR Krylov subspace method such as GMRES yields uniquely defined iterates x_n so that the nth residual $r_n = b - Ax_n$ satisfies

$$\|r_n\| = \min_{p \in \pi_n} \|p(A)r_0\|. \tag{23}$$

Similarly to the convergence analysis for normal matrices presented above, we are interested in finding a (sharp) bound on the right hand side of (23).

Eigenvalues and convergence. If A is diagonalizable, $A = V\Lambda V^{-1}$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, then the following convergence bound easily follows from (23),

$$\frac{\|r_n\|}{\|r_0\|} \;=\; \min_{p\in\pi_n} \frac{\|V p(\Lambda) V^{-1} r_0\|}{\|r_0\|} \;\leq\; \kappa(V) \min_{p\in\pi_n} \max_{k} |p(\lambda_k)|\,, \tag{24}$$

see, e.g., [62]. Here $\kappa(V) = \|V\|\,\|V^{-1}\|$ denotes the condition number of the eigenvector matrix V. A bound similar to (24) can be derived for nondiagonalizable matrices.

The bound (24) frequently is the basis for discussions of the GMRES convergence behavior. As mentioned in Section 3.1.2, this bound is sharp when A is normal. When $\kappa(V)$ is small, the right hand side of (24) typically represents a good convergence bound, and its value can be estimated using the tools described above. However, when V is far from unitary, the bound (24) may fail to provide any reasonable information about the GMRES convergence. To see this, note that when the eigenvector matrix V is ill-conditioned, then some components of the vector $V^{-1}r_0$ can be very large, potentially much larger than $\|r_0\|$. On the other hand, $\|r_n\|$ in (24) is bounded from above by $\|r_0\|$. Therefore, the linear combination $V\,[p(\Lambda)V^{-1}r_0]$ can contain a significant cancellation, which is not reflected in the minimization problem on the right hand side of (24). Apart from the fact, that the factor $\kappa(V)$ can be very large in case of ill-conditioned eigenvectors, the principal weakness of the bound (24) is that the min-max problem on the matrix eigenvalues need not have any connection with the GMRES convergence for the given nonnormal matrix. As a consequence, the curve produced by the min-max approximations on matrix eigenvalues can be substantially different from the (worst-case) GMRES convergence curve and the bound can fail to give any reasonable convergence information.

Example 3.3 For a numerical illustration consider the two N by N tridiagonal Toeplitz matrices

$$A_\lambda \;=\; \mathrm{tridiag}(-1, \lambda, -1) \quad \text{and} \quad B_\lambda \;=\; \mathrm{tridiag}(-\lambda, \lambda, -1/\lambda)\,,$$

where $\lambda \geq 2$ is a real parameter. Both A_λ and B_λ have *the same eigenvalues*, namely $\lambda - 2\cos(k\pi/(N+1))$, $k = 1, \ldots, N$. While A_λ is symmetric positive definite, B_λ is highly nonnormal (e.g. a MATLAB computation yields $\kappa(V) \approx 10^{27}$ for $N = 40$ and $\lambda = 3$). The relative GMRES residual norms for $x_0 = 0$ and the order 40 systems $A_\lambda x = [1, 0, \ldots, 0]^T$ and $B_\lambda x = [1, 0, \ldots, 0]^T$, for $\lambda = 3, 4, \ldots, 18$, are shown in Fig. 4. The relative residual norms for the systems with A_λ are plotted by solid lines (faster convergence corresponds to larger λ), and for the systems with B_λ they are plotted by dashed lines (essentially the same for all λ). We observe that the GMRES convergence speed for A_λ increases when the spectrum moves away from the origin. On the other hand, for B_λ spectral information is obviously useless for describing the GMRES convergence. In this example essentially nothing happens during the first $N - 1$ steps, and then termination occurs in the final step N. Moreover, the spectrum of B_λ gives no information about the convergence behavior after some "transient delay", which some authors attribute to the potentially large constant $\kappa(V)$ in (24). □

The above example for the matrices B_λ clearly shows that in the nonnormal case eigenvalue information is not sufficient for describing the convergence behavior of GMRES. In

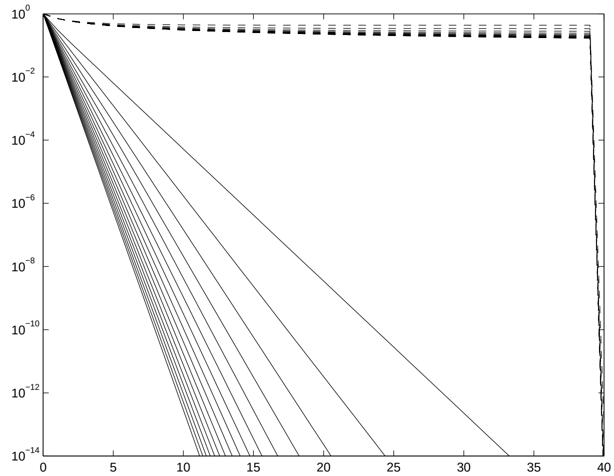

Fig. 4 Relative GMRES residual norms for the normal matrices A_λ (solid) and the nonnormal matrices B_λ (dashed) for $\lambda = 3, 4, \ldots, 18$ and $r_0 = [1, 0, \ldots, 0]^T$.

fact, in this case the eigenvalues may have nothing to do with the convergence behavior at all. As shown in [3, 32], any nonincreasing convergence curve of relative GMRES residual norms is attainable for a system matrix A having any prescribed eigenvalues. On the other hand, it needs to be stressed that from an analytic point of view the principal difficulty of nonnormality is *not* the often met belief that the convergence is slower for nonnormal than for normal matrices. This belief is incorrect because for each nonnormal matrix A there exists a normal matrix B for which the same convergence behavior can be observed (for the same initial residual r_0), cf. [33, 48]. Unfortunately, the mapping from the matrix A to the normal matrix B is highly nonlinear, and it depends strongly on r_0. Hence it is not suitable for an a priori analysis of the GMRES convergence behavior for the given A and r_0.

The idea to analyze the given nonnormal problem using a related normal problem is also used by Huhtanen and Nevanlinna [39]. They propose to split the matrix A into $A = \tilde{A} + E$, where $\tilde{A}$ is normal and E is of smallest possible rank. Using such splitting, lower bounds for the quantity $\min_{p\in\pi_n} \|p(A)\|$ (cf. (26) below) can be given in terms of certain eigenvalues of $\tilde{A}$; see [39] for details.

Worst-case GMRES analysis in the nonnormal case. It should be clear by now that in the nonnormal case the GMRES convergence behavior is *significantly more difficult to analyze* than in the normal case. A general approach to understand the worst-case GMRES convergence in the nonnormal case is to replace the complicated minimization problem (23) by another one that is easier to analyze and that, in some sense, approximates the original problem (23). Natural bounds on the GMRES residual norm arise by excluding the influence of the initial residual r_0,

$$\begin{aligned} \frac{\|r_n\|}{\|r_0\|} &= \min_{p\in\pi_n} \frac{\|p(A)r_0\|}{\|r_0\|} && \text{(GMRES)} \\ &\le \max_{\|v\|=1} \min_{p\in\pi_n} \|p(A)v\| && \text{(worst-case GMRES)} \qquad (25) \\ &\le \min_{p\in\pi_n} \|p(A)\| && \text{(ideal GMRES).} \qquad (26) \end{aligned}$$

The bound (25) corresponds to the *worst-case* GMRES behavior and represents a sharp upper bound, i.e. a bound that is attainable by the GMRES residual norm. In this sense, (25) is the best bound on $\|r_n\|/\|r_0\|$ that is independent of r_0. Despite the independence of r_0, it is not clear in general, which properties of A influence the bound (25); see, e.g., [20]. The expression (25) can be bounded by the *ideal* GMRES approximation problem (26), which was introduced by Greenbaum and Trefethen [34]. To justify the relevance of the bound (26), several researchers tried to identify cases in which (25) is equal to (26). The best known result of this type is that (25) is equal to (26) whenever A is normal [31, 40]. Despite the existence of some counterexamples [20, 67], it is still an open question whether (25) is equal or close to (26) for larger classes of nonnormal matrices. In [66] we consider this problem for a Jordan block, a representative of a nonnormal matrix, and prove equality of the expressions (25) and (26) in some steps.

A possible way to approximate the value of the matrix approximation problem (26) is to determine sets $\Omega \subset \mathbb{C}$ and $\hat{\Omega} \subset \mathbb{C}$, that are somehow associated with A, and that provide lower and upper bounds on (26),

$$c_1 \min_{p\in\pi_n} \max_{z\in\Omega} |p(z)| \;\le\; \min_{p\in\pi_n} \|p(A)\| \;\le\; c_2 \min_{p\in\pi_n} \max_{z\in\hat{\Omega}} |p(z)|.$$

Here c_1 and c_2 should be some (moderate size) constants depending on A and possibly on n. This approach represents a generalization of the idea for normal matrices, where the appropriate set associated with A is the spectrum of A and $c_1 = c_2 = 1$.

Trefethen [68] has suggested taking $\hat{\Omega}$ to be the *ϵ-pseudospectrum* of A,

$$\Lambda_\epsilon(A) = \left\{ z \in \mathbb{C} \,:\, \|(zI - A)^{-1}\| \ge \epsilon^{-1} \right\} .$$

Denoting by L the arc length of the boundary of $\Lambda_\epsilon(A)$, the following bound can be derived,

$$\min_{p\in\pi_n} \|p(A)\| \le \frac{L}{2\pi\epsilon} \min_{p\in\pi_n} \max_{z\in\Lambda_\epsilon(A)} \|p(z)\| , \qquad (27)$$

see, e.g., [53]. The parameter ϵ gives some flexibility, but choosing a good value can be tricky. Note that in order to make the right hand side of (27) reasonably small, one must choose ϵ large enough to make the constant $L/2\pi\epsilon$ small, but small enough to make the set $\Lambda_\epsilon(A)$ not too large. The bound (27) works well in some situations (see, e.g., [17]), but it is easy to construct examples for which no choice of ϵ gives a tight estimate of the ideal GMRES approximation problem (see, e.g., [33]).

Another approach is based on the *field of values* of A, cf. (7). Denote by $\nu(\mathcal{F}(A))$ the distance of $\mathcal{F}(A)$ from the origin, $\nu(\mathcal{F}(A)) = \min_{z\in\mathcal{F}(A)} |z|$, then

$$\min_{p\in\pi_n} \|p(A)\| \le \left(1 - \nu(\mathcal{F}(A))\nu(\mathcal{F}(A^{-1}))\right)^{n/2} , \qquad (28)$$

see, e.g., [14]. Suppose that $M = (A + A^H)/2$, the hermitian part of A, is positive definite. Then a special case of (28) is

$$\min_{p\in\pi_n} \|p(A)\| \leq \left(1 - \frac{\lambda_{\min}(M)}{\lambda_{\max}(A^H A)}\right)^{n/2} ,$$

which is one of the earliest convergence results for the MR Krylov subspace method [15, 16]. Since $\mathcal{F}(A)$ is a convex set that contains the convex hull of the eigenvalues of A, the requirement $0 \notin \mathcal{F}(A)$ makes the bound (28) useless in many situations. However, the field of values analysis can be very useful when the given linear system comes from the discretization of elliptic PDEs by the Galerkin finite element method. In such cases the coefficients of the N by N system matrix A are given by $A_{ij} = a(\phi_i, \phi_j)$, where $a(u, v)$ is the bilinear form from the weak formulation of the PDE, and $\phi_1, \ldots, \phi_N$ are the nodal basis functions. Let V_h denote the finite element space. Then a function $u_h \in V_h$ is represented by a vector $u_N \in \mathbb{R}^N$ that contains the values of u_h at the nodes of the triangulation. The matrix A satisfies $u_N^T A v_N = a(u_h, v_h)$ for all $u_h, v_h \in V_h$. These relations can be exploited to give bounds for the quantity $a(x - x_n, x - x_n) = (x - x_n)^T A(x - x_n)$, where x is the exact solution of the discretized PDE, and x_n is a Krylov subspace iterate. This leads naturally to bounds of the type (28) involving the smallest real parts of $\mathcal{F}(A)$ and $\mathcal{F}(A^{-1})$; see, e.g., [42, 63] for more details. Note that under the usual assumption that the bilinear form is coercive, the smallest real parts of $\mathcal{F}(A)$ and $\mathcal{F}(A^{-1})$ are both positive. In a more abstract setting, the field of values has been used in the convergence analysis by Eiermann [13].

A generalization of the field of values of A is the *polynomial numerical hull*, introduced by Nevanlinna [55], and defined as

$$\mathcal{H}_n(A) = \{z \in \mathbb{C} : \|p(A)\| \geq |p(z)| \text{ for all } p \in \mathcal{P}_n\} ,$$

where $\mathcal{P}_n$ denotes the set of polynomials of degree n or less. It can be shown that $\mathcal{F}(A) = \mathcal{H}_1(A)$. The set $\mathcal{H}_n(A)$ provides a lower bound on (26),

$$\min_{p\in\pi_n} \max_{z\in\mathcal{H}_n(A)} |p(z)| \;\leq\; \min_{p\in\pi_n} \|p(A)\|. \tag{29}$$

In some way, $\mathcal{H}_n(A)$ reflects the complicated relation between the polynomial of degree n and the matrix A, and provides often a very good estimate of the value of the ideal GMRES approximation (26). Greenbaum and her co-workers [19, 28, 29, 30] have obtained theoretical results about $\mathcal{H}_n(A)$ for Jordan blocks, banded triangular Toeplitz matrices and block diagonal matrices with triangular Toeplitz blocks. Clearly, for a larger applicability of the bound (29), the class of matrices for which $\mathcal{H}_n(A)$ is known needs to be extended. But in general, the determination of these sets represents a nontrivial open problem.

The bounds stated above are certainly useful to obtain a priori convergence estimates in terms of properties of A, and possibly to analyze the effectiveness of preconditioning techniques. However, the worst-case behavior of GMRES for nonnormal matrices is still not well understood. We again point out that the bound (26) is not sharp, and that it is in many situations unclear how closely the ideal GMRES approximates the worst-case GMRES. Moreover, none of the bounds stated above is able to *characterize* satisfactorily in terms of matrix properties, which approximation problem is solved by the worst-case GMRES in the nonnormal case.

The influence of the initial residual: A model problem. Users of Krylov subspace methods usually want to solve a particular linear system, and hence a worst-case analysis may be of lesser interest to them. In such context one needs to understand also how the convergence is influenced by the particular right hand side or initial residual r_0. It seems to be well known that the initial residual may have a significant influence on the GMRES convergence, in particular in the nonnormal case. However, no systematic study of this influence exists, and given the lack of understanding of even the worst-case behavior, it is unlikely that a complete understanding of the influence of r_0 on the convergence will be obtained in the near future.

In the context of discretized PDEs, r_0 is directly related to the boundary conditions and/or the source terms. It is of great importance to understand how such PDE data influences the convergence of an iterative solver like GMRES, as understanding of these relations will pave the way to efficient preconditioning techniques. Recently, this topic was addressed in an analysis of the GMRES convergence behavior for a well known convection-diffusion model problem [49], that was introduced in [24]. Here the convergence of GMRES applied to the discretized system is characterized by an initial phase of slow convergence, followed by a faster decrease of the residual norms. The length of the initial phase depends on the initial residual, which is determined by the boundary conditions (for simplicity, the source term in the PDE and the initial guess x_0 are chosen equal to zero in [49]). Typical examples for the convergence behavior are shown in Fig. 5. The GMRES convergence curves in this figure

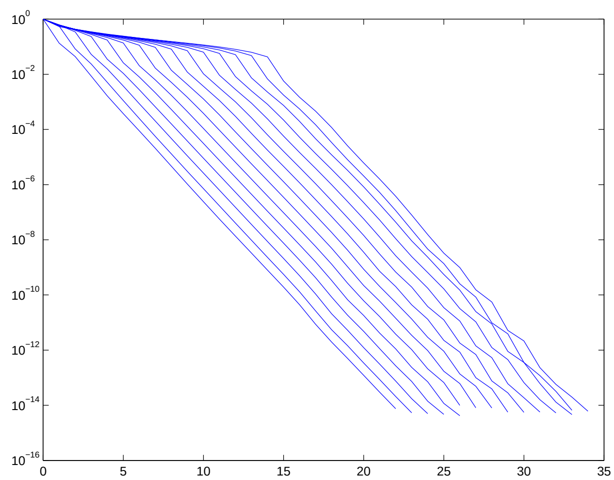

Fig. 5 Relative GMRES residual norms for the discretized convection-diffusion model problem considered in [49]. Different behavior corresponds to the same discretized operator but to different boundary conditions.

correspond to the same discretized operator but to different boundary conditions. For the considered model problem, the convergence analysis confirms an earlier conjecture of Ernst [18], that the duration of the initial phase is governed by the time it takes for boundary information to pass from the inflow boundary across the domain following the longest streamline of the velocity field. The paper [49] also discusses the question why the convergence in the second phase accelerates. Numerical results show that the speed of convergence after the initial delay is slower for larger mesh Peclet numbers, but a complete quantitative understanding of this phenomenon remains a difficult open problem.

4 Concluding remarks

The worst-case convergence behavior of many well known Krylov subspace methods (CG, MINRES, GMRES) for normal matrices is described by the min-max approximation problem on the discrete set of the matrix eigenvalues,

$$\min_{p\in\pi_n} \max_{k} |p(\lambda_k)| \,. \tag{30}$$

In this sense, the worst-case convergence behavior is well understood. Still, for a given eigenvalue distribution the min-max value is often not known, and has to be estimated. Such estimation is of course always necessary, when only a partial information about the spectrum is known. A general approach tries to find inclusion sets for (the estimate of) the spectrum, and uses (close to) optimal polynomials on these sets to approximate the min-max value. However, this approach solves a different kind of approximation problem and can provide misleading information about the convergence.

For nonnormal matrices, the situation is even less clear. To bound the worst-case GMRES residual norm, one can use the ideal GMRES approximation

$$\min_{p\in\pi_n} \|p(A)\| \,, \tag{31}$$

that represents a matrix approximation problem. Although the value (31) need not describe GMRES worst-case behavior, it can be considered as a good approximation of the worst-case approximation in many practical cases. A general approach for approximating this value consists in finding a set in the complex plain associated with the matrix A and bounding the value (31) by the min-max approximation on this set. However, theoretical results in this field are still unsatisfactory.

Finally, it is important to note that the convergence can depend strongly on the right hand side or the initial guess so that the values (30) and (31) can overestimate the actual convergence of a Krylov subspace method.

References

[1] M. Arioli, *A stopping criterion for the conjugate gradient algorithms in a finite element method framework*, Numer. Math., 97 (2004), pp. 1–24.
[2] M. Arioli, E. Noulard, and A. Russo, *Stopping criteria for iterative methods: applications to PDE's*, Calcolo, 38 (2001), pp. 97–112.
[3] M. Arioli, V. Pták, and Z. Strakoš, *Krylov sequences of maximal length and convergence of GMRES*, BIT, 38 (1998), pp. 636–643.
[4] S. F. Ashby, T. A. Manteuffel, and P. E. Saylor, *A taxonomy for conjugate gradient methods*, SIAM J. Numer. Anal., 27 (1990), pp. 1542–1568.
[5] O. Axelsson, *Iterative solution methods*, Cambridge University Press, Cambridge, 1994.
[6] O. Axelsson and I. Kaporin, *On the sublinear and superlinear rate of convergence of conjugate gradient methods*, Numer. Algorithms, 25 (2000), pp. 1–22. Mathematical journey through analysis, matrix theory and scientific computation (Kent, OH, 1999).

[7] B. BECKERMANN AND A. B. J. KUIJLAARS, *Superlinear convergence of conjugate gradients*, SIAM J. Numer. Anal., 39 (2001), pp. 300–329 (electronic).
[8] ———, *Superlinear CG convergence for special right-hand sides*, Electron. Trans. Numer. Anal., 14 (2002), pp. 1–19 (electronic). Orthogonal polynomials, approximation theory, and harmonic analysis (Inzel, 2000).
[9] C. G. BROYDEN, *A new taxonomy of conjugate gradient methods*, Comput. Math. Appl., 31 (1996), pp. 7–17. Selected topics in numerical methods (Miskolc, 1994).
[10] B. A. CIPRA, *The Best of the 20th Century: Editors Name Top 10 Algorithms*, SIAM News, 33 (2000).
[11] P. CONCUS, G. H. GOLUB, AND D. P. O'LEARY, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, in Sparse matrix computations (Proc. Sympos., Argonne Nat. Lab., Lemont, Ill., 1975), Academic Press, New York, 1976, pp. 309–332.
[12] J. CULLUM AND A. GREENBAUM, *Relations between Galerkin and norm-minimizing iterative methods for solving linear systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 223–247.
[13] M. EIERMANN, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.
[14] M. EIERMANN AND O. G. ERNST, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[15] S. C. EISENSTAT, H. C. ELMAN, AND M. H. SCHULTZ, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.
[16] H. C. ELMAN, *Iterative methods for large sparse nonsymmetric systems of linear equations.*, PhD thesis, Yale University, New Haven, 1982.
[17] M. EMBREE, *How descriptive are GMRES convergence bounds?*, Numerical Analysis Report 99/08, Oxford University Computing Laboratory, 1999.
[18] O. G. ERNST, *Residual-minimizing Krylov subspace methods for stabilized discretizations of convection-diffusion equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1079–1101 (electronic).
[19] V. FABER, A. GREENBAUM, AND D. E. MARSHALL, *The polynomial numerical hulls of Jordan blocks and related matrices*, Linear Algebra Appl., 374 (2003), pp. 231–246.
[20] V. FABER, W. JOUBERT, E. KNILL, AND T. MANTEUFFEL, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729.
[21] D. K. FADDEEV AND V. N. FADDEEVA, *Computational methods of linear algebra*, W. H. Freeman and Co., San Francisco, 1963.
[22] B. FISCHER, *Polynomial based iteration methods for symmetric linear systems*, Wiley-Teubner Series Advances in Numerical Mathematics, John Wiley & Sons Ltd., Chichester, 1996.
[23] B. FISCHER AND R. FREUND, *Chebyshev polynomials are not always optimal*, J. Approx. Theory, 65 (1991), pp. 261–272.
[24] B. FISCHER, A. RAMAGE, D. J. SILVESTER, AND A. J. WATHEN, *On parameter choice and iterative convergence for stabilised discretisations of advection-diffusion problems*, Computer methods in applied mechanics and engineering, 179 (1999), pp. 179–195.
[25] F. R. GANTMACHER, *The theory of matrices. Vols. 1, 2*, Chelsea Publishing Co., New York, 1959.
[26] A. GREENBAUM, *Comparison of splittings used with the conjugate gradient algorithm*, Numer. Math., 33 (1979), pp. 181–193.
[27] A. GREENBAUM, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[28] ———, *Generalizations of the field of values useful in the study of polynomial functions of a matrix*, Linear Algebra Appl., 347 (2002), pp. 233–249.
[29] ———, *Card shuffling and the polynomial numerical hull of degree k*, SIAM J. Sci. Comput., 25 (2003), pp. 408–416 (electronic).
[30] ———, *Some theoretical results derived from polynomial numerical hulls of Jordan blocks.*, Electron. Trans. Numer. Anal., 18 (2004), pp. 81–90 (electronic).

[31] A. GREENBAUM AND L. GURVITS, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).
[32] 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.
[33] A. GREENBAUM AND Z. STRAKOŠ, *Matrices that generate the same Krylov residual spaces*, in Recent advances in iterative methods, vol. 60 of IMA Vol. Math. Appl., Springer, New York, 1994, pp. 95–118.
[34] A. GREENBAUM AND L. N. TREFETHEN, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).
[35] W. HACKBUSCH, *Iterative solution of large sparse systems of equations*, vol. 95 of Applied Mathematical Sciences, Springer-Verlag, New York, 1994. Translated and revised from the 1991 German original.
[36] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).
[37] M. HOCHBRUCK AND C. LUBICH, *Error analysis of Krylov methods in a nutshell*, SIAM J. Sci. Comput., 19 (1998), pp. 695–701 (electronic).
[38] A. S. HOUSEHOLDER, *The theory of matrices in numerical analysis*, Dover Publications Inc., New York, 1975. Reprint of 1964 edition.
[39] M. HUHTANEN AND O. NEVANLINNA, *Minimal decompositions and iterative methods*, Numer. Math., 86 (2000), pp. 257–281.
[40] W. JOUBERT, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).
[41] I. E. KAPORIN, *New convergence results and preconditioning strategies for the conjugate gradient method*, Numer. Linear Algebra Appl., 1 (1994), pp. 179–210.
[42] A. KLAWONN AND G. STARKE, *Block triangular preconditioners for nonsymmetric saddle point problems: field-of-values analysis*, Numer. Math., 81 (1999), pp. 577–594.
[43] T. KOCH AND J. LIESEN, *The conformal "bratwurst" maps and associated Faber polynomials*, Numer. Math., 86 (2000), pp. 173–191.
[44] A. N. KRYLOV, *On the numerical solution of the equation by which the frequency of small oscillations is determined in technical problems*, Izv. Akad. Nauk SSSR Ser. Fiz.-Mat., 4 (1931), pp. 491–539. (Title translation as given in [25]).
[45] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
[46] ———, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.
[47] J. LIESEN, *Construction and analysis of polynomial iterative methods for non-hermitian systems of linear equations*, PhD thesis, Fakultät für Mathematik, Universität Bielefeld, 1998. http://archiv.ub.uni-bielefeld.de/disshabi/mathe.htm.
[48] J. LIESEN, *Computable convergence bounds for GMRES*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 882–903 (electronic).
[49] J. LIESEN AND Z. STRAKOŠ, *GMRES convergence analysis for a convection-diffusion model problem*, SIAM J. Sci. Comput., (to appear).
[50] J. LIESEN AND P. TICHÝ, *Behavior of CG and MINRES for symmetric tridiagonal Toeplitz matrices*, Preprint 34-2004, Institute of Mathematics, Technical University of Berlin, 2004.
[51] J. LIESEN AND P. TICHÝ, *The worst-case GMRES for normal matrices*, BIT Numerical Mathematics, 44 (2004), pp. 79–98.
[52] I. MORET, *A note on the superlinear convergence of GMRES*, SIAM J. Numer. Anal., 34 (1997), pp. 513–516.

[53] N. M. NACHTIGAL, S. C. REDDY, AND L. N. TREFETHEN, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795. Iterative methods in numerical linear algebra (Copper Mountain, CO, 1990).

[54] A. E. NAIMAN, I. M. BABUŠKA, AND H. C. ELMAN, *A note on conjugate gradient convergence*, Numer. Math., 76 (1997), pp. 209–230.

[55] O. NEVANLINNA, *Convergence of iterations for linear equations*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1993.

[56] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[57] A. QUARTERONI AND A. VALLI, *Numerical approximation of partial differential equations*, vol. 23 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.

[58] J. K. REID, *On the method of conjugate gradients for the solution of large sparse systems of linear equations*, in Large sparse sets of linear equations (Proc. Conf., St. Catherine's Coll., Oxford, 1970), Academic Press, London, 1971, pp. 231–254.

[59] Y. SAAD, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp., 37 (1981), pp. 105–126.

[60] ———, *The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems*, SIAM J. Numer. Anal., 19 (1982), pp. 485–506.

[61] Y. SAAD, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.

[62] 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.

[63] G. STARKE, *Field-of-values analysis of preconditioned iterative methods for nonsymmetric elliptic problems*, Numer. Math., 78 (1997), pp. 103–117.

[64] G. STARKE AND R. S. VARGA, *A hybrid Arnoldi-Faber iterative method for nonsymmetric systems of linear equations*, Numer. Math., 64 (1993), pp. 213–240.

[65] Z. STRAKOŠ AND J. LIESEN, *On numerical stability in large scale linear algebraic computations*, Z. Angew. Math. Mech., (submitted).

[66] P. TICHÝ AND J. LIESEN, *Worst-case and ideal GMRES for a Jordan block*, submitted to Linear Algebra Appl., (2004).

[67] K.-C. TOH, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.

[68] L. N. TREFETHEN, *Approximation theory and numerical linear algebra*, in Algorithms for approximation, II (Shrivenham, 1988), Chapman and Hall, London, 1990, pp. 336–360.

[69] A. VAN DER SLUIS AND H. A. VAN DER VORST, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–560.

[70] H. A. VAN DER VORST, *Iterative Krylov methods for large linear systems*, vol. 13 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2003.

[71] R. WINTHER, *Some superlinear convergence results for the conjugate gradient method*, SIAM J. Numer. Anal., 17 (1980), pp. 14–17.