MATHEMATICS OF COMPUTATION
VOLUME 56, NUMBER 194
APRIL 1991, PAGES 677–691

A CONVERGENCE ANALYSIS FOR NONSYMMETRIC LANCZOS ALGORITHMS

QIANG YE

ABSTRACT. A convergence analysis for the nonsymmetric Lanczos algorithm is presented. By using a tridiagonal structure of the algorithm, some identities concerning Ritz values and Ritz vectors are established and used to derive approximation bounds. In particular, the analysis implies the classical results for the symmetric Lanczos algorithm.

1. INTRODUCTION

Lanczos' algorithm is one of the most popular methods for computing some extreme eigenvalues of large symmetric matrices. An elegant theory and analyses of the symmetric Lanczos algorithm have been developed since the 1960's, which include error bounds of Kanial, Paige, and Saad (see [3, 7, 10] or [8]). At the same time, considerable effort has been made to generalize this work to nonsymmetric problems. The idea of tridiagonalization is naturally extended and yields a two-sided nonsymmetric algorithm (see [4, 9, 2]). However, several substantial problems, e.g., a breakdown phenomenon and a convergence analysis, remain unsolved.

The Lanczos algorithm was originally introduced as a method to tridiagonalize a general matrix [4]. Later it was found that it can be used to compute some extreme eigenvalues. It can be regarded, in particular, as a Rayleigh-Ritz projection method using Krylov subspaces; and based on this, a convergence analysis was established using the minimax theorem [3, 7, 10]. Since there is no minimax characterization for general matrices, this analysis cannot easily be generalized. Nevertheless, some approaches have been suggested in this regard. In [11] the idea using projection on Krylov subspaces is extended to show that some eigenvectors are close to the Krylov subspaces. The recent work [1] establishes some properties of the Lanczos polynomials which can be used to explain the convergence of the Lanczos algorithm. In [5, 13], by using a minimax theorem, the classical method is applied to definite matrix pencil problems. However, the result there suggests that the classical approach may not be the best for nonsymmetric problems. We therefore take another look at the Lanczos

Received January 15, 1990; revised May 18, 1990.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65F15.
Key words and phrases. Lanczos algorithm, convergence bound, tridiagonal matrices.
Research supported by a University of Calgary Research Fellowship.

0025-5718/91 $1.00 + $.25 per page

algorithm and find that it is essentially a method of approximating a tridiagonal matrix using its submatrices. It turns out that employing the tridiagonal structure of the algorithm is an appropriate approach.

We shall present a convergence analysis for the nonsymmetric Lanczos algorithms. Our analysis is based on the tridiagonal structure and is completely different from the classical approach. In particular, we shall reprove some classical results, and some of our results are even new for the classical symmetric case.

We first introduce the Lanczos algorithm in §2. Then some identities are established for tridiagonal matrices in §3. The approximation bounds are derived for Ritz values in §4 and for Ritz vectors in §5. Following that, the implications of this analysis to the symmetric Lanczos algorithm are discussed in §6. Finally, some numerical examples are presented in §7 and some remarks in §8.

Besides the standard notation in numerical analysis, we will use the following notation. I will denote an identity matrix and I_m will specify the $m \times m$ identity matrix. $e_{i,m}$ will denote the ith coordinate vector of R^m, i.e., $I_m = [e_{1,m}, \ldots, e_{m,m}]$.

2. Lanczos algorithms

In this section, we briefly introduce the Lanczos algorithms. The details can be found in [9] or [2].

Given a matrix A and two starting vectors x_1 and y_1 in C^n, the nonsymmetric Lanczos algorithm, in step m, generates sequences $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_m\}$ via a three-term recursion, so that

$$AY_m - Y_m T_m = \beta_{m+1} y_{m+1} e_{m,m}^*,$$
$$X_m^* A - T_m X_m^* = \gamma_{m+1} e_{m,m} x_{m+1}^*,$$

and

$$X_m^* Y_m = I_m,$$

where $X_m = [x_1, \ldots, x_m]$, $Y_m = [y_1, \ldots, y_m]$, and

$$T_m = \begin{pmatrix} \alpha_1 & \gamma_2 & & \\ \beta_2 & \ddots & \ddots & \\ & \ddots & \ddots & \gamma_m \\ & & \beta_m & \alpha_m \end{pmatrix}.$$

The algorithm continues until breakdown, that is, when $x_{m+1}^* y_{m+1} = 0$ at some step m. This is one of the serious problems in the nonsymmetric Lanczos algorithm (see [9] for a detailed discussion). Numerically, it is rare to have an exact breakdown. The real difficulty comes when the iteration is close to breakdown. In this theoretical analysis, we always assume that no breakdown occurs.

From the above construction, it is easy to see that

$$X_m^* A Y_m = T_m \tag{2.1}$$

and

$$X_m^* Y_m = I_m .$$

In particular, at step n, we have $X_n^* = Y_n^{-1}$. Then A is similar to T_n, and the eigenpairs of T_n give those of A. This was originally used as a method to tridiagonalize a matrix A. However, the attractive feature of the Lanczos algorithm is that it usually stops at $m \ll n$, and some extreme eigenvalues of T_m can be used to approximate some eigenvalues of A. Specifically, at step m, we find the Jordan decomposition of T_m,

$$T_m = P^* \Theta Q, \qquad P^* Q = I, \tag{2.2}$$

where Θ is the Jordan canonical form of T_m. Then the eigenvalues of T_m (or Θ) are called *Ritz values*. Further, letting

$$U = [u_1, \ldots, u_m] = X_m Q^*, \qquad V = [v_1, \ldots, v_m] = Y_m P^*, \tag{2.3}$$

we call u_i^* (resp. v_i) a *left* (resp. *right*) *Ritz vector*. We will show that some Ritz values and Ritz vectors give good approximations to the eigenpairs of A.

If A is symmetric, we take the initial vectors $x_1 = y_1$. Then the algorithm yields $X_m = Y_m$ and a symmetric T_m, which is just the classical symmetric Lanczos algorithm.

3. Tridiagonal matrices

The symmetric Lanczos algorithm has been successfully treated as a Rayleigh-Ritz projection method using Krylov subspaces. For nonsymmetric matrices, it can also be viewed as an oblique projection method (see [11]). The approach is, however, not as successful as in the symmetric case. As mentioned before, the Lanczos algorithm is closely related to its tridiagonal structure. To analyze the algorithm, we first consider tridiagonal matrices.

In the following we always denote an $n \times n$ tridiagonal matrix T_n by

$$T_n = \begin{pmatrix} \alpha_1 & \gamma_2 & & \\ \beta_2 & \ddots & \ddots & \\ & \ddots & \ddots & \gamma_n \\ & & \beta_n & \alpha_n \end{pmatrix}.$$

Lemma 3.1. *Let T_m be a tridiagonal matrix; then $e_{1,m}^* T_m^k e_{m,m} = 0$ for $k \leq m-2$ and $e_{1,m}^* T_m^{m-1} e_{m,m} = \gamma_2 \cdots \gamma_m$.*

Proof. It is easy to check that, for $k \leq m-1$,

$$e_{1,m}^* T_m^k = (*, \ldots, *, \gamma_2 \cdots \gamma_{k+1}, 0, \ldots, 0),$$

where the product of the γ's is in position $k+1$. From this, the lemma follows. □

The next two theorems establish some relations between a tridiagonal matrix and its submatrices.

Theorem 3.2. *Let T_m be the $m \times m$ leading submatrix of a tridiagonal matrix T_n. Then*

$$e_{1,n}^* T_n^k = (e_{1,m}^* T_m^k, 0) \quad \text{for } k \le m-1 \tag{3.1}$$

and

$$T_n^k e_{1,n} = \begin{pmatrix} T_m^k e_{1,m} \\ 0 \end{pmatrix}.$$

Proof. We prove (3.1) only. Let

$$T_n = \begin{pmatrix} T_m & E \\ \hat{E} & T_{n-m} \end{pmatrix},$$

where $E = \gamma_{m+1} e_{m,m} e_{1,n-m}^*$ and $\hat{E} = \beta_{m+1} e_{1,n-m} e_{m,m}^*$. If, for some $k \le m-2$,

$$e_{1,n}^* T_n^k = (e_{1,m}^* T_m^k, 0),$$

then

$$e_{1,n}^* T_n^{k+1} = (e_{1,m}^* T_m^k, 0) T_n = (e_{1,m}^* T_m^{k+1}, e_{1,m}^* T_m^k E) = (e_{1,m}^* T_m^{k+1}, 0),$$

where by Lemma 3.1, $e_{1,m}^* T_m^k E = \gamma_{m+1} e_{1,m}^* T_m^k e_{m,m} e_{1,n-m}^* = 0$. Hence the theorem follows by induction. □

Conceptually, this theorem says that T_m^k and T_n^k have essentially the same first row (column) for $k \le m-1$. Furthermore, T_m^k and T_n^k have the same $(1,1)$ element for $k \le 2m-1$, as shown in the next theorem.

Theorem 3.3. *Let T_m be the $m \times m$ leading submatrix of a tridiagonal matrix T_n. Then*

$$e_{1,n}^* T_n^k e_{1,n} = e_{1,m}^* T_m^k e_{1,m} \quad \text{for } k \le 2m-1 \tag{3.2}$$

and

$$e_{1,n}^* T_n^{2m} e_{1,n} = e_{1,m}^* T_m^{2m} e_{1,m} + \beta_2 \cdots \beta_{m+1} \gamma_2 \cdots \gamma_{m+1}. \tag{3.3}$$

Proof. We first prove by induction that, for $k \ge m+1$,

$$e_{1,n}^* T_n^k = \Big(e_{1,m}^* T_m^k + e_{1,m}^* T_m^{m-1} E \hat{E} T_m^{k-m-1} + e_{1,m}^* \sum_{i=0}^{k-m-2} G_i \hat{E} T_m^i, \; e_{1,m}^* H \Big), \tag{3.4}$$

where G_i and H are $m \times (n-m)$ matrices.

From (3.1), we obtain

$$e_{1,n}^* T_n^{m+1} = (e_{1,m}^* T_m^{m+1} + e_{1,m}^* T_m^{m-1} E \hat{E}, \; e_{1,m}^* H),$$

where $H = T_m^m E + T_m^{m-1} E T_{n-m}$. We now assume that (3.4) is true for some $k \geq m+1$. Then, for $k+1$,

$$e_{1,n}^* T_n^{k+1} = \Big(e_{1,m}^* T_m^{k+1} + e_{1,m}^* T_m^{m-1} E \hat{E} T_m^{k-m} + e_{1,m}^* \sum_{i=1}^{k-m-1} G_{i-1} \hat{E} T_m^i + e_{1,m}^* H \hat{E},\ e_{1,m}^* \hat{H} \Big),$$

where $\hat{H} = T_m^k E + T_m^{m-1} E \hat{E} T_m^{k-m-1} E + \sum_{i=0}^{k-m-2} G_i \hat{E} T_m^i E + H T_{n-m}$. Letting $\hat{G}_i = G_{i-1}$ and $\hat{G}_0 = H$, we obtain (3.4) for $k+1$. So (3.4) is proved.

Now, for $k \leq m$, (3.1) leads to (3.2). For $m+1 \leq k \leq 2m-1$ and $k = 2m$, (3.2) and (3.3) follow from (3.4) and Lemma 3.1 by a straightforward computation. □

We remark that it is possible to derive some formulae of form (3.3) with larger exponent k by using (3.4). The resulting expression, however, will be very complicated and of little use.

4. Error bounds

This section will develop error bounds for Ritz values. We define P^k to be the set of polynomials of degree not greater than k, and MP^k to be the set of monic polynomials of degree k.

Let the Lanczos algorithm be applied to a matrix A and

$$A = Z_r \Lambda Z_l^*, \qquad Z_l^* Z_r = I,$$

be the Jordan decomposition of A. Then $Z_r = [z_1^{(r)}, \ldots, z_n^{(r)}]$ (resp. $Z_l = [z_1^{(l)}, \ldots, z_n^{(l)}]$) contains the right (resp. left) eigenvectors and the generalized eigenvectors. Letting

$$X = Z_r^* X_n = (x_{ij}) \quad \text{and} \quad Y = Z_l^* Y_n = (y_{ij}), \tag{4.1}$$

with X_n, Y_n being generated by (2.1), we have

$$T_n = X^* \Lambda Y \tag{4.2}$$

and

$$X^* Y = I.$$

Note that $x_{i1} = z_i^{(r)*} x_1$ is the $z_i^{(l)}$ component of the initial vector x_1, i.e.,

$$x_1 = \sum_{i=1}^{n} x_{i1} z_i^{(l)}, \tag{4.3}$$

and $y_{i1} = z_i^{(l)*} y_1$ is the $z_i^{(r)}$ component of the initial vector y_1, i.e.,

$$y_1 = \sum_{i=1}^{n} y_{i1} z_i^{(r)}.$$

Using the properties obtained in §3, we establish the first theorem.

Theorem 4.1. *Let* $X = (x_{ij})$, $Y = (y_{ij})$ *and* $P = (p_{ij})$, $Q = (q_{ij})$ *be defined as in* (4.1) *and* (2.2), *respectively, and let*

$$\Lambda = \operatorname{diag}[\Lambda_1, \lambda_{s+1}, \ldots, \lambda_n], \qquad \Theta = \operatorname{diag}[\Theta_1, \theta_{t+1}, \ldots, \theta_m],$$

where

$$\Lambda_1 = \begin{pmatrix} \lambda_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{pmatrix} \in C^{s\times s}, \qquad \Theta_1 = \begin{pmatrix} \theta_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \theta_1 \end{pmatrix} \in C^{t\times t}.$$

Then, for any $f \in P^{2m-1}$,

$$\sum_{i=0}^{s-1} f^{(i)}(\lambda_1)\omega_i + \sum_{i=s+1}^{n} f(\lambda_i)\overline{x}_{i1}y_{i1} = \sum_{i=0}^{t-1} f^{(i)}(\theta_1)\hat{\omega}_i + \sum_{i=t+1}^{m} f(\theta_i)\overline{p}_{i1}q_{i1}, \tag{4.4}$$

and, for any $f \in MP^{2m}$,

$$\begin{aligned} &\sum_{i=0}^{s-1} f^{(i)}(\lambda_1)\omega_i + \sum_{i=s+1}^{n} f(\lambda_i)\overline{x}_{i1}y_{i1} \\ &\quad = \sum_{i=0}^{t-1} f^{(i)}(\theta_1)\hat{\omega}_i + \sum_{i=t+1}^{m} f(\theta_i)\overline{p}_{i1}q_{i1} + \prod_{j=2}^{m+1} \beta_j\gamma_j, \end{aligned} \tag{4.5}$$

where $\omega_i = \frac{1}{i!}\sum_{j=1}^{s-i} \overline{x}_{j1}y_{(j+i),1}$ *and* $\hat{\omega}_i = \frac{1}{i!}\sum_{j=1}^{t-i} \overline{p}_{j1}q_{(j+i),1}$.

Proof. Substituting (4.2) and (2.2) into (3.2) of Theorem 3.3, we obtain

$$e_{1,n}^* X^* \Lambda^k Y e_{1,n} = e_{1,m}^* P^* \Theta^k Q e_{1,m}$$

for $k \le 2m-1$. Then, for any $f \in P^{2m-1}$, we have

$$e_{1,n}^* X^* f(\Lambda) Y e_{1,n} = e_{1,m}^* P^* f(\Theta) Q e_{1,m}.$$

A straightforward computation from this proves (4.4). Similarly, we can prove (4.5), using (3.3). □

We can use this theorem to derive some identities concerning approximation errors. We state the following theorem for a diagonalizable matrix.

Theorem 4.2. *Let* A *be a matrix with* n *distinct eigenvalues and* $|\lambda_1 - \theta_k| = \min_j |\lambda_1 - \theta_j|$.

(1) *If* $k \ge 2$, *i.e.,* θ_k *is a semisimple eigenvalue, then, for any* $h \in P^{2m-2}$,

$$\begin{aligned} \lambda_1 - \theta_k = \frac{1}{h(\lambda_1)\overline{x}_{11}y_{11}} \Bigg(&- \sum_{i=2}^{n} (\lambda_i - \theta_k)h(\lambda_i)\overline{x}_{i1}y_{i1} \\ &+ \sum_{i=0}^{t-1} h^{(i)}(\theta_i)\sigma_i + \sum_{i=t+1}^{m} (\theta_i - \theta_k)h(\theta_i)\overline{p}_{i1}q_{i1} \Bigg), \end{aligned} \tag{4.6}$$

where $\sigma_i = (\theta_1 - \theta_k)\hat{\omega}_i + (i+1)\hat{\omega}_{i+1}$ *with* $\hat{\omega}_t = 0$.

(2) *If* $k = 1$, *then, for any* $h \in P^{2m-t-1}$,

$$(\lambda_1 - \theta_1)^t = \frac{1}{h(\lambda_1)\overline{x}_{11}y_{11}}\Bigg(- \sum_{i=2}^{n}(\lambda_i - \theta_1)^t h(\lambda_i)\overline{x}_{i1}y_{i1} + \sum_{i=t+1}^{m}(\theta_i - \theta_1)^t h(\theta_i)\overline{p}_{i1}q_{i1}\Bigg). \tag{4.7}$$

Proof. (1) For $k \geq 2$, we substitute $f(x) = (x - \theta_k)h(x)$ into (4.4) and obtain

$$(\lambda_1 - \theta_k)h(\lambda_1)\overline{x}_{11}y_{11} + \sum_{i=2}^{n}(\lambda_i - \theta_k)h(\lambda_i)\overline{x}_{i1}y_{i1} = \sum_{i=0}^{t-1} f^{(i)}(\theta_1)\hat{\omega}_i + \sum_{i=t+1}^{m}(\theta_i - \theta_k)h(\theta_i)\overline{p}_{i1}q_{i1}.$$

It is easy to check that

$$\sum_{i=0}^{t-1} f^{(i)}(\theta_i)\hat{\omega}_i = \sum_{i=0}^{t-1} h^{(i)}(\theta_i)\sigma_i.$$

This leads to (4.6).

(2) For $k = 1$, we substitute $f(x) = (x-\theta_1)^t h(x)$ into (4.4). Since $f^{(i)}(\theta_1) = 0$, $1 \leq i \leq t-1$, we get

$$(\lambda_1 - \theta_1)^t h(\lambda_1)\overline{x}_{11}y_{11} + \sum_{i=2}^{n}(\lambda_i - \theta_1)^t h(\lambda_i)\overline{x}_{i1}y_{i1} = \sum_{i=t+1}^{m}(\theta_i - \theta_1)^t h(\theta_i)\overline{p}_{i1}q_{i1}.$$

This leads to (4.7). □

Obviously, our method is not restricted to diagonalizable matrices. Instead, it applies to any eigenvalues of a general matrix. For instance, if λ_1 is an eigenvalue with Jordan block of size s, we can use $f(x) = (x - \lambda_1)^s h(x)$ in (4.4) and subsequently obtain similar identities. However, the results are more complicated. Such statements are therefore simply omitted.

To use the theorem, we choose for h a polynomial p so that $p(\lambda_1) = 1$ and $p(\lambda_i)$ $(i \neq 1)$, $p(\theta_i)$ $(i \neq k)$ are as small as possible. Then the right-hand side of (4.6) or (4.7) is a small number, and $\lambda_1 - \theta_k$ can be bounded by this number. Clearly, the magnitude of the bound depends on the distribution of λ_i, θ_i. On the other hand, comparison between (4.6) and (4.7) suggests that convergence of a Ritz value with Jordan block of size s is expected to slow down by an order of s.

To present some detailed bounds, we will concentrate on the case where both A and T_m are diagonalizable. Let $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $\sigma(T_m) = \{\theta_1, \ldots, \theta_m\}$ be the spectra of A and T_m, respectively, and let

$$\sigma_1 = \{\lambda_2, \ldots, \lambda_n\}, \qquad \hat{\sigma}_1 = \{\theta_2, \ldots, \theta_m\}.$$

We then define

$$\varepsilon_1^{(k)}(S) = \inf_{p\in P^k,\, p(\lambda_1)=1} \max_{x\in S} |p(x)|$$

and

$$\delta_1(S) = \max\left\{ |x-\theta_1| \prod_{\lambda\in S} \frac{|x-\lambda|}{|\lambda_1-\lambda|} : x \in \sigma_1 \cup \hat{\sigma}_1 \right\}$$

for $S \subset \sigma_1 \cup \hat{\sigma}_1$.

Corollary 4.3. *Let A and T_m be diagonalizable, with $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ and $\sigma(T_m) = \{\theta_1, \dots, \theta_m\}$. Assume that $|\lambda_1 - \theta_1| = \min_j |\lambda_1 - \theta_j|$, and let $\sigma_1 = \{\lambda_2, \dots, \lambda_n\}$, $\hat{\sigma}_1 = \{\theta_2, \dots, \theta_m\}$.*

(1) If $\sigma_1 \cup \hat{\sigma}_1 = S_1 \cup S_2$ with S_1 and S_2 disjoint, and $s = |S_2| \le 2m-2$, then

$$|\lambda_1 - \theta_1| \le \varepsilon_1^{(2m-2-s)}(S_1)\delta_1(S_2) \times \frac{(\sum_{i=2}^n |x_{i1}|^2 + \sum_{i=2}^m |p_{i1}|^2)^{1/2}}{|x_{11}|} \frac{(\sum_{i=2}^n |y_{i1}|^2 + \sum_{i=2}^m |q_{i1}|^2)^{1/2}}{|y_{11}|}. \tag{4.8}$$

(2) If $\sigma_1 = S_1 \cup S_2$ with S_1 and S_2 disjoint, and $s = |S_2| \le m-1$, then

$$|\lambda_1 - \theta_1| \le \varepsilon_1^{(m-1-s)}(S_1)\delta_1(S_2 \cup \hat{\sigma}_1) \frac{(\sum_{i=2}^n |x_{i1}|^2)^{1/2}}{|x_{11}|} \frac{(\sum_{i=2}^n |y_{i1}|^2)^{1/2}}{|y_{11}|}. \tag{4.9}$$

Proof. (1) Substituting $h(x) = p(x)\prod_{\lambda\in S_2}(x-\lambda)$ for any $p \in P^{2m-2-s}$ with $p(\lambda_1) = 1$ into (4.6), we obtain

$$\begin{aligned}
|\lambda_1 - \theta_1| &= \frac{1}{|x_{11}y_{11}|}\Bigg| - \sum_{\lambda_i\in S_1} (\lambda_i - \theta_1)p(\lambda_i)\overline{x}_{i1}y_{i1} \prod_{\lambda\in S_2} \frac{(\lambda_i-\lambda)}{(\lambda_1-\lambda)} \\
&\qquad + \sum_{\theta_i\in S_1} (\theta_i - \theta_1)p(\theta_i)\overline{p}_{i1}q_{i1} \prod_{\lambda\in S_2} \frac{(\theta_i-\lambda)}{(\lambda_1-\lambda)}\Bigg| \\
&\le \max_{x\in S_1}|p(x)|\delta_1(S_2)\left(\sum_{\lambda_i\in S_2}|x_{i1}y_{i1}| + \sum_{\theta_i\in S_2}|p_{i1}q_{i1}|\right)/|x_{11}y_{11}| \\
&\le \max_{x\in S_1}|p(x)|\delta_1(S_2)\frac{(\sum_{i=2}^n |x_{i1}|^2 + \sum_{i=2}^m |p_{i1}|^2)^{1/2}}{|x_{11}|} \\
&\qquad \times \frac{(\sum_{i=2}^n |y_{i1}|^2 + \sum_{i=2}^m |q_{i1}|^2)^{1/2}}{|y_{11}|}.
\end{aligned}$$

This proves (4.8).

(2) Substituting $h(x) = p(x)\prod_{\lambda\in S_2\cup\hat{\sigma}_1}(x-\lambda)$ for any $p \in P^{m-1-s}$ with $p(\lambda_1) = 1$ into (4.6), we obtain

$$|\lambda_1 - \theta_1| = \frac{1}{|x_{11}y_{11}|}\left|\sum_{\lambda_i\in S_1}(\lambda_i-\theta_1)p(\lambda_i)\overline{x}_{i1}y_{i1}\prod_{\lambda\in S_2\cup\hat{\sigma}_1}\frac{(\lambda_i-\lambda)}{(\lambda_1-\lambda)}\right|$$
$$\leq \max_{x\in S_1}|p(x)|\delta_1(S_2\cup\hat{\sigma}_1)\frac{(\sum_{i=2}^n|x_{i1}|^2)^{1/2}}{|x_{11}|}\frac{(\sum_{i=2}^n|y_{i1}|^2)^{1/2}}{|y_{11}|}.$$

This proves (4.9). □

We now analyze the magnitude of $\varepsilon_1^{(k)}(S_1)$. In [11], it is proved that if $k < |S_1|$, then there exists $\{\alpha_1, \ldots, \alpha_{k+1}\} \subset S_1$ such that

$$\varepsilon_1^{(k)}(S_1) = \left(\sum_{j=1}^{k+1}\prod_{i\neq j}\frac{|\alpha_i-\lambda_1|}{|\alpha_i-\lambda_j|}\right)^{-1}.$$

Furthermore, it is shown from this that $\varepsilon_1^{(k)}(S_1)$ is small when λ_1 is well separated from $\{\alpha_1, \ldots, \alpha_{k+1}\}$. For details see [11]. Another analysis can be conducted using Chebyshev polynomials [11, 6]. Let S_1 lie inside an ellipse and λ_1 lie outside of it and on the major axis. More specifically, by a shift and rotation we can assume that $\lambda_1 = 0$ and S_1 lies inside of the ellipse E which is centered at d and has foci at $d+c$ and $d-c$ and semimajor axis a with $0 < c \leq a \leq d$ (i.e., the real axis is the major axis of E and the origin lies outside of E). Let $T_k(x)$ denote the Chebyshev polynomial of degree k on the interval $[-1, 1]$ (see [6] or [8] for a detailed definition). Then

$$\min_{p\in P^k, p(\lambda_1)=1}\max_{x\in E}|p(x)| = \max_{x\in E}|p_k(x)| = T_k\left(\frac{a}{c}\right)/T_k\left(\frac{d}{c}\right),$$

where $p_k(x) = T_k(\frac{d-x}{c})/T_k(\frac{d}{c})$ (see [6] and references therein). Hence,

$$\varepsilon_1^{(k)}(S_1) \leq T_k\left(\frac{a}{c}\right)/T_k\left(\frac{d}{c}\right). \tag{4.10}$$

Since $1 \leq \frac{a}{c} \leq \frac{d}{c}$, we have $T_k(\frac{a}{c}) \leq T_k(\frac{d}{c})$. Furthermore, the bigger the difference between $\frac{a}{c}$ and $\frac{d}{c}$, the smaller is the bound of (4.10). Note that $\frac{d}{c}$ is a measure of separation of λ_1 from E, and $\frac{a}{c}$ is a measure of flatness of the ellipse E.

On the other hand, if $s = |S_2|$ is small, $\delta_1(S_2)$ is a bounded number. If $s = |S_2|$ is large and, in addition, $|x-\lambda| < |\lambda_i - \lambda|$ for most $\lambda \in S_2$ and any $x \in S_1$, then $\delta_1(S_2)$ is a product of s numbers, most of which are less than one. Hence it is a small number.

Thus, for an extreme eigenvalue λ_1, we can partition $\sigma_1 \cup \hat{\sigma}_1$ into a union of S_1 and S_2, so that S_1 lies in a flat ellipse well separated from λ_1 and $t = |S_2|$ is small. Then Corollary 4.3 says that we can expect a good approximation bound for λ_1.

An alternative partition is by taking S_1 and $S_2 \subset \sigma_1 \cup \hat{\sigma}_1$ in (4.8) (or $S_2 \subset \hat{\sigma}_1$ in (4.9), respectively), with $|S_2| = 2m-2$ (resp. $|S_2| = m-1$). Since $\varepsilon_1^{(0)}(S_1) = 1$, we have the following

Corollary 4.4. *Under the hypotheses of Corollary* 4.3, *we have*

$$|\lambda_1 - \theta_1| \le \min_{S \subset \sigma_1 \cup \hat{\sigma}_1, |S|=2m-2} \delta_1(S) \frac{(\sum_{i=2}^n |x_{i1}|^2 + \sum_{i=2}^m |p_{i1}|^2)^{1/2}}{|x_{11}|} \times \frac{(\sum_{i=2}^n |y_{i1}|^2 + \sum_{i=2}^m |q_{i1}|^2)^{1/2}}{|y_{11}|}$$

and

$$|\lambda_1 - \theta_1| \le \min_{S \subset \sigma_1, |S|=m-1} \delta_1(S \cup \hat{\sigma}_1) \frac{(\sum_{i=2}^n |x_{i1}|^2)^{1/2}}{|x_{11}|} \frac{(\sum_{i=2}^n |y_{i1}|^2)^{1/2}}{|y_{11}|}. \tag{4.11}$$

Finally, we remark that analogues of Theorem 4.2 and Corollaries 4.3 and 4.4 can be obtained by using (4.5) with polynomials of higher degree. However, this increase in degree is not significant, and all such statements are therefore omitted.

5. Ritz vectors

In the previous section, we only considered the convergence of Ritz values. We note that the behavior of Ritz vectors could be quite different from that of Ritz values. In some cases it is more appropriate to consider Ritz vectors and invariant subspaces, e.g., when there are some close eigenvalues, or eigenvalues with Jordan block of size greater than two.

In this section, we give an analysis for Ritz vectors. Again, we use the properties of tridiagonal matrices developed in §3. For simplicity, we always discuss the case where both A and T_m are diagonalizable.

Theorem 5.1. *Assume that* A *and* T_m *are diagonalizable. Then, for any* $f \in P^{m-1}$,

$$\sum_{i=1}^n z_i^{(r)} f(\lambda_i) y_{i1} = \sum_{i=1}^m v_i f(\theta_i) q_{i1}, \tag{5.1}$$

where $z_i^{(r)}$ *are the right eigenvectors and* v_i *the right Ritz vectors.*

Proof. By Theorem 3.1, for any $f \in P^{m-1}$,

$$f(T_n) e_{1,n} = \begin{pmatrix} f(T_m) e_{1,m} \\ 0 \end{pmatrix}.$$

By (4.2) and (2.2), we have

$$X^* f(\Lambda) Y e_{1,n} = \begin{pmatrix} P^* f(\Theta) Q e_{1,m} \\ 0 \end{pmatrix},$$

or

$$Z_r f(\Lambda) Y e_{1,n} = Y_n \begin{pmatrix} P^* f(\Theta) Q e_{1,m} \\ 0 \end{pmatrix} = V f(\Theta) Q e_{1,m}.$$

Expanding this, we obtain (5.1). □

For some λ_k and θ_l, which are close to each other, we choose f so that $f(\lambda_i)$ $(i \neq k)$ and $f(\theta_i)$ $(i \neq l)$ are small and $f(\lambda_k) = 1$. Assume that $f(\theta_l) q_{l1} \neq 0$. Then

$$v_l = \alpha(z_k^{(r)} + w_\varepsilon),$$

where α is a constant and

$$w_\varepsilon = \sum_{i \neq k} z_i^{(r)} f(\lambda_i) \frac{y_{i1}}{y_{k1}} - \sum_{i \neq l} v_i f(\theta_i) \frac{q_{i1}}{y_{k1}}$$

is a small vector compared to $z_k^{(r)}$. A particular choice of $g_l(x) = (x - \theta_1) \cdots (x - \theta_{l-1})(x - \theta_{l+1}) \cdots (x - \theta_m)$ yields an expression of v_l in terms of $z_k^{(r)}$.

Corollary 5.2. *Let* $g_l(x) = (x - \theta_1) \cdots (x - \theta_{l-1})(x - \theta_{l+1}) \cdots (x - \theta_m)$. *Under the hypotheses of Theorem* 5.1 *and* $q_{l1} \neq 0$, *we have*

$$v_l = \alpha \sum_{i=1}^{n} z_i^{(r)} g_l(\lambda_i) y_{i1}$$

for some constant α.

For a fixed l, there are some λ_i close to θ_j $(j \neq l)$, in which case $g_l(\lambda_i)$ is relatively small. Hence, v_l is close to some spectral subspace, though θ_l may not be close to any eigenvalue. So in this case, v_l can make a good starting vector to find the remaining eigenvalues.

6. The symmetric case

In this section we apply our techniques to the classical symmetric case. In particular, we are going to derive some generalizations of the classical bound.

For a symmetric matrix A, all the eigenvalues λ_i and the Ritz values are real. Let $\lambda_1 \leq \cdots \leq \lambda_n$ and $\theta_1 \leq \cdots \leq \theta_m$. Then $\lambda_i \leq \theta_i$ for $i = 1, \ldots, m$. The classical convergence analysis compares λ_i with θ_i, which is not necessarily the best approximation to λ_i. For example, if the initial vector x_1 has a significantly small component in the direction of the eigenvector associated with λ_1, then θ_1 will converge to λ_2 first. In such a case, a bound on $|\lambda_1 - \theta_1|$ is irrelevant. Without using the minimax theorem, our method does not require this match in ordering. This allows us to compare an eigenvalue with the Ritz value that is closest to it.

When considering the left end of the eigenvalues, we note that θ_i decreases as m increases. Then an approximation of θ_l to λ_k can only be improved if $\theta_l > \lambda_k$. Otherwise, θ_l will depart from λ_k and approach λ_{k-1}. From this point of view, we naturally consider λ_k approximated by some $\theta_l > \lambda_k$.

We have seen in §2 that the symmetric Lanczos algorithm can be obtained by taking $x_1 = y_1$ in the nonsymmetric case. Furthermore, we have $X_n = Y_n$,

$X = Y$, and $P = Q$. Combining this with the previous results, we obtain the following

Theorem 6.1. *For a fixed* λ_k, *let* $\theta_{l-1} - \lambda_k < 0 < \theta_l - \lambda_k$ (*where* $\theta_0 = -\infty$). *Then*

$$|\lambda_k - \theta_l| \le |\lambda_n - \lambda_k| \frac{\sum_{i=k+1}^{n} |x_{i1}|^2}{|x_{k1}|^2} \inf \left\{ \max_{k+1 \le i \le n} |h(\lambda_i)| : h \in \Phi \right\},$$

where $\Phi = \{h \in P^{2(m-1)} : h(\lambda_k) = 1,\ h(\theta_i) \le 0 \ \ (1 \le i \le l-1)$ *and* $h(\theta_i) \ge 0$ $(l+1 \le i \le m)\}$.

Proof. By Theorem 4.2, for $h \in \Phi$, we have

$$\begin{aligned} 0 \le -\lambda_k + \theta_l &= \frac{\sum_{i \ne k} (\lambda_i - \theta_l) h(\lambda_i) |x_{i1}|^2 - \sum_{i \ne l} (\theta_i - \theta_l) h(\theta_i) |p_{i1}|^2}{h(\lambda_k) |x_{i1}|^2} \\ &\le \frac{\sum_{i=k+1}^{n} (\lambda_i - \theta_l) h(\lambda_i) |x_{i1}|^2}{h(\lambda_k) |x_{i1}|^2} \\ &\le |\lambda_n - \lambda_k| \frac{\sum_{i=k+1}^{n} |x_{i1}|^2}{|x_{k1}|^2} \max_{k+1 \le i \le n} |h(\lambda_i)|. \end{aligned}$$

This proves the theorem. □

Now consider the polynomial

$$h(x) = (x - \theta_1)^2 \cdots (x - \theta_{l-1})^2 T_{m-l}^2 \left(\frac{2x - \lambda_{k+1} - \lambda_n}{\lambda_n - \lambda_{k+1}} \right).$$

It is easy to see that $h(x)/h(\lambda_k) \in \Phi$. Hence, we obtain the following more general form of the classical bound (see [8]).

Corollary 6.2. *Under the hypotheses of Theorem* 6.1,

$$|\lambda_k - \theta_l| \le |\lambda_n - \lambda_k| \frac{\sum_{i=k+1}^{n} |x_{i1}|^2}{|x_{k1}|^2} \prod_{i=1}^{l-1} \frac{(\lambda_n - \theta_i)^2}{(\lambda_k - \theta_i)^2} \Big/ T_{m-l}^2 \left(\frac{2\lambda_k - \lambda_{k+1} - \lambda_n}{\lambda_n - \lambda_{k+1}} \right).$$

For Ritz vectors, Theorem 5.1 and Corollary 5.2 apply and give two new results. In particular, we notice that $\sum_{i=1}^{m} v_i f(\theta_i) q_{i1}$ lies in the Krylov subspace. Taking

$$h(x) = (x - \theta_1) \cdots (x - \theta_{l-1}) T_{m-l} \left(\frac{2x - \lambda_{k+1} - \lambda_n}{\lambda_n - \lambda_{k+1}} \right)$$

in Theorem 5.1, we can obtain the bound of Saad concerning eigenvectors. We will not state this result, but refer to [10] or [8] for the details.

7. Examples

We present two simple examples in this section. Both are taken from the nonsymmetric examples of [1]. For the sake of computational convenience, we have used bound (4.11) in our calculation.

Example 1. Let the matrix be

$$\begin{aligned} A &= \operatorname{diag}[70-70i,\ -40+80i,\ 8-7i,\ -1-5i,\ 8] \\ &= \operatorname{diag}[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5], \end{aligned}$$

and the initial vectors be $x_1 = y_1 = [1, 1, 1, 1, 1]^*$. Then for $m = 3$, the three Ritz values are

$$\theta_1 = 69.9 - 70.0i, \qquad \theta_2 = -40.0 + 80.1i, \qquad \theta_3 = 4.9 - 4.0i$$

to one decimal digit of accuracy. By taking $S = \{\lambda_2, \lambda_5\}$ in (4.11), we obtain the bound 0.346 for $\lambda_1 - \theta_1$. By taking $S = \{\lambda_1, \lambda_5\}$, we obtain the bound 0.348 for $\lambda_2 - \theta_2$.

Example 2. In this example,

$$\begin{aligned} A &= U \operatorname{diag}[10,\ 13-4i,\ 7+3i,\ -80i,\ -20+90i]U^{-1} \\ &= U \operatorname{diag}[\lambda_5, \lambda_4, \lambda_3, \lambda_2, \lambda_1]U^{-1}, \end{aligned}$$

with U and the initial vectors chosen randomly as

$$U = \begin{pmatrix} 2113+2922i & 6284+5015i & 5608+9185i & 2321+2860i & 3076+6857i \\ 7560+5664i & 8497+4369i & 6624+0437i & 2312+1280i & 9330+1531i \\ 2+4826i & 6857+2693i & 7264+4819i & 2165+7783i & 2146+6971i \\ 3303+3322i & 8782+6326i & 1985+2640i & 8834+2119i & 3126+8416i \\ 6654+5935i & 684+4052i & 5443+4148i & 6525+1121i & 3616+4062i \end{pmatrix}$$

and

$$x_1 = y_1 = \begin{pmatrix} 1.0942+0.2736i \\ 0.6524-0.1925i \\ 0.6630-0.7264i \\ 0.4387+0.5967i \\ 0.0888+0.5170i \end{pmatrix}.$$

For $m = 3$, the three Ritz values are

$$\theta_1 = -19.9524+90.2763i, \quad \theta_2 = 0.1894-79.8844i, \quad \theta_3 = 11.6666+0.9774i.$$

By taking $S = \{\lambda_2, \lambda_4\}$ in (4.11), the bound for $\lambda_1 - \theta_1$ is 0.435. By taking $S = \{\lambda_1, \lambda_3\}$ the bound for $\lambda_2 - \theta_2$ is 1.62.

8. Conclusion

The analysis of the nonsymmetric Lanczos algorithm is considerably more complicated than that of the symmetric one. In this paper, we have developed a convergence analysis which leads to the classical results for symmetric matrices. To our knowledge, it is the only analysis of this kind. Furthermore, the analysis

of Ritz vectors is new and demonstrates that some Ritz vectors which do not give good approximations of eigenvectors may still be close to some small spectral subspaces.

We remark that all the approximation bounds derived in this paper are not intended to provide a practical computable estimation of the number of iterations needed, but rather to demonstrate the convergence of the Lanczos algorithm.

As is known, the error bound for the largest or smallest eigenvalue in the symmetric Lanczos algorithm depends only on the matrix A and the initial vector x_1. In contrast to this, the bounds for all eigenvalues in the nonsymmetric Lanczos algorithm depend on Ritz values as well. Unfortunately, there is no guarantee that Ritz values will be well distributed. Indeed for nonsymmetric matrices, the Ritz values can be anywhere in C. Then there could be no convergence at all. At this point we should mention that when we talk about convergence of the Lanczos algorithms, it is not strictly in the sense of mathematical convergence, even for the symmetric Lanczos algorithm. As is shown in [12], there are always contrived choices of initial vectors that give rise to nonconvergence of the Lanczos process. From this point of view, it is not reasonable to expect a bound that guarantees convergence all the time, but only a bound that reveals the convergence behavior. Of course, a bound depending only on the matrix and the initial vectors would be most desirable.

Acknowledgment

I would like to thank Dr. P. Lancaster for his interest and comments. I would also like to thank Dr. Y. Huang for his help.

Bibliography

1. G. Cybenko, *An explicit formula for Lanczos polynomials*, Linear Algebra Appl. **88/89** (1987), 99–115.
2. G. H. Golub and C. F. Van Loan, *Matrix computations*, The John Hopkins University Press, Baltimore, 1983.
3. S. Kaniel, *Estimate for some computational techniques in linear algebra*, Math. Comp. **20** (1966), 369–378.
4. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards Sect. B **45** (1950), 225–280.
5. P. Lancaster and Q. Ye, *Rayleigh-Ritz and Lanczos methods for symmetric matrix pencils* (submitted).
6. T. A. Manteuffel, *The Tchebychev iteration for nonsymmetric linear systems*, Numer. Math. **28** (1977), 307–327.
7. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, Ph.D. dissertation, University of London, 1971.
8. B. N. Parlett, *The symmetric eigenvalue problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
9. B. N. Parlett, D. R. Taylor, and Z. A. Liu, *A look-ahead Lanczos algorithm for unsymmetric matrices*, Math. Comp. **44** (1985), 105–124.

10. Y. Saad, *On the rate of convergence of the Lanczos and block Lanczos methods*, SIAM J. Numer. Anal. **17** (1980), 687–706.
11. ____, *Projection methods for solving large sparse eigenvalue problems*, Matrix Pencils (A. Ruhe and B. Kagstrom, eds.), Lecture Notes in Math., vol. 973, Springer, New York, 1983, pp. 221–244.
12. D. S. Scott, *How to make the Lanczos algorithm converge slowly*, Math. Comp. **33** (1979), 239–247.
13. Q. Ye, *Variational principles and numerical algorithms for symmetric matrix pencils*, Ph.D. thesis, University of Calgary, 1989.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CALGARY, CALGARY, ALBERTA T2N 1N4, CANADA
E-mail address: ye@uncamult.bitnet