Applied Mathematics and Computation 144 (2003) 441–455
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

Analysis of peaks and plateaus in a Galerkin/minimal residual pair of methods for solving $Ax = b$ ☆

Yuming Shen [a,b], Jinxi Zhao [a,*]

[a] *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, PR China*

[b] *Department of Mathematics, Nanjing University, Nanjing 210093, PR China*

Abstract

Irregular peaks often appear if we use Galerkin methods for solving linear systems of equations $Ax = b$. These peaks bring about too difficult to identify convergence. To remedy this disadvantage, we have to spend more work and memory, that is we use norm minimizing methods for solving $Ax = b$. However, plateaus cannot be avoided. In this paper we give a sufficient and necessary condition for occurring of peaks. Also we present some related factors for this behavior.

1. Introduction and definitions

For solving linear systems of equations

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

there are two classes of iterative methods commonly used. One is Galerkin methods such as Lanczos [10], BCG [5] and FOM [11]. The other is norm minimizing methods such as MINRES [12], QMR [6] and GMRES [13].

☆ Supported by the State 863-plan High Science and Technology of China.

* Corresponding author. Address: Department of Computer Science and Technology, 210008 Nanjing, China.

E-mail address: jxzhao@nju.edu.cn (J. Zhao).

0096-3003/02/$ - see front matter
doi:10.1016/S0096-3003(02)00419-8

The convergence of any iterative method is said to have occurred at iteration k if for some specified convergence tolerance ε,

$$\|r_k\|/\|r_0\| \leqslant \varepsilon \tag{1.2}$$

where r_0 is the initial residual, $r_k = -Ax_k + b$, and x_k is the kth iteration. Without any loss of generality, we will assume that A is real and nonsingular. The initial guess $x_0 = 0$ so that $r_0 = b$. In this paper we focus on a pair of Lanczos/MINRES methods for solving Eq. (1.1) when A is an $n \times n$ symmetric matrix. See [2] for results of the pairs GMRES/FOM and QMR/BCG and for details of theorems and proofs are not include in this paper. It is shown that using Galerkin methods for solving linear system (1.1) with A either real symmetric or nonsymmetric the residual norms, $\|r_k\|$, $k = 1, 2, \ldots,$ are not always monotonically decreasing as a function of the iteration number. Irregular peaks can appear in such curves, making it difficult to identify convergence, and making the user feel insecure about using the method. See for example Fig. 1.

One way of copying with this problem is to use norm minimizing methods. Since the Krylov subspaces generated are nested, the residual norm $\|r_k\|$ must be a monotone decreasing function of the iteration number k. However, these

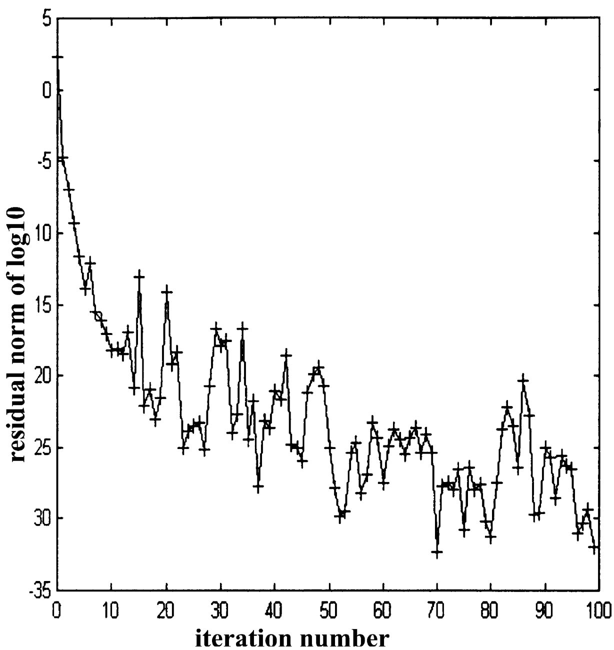

Fig. 1. The convergence curves of Lanczos.

methods have been devised that require a great deal of work and memory. Moreover, plateaus can appear in such plots, intervals of iterations over which the norm of the residual decrease at an unacceptably slow rate of change. See for example Fig. 2.

In this paper we examine, both experimentally and theoretically, peak and plateau formation generated by the Lanczos/MINRES. In Section 2 we present relationships between peaks and plateaus. In Section 3 we identify some factors which initiate peak formations in the Lanczos residual norm plots. In Section 4 we give some numerical experiments to examine our conclusions.

The norm $\|.\|$ is the Euclidean two norm, or spectral norm. The r_k^{LR} and r_k^{MR} denote the Lanczos residual and MINRES residual, respectively.

The Krylov subspace is defined by

$$K_k :\equiv K_k(A; r_0) \equiv \mathrm{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}, \quad k \geqslant 1$$

and the corresponding Krylov matrix is

$$K_k \equiv (r_0, Ar_0, \ldots, A^{k-1}r_0), \quad k \geqslant 1$$

Throughout the paper we refer to peaks and plateaus in residual norm plots as follows.

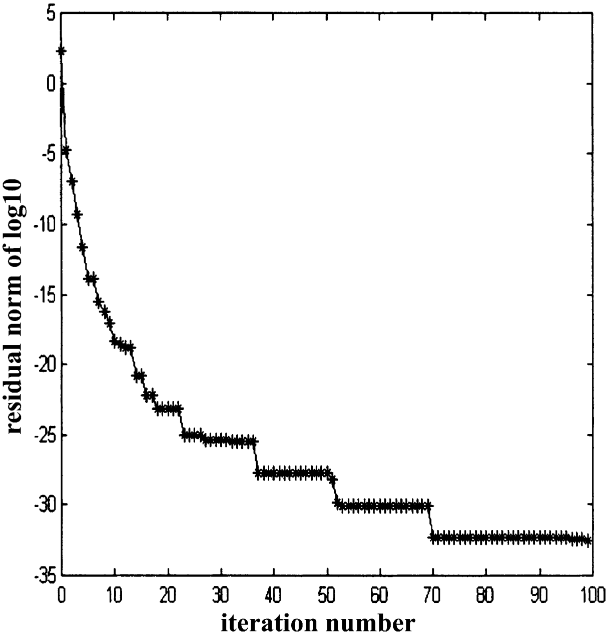

Fig. 2. The convergence curves of MINRES.

Definition 1.1 (*Cullum*, 1995 [1]). A peak is any consecutive section of a residual norm plot during which the residual norms increase to a local maximum and then decrease to a local minimum.

Definition 1.2 (*Cullum*, 1995 [1]). A plateau is any consecutive section of a residual norm plot during which the norm of the residual decrease at an unacceptably slow rate of change.

2. Peaks, plateaus and angles between subspaces

Thanks to J. Cullum and A. Greenbaum, in [1,2] they indicate a correlation between peaks and plateaus. Whenever a peak occurs there is a plateau under it. The converse however may not be true. It is possible for a plateau to occur in a MINRES residual norm plot without a visible corresponding peak in the corresponding Lanczos residual norm plot. They also indicate that whenever the residual norm plot for the MINRES is decreasing rapidly the corresponding residual norm plot for the Lanczos iterates is also decreasing rapidly. The corresponding residual norm plots appear to track each other. In this section we consider the same problem in another way. We recall that in many MINRES implementations a least squares problems is solved in each iteration by reducing a Hessenberg matrix to upper triangular form via Givens rotation [11]. At iteration k a Givens rotation,

$$G_k = \begin{pmatrix} 1 & & & & & & & 0 \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & c_k & s_k & & & \\ & & & -s_k & c_k & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ 0 & & & & & & & 1 \end{pmatrix} \tag{2.1}$$

is generated to eliminate the trailing element of the Hessenberg matrix. Notice that these s_k and c_k are not merely artifacts of the computational scheme but are the sines and cosines of the angles AK_k and K_k.

The following relations are fundamental for our later investigations (see [4] Section 2).

Theorem 2.1 (Eiermann and Ernst, 2001 [4]). *For* $k = 1, 2, \ldots, L-1$, *there holds*

$$s_k = \sin\angle(K_k, AK_k) \quad \textit{and} \quad c_k = \cos\angle(K_k, AK_k) \tag{2.2}$$

where $\angle(K_k, AK_k)$ *denotes the largest canonical angle between the spaces* K_k *and* AK_k.[1] *The quantities* c_k *and* s_k *are given by the Givens rotations* G_k *of* (2.1) $L := \min\{k : x_k^{\mathrm{MR}} = A^{-1}b\} = \min\{k : x_k^{\mathrm{LR}} = A^{-1}b\}$.

Theorem 2.2 (Eiermann and Ernst, 2001 [4]). *With* $s_k = \sin\angle(K_k, AK_k)$ *and* $c_k = \cos\angle(K_k, AK_k)$ *the MINRES residual and Lanczos residual approximations with respect to the Krylov subspaces* K_k *satisfy*

$$\|r_k^{\mathrm{MR}}\| = c_k \|r_k^{\mathrm{LR}}\| \tag{2.3}$$

$$\|r_k^{\mathrm{MR}}\| = s_k \|r_{k-1}^{\mathrm{MR}}\| = s_1 s_2 \dots s_k \|r_0\| \tag{2.4}$$

$$r_k^{\mathrm{MR}} = s_k^2 r_{k-1}^{\mathrm{MR}} + c_k^2 r_k^{\mathrm{LR}} \tag{2.5}$$

In view of $s_k = \|r_k^{\mathrm{MR}}\| / \|r_{k-1}^{\mathrm{MR}}\|$, i.e., $c_k^2 = 1 - \|r_k^{\mathrm{MR}}\|^2 / \|r_{k-1}^{\mathrm{MR}}\|^2$ we obtain the following theorem.

Theorem 2.3 (Eiermann and Ernst, 2001 [4]). *In exact arithmetic, if* $c_k \neq 0$ *at each iteration* k, *then the Lanczos residuals and MINRES residuals are related by*

$$\|r_k^{\mathrm{LR}}\| = \|r_k^{\mathrm{MR}}\| / \sqrt{1 - \|r_k^{\mathrm{MR}}\|^2 / \|r_{k-1}^{\mathrm{MR}}\|^2} \tag{2.6}$$

Using the expression of $\sin\angle(K_k, AK_k) = \|r_k^{\mathrm{MR}}\| / \|r_{k-1}^{\mathrm{MR}}\|$, we can rewrite (2.6) as follows:

$$\|r_k^{\mathrm{LR}}\| = \|r_k^{\mathrm{MR}}\| / \sqrt{1 - (\sin\angle(K_k, AK_k))^2} \tag{2.7}$$

(2.7) shows that if $\angle(K_k, AK_k)$ is reduced by a significant factor at step k, then the Lanczos residual norm will be approximately equal to the MINRES residual norm at step k, since the denominator in the right-hand side of (2.7) will be close to 1. If the $\angle(K_k, AK_k)$ remains almost $\pi/2$, however, then the denominator in the right-hand side of (2.7) is close to 0 and the Lanczos residual norm will be much larger.

As shown before the behavior of the angles $\angle(K_k, AK_k)$ as k approaches ∞ play a crucial role in the convergence of the MR and LR approximates. If the angles actually tend to zeros rapidly, in view of (2.4), implies superlinear convergence. Based on the Theorem 2.3 we can prove the following two propositions.

[1] Given orthogonal bases $\{v_j\}_{j=1}^m$ and $\{w_j\}_{j=1}^m$ of two m-dimensional subspaces V and W, then the cosines of the canonical angles between V and W are the singular values of the matrix of inner products $[(v_j, w_k)] \in R^{m\times m}$. We remark that the sine of the largest canonical angle between the spaces V and W of equal dimension is given by $\|(I - P_v)P_w\|$ [14, Theorem 4.37].

Proposition 2.1. *If, under the assumptions of Theorem* 2.3, *there exist iterations* $K1 \leqslant k \leqslant K2$, $0 < \theta_k < \pi/2$, *such that* $\angle(K_k, AK_k) \leqslant \theta_k$, *then for* $K1 \leqslant k \leqslant K2$

$$\|r_k^{\mathrm{MR}}\| \leqslant \|r_k^{\mathrm{LR}}\| \leqslant \|r_k^{\mathrm{MR}}\|/\sqrt{1-(\sin\theta_k)^2} \tag{2.8}$$

Proof. Since $\angle(K_k, AK_k) \leqslant \theta_k$ and $0 < \theta_k < \pi/2$ then

$$\sin\angle(K_k, AK_k) \leqslant \sin\theta_k < 1$$

From Theorems 2.1 and 2.2

$$s_k = \sin\angle(K_k, AK_k) \leqslant \sin\theta_k < 1$$

Notice that

$$\|r_k^{\mathrm{MR}}\|/\|r_{k-1}^{\mathrm{MR}}\| = s_k$$

From Theorem 2.3 we get

$$\|r_k^{\mathrm{MR}}\| \leqslant \|r_k^{\mathrm{LR}}\| \leqslant \|r_k^{\mathrm{MR}}\|/\sqrt{1-(\sin\theta_k)^2}$$

Proposition 2.1 shows that if for a given interval of iterations the angles $\angle(K_k, AK_k) \leqslant \theta_k$, $0 < \theta_k < \pi/2$ then during those iterations the corresponding Lanczos residual norm plot is trapped between small multiples of the MINRES curve. □

Proposition 2.2. *If under the assumptions of Theorem* 2.3, *there exist iterations* $K1 \leqslant k \leqslant K2, \gamma > 1$ *with* $\|r_k^{\mathrm{LR}}\| \geqslant \gamma\|r_{k-1}^{\mathrm{LR}}\|$, *then if there exists* $0 < \theta < \pi/2$ *such that* $\angle(K_k, AK_k) < \theta$ *then*

$$\theta > \arctan\gamma \tag{2.9}$$

Proof. From Theorem 2.3 and the inequality on $\|r_k^{\mathrm{LR}}\|$ we have that

$$\begin{aligned}\|r_k^{\mathrm{LR}}\| &= \|r_k^{\mathrm{MR}}\|/\sqrt{1-\|r_k^{\mathrm{MR}}\|^2/\|r_{k-1}^{\mathrm{MR}}\|^2} \geqslant \gamma\|r_{k-1}^{\mathrm{LR}}\| \\ &= \gamma\left(\|r_{k-1}^{\mathrm{MR}}\|/\sqrt{1-\|r_{k-1}^{\mathrm{MR}}\|^2/\|r_{k-2}^{\mathrm{MR}}\|^2}\right)\end{aligned}$$

Since

$$\|r_k^{\mathrm{MR}}\|/\|r_{k-1}^{\mathrm{MR}}\| = \sin\angle(K_k, AK_k) < \sin\theta$$

then

$$\sin\theta/\sqrt{1-(\sin\theta)^2} > \gamma, \quad 0 < \theta < \pi/2, \;\; \gamma > 1$$

i.e.,

$$\theta > \arctan \gamma$$

Proposition 2.2 shows that if over some interval of iterations residual norms generated by Lanczos are increasing at least as a specified rate γ, then the angle $\angle(K_k, AK_k)$ cannot decrease at a rate faster than the bound on θ given in (2.9), i.e., the corresponding MINRES residual norms cannot decrease at a rate faster than the bound on $\sin\theta$. For example, if $\gamma > 2$ then $\theta > \arctan 2 \approx 63.4349$, $\sin\theta \approx 0.89442$. □

3. Peaks and some related factors

The reason why the residual norm is not always minimized is more interesting, for it touches a deeper issue and is a topic of current concern and imperfect understanding [3]. In this section we consider four factors which are related to the Lanczos peaks.

I. *Numerical Instabilities*. What role do numerical instabilities play in the generation of the peak formations observed in the Lanczos residual norm plots? In [1], we know that if the linear system (1.1) is sufficiently well conditioned [1, Definition 3.3], then numerical instabilities have no role in any observed peak formations.

II. *Finite precision arithmetic*. Are the peaks and plateaus artifacts of the finite precision arithmetic? See [1], we know that peaks and plateaus are not artifacts of the finite precision arithmetic. Peaks and plateaus can also occur when the arithmetic is exact. However, more peaks or plateaus will occur in finite precision arithmetic than would occur if the computations were exact. Moreover, the effect of finite precision arithmetic is an open problem.

III. *Angle between subspaces*. Based on properties of angle between subspaces and use the relationship between the orthogonal residual norm and the minimal residual norm, we obtain a sufficient and necessary condition for occurring of peaks.

Theorem 3.1. *In the exact arithmetic, for* $c_k \neq 0, k = 1, 2, \ldots, L-1$ *the condition*

$$\frac{1}{F_k^2} + \frac{F_{k-1}^2}{\beta^2} < \frac{1}{\beta^2} + 1, \quad \beta \geqslant 1 \tag{3.1}$$

is satisfied if and only if

$$\frac{\|r_k^{\mathrm{LR}}\|}{\|r_{k-1}^{\mathrm{LR}}\|} > \beta \tag{3.2}$$

where

$$F_k = \sin\angle(K_k, AK_k)$$

Proof. Making use of the relation between the Lanczos residual norm and MINRES residual norm $\|r_k^{\mathrm{MR}}\| = c_k\|r_k^{\mathrm{LR}}\|$ we obtain

$$\frac{\|r_k^{\mathrm{LR}}\|}{\|r_{k-1}^{\mathrm{LR}}\|} = \frac{c_{k-1}}{c_k}\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} = \frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|}\frac{\sqrt{1-s_{k-1}^2}}{\sqrt{1-s_k^2}}$$

Notice that

$$s_k = \frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} = \sin\angle(K_k, AK_k) = F_k$$

then

$$\frac{\|r_k^{\mathrm{LR}}\|}{\|r_{k-1}^{LR}\|} = \left(\frac{1-F_{k-1}^2}{\frac{1}{F_k^2}-1}\right)^{1/2}$$

If there exists $\beta \geqslant 1$ such that the condition (3.1) is satisfied then

$$\frac{1}{F_k^2} - 1 < \frac{1-F_{k-1}^2}{\beta^2}$$

That is

$$\frac{\|r_k^{\mathrm{LR}}\|}{\|r_{k-1}^{\mathrm{LR}}\|} > \beta$$

If the Lanczos residual norm increases, i.e.,

$$\frac{\|r_k^{\mathrm{LR}}\|}{\|r_{k-1}^{\mathrm{LR}}\|} = \left(\frac{1-F_{k-1}^2}{\frac{1}{F_k^2}-1}\right)^{1/2} > \beta, \quad \beta \geqslant 1$$

which implies

$$1 - F_{k-1}^2 > \beta^2\left(\frac{1}{F_k^2} - 1\right) \quad \text{i.e.,} \quad \frac{1}{F_k^2} + \frac{F_{k-1}^2}{\beta^2} < \frac{1}{\beta^2} + 1$$

Theorem 3.1 shows that if the sines of $\angle(K_k, AK_k)$ satisfy the condition (3.1) then the Lanczos residual norm increases. It also explains why a plateau to occur without a visible corresponding peak, since the condition (3.1) is not satisfied. □

Corollary 3.1. *If the condition* (3.1) *is satisfied and* $c_k \neq 0, \quad k = 1, 2, \ldots, L-1$ *then*

$$\frac{\sqrt{2}}{2} < F_k < 1 \quad \textit{and} \quad F_k > F_{k-1},$$

i.e.,

$$\pi/4 < \angle(K_k, AK_k) < \pi/2 \quad \textit{and} \quad \angle(K_k, AK_k) > \angle(K_{k-1}, AK_{k-1})$$

Proof. Since the condition (3.1) is satisfied then

$$F_k^2 > \frac{1}{1 + \frac{1-F_{k-1}^2}{\beta^2}}, \quad \beta \geqslant 1$$

Because $c_k \neq 0$ we get

$$0 < F_{k-1} < 1 \quad \text{and} \quad 0 < F_k < 1$$

which implies

$$\frac{\sqrt{2}}{2} < F_k < 1$$

Suppose that

$$F_k \leqslant F_{k-1}$$

then

$$\frac{1}{F_k^2} + F_{k-1}^2 \geqslant \frac{1}{F_{k-1}^2} + F_{k-1}^2 > 2 \quad \text{contradiction}$$

This means during iterations, which the Lanczos residual norms are increasing, the angle between the present subspaces is larger than the angle between the prior subspaces and these angle $\angle(K_k, AK_k) \in (\pi/4, \pi/2)$. These are two necessary conditions for occurring of peaks. For the special case $\beta = 1$, we get a sufficient condition for the Lanczos residual norm increases. □

Corollary 3.2. *If* $F_k^2 \geqslant \frac{1}{2}(F_{k-1}^2 + 1)$ *and* $c_k \neq 0,\ k = 1, 2, \ldots L - 1$ *then*

$$\frac{1}{F_k^2} + F_{k-1}^2 < 2$$

i.e., the Lanczos residual norm increases.

Proof. Since $F_k^2 \geqslant \frac{1}{2}(F_{k-1}^2 + 1)$ then

$$\frac{1}{F_k^2} + F_{k-1}^2 \leqslant \frac{2}{(F_{k-1}^2 + 1)} + F_{k-1}^2 = \frac{(2 + F_{k-1}^4 + F_{k-1}^2)}{(F_{k-1}^2 + 1)} = 1 + \frac{(F_{k-1}^4 + 1)}{(F_{k-1}^2 + 1)}$$

Since $c_k \neq 0$ we get $0 < F_{k-1} < 1$ and $F_{k-1}^2 > F_{k-1}^4$ which implies

$$\frac{1}{F_k^2} + F_{k-1}^2 < 2$$

Theorem 3.1 and Corollary 3.1 clearly indicate that the angle between subspaces plays a crucial role in the Lanczos residual norm plot increases.

However, the condition (3.1) is an abstract inequality, it can not clearly examine different eigenvalue distributions. From [7, Theorem 4.1] we can see that if A is normal and $1 \leqslant k \leqslant L-1$ then

$$\min_{z \in K_k} \frac{\|b - Az\|}{\|b\|} = m_{k+1} \min_{1 \leqslant j \leqslant k+1} \left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{k+1} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\} \tag{3.3}$$

where β_j is the norm of the orthogonal projection of b onto the eigenspace associated with λ_j, $1/\sqrt{k+1} \leqslant m_{k+1} \leqslant \sqrt{(k+1)(L-k)}$ and $\lambda_1, \ldots, \lambda_{k+1}$ are $k+1$ distinct eigenvalues of A that maximize $\prod_{j=1}^{k+1} \beta_j \prod_{l=j+1}^{k+1} |\lambda_l - \lambda_j|$. Hence we can obtain

$$F_k = \frac{m_{k+1} \min_{1 \leqslant j \leqslant k+1} \left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{k+1} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\}}{m_k \min_{1 \leqslant j \leqslant k} \left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{k} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\}}$$

Suppose there exists index s such that: minimum both

$$\left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{k+1} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\} \quad \text{and} \quad \left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{k} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\}$$

then

$$F_k = \frac{m_{k+1}}{m_k} \frac{|\lambda_{k+1} - \lambda_s|}{|\lambda_{k+1}|}.$$

For this particular case the following examples illustrate how to interpret the condition (3.1) in Theorem 3.1 for different eigenvalue distributions. We cite [7, Example 4.1], matrix A has one cluster of eigenvalues centered at a point c in the complex plane with radius $\epsilon > 0$, and a single outlier $c + \delta$. Then $|\delta|$ is the absolute distance between cluster and outlier. We make three assumptions: first the absolute separation between cluster and outlier is much larger than the absolute cluster radius, $|\delta| \gg \epsilon$; second, the relative cluster radius is small, $\epsilon/|c| < 1$; and third, the outlier is farther away from zero than the cluster, $|c + \delta| \geqslant |c|$. Then one can show [8, Section 5.2] that in iteration k,

$$\min_{z \in K_k} \frac{\|b - Az\|}{\|b\|} \approx \left| \frac{\delta}{c + \delta} \right| \left(\frac{\epsilon}{|c|} \right)^{k-1}$$

If we examine the condition (3.1) in Theorem 3.1, we obtain $F_k \approx \epsilon/|c|$. Since $\epsilon/|c| < 1$ then we get $(1/F_k^2) + F_{k-1}^2 > 2$. This suggests in the exact arithmetic we can say there is no peaks in the Lanczos residual norm plot for this particular case. The result is same for [7] Example 4.2. Here the matrix A has one cluster of eigenvalues centered at c and a second cluster at $c + \delta$. The two clusters have the same number of eigenvalues and the same absolute cluster radius $\epsilon > 0$. The absolute cluster separation is $|\delta|$. $\square$

IV. *The convergence of Ritz value*. Also in [1], J. Cullum presented empirical evidence that the formation of peaks in the Lanczos residual norm plot correlated with the stabilization of the Ritz values [11]. From this he inferred that peaks formations correspond to the identification or re-identification of certain portions of the solution space, as indicated by the Ritz values convergence to the eigenvalues of A.

In his numerical experiments, he observed that the convergence of each Ritz value approximation appears to initiate the formation of a peak. The subsequent pictorial convergence of such an approximation appears to correspond to the down side of such a peak. However, this is not always true. In our numerical experiment, we observe such a phenomenon: when the Ritz values converge to the eigenvalues of A, there is no peaks in the Lanczos residual norm plot. (our numerical experiment is in Section 4). How to explain the two different phenomena?

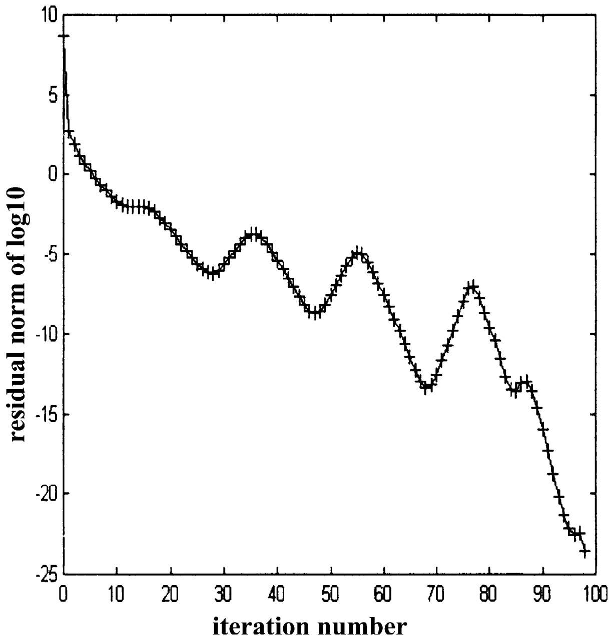

Fig. 3. The convergence curves of Lanczos.

In my opinion, the Ritz value closes to a right or wrong eigenvalue is an important factor. When the Ritz value is close to a wrong eigenvalue in passing, it appears to initiate the formation of a peak. Moreover, which eigenvalues of A are approximated by the Ritz values? [9] contain related results.

4. Numerical experiments

In this Section we give some numerical experiments for solving linear systems $Ax = b$ to examine our conclusions. We specified a diagonal matrix Σ by specifying a few small eigenvalues and spacing the remaining eigenvalues equidistantly situated within a specified interval. In each example the convergence tolerance was $\epsilon = 10^{-12}$.

Table 1
The numerical results of Example 1

k	F_k	$1/F_k^2 + F_{k-1}^2$
28	0.8308	2.0084
29	0.9084	1.9022
30	0.9582	1.9144
31	0.9827	1.9537
32	0.9930	1.9797
33	0.9971	1.9919
34	0.9986	1.9970
35	0.9991	1.9990
86	0.5962	3.0126
87	0.8971	1.5981
88	0.9614	1.8867
47	0.7449	2.2229
48	0.8681	1.8819
49	0.9501	1.8615
50	0.9842	1.9350
51	0.9954	1.9780
52	0.9987	1.9934
53	0.9996	1.9981
54	0.9999	1.9995
55	0.9999	1.9999
68	0.6064	2.9473
69	0.8317	1.8134
70	0.9608	1.7750
71	0.9931	1.9371
72	0.9989	1.9884
73	0.9998	1.9981
74	1.0000	1.9997
75	1.0000	2.0000
76	1.0000	2.0000
77	1.0000	2.0000

Example 1. We denote the 100×100 matrix as $A(100, 100)$, which has four small eigenvalues $\lambda_1 = 10^{-6}$, $\lambda_2 = 10^{-4}$, $\lambda_3 = 10^{-3}$, $\lambda_4 = 10^{-1}$ and $\lambda_k = 0.1 + (k-1) \times d$, $d = 0.1 (5 \leqslant k \leqslant 100)$. Let the right-side vector $b = \mathrm{diag}(A)$. The result and the graph of convergence are shown in Fig. 3 and Table 1.

In Fig. 3 we observe that over iterations 28–35, 47–55, 68–77 and 86–88 the Lanczos residual norms are increasing. For special case $\beta = 1$, F_k values are listed in Table 1.

Remark. From Table 1 we also observe over iterations 28–35, 48–55, 69–73 the values of F_k are larger than $\sqrt{2}/2$ and $F_k > F_{k-1}$. However, there is a problem which cannot be ignored. From iterations 74–77 the Lanczos residual norms are increasing, but the values of F_k are equal to 1, i.e., $c_k = 0(\angle(K_k, AK_k) = \pi/2)$.

Example 2. Denote the matrix by $A(100, 100)$, its eigenvalues $\lambda_k = 0.1 + (k-1)0.1$, $k = 1, 2, \ldots, 100$. Let the right-side vector $b = \mathrm{diag}(A)$. The graph of convergence is shown in Fig. 4.

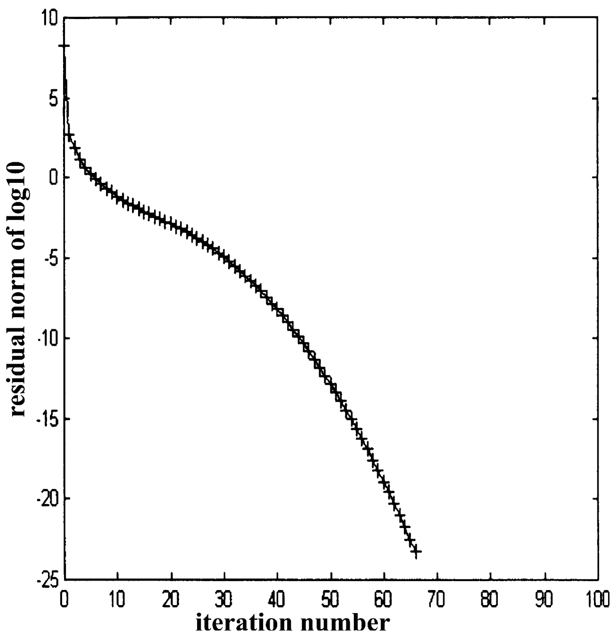

Fig. 4. The convergence curves of Lanczos.

In Fig. 4 the residual norm plot is smooth and we cannot see any peaks in the Lanczos residual norm plot. We further observe in Fig. 4 at iteration 36, $\theta_1^{(36)} = 0.1001$, i.e. the first Ritz value is close to the right eigenvalue λ_1. Consider Fig. 3 for Example 1, peaks occur at iterations 35, 55, 77, and 87. At iteration 24, $\theta_1^{(24)} = 0.10002278$, i.e., the first Ritz value is close to the wrong eigenvalue λ_4. At iteration 40, $\theta_1^{(40)} = 0.0010$, i.e., the first Ritz value is close to another wrong eigenvalue λ_3. At iteration 62, the first Ritz value is close to λ_2, $\theta_1^{(62)} = 0.0001$. At iteration 88, $\theta_1^{(88)} = 0.000001$, now the first Ritz value finds its way to the right eigenvalue λ_1. Figs. 3 and 4 indicate that there may be a certain relationship between the formation of peaks and the convergence of Ritz values. However, the qualitative analysis of this relation we do not know. Perhaps, this is still an open problem.

5. Summary

The theorems and experiments described in the preceding sections provide some insight into the behavior of residual norm plots for both Lanczos and MINRES. Using the relationship between the orthogonal residual norm and the minimal residual norm, we provide a plausible explanation for the erratic behavior of typical residual norm plots generated by Lanczos. And we also observe that this erratic behavior is due to a Ritz value in passing closes to a wrong eigenvalue.

References

[1] J. Cullum, Peaks, plateaus, numerical instabilities in a Galerkin/minimal residual pair of methods for solving $Ax = b$, Appl. Numer. Math. 19 (1995) 255–278.
[2] J. Cullum, A. Greenbaum, Residual relationships within three pairs of iterative algorithms for solving $Ax = b$, SIAM J. Matrix Anal. Appl. 17 (1996) 223–247.
[3] T.A. Driscoll, K.C. Toh, L.N. Trefethen, From potential theory to matrix iterations in six steps, SIAM Rev. 40 (1998) 547–578.
[4] M. Eiermann, O.G. Ernst, Geometric aspects in the theory of Krylov subspace methods, Acta Numer. 10 (2001) 251–312.
[5] R. Fletcher, Conjugate gradient methods for indefinite systems, in: G.A. Watson (Ed.), Proceedings of the Dundee Biennial Conference on Numerical Analysis, Spring-Verlag, Berlin, 1975.
[6] R. Freund, N. Nachtigal, QMR: a quasi-minimal residual method for non-Hermitian linear systems, Numer. Math. 60 (1991) 315–339.
[7] I.C.F. Ipsen, Expressions and bounds for the GMRES residual, BIT 40 (3) (2000) 524–535.
[8] I.C.F. Ipsen, A different approach to bounding the minimal residual norm in the Krylov methods, Tech. Rep. CRSC-TR98-19, Center for Research in Scientific Computation, Department of Mathematics, North Carolina State University, 1998. http://www4.ncsu.edu/~ipsen/linsys.html.

[9] A.B.J. KuijLaars, Which eigenvalues are found by the Lanczos method? SIAM J. Matrix Anal. Appl. 22 (1) (2000) 306–321.
[10] C. Lanczos, Solution of systems of linear equations by minimized iterations, Res. Nat. Bur. Standards 49 (1952) 33–53.
[11] G. Meurant, Computer Solution of Large Linear Systems, first ed., Elesvier Science B.V., 1999.
[12] C.C. Paige, M.A. Saunders, Solution of sparse indefinite systems of linear equation, SIAM J. Numer. Anal. 12 (1975) 617–629.
[13] Y. Saad, M.H. Schultz, GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear system, SIAM J. Sci. Stat. Comput. 7 (1986) 856–869.
[14] G.W. Stewart, Matrix Algorithms, vol. I, Basic Decompositions, SIAM philadelphia, PA, 1998.