Chapter 1
On the Superlinear Convergence of MINRES*

Valeria Simoncini** and Daniel B. Szyld***

Abstract Quantitative bounds are presented for the superlinear convergence of the MINRES method of Paige and Saunders [*SIAM J. Numer. Anal.*, 1975] for the solution of sparse linear systems $Ax = b$, with A symmetric and indefinite. It is shown that the superlinear convergence is observed as soon as the harmonic Ritz values approximate well the eigenvalues of A that are either closest to zero or farthest from zero. This generalizes a well-known corresponding result obtained by van der Sluis and van der Vorst with respect to the Conjugate Gradients method, for A symmetric and positive definite.

1.1 Introduction

The MINRES method is a short-term recurrence Krylov subspace method developed by Paige and Saunders [8] for the solution of large and sparse linear systems of equations of the form

$$Ax = b, \tag{1.1}$$

where the $n \times n$ matrix A is symmetric and indefinite. MINRES is in fact very popular for solving *indefinite* linear systems, and it has become the leading solver for symmetric saddle point linear systems, for which spectral

* This version 11 January 2012

** Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy; and also CIRSA, Ravenna, Italy (`valeria.simoncini@unibo.it`).

*** Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, Pennsylvania 19122-6094, USA (`szyld@temple.edu`). This research is supported in part by the U.S. National Science Foundation under grant DMS-1115520 and by the U.S. Department of Energy under grant DE-FG02-05ER25672.

information can often be obtained from the application problem; see, e.g., [2] and references therein. It is well-known that MINRES exhibits superlinear convergence, i.e., that the norm of the residuals decreases linearly at first, but then, as the iterations progress the linear rate accelerates (cf. Figure 1.2). The motivation of this paper is to explain this observed phenomenon. We show that the superlinear convergence behavior (i.e., the change of the linear rate) occurs when the harmonic Ritz values approximate well the eigenvalues of the matrix A that are closest to or farthest away from the origin. This is consistent with the exposition in [7, §7], and with the comments found in [16, p. 78]. We are interested in describing a quantitative bound explaining more precisely these observations.

After a brief description of the algorithm in Section 1.2, we collect different results on the convergence of MINRES available in the literature (Section 1.3) and, inspired by ideas from other contexts, we develop a quantitative bound for its superlinear convergence (Section 1.4).

Throughout the paper exact arithmetic is assumed.

1.2 Review and preliminaries

We review here some concepts which we use throughout the paper. Given a first approximation x_0 to the solution of (1.1), and the corresponding initial residual $r_0 = b - Ax_0$, the Krylov subspace of dimension m defined by A and r_0 is given by

$$\mathcal{K}_m = \mathcal{K}_m(A, r_0) = \mathrm{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}. \tag{1.2}$$

An orthonormal basis $\{v_1, \ldots, v_m\}$ of $\mathcal{K}_m$ can be built by means of the Lanczos method. Let $V_m = [v_1, \ldots, v_m]$ collect these vectors, and observe that the matrix $T_m = V_m^T A V_m$ is symmetric and tridiagonal; the latter property is a consequence of the three-term recurrence from the Lanczos process. For details, see, e.g., [4], [8], [9], [13].

Like many other projection type approaches, at the mth step an approximation to the solution of (1.1), x_m, can be obtained in $x_0 + \mathcal{K}_m$, by imposing some additional condition. In MINRES, this approximation is found by requiring that the norm of the corresponding residual $r_m = b - Ax_m$ is minimized over all possible vectors of the form $x_m = x_0 + z$, with $z \in \mathcal{K}_m$; here and in the following we shall only consider the Euclidean norm, although the use of other norms has been analyzed in the literature; see, e.g., [10]. Thus, this approximation is of the form $x_m = x_0 + q_{m-1}(A) r_0$, where q_{m-1} is a polynomial of degree at most $m-1$. This implies that the residual $r_m = b - Ax_m$ is associated with the so-called *residual polynomial* $p_m(t)$ of degree at most m with $p_m(0) = 1$, since $r_m = b - Ax_m = r_0 - Aq_{m-1}(A) r_0 = p_m(A) r_0$. We recall two sets of scalars approximating the eigenvalues of the matrix A, as the iteration progresses: The *Ritz values* (with respect to $\mathcal{K}_m$), which are the eigenvalues of T_m, and the *harmonic Ritz values*, which instead are the roots of the residual polynomial $p_m(t)$, and are denoted by $\theta_1^{(m)}, \ldots, \theta_m^{(m)}$, i.e.,

$$p_m(t) = \frac{(\theta_1^{(m)} - t) \cdots (\theta_m^{(m)} - t)}{\theta_1^{(m)} \cdots \theta_m^{(m)}} .$$

The harmonic Ritz values can be equivalently characterized as the Ritz values of A^{-1} with respect to $A\mathcal{K}_m$; see, e.g., [5]. From a computational view point, the harmonic Ritz values can be obtained as the eigenvalues of the pencil $(\underline{T_m}^T \underline{T_m}, T_m)$, where $\underline{T_m} = V_{m+1}^T A V_m$; see [7] and references therein. As a special feature, we also notice that harmonic Ritz values approximate the eigenvalues from the interior of the spectral intervals of A. Therefore, any interval around the origin that is free of eigenvalues of A is also free of harmonic Ritz values [7]. This ensures that the approximation, say, to the smallest positive eigenvalues is genuine, and it is not incidental, since no harmonic Ritz value will cross the origin to approximate the negative eigenvalues as m increases, the way Ritz values would do, on indefinite matrices. We also mention that harmonic Ritz values may play an important role in practical circumstances, such as the approximation of interior eigenvalues, see, e.g., [5], and for devising problem-dependent stopping criteria [11].

1.3 Known bounds for the residual norm

Let $\Lambda(A) = \{\lambda_1, \ldots, \lambda_n\}$ be the set of eigenvalues of A, with the eigenvalues ordered increasingly, and let $\mathcal{P}_m$ be the set of all polynomials p of degree at most m such that $p(0) = 1$.

From $r_m = p_m(A)r_0$, we have the following standard bound

$$\|r_m\| = \|p_m(A)r_0\| \leq \min_{p \in \mathcal{P}_m} \max_{i=1,\ldots,n} |p(\lambda_i)| \, \|r_0\|. \quad (1.3)$$

Therefore, it is useful to find appropriate bounds for

$$E_m(\Lambda(A)) = \min_{p \in \mathcal{P}_m} \max_{\lambda \in \Lambda(A)} |p(\lambda)|,$$

and these will depend of course on the form of the set of eigenvalues $\Lambda(A)$. One such bound was developed for the case where $\Lambda(A) \subset [a, b] \cup [c, d]$, where $a < b < 0 < c < d$, under the constraint that $|b - a| = |d - c|$, that is, the two intervals have equal length. In this case, using an appropriate transformation of the intervals and bounds on Chebychev polynomials, the following bound holds:

$$\frac{\|r_m\|}{\|r_0\|} \le 2 \left(\frac{\sqrt{|ad|} - \sqrt{|bc|}}{\sqrt{|ad|} + \sqrt{|bc|}} \right)^{[m/2]},$$

where $[m/2]$ is the integer part of $m/2$; see [3, Ch. 3], or [4, §3.1], for details.

Bounds for the asymptotic convergence factor $\lim_{m\to\infty} e_m^{\frac{1}{m}}$ with $e_m = E_m([a,b] \cup [c,d])$, were proposed in [18], where the role of $\sqrt{bc/ad}$ was also emphasized.

For the special case where the number of negative (or positive) eigenvalues is relatively small, say k, we can use the technique in [14, Theo. 4.4] to provide a more descriptive bound as follows.

Proposition 1. *Let $\Lambda(A) \subset \{\lambda_1, \ldots, \lambda_k\} \cup [c,d]$, with $\lambda_1, \ldots, \lambda_k$ negative, and $0 < c \le d$. Then, for $m > k$,*

$$\frac{\|r_m\|}{\|r_0\|} \le \Omega_k \frac{2}{\rho^{k-m} + \rho^{m-k}},$$

where $\rho = \dfrac{\sqrt{\tilde\kappa}+1}{\sqrt{\tilde\kappa}-1}$, $\tilde\kappa = \frac{d}{c}$, and $\Omega_k = \prod_{j=1}^{k} \left(1 - \dfrac{d}{\lambda_j}\right)$ is independent of m.

Beckermann and Kuijlaars [1] developed bounds for the quantities $E_m(S)$ for very specific sets S containing the spectrum of positive definite matrices A. These bounds were useful to follow the superlinear convergence of Conjugate Gradients (CG). For description of Conjugate Gradients, or other Krylov subspace methods, see, e.g., [4], [9], [13]. Beckermann and Kuijlaars further indicated that the general results they proved would be applicable to MINRES as well, but for this one needs to build the appropriate sets S containing $\Lambda(A)$ now having negative and positive elements. Calculating these sets "is a problem in itself," and this was not developed in [1].

We note in passing that the *a posteriori* convergence bounds developed in [12] can also apply to MINRES. They are based on how close invariant subspaces of A are to the Krylov subspace; in the present context, this reduces to the angle between eigenvectors and the Krylov subspace.

1.4 A new a-posteriori bound

As opposed to most a-priori estimates recalled in the previous section, here we describe a new *a-posteriori* bound that aims to describe the possibly abrupt steepness change in the linear convergence rate that is often encountered when using MINRES. Detecting and understanding this behavior may help devise an improved method, or an improved preconditioner, that allow the method to immediately enter the superlinear convergence stage without the initial slower phase; see, e.g., [6].

We show that after a sufficient number of iterations have been performed, the method behaves as if the eigencomponents corresponding to the smallest eigenvalues (in modulo) had been removed. Since the (worst case) rate of convergence depends on the spectral interval, the method behaves as if the matrix had a reduced spectral interval, hence improving its convergence rate. This phenomenon is well known for CG, and it was completely uncovered by van der Sluis and van der Vorst in their 1986 paper [15]. We essentially take their proof for CG, which uses Ritz values, and obtain a similar result for MINRES using harmonic Ritz values. We use the same polynomial for the bound, which is also used in [17, Lemma 1.5] for the nonsymmetric case. We should also add that Van der Vorst in [16, p. 78] already mentions the possibility of developing this bound in this form. Here we present it in detail.

Let (λ_k, z_k), $k = 1, \ldots, n$ be the eigenpairs of A, with λ_k, $k = 1, \ldots, n$ sorted in increasing absolute value, and assume that λ_1 is simple.

Theorem 1. *Let r_m be the residual after m MINRES iterations with starting residual r_0, so that in particular $r_m = b - Ax_m$ with $x_m \in x_0 + \mathcal{K}_m(A, r_0)$. Let us write $r_m = \bar{r}_0 + s^{(1)}$, with $\bar{r}_0 \perp z_1$, and let $\bar{r}_j$ be the MINRES residual after j iterations in $\mathcal{K}_j(A, \bar{r}_0)$. Then after $m + j$ MINRES iterations with starting residual r_0 we obtain*

$$\|r_{m+j}\| \leq F_m \|\bar{r}_j\|, \quad \text{where} \quad F_m = \max_{k \geq 2} \frac{|\theta_1^{(m)}|}{|\lambda_1|} \frac{|\lambda_1 - \lambda_k|}{|\theta_1^{(m)} - \lambda_k|}$$

and $\theta_1^{(m)}$ is the harmonic Ritz value closest to λ_1 in $\mathcal{K}_m(A, r_0)$.

Proof. Let $p_m, \bar{q}_j$ be the MINRES residual polynomials in $\mathcal{K}_m(A, r_0)$ and $\mathcal{K}_j(A, \bar{r}_0)$, respectively. We write $r_0 = \sum_{k=1}^{n} \gamma_k z_k$ so that

$$r_m = p_m(A) r_0 = \sum_{k=1}^{n} p_m(\lambda_k) \gamma_k z_k, \quad \bar{r}_0 = \sum_{k=2}^{n} p_m(\lambda_k) \gamma_k z_k.$$

Moreover, $\bar{r}_j = \bar{q}_j(A) \bar{r}_0 = \sum_{k=2}^{n} \bar{q}_j(\lambda_k) p_m(\lambda_k) \gamma_k z_k$. Let

$$\phi_m(\lambda) = \frac{\theta_1^{(m)}}{\lambda_1} \frac{\lambda_1 - \lambda}{\theta_1^{(m)} - \lambda} p_m(\lambda),$$

and notice that $\phi_m(\lambda_1) = 0$.

Since the MINRES polynomial is a minimizing polynomial, we obtain

$$\begin{aligned}\|r_{m+j}\|^2 = \|p_{m+j}(A) r_0\|^2 \leq \|\phi_m(A) \bar{q}_j(A) r_0\|^2 &= \sum_{k=2}^{n} \phi_m(\lambda_k)^2 \bar{q}_j(\lambda_k)^2 \gamma_k^2 \\ &\leq F_m^2 \sum_{k=2}^{n} p_m(\lambda_k)^2 \bar{q}_j(\lambda_k)^2 \gamma_k^2 = F_m^2 \|\bar{r}_j\|^2. \; \square\end{aligned}$$

The bound for $\|r_{m+j}\|$ shows that the residual norm can be bounded by the norm of the residual deflated of the eigenvector component corresponding to λ_1. If one of the harmonic Ritz values is a good approximation to λ_1, then the factor F_m will be very close to one. Therefore, in this case the behavior of the residual norm $\|r_{m+j}\|$ is well represented by that of $\bar{r}_j$, which has no eigencomponent onto z_1.

The result can be easily generalized to a group of eigenvalues, the only technical change would be the use of more orthogonality conditions to define $\bar{r}_0$. Nowhere in the proof we used the fact that λ_1 is the eigenvalue closest to the origin. In fact, the proof holds for any simple eigenvalue of A, and in particular for those farthest from the origin.

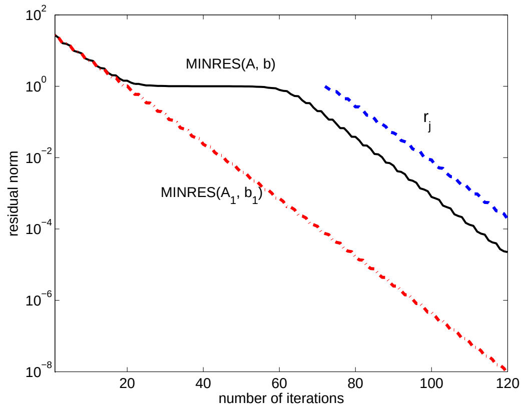

Fig. 1.1 Example 1. Convergence history of MINRES on $Ax = b$ and $A_1x_1 = b_1$.

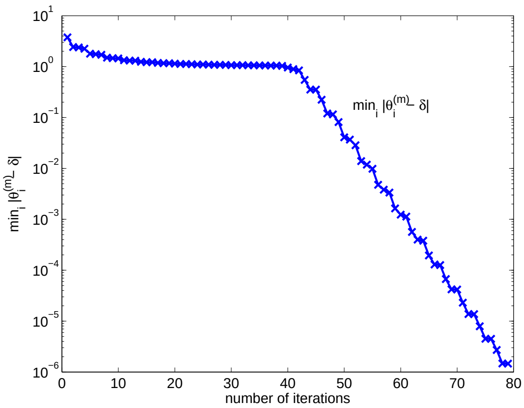

Fig. 1.2 Example 1. Convergence history of the harmonic Ritz value closest to δ.

Example 1. We consider the following data:

$$A = \begin{bmatrix} A_- & & \\ & \delta & \\ & & A_+ \end{bmatrix}, \qquad b = \mathbf{1},$$

where $\delta = -10^{-3}$, and A_+, A_- are diagonal matrices with values logarithmically distributed in $[10^0, 10^{0.5}]$ and $[-10^1, -10^0]$, respectively. The dimension of A is $n = 2 \cdot 399 + 1 = 799$.

The convergence history of MINRES on $Ax = b$ shows a long plateau, with an almost complete stagnation (cf. Figure 1.1), corresponding to the effort the method is making in approximating the interior eigenvalue δ, once it discovers there is one. This fact can be clearly observed in Figure 1.2, where the values $\min_i |\theta_i^{(m)} - \delta|$ are reported, where $\theta_i^{(m)}$ $i = 1, \ldots, m$ are the harmonic Ritz values at the mth iteration.

Let r_{70} be the residual of MINRES on $Ax = b$ after 70 iterations. The dashed curve in Figure 1.1 reports the convergence history of a MINRES process started with r_{70} as initial residual. Its convergence rate matches quite well that of the original MINRES after the smallest eigenvalue is singled out. For the sake of completeness, in Figure 1.1 we also report the convergence history of MINRES applied to the companion problem $A_1 x_1 = b_1$ where the row and column corresponding to δ are removed. The plot shows that the convergence delay is only due to the isolated small eigenvalue.

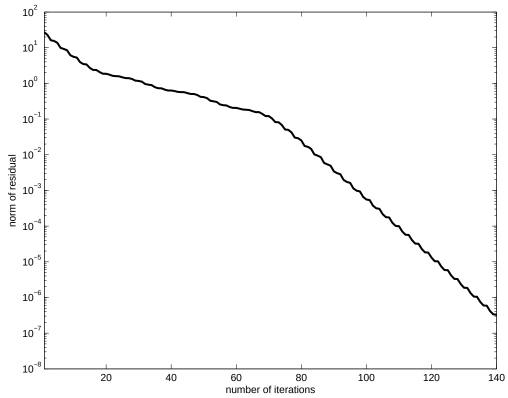

Fig. 1.3 Example 2. Convergence history of MINRES on $Ax = b$.

Example 2. We next consider a spectral distribution that is possibly more common in practice, and in which the picture of superlinear convergence rate is more typical. We consider a variant of the previous example, where now $\delta = \mathrm{diag}(-10^{-1}, -3 \cdot 10^{-1}, -2 \cdot 10^{-1})$, so that the matrix A has size $n = 801$;

the right-hand side is $b = \mathbf{1}$, as in Example 1. The small negative eigenvalues are now less isolated, and their approximation during the MINRES process is more effective (cf. Figure 1.3). Nonetheless, as soon as the Krylov space captures the small eigenvalues - after about 70 iterations - the MINRES convergence rate changes, showing superlinear convergence.

References

1. Bernhard Beckermann and Arno B.J. Kuijlaars. Superlinear convergence of Conjugate Gradients. *SIAM Journal on Numerical Analysis*, 39:300–329, 2001.
2. Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.
3. Bernd Fischer. *Polynomial Based Iteration Methods for Symmetric Linear Systems*. Willey and Teubner, Chichester, West Essex, England, and Stuttgart, 1996.
4. Anne Greenbaum. *Iterative Methods for Solving Linear Systems*, volume 17 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1997.
5. Ronald B. Morgan. Computing interior eigenvalues of large matrices. *Linear Algebra and its Applications*, pages 289–309, 1991.
6. Maxim A. Olshanskii and Valeria Simoncini. Acquired clustering properties and solution of certain saddle point systems. *SIAM Journal on Matrix Analysis and Applications*, 31:2754–2768, 2010.
7. Christopher C. Paige, Beresford N. Parlett, and Henk A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2:115–134, 1995.
8. Christopher C. Paige and Michael A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12:617–629, 1975.
9. Yousef Saad. *Iterative Methods for Sparse Linear Systems*. The PWS Publishing Company, Boston, 1996. Second edition, SIAM, Philadelphia, 2003.
10. Marcus Sarkis and Daniel B. Szyld. Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. *Computer Methods in Applied Mechanics and Engineering*, 196:1612–1621, 2007.
11. David J. Silvester and Valeria Simoncini. An optimal iterative solver for symmetric indefinite systems stemming from mixed approximation. *ACM Transactions on Mathematical Software*, 37:42:1–42:22, 2011.
12. Valeria Simoncini and Daniel B. Szyld. On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods. *SIAM Review*, 47:247–272, 2005.
13. Valeria Simoncini and Daniel B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14:1–59, 2007.
14. Daniel B. Szyld and Olof B. Widlund. Variational analysis of some Conjugate Gradient methods. *East-West Journal of Numerical Mathematics*, 1:51–74, 1993.
15. Abraham van der Sluis and Henk A. van der Vorst. The rate of convergence of Conjugate Gradients. *Numerische Mathematik*, 48:543–560, 1986.
16. Henk A. van der Vorst. *Iterative Krylov Methods for Large Linear Systems*. Cambridge University Press, Cambridge, New York, and Melbourne, 2003.
17. Henk A. van der Vorst and C. (Kees) Vuik. The superlinear convergence behaviour of GMRES. *Journal of Computational and Applied Mathematics*, 48:327–341, 1993.
18. Andrew J. Wathen, Bernd Fischer, and David J. Silvester. The convergence rate of the minimal residual method for the Stokes problem. *Numerische Mathematik*, 71:121–134, 1995.