ELSEVIER

Applied Numerical Mathematics 19 (1995) 255-278

APPLIED NUMERICAL MATHEMATICS

Peaks, plateaus, numerical instabilities in a Galerkin minimal residual pair of methods for solving $Ax = b$ [☆]

Jane K. Cullum [*]

Mathematical Sciences Department, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Abstract

Using a Galerkin/minimal residual pair of bidiagonalization methods for solving nonsymmetric systems of equations $Ax = b$, we examine the behavior of residual norm plots generated by these methods and present explanations for this behavior.

1. Introduction

The convergence of any iterative method for solving a linear system of equations

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

is normally monitored by tracking the sizes of the associated residual norms, $\|r_k\|/\|r_0\|$ [6], where r_0 is the initial residual, $r_k = -Ax_k + b$, and x_k is the kth iterate. Convergence is said to have occurred at iteration k if for some prespecified convergence tolerance ε,

$$\|r_k\|/\|r_0\| \leqslant \varepsilon. \tag{2}$$

In practice residual norms may not be computed on every iteration and the convergence test may be applied to estimates of these norms which are computed at every iteration. In the discussions and without any loss of generality, we will assume that A is real and nonsingular and that $x_0 = 0$ so that $r_0 = b$.

In this paper we focus on a pair of bidiagonalization methods for solving equation (1) when A is a nonsymmetric $n \times n$ matrix, [8,14,15]. We denote these methods by BLanczos and BMinres. See

[☆] Part of this work was supported by NSF grant GER-9450081 while the author was visiting the Department of Computer Science, University of Maryland, College Park, MD.

[*] Telephone: 914-945-2227. Fax: 914-945-3434. E-mail: cullumj@watson.ibm.com.

SSDI 0168-9274(95)00086-0

[3] for results for the pairs GMRES/FOM [16] and QMR/BCG [7], and for details of theorems and proofs not included in this paper. Each of these pairs of methods utilizes recursions to generate simultaneously bases for one or more families of nested Krylov subspaces and corresponding iteration matrices which represent orthogonal or bi-orthogonal projections of A or of a matrix associated with A onto those subspaces. BMinres and BLanczos are based upon the real symmetric Lanczos recursions. See Section 2. The BLanczos iteration matrices are real, symmetric, and tridiagonal. GMRES and FOM are based upon the Arnoldi recursions which generate orthogonal bases for Krylov subspaces corresponding to A and b. The FOM iteration matrices are Hessenberg matrices which represent orthogonal projections of A onto those subspaces. QMR (quasi-minimal residuals) and BCG (biconjugate gradients) use nonsymmetric Lanczos recursions which, in exact arithmetic, generate bi-orthogonal bases for Krylov subspaces corresponding to A and to A^{T}. The BCG iteration matrices are tridiagonal and, in exact arithmetic, represent bi-orthogonal projections of A onto these Krylov subspaces. Details of implementations of the GMRES/FOM methods and of the QMR/BCG methods can be found in [16,7]. Both members of each pair of methods select their iterates from the same subspaces.

BLanczos, FOM, and BCG are Galerkin methods. At each iteration of any of these methods and in exact arithmetic, the iterate x_k is chosen such that its residual is orthogonal to a certain subspace. Using the recursion for each method which generates the iteration matrices, it can be shown that computing the Galerkin iterates is equivalent to solving systems of equations involving these matrices. BMinres, GMRES, and QMR are norm minimizing methods. At each iteration of BMinres and GMRES and in exact arithmetic, x_k is chosen such that its residual has minimal norm in a certain subspace. At each iteration of QMR, x_k is chosen such that the norm of a related quasi-residual vector is minimized. Using the recursion for each method which generates the iteration matrices, it can be shown that computing the minimal or quasi-minimal residual iterates is equivalent to solving full rank least squares problems involving simple modifications of these iteration matrices.

If residual norms, $\|r_k\|$, $k = 1,2,\ldots$, generated by the application of a Galerkin method to equation (1) with A either real symmetric or nonsymmetric are plotted, typically the resulting curve is not 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. In addition, a Galerkin iteration matrix may be indefinite, ill-conditioned or even singular, in which case the Galerkin iterate is not well defined.

Minimal and quasi-minimal residual variants of the Galerkin methods were proposed to eliminate both of these problems. Each minimal and quasi-minimal residual iterate is obtained by solving a full rank least squares problem, so that these iterates are always well defined. Since the Krylov subspaces generated are nested, the $\|r_k\|$ generated by any minimal residual method must be a monotone decreasing function of the iteration number k. In a quasi-minimal residual method such as QMR the norms of the quasi-residual vectors are a monotone decreasing function of the iteration number. Minimal and quasi-minimal residual norm plots however, often exhibit a different type of irregular behavior. Plateaus can appear in such plots, intervals of iterations over which there exist unacceptably small improvements in the residual norm. See for example Fig. 2.

In this paper we examine, both experimentally and theoretically, peak and plateau formations generated by the bidiagonalization pair BLanczos/BMinres. There are several reasons why we

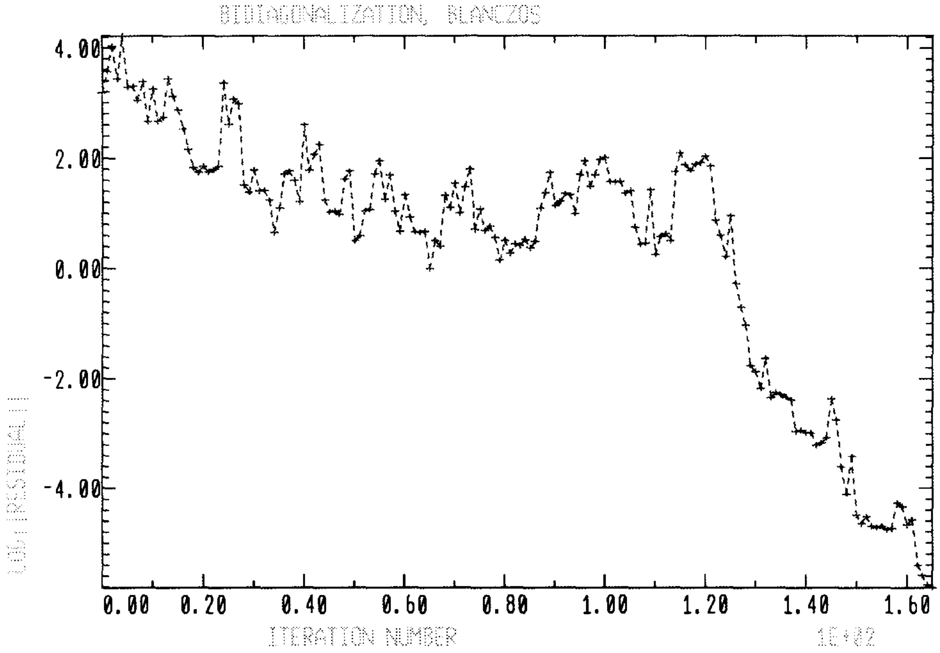

Fig. 1. Typical Galerkin plot, $\log_{10}(\|r_k\|)$ versus iteration number k.

consider these methods in detail. They are based upon the well-defined, real symmetric Lanczos recursions, and the iteration matrices are real, symmetric, and tridiagonal. The effects of finite precision arithmetic on these recursions are understood and can be used to derive theorems about the behavior of these methods in finite precision arithmetic. Moreover, in exact arithmetic, see Theorems 5.1 and 5.3, it is easy to demonstrate that the set of all positive definite diagonal matrices generates all of the possible residual norm plots which can be generated for arbitrary A and b in equation (1), whereas suitable sets of test matrices for the other pairs of methods GMRES/FOM and QMR/BCG are not known. Finally, we expected and [3] provides partial confirmation that the residual norm behavior observed with the simple bidiagonalization methods has related but more complicated analogs in the more popular pairs of methods GMRES/FOM and QMR/BCG.

We consider the following four questions, We provide answers to each question. For Questions 1, 2, and 3 we incorporate the effects of finite precision arithmetic.

Question 1. What role do numerical instabilities play in the generation of the peak formations observed in residual norm plots generated by the Galerkin method BLanczos?

Question 2. Are these peaks and plateaus artifacts of the finite precision arithmetic?

Question 3. For a specified problem $Ax = b$, are there correlations between the residual norms generated by the Galerkin method BLanczos and the residual norms generated by the minimal residual method BMinres?

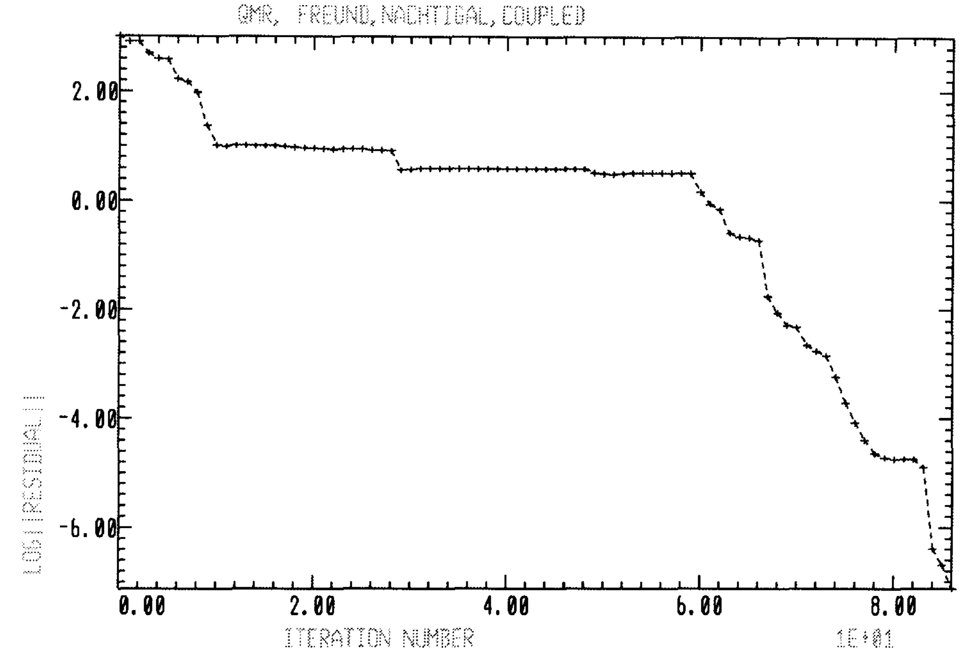

Fig. 2. Typical minimal residual plot, $\log_{10}(\|r_k\|)$ versus iteration number k.

Question 4. Can we identify any other factors which initiate peak and plateau formations in these residual norm plots?

Answers to each question are provided in Sections 3, 4, 6, and 7 of this paper. In Section 2 we briefly describe the bidiagonalization methods. In Section 3 we consider Question 1. Theorem 3.1 indicates that the BLanczos (Galerkin) iteration matrices are essentially as well-conditioned as the matrix A in equation (1). Therefore, if A is sufficiently well-conditioned and the bidiagonalization methods are implemented properly, numerical instabilities do not play a role in the peak and the plateau generation.

In Section 4 we consider Question 2. Theorem 4.1 states that any significant residual norm behavior which can be observed in finite precision arithmetic can also be observed in exact arithmetic but typically on a different problem. Therefore, peaks and plateaus occur even in exact arithmetic and are not merely artifacts of the finite precision arithmetic. We note, however, that for a given problem the number of peaks or plateaus generated when the precision is finite is typically greater than the number that would be generated in exact arithmetic. See Section 7. In Section 5 we consider the question of suitable test problems.

In Section 6 we consider Question 3. We first present representative results from numerical experiments on a sequence of well-chosen test matrices. Residual norm plots generated by BLanczos and BMinres indicate correlations between peak formations in a BLanczos plot and plateaus formations in the corresponding BMinres plot. They also indicate correlations between portions of the curves corresponding to intervals of iterations during which the residual norms generated by BMinres are decreasing rapidly. See Figs. 3 and 4. These experiments led to Theorem 6.1 [3] which exhibits a direct correlation between residual norms generated by BLanczos and the corresponding residual norms generated by BMinres. We then use Theorem 6.1 to obtain specific relationships between peaks and plateaus, and between the two residual norm plots when the plot for BMinres is decreasing rapidly. We then infer that if the minimum residual variant BMinres converges well, then the Galerkin variant BLanczos must also converge well.

In Section 7 we consider Question 4. Theorem 3.1 demonstrates that numerical instabilities do not play a role in peak and plateau formation if A is sufficiently well-conditioned. We attempt to find other causes for peak formations. We present representative results from a set of numerical experiments which demonstrate an apparent correlation between peak formations in residual norm plots generated by BLanczos and the appearance in the spectra of the corresponding iteration matrices of approximations to certain singular values of A. The numerical experiments indicate that low frequency peaks in such plots correspond to the initial appearance or the reappearance of approximations to small singular values of A and that high frequency peaks correspond to the initial appearance or reappearance of approximations to singular values of A which are readily approximated by eigenvalues of the iteration matrices. Typically, these are large singular values of A. High frequency peaks can make the behavior of the residual norms appear to be very erratic. See for example, Fig. 6 versus Fig. 5. In practice, a typical Galerkin residual norm plot is a composite of overlapping peaks and of strongly monotone decreasing sections.

The experiments in Section 7 together with the theorems in Section 6 indicate that peaks and plateaus probably correspond to intervals of iterations during which both BLanczos and BMinres are attempting to identify or to re-identify critical portions of the solution space and that until such an identification is achieved, significant improvements cannot be made in the residual norms in either method. We use the following notation and definitions.

1.1. Notation

- $A = (a_{ij})$, $1 \leqslant i, j \leqslant n$, $n \times n$ real matrix,
- $A^{\mathrm{T}} = (a_{ji})$, $1 \leqslant i, j \leqslant n$, transpose of A,
- $D = \operatorname{diag}\{d_1, \ldots, d_n\}$, $n \times n$ diagonal matrix,
- $\lambda_j(A)$, $1 \leqslant j \leqslant n$, eigenvalues of A,
- $\sigma_j(A)$, $1 \leqslant j \leqslant n$, singular values of A where $\sigma_1 \geqslant \cdots \geqslant \sigma_n$,
- $\Sigma = \operatorname{diag}\{\sigma_1, \ldots, \sigma_n\}$,
- $\mathcal{K}_j(b, A) = \operatorname{span}\{b, Ab, \ldots, A^{j-1}b\}$, jth Krylov subspace generated by A, b,
- $\kappa(A) = \sigma_{\max}(A)/\sigma_{\min}(A)$, condition number of A,
- $\|A\|_2 = \sigma_{\max}(A)$, $\|x\|_2 = \sqrt{\Sigma x_j^2}$,
- x_k, kth iterate of any iterative method, $X_k = \{x_1, \ldots, x_k\}$,
- $r_k = -Ax_k + b$, kth residual of any iterative method,
- r_k^{BM}, r_k^{BL}, kth residual in BMinres, BLanczos,
- R^m, m-dimensional Euclidean space,
- e_j, jth coordinate vector in R^m where m is specified in the context,
- $\hat{e}_j$, jth coordinate vector in R^{m+1} where m is specified in the context,
- I_j, $j \times j$ identity matrix.

1.2. Definitions

Throughout the paper we refer to *peaks* and *plateaus* in residual norm plots. It is difficult to assign precise definitions to these features because we do not have a characterization for the initiation and termination points of either of these. However, the following attempts at definitions may be useful to the reader. Typical residual norm plots are composites of peaks (plateaus) with strongly monotone sections and frequently it is impossible to distinquish visually individual peaks (plateaus) in these plots.

Definition 1.1. 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. A *plateau* is any consecutive section of a residual norm plot during which the norms of the residuals decrease at an unacceptably slow rate of change.

2. Bidiagonalization methods

We replace equation (1) by the following $2n \times 2n$, real symmetric but indefinite system of equations whose solution contains the desired solution.

$$B\bar{x} = \bar{b} \quad \text{where } B \equiv \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{pmatrix}, \quad \bar{x} \equiv \begin{pmatrix} y \\ x \end{pmatrix}, \quad \bar{b} \equiv \begin{pmatrix} b \\ 0 \end{pmatrix}. \tag{3}$$

See [15,14,8]. If A were complex, then we would use A^{H} instead of A^{T}. The discussion is easily extended to that case. Lemma 2.1 states that the eigenvalues of B are $\pm\sigma_j(A)$, and that the eigenvectors of B are simple concatenations of left and right singular vectors of A.

Lemma 2.1 ([4]). *Let A be any real nonsymmetric $n \times n$ matrix with singular value decomposition $A = X\Sigma Y^{\mathrm{T}}$ where $\Sigma = \operatorname{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ and $Y^{\mathrm{T}}Y = X^{\mathrm{T}}X = I$. Then*

$$BZ = Z \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \quad \textit{where } Z = \frac{1}{\sqrt{2}} \begin{pmatrix} X & X \\ Y & -Y \end{pmatrix}. \tag{4}$$

If we apply the real symmetric Lanczos recursions to B with starting vector $\|b\|w_1 = (b^{\mathrm{T}}, 0)^{\mathrm{T}}$, then we obtain the matrix recursion,

$$BW_k = W_kT_k + \beta_{k+1}w_{k+1}e_k^{\mathrm{T}} + E_k \tag{5}$$

where the real symmetric tridiagonal matrix

$$T_k = \begin{pmatrix} 0 & \beta_2 & & & \\ \beta_2 & 0 & \beta_3 & & \\ & \beta_3 & 0 & \ddots & \\ & & \ddots & \ddots & \beta_k \\ & & & \beta_k & 0 \end{pmatrix} \tag{6}$$

and E_k denotes the errors in the recursion resulting from the finite precision arithmetic [13]. The entries in T_k, or equivalently the coefficients in equation (5), are chosen so that theoretically, the $W_k \equiv \{w_1, \ldots, w_k\}$ are orthonormal and therefore $T_k = W_k^{\mathrm{T}} B W_k$ is an orthogonal projection of B onto the Krylov subspace $\mathcal{K}_k(w_1, B)$. For reasonable size k, $\|E_k\|$ is small [13,12]. For details see [4,13,12]. If $\beta_j \neq 0$, $2 \leqslant j \leqslant 2k$, then T_{2k} is nonsingular and its eigenvalues occur in ± pairs. Each T_{2k-1} is singular. See for example [5] for a proof. Therefore, BLanczos and BMinres will use only the even-ordered Lanczos matrices T_{2k}.

We define the BLanczos and BMinres iterates in terms of the iterates obtained by applying Galerkin and minimal residual real symmetric Lanczos procedures to equation (3) but using only the even-ordered tridiagonal matrices. We denote the resulting Galerkin and minimal residual iterates by $\bar{x}_k^{\mathrm{G}}$ and $\bar{x}_k^{\mathrm{M}}$ and the corresponding residuals by $\bar{r}_k^{\mathrm{G}}$ and $\bar{r}_k^{\mathrm{M}}$. We assume $\bar{x}_0 = 0$ and therefore, $\bar{r}_0 = (b^{\mathrm{T}}, 0)^{\mathrm{T}}$. Each Galerkin Lanczos iterate $\bar{x}_k^{\mathrm{G}}$ must satisfy the Galerkin conditions $\bar{x}_k^{\mathrm{G}} \in \mathcal{K}_{2k}(\bar{r}_0, B)$ and $\bar{r}_k^{\mathrm{G}} \perp \mathcal{K}_{2k}(\bar{r}_0, B)$. From equation (5) and the orthogonality of the W_{2k} vectors, we have that in exact arithmetic the kth Galerkin Lanczos iterate can be obtained by solving

$$T_{2k}\bar{z} = \|\bar{r}_0\| e_1 \text{ and forming } \bar{x}_k^{\mathrm{G}} = W_{2k}\bar{z}. \tag{7}$$

Similarly, we have that each minimal residual Lanczos iterate $\bar{x}_k^{\mathrm{M}}$, which must have minimal residual norm in $\mathcal{K}_{2k}(\bar{r}_0, B)$, can be obtained by solving the following least squares problem

$$\min \|\hat{T}_{2k}\bar{z} - \|\bar{r}_0\|\hat{e}_1\| \text{ with } \hat{T}_{2k} = \begin{pmatrix} T_{2k} \\ \beta_{2k+1} e_{2k}^{\mathrm{T}} \end{pmatrix} \text{ and forming } \bar{x}_k^{\mathrm{M}} = W_{2k}\bar{z}. \tag{8}$$

We define the corresponding BLanczos and BMinres iterates in Lemmas 2.2 and 2.3. We set $\beta_1 \equiv \|\bar{r}_0\| = \|r_0\| = \|b\|$.

Lemma 2.2 ([4]). *If we apply the Galerkin real symmetric Lanczos method to equation* (3) *and all* β_j, $1 \leqslant j \leqslant 2k$ *are nonzero, then* T_{2k} *is nonsingular and all of the odd-numbered components of* $\bar{z}$ *in equation* (7) *are zero. Furthermore, if we specify the kth BLanczos iterate* x_k^{BL} *by*

$$\bar{x}_k^{\mathrm{G}} = \begin{pmatrix} 0 \\ x_k^{\mathrm{BL}} \end{pmatrix}, \quad \textit{then } \bar{r}_k^{\mathrm{G}} = \begin{pmatrix} r_k^{\mathrm{BL}} \\ 0 \end{pmatrix} \textit{ and } x_k^{\mathrm{BL}} = V_k\bar{z}^* \tag{9}$$

where $*$ *denotes the even-numbered components of* $\bar{z}$, $V_k \equiv \{v_1, \ldots, v_k\}$, *and each* v_j *consists of the last n components of each* w_{2j}. *In addition, in exact arithmetic,*

$$x_k^{\mathrm{BL}} = x_{k-1}^{\mathrm{BL}} + \bar{z}(2k)v_k \quad \textit{and} \quad \|r_k^{\mathrm{BL}}\| = |\beta_{2k+1}\bar{z}(2k)| \tag{10}$$

$$\textit{where } \bar{z}(2k) = (-1)^{k+1}\prod_{j=1}^{k}\beta_{2j-1} / \prod_{j=1}^{k}\beta_{2j}.$$

Now consider the BMinres iterates, x_k^{BM}. The least squares problem in equation (8) can be solved by successively applying Givens transformations G_j to $\hat{T}_{2k}$ for $j = 1, 2k$ where

$$G_j = \begin{pmatrix} I_{j-1} & & & \\ & c_j & s_j & \\ & -s_j & c_j & \\ & & & I_{2k-j} \end{pmatrix} \quad \text{with } c_j^2 + s_j^2 = 1 \tag{11}$$

to obtain a rectangular triangular matrix. Specifically,

$$\hat{T}_{2k} = Q_{2k}\hat{R}_{2k} \quad \text{where } Q_{2k} = G_1^{\mathrm{T}} \cdots G_{2k}^{\mathrm{T}},\ Q_{2k}^{\mathrm{T}}Q_{2k} = I_{2k+1},\ \hat{R}_{2k} = \begin{pmatrix} \tilde{R}_{2k} \\ 0 \end{pmatrix}, \tag{12}$$

and $\tilde{R}_{2k}$ is $2k \times 2k$ and upper triangular. At each stage in this reduction c_j, s_j are chosen to map the $(j+1, j)$th element of the current modified matrix into zero. An induction argument demonstrates that for $j = 1, \ldots, k$,

$$G_{2j-1} = \begin{pmatrix} I_{2j-2} & & & \\ & 0 & 1 & \\ & -1 & 0 & \\ & & & I_{2k-2j+1} \end{pmatrix}, \qquad \tilde{R}_{2k} = \begin{bmatrix} \delta_1 & 0 & \theta_3 & & & 0 \\ & \delta_2 & 0 & \theta_4 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \theta_{2k} \\ & & & & & 0 \\ 0 & & & & & \delta_{2k} \end{bmatrix}. \tag{13}$$

Define

$$\bar{P}_{2k} = W_{2k}\tilde{R}_{2k}^{-1}, \quad \text{then } p_k = \left[v_k - \beta_{2k-1}\beta_{2k}\delta_{2k-2}^{-1}p_{k-1}\right]\delta_{2k}^{-1} \tag{14}$$

where p_k is the n-vector consisting of the $n+1$ to $2n$ components of the $(2k)$th column of $\bar{P}_{2k}$.

Lemma 2.3. *If we apply the minimal residual real symmetric Lanczos method to equation* (3), *then all of the odd-numbered components of $\tilde{z}$ in* (8) *are zero. Furthermore, if we specify the kth BMinres iterate x_k^{BM} by*

$$\bar{x}_k^{\mathrm{M}} = \begin{pmatrix} 0 \\ x_k^{\mathrm{BM}} \end{pmatrix}, \quad \textit{then } \bar{r}_k^{\mathrm{M}} = \begin{pmatrix} r_k^{\mathrm{BM}} \\ 0 \end{pmatrix} \textit{ and } x_k^{\mathrm{BM}} = P_k\tilde{z}^* \tag{15}$$

where $*$ denotes the even-numbered components of $\tilde{z}$, $P_k \equiv \{p_1, \ldots, p_k\}$, and p_j consists of the last n components of $\bar{p}_{2j}$. In addition, in exact arithmetic,

$$x_k^{\mathrm{BM}} = x_{k-1}^{\mathrm{BM}} + c_k\|r_{k-1}^{\mathrm{BM}}\|p_k \quad \textit{where } \|r_{k-1}^{\mathrm{BM}}\| = \min\|\hat{T}_{2k-2}\tilde{z} - \|\bar{r}_0\|e_1\| = |\prod_{j=1}^{k-1} s_j|\|\bar{r}_0\| \tag{16}$$

and c_j, s_j define the $(2j)$th Givens transformations which were used in the factorization of $\hat{T}_{2k}$.

The vector $\hat{T}_{2k-2}\tilde{z} - \|\bar{r}_0\|e_1$ corresponding to the minimum norm in equation (16) is called the $(k-1)$th quasi-minimal residual vector. Proofs of Lemmas 2.2 and 2.3 are given for example, in [2]. BMinres is a real symmetric version of the QMR method [7] applied to equation (3) and in exact arithmetic of the GMRES method [16] applied to equation (3). If we were defining only BMinres, then there is no apparent reason not to consider the odd-ordered $\hat{T}_{2k-1}$. However from Brown [1] we know that since each T_{2k-1} is singular, the minimization problems in equation (16) for any $\hat{T}_{2k-1}$ and $\hat{T}_{2k}$ yield the same BMinres iterate.

3. Numerical instabilities

We consider Question 1: What role do numerical instabilities play in the generation of the peak formations observed in residual norm plots generated by the Galerkin method BLanczos? We demonstrate that if equation (1) is sufficiently well-conditioned, then numerical instabilities have no role in any observed peak formations.

We need the following two theorems and definition. We can summarize Theorem 3.1 as follows. Given any real symmetric matrix C, let T_j for $j = 1, 2, \dots, K$, be the Lanczos matrices generated by applying the real symmetric Lanczos recursions to C in finite precision arithmetic. Theorem 3.1 states that it is possible to construct a larger matrix $\bar{C}$ which depends upon K such that the eigenvalues of $\bar{C}$ are in small intervals around the eigenvalues of C and such that for $1 \leqslant j \leqslant K$ the tridiagonal matrices generated by applying the real symmetric Lanczos recursions to $\bar{C}$ in exact arithmetic are identical to the T_j obtained in the finite precision computation on C. Theorem 3.1 allows us to ask Question 1 for a related exact arithmetic computation.

Theorem 3.2 is a variant of the standard interlacing theorem for real symmetric matrices. We know that the eigenvalues of the successive principal submatrices of any real symmetric tridiagonal matrix must interlace. Theorem 3.2 relates the sets of intervals defined by these eigenvalues to the eigenvalues of the original tridiagonal matrix. Theorem 3.2 states that each such subinterval must contain an eigenvalue of the original matrix. We use Theorem 3.2 in a proof of Theorem 3.4 which gives us bounds on the condition numbers of the Galerkin matrices generated by BLanczos.

We use Theorems 3.1 and 3.2 in the proof of Theorem 3.4 which states that if equation (1) is sufficiently well-conditioned, then the BLanczos Galerkin equations which are used to define the BLanczos iterates are also well-conditioned. Therefore, in such a situation, assuming BLanczos is implemented properly, numerical instabilities do not play a role in the generation of peak formations in residual norm plots generated by BLanczos. In our numerical experiments we tracked the condition numbers of the BLanczos Galerkin matrices and observed that the condition numbers of these matrices were typically not significantly worse than the condition number of A even when A was very ill-conditioned.

Theorem 3.1 ([10]). *Given any $n \times n$ real symmetric matrix C and unit vector w_1, let T_j^C, $j = 1, 2, \dots$ be the matrices generated by applying the real symmetric Lanczos recursion to C and w_1 in finite precision arithmetic. For any K, there exists an $N \times N$ matrix $\bar{C}$, a unit N-vector $\bar{w}_1$ with $N \geqslant n$, and a small $\varepsilon_K > 0$ such that the eigenvalues of $\bar{C}$ lie in ε_K-intervals about the eigenvalues of C and the matrices $T_j^{\bar{C}}$ generated by applying the real symmetric Lanczos recursion to $\bar{C}$ and a particular $\bar{w}_1$ in exact arithmetic satisfy $T_j^{\bar{C}} = T_j^C$ for $1 \leqslant j \leqslant K$.*

Theorem 3.2 ([11]). *Let T be a real symmetric, irreducible tridiagonal matrix of order K. Let T_k be the kth principal minor of T. Let μ_i^k for $1 \leqslant i \leqslant k < K$ be the eigenvalues of T_k ordered such that for each k, $\mu_i^k \leqslant \mu_{i+1}^k$ for $1 \leqslant i \leqslant k$. If we set $\mu_0^k = -\infty$ and $\mu_{k+1}^k = \infty$, then for any $k < K$ and $0 \leqslant i \leqslant k$ each open interval (μ_i^k, μ_{i+1}^k) contains an eigenvalue of T.*

Definition 3.3. Equation (1) is said to be *sufficiently well-conditioned for some* $2K$, if the ε_{2K} defined by Theorem 3.1 for the corresponding problem equation (3) satisfies $\sigma_n - \varepsilon_{2K} > 0$ where σ_n is the smallest singular value of A.

Theorem 3.4. *If for a specified $2K$, equation (1) is sufficiently well-conditioned, then for $1 \leqslant k \leqslant K$, the condition numbers $\kappa(T_{2k})$ of the even-ordered Galerkin matrices defined by equations (5) and (6) and used to compute the BLanczos iterates corresponding to equation (1) satisfy*

$$\kappa(T_{2k}) \leqslant (\sigma_1 + \delta_k)/(\sigma_n - \varepsilon_{2K}) \tag{17}$$

where σ_n and σ_1 denote respectively the smallest and the largest singular values of A,

$$0 < \delta_k \leqslant 2\sqrt{2}(2K)^{5/2}\|A\| \max(6\varepsilon_0, \varepsilon_1) \quad \textit{where } \varepsilon_0 \equiv (n+4)\varepsilon^*, \ \varepsilon_1 \equiv (7 + n_z\alpha)\varepsilon^*, \tag{18}$$

ε^* is the machine epsilon, n_z is the average number of nonzeros in each row of A, and $\alpha \equiv \||A|\|/\|A\|$.

Proof. From Theorem 3.1 we know that there exists an $N \times N$ matrix $\bar{C}$ and a corresponding starting vector $\bar{c}$ such that if the real symmetric Lanczos recursions are applied to $\bar{C}$ and $\bar{c}$, in exact arithmetic, $\bar{T}_j = T_j$ for $j = 1, \ldots, 2K$. Furthermore, we know that $\bar{C}$ can be chosen such that its eigenvalues lie in ε_{2K}-intervals about the eigenvalues of B in equation (3). Since equation (1) is sufficiently well-conditioned at $2K$, we have $\sigma_n - \varepsilon_{2K} > 0$. Therefore, there are no eigenvalues of $\bar{C}$ in the interval $\mathcal{I}_\varepsilon \equiv (-\sigma_n + \varepsilon_{2K}, \sigma_n - \varepsilon_{2K})$. Moreover for any $k < K$ there are no eigenvalues of any T_k in $\mathcal{I}_\varepsilon$. The proof is by contradiction. Suppose for some $k \leqslant K$ there is an eigenvalue μ_j^{2k} of T_{2k} in $\mathcal{I}_\varepsilon$. Since the eigenvalues of T_{2k} occur in ± pairs, $-\mu_j^{2k}$ is also an eigenvalue of T_{2k}. Since the arithmetic for $\bar{C}$ is exact, for some J with $K \leqslant J \leqslant N$ the eigenvalues of $\bar{T}_J$ are the same as the distinct eigenvalues of $\bar{C}$. Moreover, since $T_{2k} = \bar{T}_{2k}$ for $1 \leqslant k \leqslant K$, we have from Theorem 3.2 that there must be an eigenvalue of B in the interval $(-\mu_j^{2k}, \mu_j^{2k})$. But, the interval $(-\mu_j^{2k}, \mu_j^{2k})$ is contained in $\mathcal{I}_\varepsilon$ which by construction does not contain any eigenvalues of B. This is a contradiction. Therefore, the smallest singular value of any such T_{2k} is at least as large as $\sigma_n - \varepsilon_{2K}$. The upper bound on the largest singular value of each T_{2k} is derived in Paige [12]. □

4. Peaks and plateaus in exact arithmetic

In this section we consider Question 2: Are the peaks and plateaus artifacts of the finite precision arithmetic? We demonstrate that the answer is no. Peaks and plateaus can also occur when the arithmetic is exact. However, the experiments in Section 7 clearly demonstrate that for Galerkin/minimal residual methods which do not enforce the theoretical orthogonality or bi-orthogonality, more peaks or plateaus will occur in finite precision arithmetic than would occur if the computations were exact. In any theorem or lemma which involves Galerkin iterates we are implicitly assuming that these iterates are well defined.

Specifically, Theorem 4.6 states that any significant peak (plateau) behavior which occurs when the bidiagonalization method BLanczos (BMinres) is applied to equation (1) in finite precision arithmetic can also be observed in exact arithmetic but usually on a different problem. Theorem 4.6 is an extension of Theorems 4.1 and 4.2 which apply to the Galerkin and minimal residual real symmetric Lanczos methods to the bidiagonalization methods BLanczos and BMinres. Detailed statements and proofs of Theorems 4.1 and 4.2 are given in [3].

Theorem 4.1 ([3]). *Let C be a real symmetric $n \times n$ matrix and c be an n-vector. For $k = 1, 2, \ldots,$ let r_k^{G} be the kth residual vector generated by applying the Galerkin real symmetric Lanczos method* *to $Cx = c$. For some K let $\bar{C}$ and $\bar{c}$ be a corresponding matrix and vector defined by Theorem* 3.1, *and let $\bar{r}_k^{\mathrm{G}}$ be the corresponding residuals obtained by applying this method to $\bar{C}\bar{x} = \bar{c}$ in exact arithmetic. Then for $k = 1, \ldots, K$,*

$$\|r_k^{\mathrm{G}}\| = \|\bar{r}_k^{\mathrm{G}}\| + h_k^{\mathrm{G}}, \quad \textit{where } |h_k^{\mathrm{G}}| \leqslant \|E_k\|\|y_k^{\mathrm{G}}\| + \|C\|\|g_k^{\mathrm{G}}\| \tag{19}$$

where y_k^{G} is the exact solution of the Galerkin iteration equations, E_k is the error in the Lanczos recursions, and g_k^{G} is the error in computing the Galerkin iterate.

Theorem 4.2 ([3]). *Let C be a real symmetric $n \times n$ matrix and c be an n-vector. For $k = 1, 2, \ldots$, let r_k^{M} be the kth residual vector generated by applying the minimal residual real symmetric Lanczos method to $Cx = c$. For some K let $\bar{C}$ and $\bar{c}$ be a corresponding matrix and vector defined by Theorem* 3.1, *and let $\bar{r}_k^{\mathrm{M}}$ be the corresponding residuals obtained by applying this method to $\bar{C}\bar{x} = \bar{c}$ in exact arithmetic. Then for $k = 1, \ldots, K$,*

$$\|r_k^{\mathrm{M}}\| = \|\bar{r}_k^{\mathrm{M}}\| + h_k^{\mathrm{M}}, \quad \textit{where } |h_k^{\mathrm{M}}| \leqslant \|\bar{r}_k^{\mathrm{M}}\|\phi \tag{20}$$

where ϕ is a complicated function of the differences between the eigenvalues of $\bar{C}$ and C, of $\|r_k^{\mathrm{M}}\|$, of n, of the number of iterations k, of the eigenvalues of C with the largest and smallest magnitudes, of appropriately defined differences between the corresponding Lanczos vectors for these matrices, and of $\eta \equiv (E_{k+1}z_k^{\mathrm{M}} - CE_k y_k^{\mathrm{M}} - C^2 g_k^{\mathrm{M}})$ where z_k^{M} is the quasi-minimal residual vector obtained in the least squares problem, y_k^{M} is the exact solution of the associated least squares problem, g_k^{M} is the error incurred in computing the minimal residual iterate, and E_j denotes the errors in the Lanczos recursions.

Theorems 4.1 and 4.2 state that as long as the corresponding error terms h_k^{G} or h_k^{M} are sufficiently small the residual norm plots generated for the finite precision computation and those generated by a corresponding exact arithmetic computation are essentially the same. In Theorem 4.6 we extend these results to the bidiagonalization methods.

The following lemmas are used in the proof of Theorem 4.6. To simplify the discussion we use, for example, $\|r_k^{\mathrm{G}}(C, c)\|$ to denote the kth residual norm generated by applying the Galerkin real symmetric Lanczos method to the system $Cx = c$. If the arithmetic is exact, we will use $\bar{r}$. Moreover, since Galerkin iterates for systems of the form equation (3) are only defined for even-ordered Galerkin matrices, the kth iterate corresponds to the Galerkin system of size $2k$.

Lemma 4.3 ([10]). *Let $T = S\Theta S^{\mathrm{T}}$ be a real symmetric $m \times m$ tridiagonal matrix where $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_m)$, $S = (s_{ij})$ with $1 \leqslant i, j \leqslant m$, and $S^{\mathrm{T}}S = I$. Define $z \equiv (s_{11}, \ldots, s_{1m})$. Then if we apply m steps of the real symmetric Lanczos recursions to Θ in exact arithmetic with the starting vector z, we obtain*

$$\Theta S^{\mathrm{T}} = S^{\mathrm{T}}T. \tag{21}$$

Lemma 4.3 states that if we apply the Lanczos recursion to Θ with the starting vector defined by the first components of eigenvectors of T, then the Lanczos vectors generated are the columns of the transpose of the corresponding normalized eigenvector matrix of T. We also need the following trivial lemmas.

Lemma 4.4. *Let T_{2j} be any BLanczos iteration matrix with eigenvalues $\Theta_{2j} = \mathrm{diag}(\theta_1, \ldots, \theta_j, -\theta_1, \ldots, -\theta_j)$ where θ_i denote the positive eigenvalues of T_{2j}. Then if u_i is an eigenvector corresponding to θ_i, the vector obtained from u_i by reversing the signs of the odd-numbered components of u_i is an eigenvector corresponding to $-\theta_i$.*

Lemma 4.5. *Let T_{2j} be any BLanczos iteration matrix with eigenvalues $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_j, -\theta_1, \ldots, -\theta_j)$ where θ_i denote the positive eigenvalues of T_{2j}. Define $\Theta^+ = \mathrm{diag}(\theta_1, \ldots, \theta_j)$,*

$$\Theta^* \equiv \begin{pmatrix} 0 & \Theta^+ \\ \Theta^+ & 0 \end{pmatrix} \text{ and } \sqrt{2}U \equiv \begin{pmatrix} I_j & -I_j \\ I_j & I_j \end{pmatrix} \text{ then } \Theta^* = U\Theta U^{\mathrm{T}}. \tag{22}$$

Theorem 4.6. *Apply BLanczos and BMinres to equation (1). Let r_k^{BL} and r_k^{BM} denote the corresponding residuals obtained. Then for any given K, there exists a diagonal matrix D and a vector d such that if BLanczos and BMinres are applied to $Dx = d$ in exact arithmetic, for $k = 1, 2, \ldots, K$, then the corresponding residuals $\bar{r}_k^{\mathrm{BL}}$ and $\bar{r}_k^{\mathrm{BM}}$ satisfy*

$$\|r_k^{\mathrm{BL}}\| = \|\bar{r}_k^{\mathrm{BL}}\| + h_k^{\mathrm{BL}} \quad \text{where } |h_k^{\mathrm{BL}}| \leqslant \|E_{2k}\|\|y_{2k}^{\mathrm{G}}\| + \|B\|\|g_{2k}^{\mathrm{G}}\| \tag{23}$$

and

$$\|r_k^{\mathrm{BM}}\| = \|\bar{r}_k^{\mathrm{BM}}\| + h_k^{\mathrm{BM}} \quad \text{where } |h_k^{\mathrm{BM}}| \leqslant \|r_k^{\mathrm{M}}(\bar{C}, \bar{c})\|\phi \tag{24}$$

where in equation (23), y_{2k}^{G} is the exact solution of the Galerkin iteration equations for equation (3), E_{2k} is the error in the Lanczos recursions for equation (3), and g_{2k}^{G} is the error in computing the Galerkin iterate for equation (3) where we have numbered the Galerkin iterates for each even-ordered T_{2k}; and in equation (24), ϕ is a complicated function of the differences between the eigenvalues of $\bar{C}$ (defined in Theorem 3.1 for B in equation (3)) and the eigenvalues of B, of appropriately defined differences between corresponding Lanczos vectors for these matrices, of $\eta \equiv (E_{2k+1}z_{2k}^{\mathrm{M}} - BE_{2k}y_{2k}^{\mathrm{M}} - B^2 g_{2k}^{\mathrm{M}})$ where z_{2k}^{M} is the quasi-minimal residual vector obtained in the least squares problem in equation (16), y_{2k}^{M} is the exact solution of that associated least squares problem, g_{2k}^{M} is the error incurred in computing that minimal residual iterate, and of $\|r_k^{\mathrm{M}}(B, \bar{b})\|$, of n, of the number of iterations k, and of the largest and smallest singular values of A.

Proof. We consider BLanczos. The equivalences demonstrated apply directly to a similar argument for BMinres. Theorem 4.1 states that given equation (3) and a number of iterations K, there is a real symmetric system $\bar{C}\bar{x} = \bar{c}$ such that in exact arithmetic, the residual norms generated when the Galerkin real symmetric Lanczos method is applied to this system satisfy equation (19).

$$\|r_k^{\mathrm{BL}}(A, b)\| = \|r_k^{\mathrm{G}}(B, \bar{b})\| = \|\bar{r}_k^{\mathrm{G}}(\bar{C}, \bar{c})\| + h_{2k}^{\mathrm{G}} \tag{25}$$

where h_{2k}^{G} satisfies the conditions stated in Theorem 4.1. The following argument demonstrates that for the given K and in exact arithmetic there exists a matrix D and a right-hand side d such that when BLanczos is applied to this system the residual norms $\|\bar{r}_k^{\mathrm{BL}}(D, d)\| = \|\bar{r}_k^{\mathrm{G}}(\bar{C}, \bar{c})\|$.

Let $T_{2K} = S\Theta_{2K}S^{\mathrm{T}}$ be an eigenvector decomposition of T_{2K} where the eigenvalues are ordered as in Lemma 4.5. If we apply the Galerkin real symmetric Lanczos method in exact arithmetic to $\Theta_{2K}y = \|b\|z$ where z is defined as in Lemma 4.3, we get that

$$\|\bar{r}_k^{\mathrm{G}}(\Theta_{2K}, \|b\|z)\| = \|\bar{r}_k^{\mathrm{G}}(\bar{C}, \bar{c})\| \tag{26}$$

since the norms of the residuals are completely defined by the size of the starting residual and the iteration matrices T_{2k}. But from Lemma 4.5 we have

$$\Theta^*_{2K}(US^{\mathrm{T}}) = (US^{\mathrm{T}})T_{2K}. \tag{27}$$

Therefore, if we apply the Galerkin real symmetric Lanczos method in exact arithmetic to $\Theta^*_{2K}y = \|b\|Uz$, we get

$$\|\bar{r}^{\mathrm{G}}_k(\Theta^*_{2K}, \|b\|Uz)\| = \|\bar{r}^{\mathrm{G}}_k(\Theta_{2K}, \|b\|z)\|. \tag{28}$$

Moreover from Lemma 4.4 we have that the last K components of the vector Uz are 0. Therefore if we define d equal to the first K components of $\|b\|Uz$ and $D = \Theta^+$ and apply BLanczos to $Dx = d$ in exact arithmetic, we obtain

$$\|\bar{r}^{\mathrm{BL}}_k(D, d)\| = \|\bar{r}^{\mathrm{G}}_k(\Theta^*_{2k}, \|b\|Uz)\| \tag{29}$$

which yields the desired result. □

5. Test matrices

To gain insight into the behavior of the residual norms of iterative methods it is necessary to perform numerical experiments with these methods. We would like to have a set of test problems which would generate all possible residual norm plots for the bidiagonalization methods. We can identify such a set when the arithmetic is exact. We have the following theorems which characterize this set of matrices. In this section we deviate from the notation used in earlier sections and use a superscript C to denote iterates corresponding to a matrix C.

Theorem 5.1. *For $k = 1, \ldots, K$, let T^C_k, T^Λ_k and W^C_k, W^Λ_k denote respectively, Lanczos matrices and Lanczos vectors, obtained by applying the real symmetric Lanczos recursions in exact arithmetic to real symmetric matrices C and Λ. If $C = U\Lambda U^{\mathrm{T}}$ where $U^{\mathrm{T}}U = I$, and the starting vector for C is $w_1 = b/\|b\|$ and for Λ is $U^{\mathrm{T}}w_1$, then for $k = 1, \ldots, K$, $T^C_k = T^\Lambda_k$ and $W^C_k = UW^\Lambda_k$. Furthermore, for any c and exact arithmetic, if we apply the Galerkin real symmetric Lanczos method or the minimal residual real symmetric Lanczos method to solve $Cx = c$ and $\Lambda x = U^{\mathrm{T}}c$, and assume that the Galerkin iterates are defined on each iteration, then for either method and on each iteration, $x^C_k = Ux^\Lambda_k$, $r^C_k = Ur^\Lambda_k$, and $\|r^C_k\| = \|r^\Lambda_k\|$.*

Proof. In exact arithmetic for each j,

$$CW^C_j = W^C_jT^C_j + \beta_{j+1}w^C_{j+1}e^{\mathrm{T}}_j \quad \text{so that } \Lambda(U^{\mathrm{T}}W^C_j) = (U^{\mathrm{T}}W^C_j)T^C_j + \beta_{j+1}U^{\mathrm{T}}w^C_{j+1}e^{\mathrm{T}}_j. \tag{30}$$

Therefore, for each j, $W^\Lambda_j = U^{\mathrm{T}}W^C_j$ and $T^\Lambda_j = T^C_j$. We consider the Galerkin real symmetric Lanczos method to solve $Cx = c$ and $\Lambda y = U^{\mathrm{T}}c$. A similar argument applies to the corresponding minimal residual method. For both equations on each iteration we solve $T_ku_k = \|c\|e_1$ and compute

$$x^C_k = W^C_ku^C_k = UW^\Lambda_ku^\Lambda_k = Ux^\Lambda_k \quad \text{and} \quad r^C_k = -Cx^C_k + c = -U\Lambda U^{\mathrm{T}}x^C_k + c = Ur^\Lambda_k. \qquad \square \tag{31}$$

Corollary 5.2. *If A and B are similar real symmetric matrices, then there exist orthogonal matrices U such that $A = UBU^{\mathrm{T}}$ and for any b and in exact arithmetic, the Galerkin (minimal residual) real* *symmetric Lanczos method applied to $Ax = b$ and to $Bx = U^{\mathrm{T}}b$ generate identical residual norm plots.*

Theorem 5.3. *For either bidiagonalization method, BLanczos or BMinres, and in exact arithmetic, the set of all positive definite $n \times n$ positive diagonal matrices Σ together with all possible vectors b in equations (1) will generate the set of all possible residual norm plots that can be generated by applying BLanczos or BMinres to equation (1) with arbitrary $n \times n$ matrices A and arbitrary vectors b of size n.*

Proof. Let $A = U\Sigma V^{\mathrm{T}}$ where $U^{\mathrm{T}}U = I$, $V^{\mathrm{T}}V = I$, and Σ is a positive definite diagonal matrix of singular values of A. We have the following orthogonal equivalence.

$$\tilde{D} \equiv \begin{pmatrix} 0 & \Sigma \\ \Sigma & 0 \end{pmatrix} = W^{\mathrm{T}}BW \quad \text{where } B \equiv \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{pmatrix} \text{ and } W = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}. \tag{32}$$

Therefore from Lemmas 2.2 and 2.3 and Theorem 5.1, and for any b, we have that for each k, $\|\bar{r}_k^{\mathrm{BL}}(\Sigma, U^{\mathrm{T}}b)\| = \|\bar{r}_k^{\mathrm{BL}}(A, b)\|$. □

In each test discussed, as well as in the tests in [2], we specified a diagonal matrix Σ by specifying a few large singular values, a few small singular values, and spacing the remaining singular values uniformly within a specified interval. Experiments were also run using matrices with more than one interval of uniformly spaced or randomly spaced singular values. No significant changes in behavior were observed with the larger set of test problems.

We include results for two choices of right-hand sides b in equation (1). In Figs. 3-6, $b = (\sigma_1, \ldots, \sigma_n)$ and the true solution, $x_{\text{true}} = (1, 1, \ldots, 1)$, where σ_j, $1 \leqslant j \leqslant n$, are the specified diagonal entries of the $n \times n$ test matrix Σ. In Fig. 7, $b = (1, 1, \ldots, 1)$ and $x_{\text{true}} = (1/\sigma_1, \ldots, 1/\sigma_n)$. The basic properties of the peak and the plateau formations we discuss were not dependent upon the choice of b as long as b had a projection on each eigenvector, although the residual norms plots varied considerably as b was varied. See for example, Fig. 7 versus Fig. 6.

Residual norms were computed but not used in the convergence tests, equation (2). Estimates of the residuals defined by equations (11) and (16) in Lemmas 2.2 and 2.3 were used to determine convergence. In each case the convergence tolerance was $\varepsilon = 10^{-13}$. However, the values plotted in each figure are the true residual norms. For these methods, the estimates seem to accurately reflect the actual values until the effects of the finite precision arithmetic begin to dominate the computations. No experiments were run in exact arithmetic and the Lanczos vectors generated by equations (5) and (6) were not reorthogonalized.

6. Relationships between peaks and plateaus

In this section we consider Question 3 in the context of the two bidiagonalization methods, BLanczos and BMinres. Is there a correlation between the residual norms generated when these two methods are applied to a given problem $Ax = b$? In particular are there correlations between peak formations and plateaus? We first present some experimental results and then give a theoretical result which is proved in [3], along with analogous results for GMRES/FOM and QMR/BCG. We then use

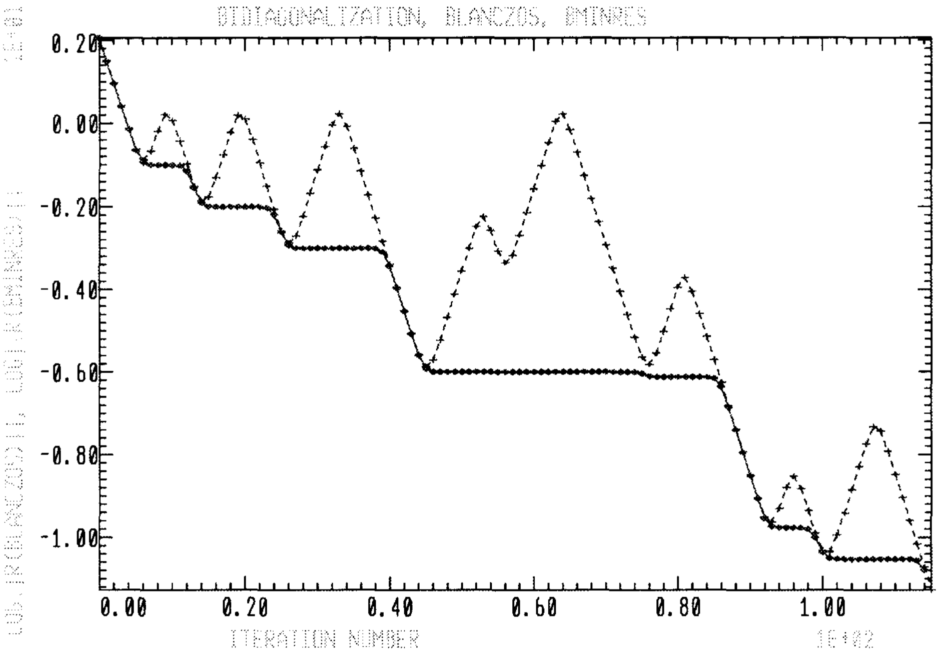

Fig. 3. Mat1, $\log_{10}(\|r_k^{\text{BL}}\|)$, $\log_{10}(\|r_k^{\text{BM}}\|)$ versus k.

this result to obtain two lemmas, one relating portions of these curves during intervals of iterations over which the residual norms generated by BMinres are decreasing rapidly, and the other lemma relating peaks and plateaus. A consequence of these correspondences is that when the iterates for both methods are well defined, the rates of convergence for these two methods are similar. Thus, minimizing the residual norm does not accelerate the convergence. It does, however, provide the user with a residual norm curve that is easier to monitor for convergence than the typical Galerkin residual norm curve is.

We applied BLanczos and BMinres to a sequence of increasing complicated "diagonal" test problems. The results of two experiments which are representative of the results obtained from these tests are depicted in Figs. 3 and 4. In Figs. 3 and 4 we use + to denote the values of the residual norms generated by BLanczos, and ◇ to denote the corresponding values for BMinres. We plot the base 10 logarithms of these values versus the iteration number. Fig. 3 corresponds to a test matrix, Mat1, which has four *small* singular values 10^{-6}, 10^{-3}, 10^{-2} and 10^{-1}, no large singular values, and the remaining singular values spaced uniformly in an interval near 2.6. Fig. 4 corresponds to a test matrix, Mat2, with the same small singular values as Mat1, essentially the same uniform portion, but with two *large* singular values 10.0 and 100.0. For both matrices $n = 1000$.

Both residual norm plots in Fig. 3 are smooth. Peaks occur in the BLanczos residual norm plot at iterations 9, 19, 33, 53, 64, 81, 96 and 107. Plateaus occur in the BMinres residual norm plot approximately over iterations 7–11, 15–23, 27–38, 46–75, 76–85, 93–98, and 101–113. Each peak in the BLanczos norm plot corresponds to a plateau in the BMinres norm plot, with the double peak corresponding to the long plateau across iterations 46–75. Both methods required 115 iterations for

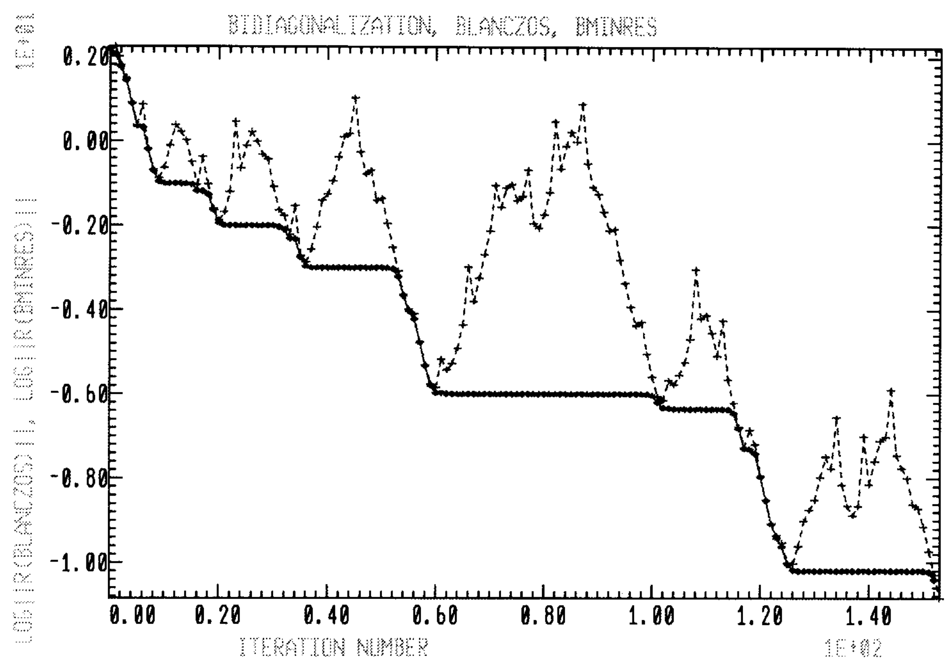

Fig. 4. Mat2, $\log_{10}(\|r_k^{BL}\|)$, $\log_{10}(\|r_k^{BM}\|)$ versus k.

convergence.

The BLanczos residual norm plot in Fig. 4 is more a representative of plots seen in practice. This plot is not smooth. It contains 28 clearly defined peaks plus eight wider but less well-defined peaks. In the BMinres norm plot significant plateaus occur approximately at iterations 9–15, 20–31, 36–52, 60–100, 102–114, and 126–151. In addition there are short plateaus at 5–6, 16–17, 33–34, and 117–118. Each peak in the BLanczos norm plot corresponds to a plateau in the BMinres norm plot. High frequency peaks which are not superimposed upon low frequency peaks correspond to short plateaus. Both methods required 153 iterations for convergence.

Both figures 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 BMinres residual norm plot without a visible corresponding peak in the corresponding BLanczos residual norm plot. Also several peaks may sit on top of what appears to be a single plateau. These plots also indicate that whenever the residual norm plot for the BMinres iterates is decreasing rapidly, the corresponding residual norm plot for the BLanczos iterates is also decreasing rapidly. Thus, corresponding residual norm plots appear to track each other. Figs. 3 and 4 are similar in terms of the numbers of wide peaks and plateaus observed. However, the BLanczos norm plot in Fig. 3 is smooth and the corresponding plot in Fig. 4 is very irregular. Moreover, more iterations are required for full convergence with Mat2 than with Mat1. In Section 7 we propose a plausible explanation for these differences. The results of additional tests are given in [2].

In [2] we also considered the following relationship which is proved in [1] for the GMRES/FOM residuals in exact arithmetic.

$$\|r_k^{\mathrm{BM}}\| = |c_k|\|r_k^{\mathrm{BL}}\|. \tag{33}$$

In exact arithmetic if BLanczos/BMinres and GMRES/FOM are applied to equation (3), they generate the same iterates. In equation (33), c_k denotes the cosine of the $(2k)$th Givens transformation used in the QR factorization of $\hat{T}_{2k}$. For each of the test problems considered in [2] we plotted

$$\log_{10}\left(\|r_k^{\mathrm{BL}}\|/\|r_k^{\mathrm{BM}}\|\right) \quad \text{and} \quad \log_{10}|c_k| \tag{34}$$

versus the iteration number k and observed that even in finite precision arithmetic these two curves were essentially mirror images. This combination of experiments led to the following theorem which is proved in [3] along with related results for the QMR/BCG and GMRES/FOM methods.

Theorem 6.1 ([3]). *If for any A and b we apply the bidiagonalization methods BLanczos and BMinres to equations* (1) *in finite precision arithmetic, then at each iteration k,*

$$\|r_k^{\mathrm{BL}}\| = \|r_k^{\mathrm{BM}}\|/\sqrt{1-\left(\|r_k^{\mathrm{BM}}\|/\|r_{k-1}^{\mathrm{BM}}\|\right)^2} + \omega_k \tag{35}$$

where the error terms $\omega_k = \omega_k(h_k^{\mathrm{BM}}, h_{k-1}^{\mathrm{BM}}, h_k^{\mathrm{BL}}, \|r_k^{\mathrm{BM}}\|, \|r_{k-1}^{\mathrm{BM}}\|)$ are small if the error terms defined in Theorem 4.6, *h_k^{BM}, h_{k-1}^{BM}, and h_k^{BL} are smaller than $\|r_k^{\mathrm{BM}}\|$ and the ratio $\|r_k^{\mathrm{BM}}\|/\|r_{k-1}^{\mathrm{BM}}\|$ is less than but not too close to* 1.

In Lemma 6.2 we consider strongly monotone sections and prove that if for a given interval of iterations the BMinres residual norm plot is strongly monotone decreasing, then during those iterations the corresponding BLanczos residual norm plot is trapped between small multiples of the BMinres curve. In Lemma 6.3 we consider a peak and plateau correspondence and prove that if over some interval of iterations residual norms generated by BLanczos are increasing at least as fast as a specified rate γ, then the corresponding BMinres residual norms cannot decrease at a rate faster than the bound on θ given in equation (39). For example, if $\gamma = 3$ and $\delta = .01$, then Lemma 6.3 implies that $\theta > .947$. This indicates that peaks and plateaus are different manifestations of some underlying phenomena. We discuss this possibility in Section 7.

Lemma 6.2. *If, under the assumptions of Theorem* 6.1, *there exist iterations $K_1 \leqslant k \leqslant K_2$, $0 < \gamma < 1$ such that $\|r_k^{\mathrm{BM}}\| \leqslant \gamma\|r_{k-1}^{\mathrm{BM}}\|$, and $0 < \delta < 1/2$ such that*

$$\max\left(|\omega_k|, |h_k^{\mathrm{BM}}|, |h_k^{\mathrm{BL}}|\right) \leqslant \delta \min\left(\|r_k^{\mathrm{BM}}\|, \|r_k^{\mathrm{BL}}\|\right), \tag{36}$$

then for $K_1 \leqslant k \leqslant K_2$,

$$(1-2\delta)\|r_k^{\mathrm{BM}}\| \leqslant \|r_k^{\mathrm{BL}}\| \leqslant \|r_k^{\mathrm{BM}}\|\left(1/\sqrt{1-\gamma^2}+\delta\right). \tag{37}$$

Proof. The left-hand portion of the inequality in equation (37) follows from Theorem 4.6 and the fact that in exact arithmetic the norms of the residuals generated by BMinres must be monotone decreasing. Specifically,

$$\|r_k^{\mathrm{BL}}\| = \|\bar{r}_k^{\mathrm{BL}}\| + h_k^{\mathrm{BL}} \geqslant \|\bar{r}_k^{\mathrm{BM}}\| + h_k^{\mathrm{BL}} \geqslant \|r_k^{\mathrm{BM}}\|(1-2\delta). \tag{38}$$

An application of Theorem 6.1, the assumption on γ, and the bound on ω_k yield the other half of this inequality. □

Lemma 6.3. *If under the assumptions of Theorem* 6.1, *there exist iterations* $K_1 \leqslant k \leqslant K_2$, $\gamma > 1$ *with* $\|r_k^{\mathrm{BL}}\| \geqslant \gamma \|r_{k-1}^{\mathrm{BL}}\|$ *and* $\delta < 1/2$ *such that* $\max(|\omega_k|) \leqslant \delta(\|r_k^{\mathrm{BM}}\|)$, *then if there exists a* $0 < \theta < 1$ *such that*

$$\|r_k^{\mathrm{BM}}\| < \theta \|r_{k-1}^{\mathrm{BM}}\| \quad \textit{then } \theta > \gamma(1-2\delta)/\sqrt{1+\gamma^2(1-2\delta)^2}. \tag{39}$$

Proof. From Theorem 6.1 and the inequality on $\|r_k^{\mathrm{BL}}\|$ we have that

$$\begin{aligned} \|r_k^{\mathrm{BL}}\| &= \|r_k^{\mathrm{BM}}\| / \sqrt{\left(1 - \left(\|r_k^{\mathrm{BM}}\| / \|r_{k-1}^{\mathrm{BM}}\|\right)^2\right)} + \omega_k \geqslant \gamma \|r_{k-1}^{\mathrm{BL}}\| \\ &= \gamma \left(\|r_{k-1}^{\mathrm{BM}}\| / \sqrt{1 - \left(\|r_{k-1}^{\mathrm{BM}}\| / \|r_{k-2}^{\mathrm{BM}}\|\right)^2} + \omega_{k-1} \right). \end{aligned} \tag{40}$$

If we use the assumptions on θ, γ, δ and ω, we obtain the weaker inequality,

$$\theta \left(1/\sqrt{1-\theta^2} + \delta \right) > \gamma(1-\delta). \tag{41}$$

Rearranging we obtain

$$\theta/\sqrt{1-\theta^2} > \gamma(1-\delta) - \theta\delta. \tag{42}$$

Using the fact that $\theta < 1$ and $\gamma > 1$, and then squaring both sides and rearranging the terms, we obtain equation (39). □

7. Peaks and spectral convergence

Therefore, peaks and plateaus can occur even when the Galerkin iteration matrices are well-conditioned, the computations are numerically stable, and the arithmetic is exact. Moreover, if $Ax = b$ is sufficiently well-conditioned, then the corresponding BLanczos iteration matrices T_{2k} are well-conditioned, even in finite precision arithmetic. In this situation BLanczos and BMinres are numerically stable. In addition peak and plateau formations are correlated.

In this section we consider Question 4: Can we identify any other factors which initiate peak and plateau formations in these residual norm plots?

We present empirical evidence that the formation of peaks in BLanczos residual norm plots is correlated with the stabilization of eigenvalues of the associated Galerkin iteration matrices T_{2k} defined by equations (5) and (6). From this we infer that peak formations in residual norm plots generated by BLanczos correspond to the identification or re-identification of certain portions of the solution space, as indicated by the convergence of eigenvalues of the T_{2k} to singular values of A. From Theorem 6.1 we can then infer that plateau formations in residual norm plots generated by BMinres also correspond to the identification or re-identification of certain portions of the solution space.

[17,18] contain related results. [17] examines the conjugate gradient method for real symmetric positive definite A, and [18] examines the GMRES method [16]. Both of these papers address the question of speedups in the instantaneous rate of convergence when some eigenvalue in the spectra of the corresponding iteration matrices converges. Such a speedup almost surely corresponds to a strongly monotone portion of a residual norm curve as seen in Figs. 3–7. The experiments and results

in [17,18] however differ in terms of the questions being asked, the information being plotted, the conclusions, and the construction of the examples. We have focused upon peak formations and correlations between pairs of Galerkin/minimal residual methods.

We present the results of three experiments in Figs. 5, 6 and 7 obtained using Mat1 and Mat2. Fig. 7 corresponds to Mat2 with $b = (1, 1, \ldots, 1)$. We discuss the differences between Figs. 6 and 7. Details of additional experiments are included in [2]. For each test, we monitored estimates of the norms of the residuals $\|r_k\|$, the norms of the errors in the approximate solutions $\|x_k^e\|$, the coefficients and the residual norm estimates defined in equations (11) and (16), the cosines of the Givens transformations c_k used to compute the QR factorizations of $\hat{T}_{2k}$, and the eigenvalues of each T_{2k}. In this section we focus only on plots of the base 10 logarithms of residual norms generated by BLanczos overlaid with plots of the base 10 logarithms of the magnitudes of the eigenvalues of each T_{2k} and plotted versus the iteration number k. Since the eigenvalues of each T_{2k} occur in ± pairs, in these plots each pair appears as a single point.

Other types of plots are contained in [2]. On each vertical line in Figs. 5, 6, and 7, $+$ symbols correspond to eigenvalues. Values of residual norms are denoted by χ. The dark band of $+$ symbols in each figure corresponds to logarithms of the magnitudes of eigenvalues in the uniform portion of the singular value spectrum of the test matrix A.

These figures do not show enough detail to determine accurately the relationships between peak formations and spectral convergence. They are, however, sufficient to give accurate indications of the apparent effects of the slowly varying portions of that convergence. Each comment which we make relative to the peak formations and to the approximation and convergence of eigenvalues in the spectra of the T_{2k} to singular values of A is based upon detailed listings of the eigenvalues of these matrices and of the norms of the residuals in each test.

First consider Fig. 5 for Mat1. Mat1 has a uniform spectrum plus three *not so small* small singular values, 10^{-1}, 10^{-2}, and 10^{-3}, and one small singular value 10^{-6}. As we saw in Section 6, the BLanczos residual norm plot is smooth. If we track each of the $+$ symbols which emerges from the solid band of eigenvalues of T_{2j} in Fig. 5 as we move from iteration 1 to iteration k, we can observe the initial convergence of approximations to the singular values 10^{-1}, 10^{-2}, 10^{-3}, and 10^{-6} during iterations 1 to approximately iteration 66. We further observe that the convergence of each approximation appears to be correlated with the formation of peaks in the overlaid residual norm plot. The emergence of each eigenvalue 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.

In particular, peaks occur at iterations 9, 19, 33, 53, 64, 81, 96, and 107. If we look at the lists of eigenvalues of each of the Lanczos matrices T_{2k} for $k = 1, 2, \ldots, 115$, at iteration 7 we can identify an eigenvalue which eventually converges to 10^{-1}. At iteration 15 an eigenvalue emerges which eventually converges to 10^{-2}. At iteration 27 an initial approximation to 10^{-3} appears. Finally at iteration 46 we get a fourth eigenvalue outside the uniform spectrum which eventually becomes 10^{-6}.

Because there is no reorthogonalization of the Lanczos vectors generated in equations (5), replication of the approximations to singular values of A in the spectra of the T_{2k} can occur. We observe empirically that as soon as an eigenvalue of an iteration matrix stabilizes to an accurate approximation to some eigenvalue of the given matrix (in this case B in equation (3)), the procedure forgets it has successfully approximated that eigenvalue. If the recursion is continued long enough a copy of that converged approximation will appear in the spectra of the Galerkin matrices. How rapidly copies of

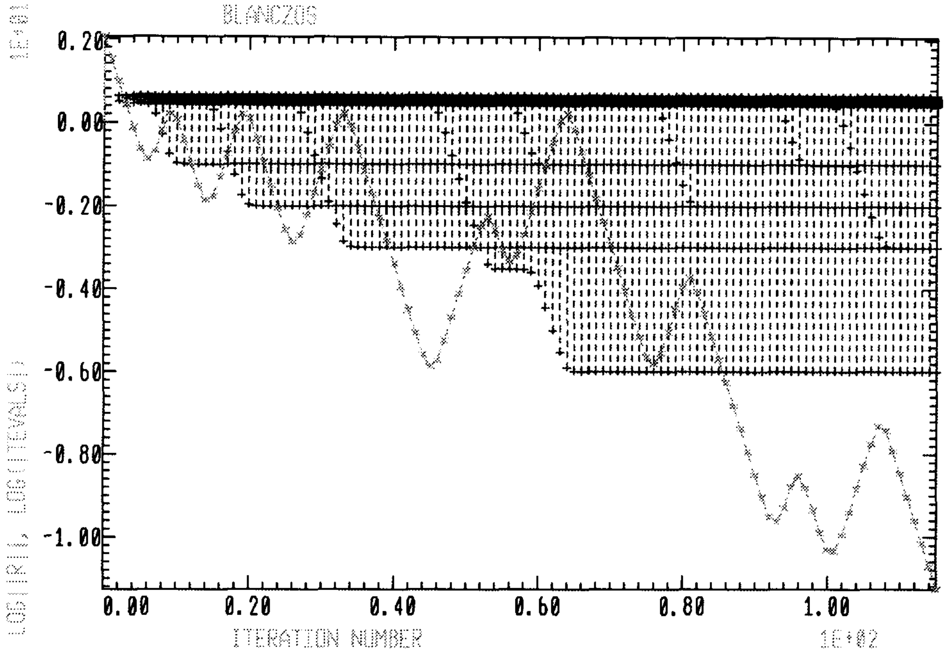

Fig. 5. Mat1, $\log_{10}(|\mu_j^k|)$, $1 \leqslant j \leqslant k$, $\log_{10}(\|r_k^{\mathrm{BL}}\|)$ versus k.

a particular eigenvalue will be generated depends principally upon the location of that eigenvalue in the spectrum of B, the gap between it and its nearest neighbors, its magnitude, and the size of its gap relative to the size of the biggest gaps between any other two eigenvalues.

For Mat1 we observe that at iteration 57 an eigenvalue appears which by iteration 61 becomes (visually at least) a copy of 10^{-1}. Its convergence interferes with the convergence of the approximations to 10^{-6} and a secondary peak appears which apparently corresponds to the temporary stabilization of the approximation to 10^{-6} as the approximation to 10^{-1} is initiated and emerges from the uniform band. Thus, the double peak from iterations 46–76 appears to correspond to the convergence of approximations to 10^{-6} and to a copy of 10^{-1}. At iteration 77 another copy starts to emerge. This eigenvalue becomes a copy of 10^{-2}. At iteration 93 a third copy of 10^{-1} starts to appear. At iteration 101 a copy of 10^{-3} starts to emerge. During the 115 iterations, eight peaks occur.

Mat2 has two *large* but *not very large* singular values, 10., 100., and the four small singular values in Mat1. The residual norm plot in Fig. 6 contains 28 clearly defined peaks plus 8 lower frequency, less well-defined peaks. Each of these wider peaks appears to correspond to the convergence of an eigenvalue of T_{2k} to one of the small singular values of Mat2. There are 8 peaks apparently because as we increase k, two copies of 10^{-1} and single copies of 10^{-2} and 10^{-3} appear in the spectra of the T_{2k}.

A detailed examination of the list of all the eigenvalues of the iteration matrices T_{2k} for $k = 1, \ldots, 153$ indicates a correlation between the appearance of each of the 28 smaller higher frequency peaks and the repeated reappearance of 100. and 10. in the spectra of the T_{2k}. In some cases approximations to both eigenvalues were obtained more or less simultaneously. When this occurs the

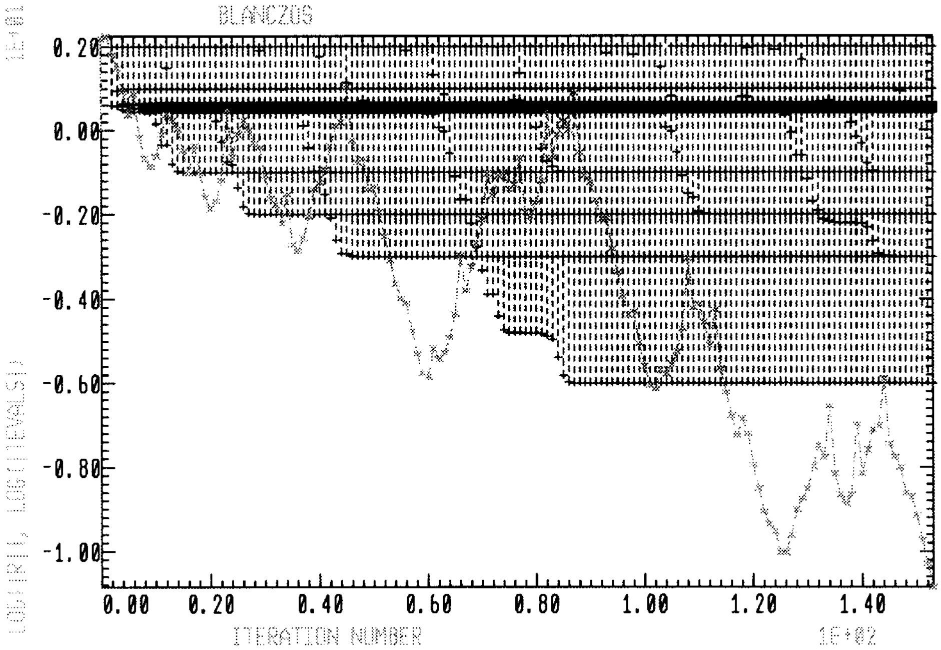

Fig. 6. Mat2, $\log_{10}(|\mu_j^k|)$, $1 \leqslant j \leqslant k$, $\log_{10}(\|r_k^{\mathrm{BL}}\|)$ versus k.

peaks can be distorted. In some cases, peaks can nullify each other.

For Mat2 and any k there were approximately 3 times as many replications of 100. as there were of 10.. Replications slow the convergence, as compared with procedures which enforce orthogonality or bi-orthogonality, but at least for the bidiagonalization procedures, replications do not seem to adversely affect the achievable accuracy. We can also observe that the amplitudes of the peaks which apparently correspond to the *large* singular values appear to correlate with the sizes of these large singular values. Other examples in [2] support this comment.

To determine the effect of the choice of b on the convergence, we ran tests varying b. Results corresponding to Mat2 with $b = (1, \ldots, 1)$ are contained in Fig. 7. We observe that this plot differs considerably from Fig. 6. However, it is also very similar. We observe 8 wide but irregular peaks occurring approximately at iterations 7, 15, 27, 46, 57, 76, 93, and 101.

The first peak in Fig. 7 appears to correspond to the stabilization of an eigenvalue of the T_{2k} to a value between 10^{-1} and 10^{-2}. We observe empirically that stabilizations at "average" values can occur when several approximations are converging simultaneously. We also observe empirically that stabilizations at "averages" can yield some improvement in the residual norms. The second peak appears to correspond to the convergence of an approximation to 10^{-1} (and a similar stabilization between 10^{-2} and 10^{-3}). Similarly, the third peak appears to correspond to the convergence of an approximation to 10^{-2} (and a similar stabilization between 10^{-3} and 10^{-6}). During the fourth peak approximations to both 10^{-3} and 10^{-6} appear and converge. The fifth peak appears to correspond to the appearance and convergence of a copy of 10^{-1}. Similarly, the sixth peak appears to correspond to a copy of 10^{-2}. The last two peaks appear to correspond to the overlapping convergence of copies

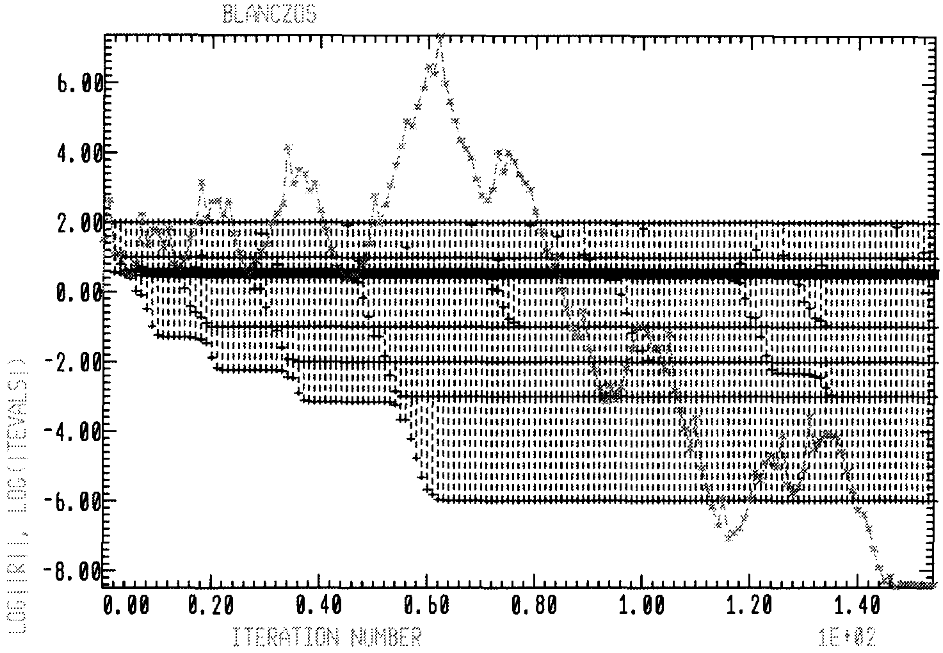

Fig. 7. Mat2, $\log_{10}(|\mu_j^k|)$, $1 \leqslant j \leqslant k$, $\log_{10}(\|r_k^{\mathrm{BL}}\|)$ versus k.

of approximations to 10^{-1} and 10^{-3}.

There are an additional 31 high frequency peaks. Using a detailed listing of all the eigenvalues of all the T_{2k} we can match each of these peaks with the reappearance in these spectra of approximations to either 10. or 100. or to both simultaneously. Thus, as in Figs. 5 and 6, each peak appears to correlate with a stabilization of eigenvalues in the spectra of the T_{2k} matrices. We also observe that approximately the same number of iterations, 153, were required for full convergence. These and other tests indicate that the fundamental characteristics of these plots are not dependent upon the vector b in equation (1), as long as that vector contains a projection on each of the eigenvectors of the corresponding B in equation (3).

These plots indicate correlations between spectral stabilization and peak formations in the residual norm plots for BLanczos. They also indicate that the peaks observed in practice are composites of individual peaks corresponding to various stabilized eigenvalues in the spectra of the T_{2k}. If the convergence of several approximations overlap, one or more approximations may stagnate until the newly generated approximation begins to converge. This appears to be reflected in the shapes of the corresponding peaks generated. If we were to overlay the corresponding BMinres residual norm plots onto Figs. 5-7 we would observe that strongly monotone decreasing sections of the BMinres residual norm plots appear to correspond to intervals of iterations during which there is no overt spectral convergence. We note also that in practice, see Figs. 6 and 7, that tiny plateaus corresponding to the reappearance of typically large singular values typically occur in the middle of strongly monotone sections of a BMinres curve.

Therefore, it appears that we have the following behavior. The BLanczos residual norm plots for

Mat1 are smooth because there are no singular values of Mat1 whose approximations replicate easily in the spectra of the associated T_{2k}. The BLanczos residual norm curves for Mat2 are very irregular because copies of the large singular values appear repeatedly in the spectra of the T_{2k}. The composite of these corresponding high frequency peaks with the underlying low frequency peaks creates the illusion that the residual norms are behaving very erratically. These experiments, however, indicate that the behavior may be predictable, in a global sense, if the singular value spectrum of A is known. We also note that the widths of the peaks appear to be correlated with the degree of difficulty in computing approximations to a corresponding singular value. Widths and heights appear, however, to also be affected by the interactions between peaks when more than one approximation is stabilizing.

The fact that eigenvalues of the T_{2k} approximate key singular values of A can be used in practice to check for adequate convergence. For example, if the convergence tolerance $\varepsilon = 10^{-6}$, then BMinres on Mat2 (see Fig. 6) would have terminated at iteration 55 where the norm of the true error is .99999899. However, the condition number of T_{110} is 10^5, from which we can surmise that satisfying $\|r_{55}^{\mathrm{BL}}\|/\|r^0\| < 10^{-6}$ is not sufficient to yield an accurate approximate solution.

8. Summary

The theorems and experiments described in the preceding sections provide some insight into the behavior of residual norm plots for both BLanczos and BMinres. They do not, however, suggest modifications to improve these methods nor were they intended to demonstrate that one of these two methods is better than the other one. We also want to emphasize that we are not advocating the use of bidiagonalization procedures over other nonsymmetric procedures such as GMRES/FOM or QMR/BCG. BLanczos and BMinres are however very robust. Their convergence behavior depends solely upon the singular values of the matrix A. No other properties of A influence their behavior.

The proof in Section 3 that numerical instabilities play no role in peak (or plateau) generation in BLanczos (BMinres) when A is sufficiently well-conditioned, allowed us to look for other factors which appear to influence the formation of peaks and plateaus in such methods. That proof cannot, however, be extended to either QMR/BCG or GMRES/FOM. We know in fact that in any analogous study for GMRES/FOM or QMR/BCG we must include the effects of possible numerical instabilities. In addition for QMR/BCG we must also include the possibility of breakdowns or near breakdowns in the nonsymmetric Lanczos recursions which define these procedures [7].

Our experiments in Section 7 and [2] provide a plausible deterministic explanation for the seemingly erratic behavior of typical residual norm plots generated by BLanczos. Theorem 6.1 indicates that this explanation is also applicable to the puzzling appearance of plateaus observed in the corresponding BMinres residual norm plots. Theorem 6.1 also demonstrates that if for some equation (1), the BMinres residual norms converge well, and the corresponding BLanczos iterates are well defined, then the corresponding BLanczos residual norms also converge well. Thus, for these procedures, the minimal residual method does not converge more rapidly than the Galerkin method. We are not, however, suggesting that it is irrelevant which variant of a given pair of methods is used in practice.

[3] contains extensions of Theorem 6.1 to each of the pairs of methods, GMRES/FOM and QMR/BCG. Moreover, the results of initial experiments indicate that most of the peak and plateau formations in residual norm plots generated by QMR/BCG and GMRES/FOM, when they are applied to well-conditioned matrices, also appear to be correlated with the convergence or stabilizations of

eigenvalues in the spectra of the associated Galerkin matrices.

References

[1] P. Brown, A theoretical comparison of the Arnoldi and GMRES algorithms, *SIAM J. Sci. Statist. Comput.* 20 (1991) 58–78.

[2] J. Cullum, Peaks and plateaus in Lanczos methods for solving nonsymmetric systems of equations $Ax = b$, IBM Research Report RC 18084, IBM T.J. Watson Research Center, Yorktown Heights, NY (1992).

[3] J. Cullum and A. Greenbaum, Residual relationships within three pairs of iterative algorithms for solving $Ax = b$, *SIAM J. Matrix Anal. Appl.* 19 (1996).

[4] J. Cullum and R.A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations Vol. 1: Theory*, in: S. Abarbanel et al., eds., Progress in Scientific Computing 3 (Birkhaüser, Basel, 1985).

[5] J. Cullum, R.A. Willoughby and M. Lake, A Lanczos algorithm for computing singular values and vectors of large matrices, *SIAM J. Sci. Statist. Comput.* 4 (1983) 197–215.

[6] R.W. Freund, G.H. Golub and N. Nachtigal, Iterative solution of linear systems, *Acta Numer.* 1 (1992) 57–100.

[7] R. Freund and N. Nachtigal, QMR: a quasi-minimal residual method for non-Hermitian linear systems, *Numer. Math.* 8 (1992) 43–71.

[8] G. Golub and W. Kahan, Calculating the singular values and pseudoinverse of a matrix, *SIAM J. Numer. Anal.* 2 (1965) 197–209.

[9] G. Golub and C. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, MD, 1989).

[10] A. Greenbaum, Behavior of slightly perturbed Lanczos and conjugate gradient recurrences, *Linear Algebra Appl.* 113 (1989) 7–63.

[11] R.O. Hill and B.N. Parlett, Refined interlacing properties, *SIAM J. Matrix Anal. Appl.* 13 (1992) 239–247.

[12] C.C. Paige, Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem, *Linear Algebra Appl.* 34 (1980) 235–258.

[13] C.C. Paige, Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix, *J. Inst. Math. Appl.* 18 (1976) 341–349.

[14] C.C. Paige, Bidiagonalization of matrices and solution of linear equations, *SIAM J. Numer. Anal.* 11 (1974) 197–209.

[15] C.C. Paige and M.A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software* 8 (1982) 43–71.

[16] Y. Saad and M.H. Schultz, GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Statist. Comput.* 7 (1986) 856–869.

[17] A. van der Sluis and H.A. van der Vorst, The convergence behavior of Ritz values in the presence of close eigenvalues, *Linear Algebra Appl.* 88/89 (1987) 651–694.

[18] H.A. van der Vorst and C. Vuik, The superlinear convergence behavior of GMRES, *J. Comput. Appl. Math.* 48 (1993) 327–341.

[19] H.F. Walker, Residual smoothing and peak/plateau behavior in Krylov subspace methods, *Appl. Numer. Math.* 19 (1995) 279–286.

[20] Lu Zhou and H.F. Walker, Residual smoothing techniques for iterative methods, *SIAM J. Sci. Comput.* 15 (1994) 297–312.