Accuracy and Effectiveness of the Lanczos Algorithm for the Symmetric Eigenproblem

C. C. Paige
School of Computer Science
McGill University
Montreal, Quebec, Canada

Submitted by Å. Björk

ABSTRACT

Eigenvalues and eigenvectors of a large sparse symmetric matrix A can be found accurately and often very quickly using the Lanczos algorithm without reorthogonalization. The algorithm gives essentially correct information on the eigensystem of A, although it does not necessarily give the correct multiplicity of multiple, or even single, eigenvalues. It is straightforward to determine a useful bound on the accuracy of every eigenvalue given by the algorithm. The initial behavior of the algorithm is surprisingly good: it produces vectors spanning the Krylov subspace of a matrix very close to A until this subspace contains an exact eigenvector of a matrix very close to A, and up to this point the effective behavior of the algorithm for the eigenproblem is very like that of the Lanczos algorithm using full reorthogonalization. This helps to explain the remarkable behavior of the basic Lanczos algorithm.

1. INTRODUCTION

The effectiveness of the Lanczos algorithm [1] in computing eigenvalues, and perhaps eigenvectors, of large symmetric matrices has led to considerable interest in the algorithm and its properties; see for example [2] to [7], [10] to [16]. Here we will extend the work in [2] and [4] to derive some properties of the algorithm which may help in the development and use of reliable computer programs, and also give more understanding of what is actually happening in such a computation and what can be expected, as well as more confidence in the reliability of the computed results. We will only consider the basic Lanczos algorithm, as opposed to any block version such as in [7], and we will not include any reorthogonalization, such as in [6].

The Lanczos algorithm can be considered as a way of producing a symmetric tridiagonal matrix and a set of vectors related to our given symmetric $n \times n$ matrix A. These can then be used to solve several different

52 Vanderbilt Ave., New York, NY 10017 0024-3795/80/060235+24$01.75

problems. It was shown in [3] that some variants of the basic Lanczos algorithm do not work at all well, and that we need only consider the variants labeled A1 and A2 in [4]. A rounding error analysis of A2 was given in [2], and the analysis for A1 was given in [4], together with a summary of the results for A2. Although the bounds for A1 were superior, the author had not then encountered computations showing any significant difference between A1 and A2. Much more testing and thorough work by Lewis ([5] and later personal communications) showed that A1 was clearly superior in some very large problems, and from now on we will concentrate our attention on this variant.

In Sec. 2 we will present the basic algorithm A1 and summarize the results of the rounding error analysis given in [4]. In [3] it was suggested that the last elements of the eigenvectors of the tridiagonal matrix could be used to indicate how well the eigenelements of A were approximated, and in Sec. 3 we will show that this approach always gives reliable information. As a result we can always obtain useful intervals containing eigenvalues of A, though at present we cannot determine multiplicities of eigenvalues without computing eigenvectors. An approach to determining multiplicities was suggested in [3], but the method of [6] would seem preferable. In Sec. 3 it will also be shown in what way the computed eigenvalues and eigenvectors from the Lanczos algorithm correspond to Rayleigh quotients of A; in particular it will be shown that the computed eigenvalues always lie between the extreme eigenvalues of A to within close to machine accuracy.

In Sec. 4 we will indicate just how quickly the first few eigenvalues of A can be found. First we will show that we must obtain at least one very small interval containing an eigenvalue of A by the nth step. Next we will show that for the eigenproblem the algorithm behaves remarkably like the Lanczos algorithm using full reorthogonalization, at least until a very small eigenvalue interval is found. At this point several eigenvalues and eigenvectors of A are usually represented very accurately, and so the initial group of eigenvalues is found to near machine precision after about the same number of steps that the algorithm with reorthogonalization would take. This result is based on Krylov sequences and says little about orthogonality, and of course the algorithms with and without reorthogonalization behave quite differently as producers of orthogonal vectors.

It has not been proven that all eigenvalues of A will eventually be given, whereas such knowledge would be important for some problems (see for example [13]). W. Kahan [14] has encountered cases where the initial vector is orthogonal to some eigenvectors, and even with rounding errors all computed vectors remain orthogonal to these, so the best we could hope to prove is that an eigenvector of A with a nonnegligible component in the initial vector will always eventually be found by the Lanczos process, with

perhaps some reasonable restriction on the dimension of the problem. Computations in [4], [13], and elsewhere in the literature support this possibility, as do the computations mentioned at the end of Sec. 4.

2. THE ALGORITHM AND ITS ROUNDING ERRORS

For a given $n \times n$ symmetric matrix A and nonzero n-vector b, k steps of the variant of the Lanczos algorithm we will consider here can be described as follows:

$$\beta := +(b^T b)^{1/2}, \tag{2.1}$$

$$v_1 := b/\beta, \tag{2.2}$$

$$u_1 := Av_1, \tag{2.3}$$

and for $j = 1, 2, \ldots, k$ repeat steps (2.4) to (2.8):

$$\alpha_j := v_j^T u_j \tag{2.4}$$

$$w_j := u_j - \alpha_j v_j \tag{2.5}$$

$$\beta_{j+1} := +\left(w_j^T w_j\right)^{1/2},$$

$$\text{if} \quad \beta_{j+1} = 0 \quad \text{then STOP}, \tag{2.6}$$

$$v_{j+1} := w_j/\beta_{j+1} \tag{2.7}$$

$$u_{j+1} := Av_{j+1} - \beta_{j+1} v_j. \tag{2.8}$$

This algorithm is simple, elegant, and easy to program, and produces a sequence of vectors $v_1, \ldots, v_k$ and a symmetric tridiagonal matrix T_k, where we write

$$V_k \triangleq [v_1, \ldots, v_k], \qquad T_k \triangleq \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \cdot & \\ & \cdot & \cdot & \beta_k \\ & & \beta_k & \alpha_k \end{bmatrix} \tag{2.9}$$

In theory $V_k^T V_k = I$ and $\beta_{k+1} = 0$ for some $k \leqslant n$, in which case each eigenvalue of T_k is an eigenvalue of A. But on a computer orthogonality is usually completely lost, and there need not even be a small β_{j+1} in (2.6). Even so, the algorithm is amazingly effective, and it is our purpose here to show why this is so.

A rounding error analysis of (2.1) to (2.8) was given in [4] for a floating point digital computer with relative precision ε. The results will be used extensively here and for easy reference will now be summarized. Suppose A has at most m nonzero elements in any row and

$$\|A\| \triangleq \sigma, \qquad \||A|\| \triangleq \alpha\sigma, \tag{2.10}$$

where $\|\cdot\|$ without a subscript will always represent the 2-norm and $|A|$ has elements $|\alpha_{ij}|$, α_{ij} being the elements of A. The computed results were shown to satisfy

$$AV_k = V_k T_k + \beta_{k+1} v_{K+1} e_k^T + \delta V_k, \tag{2.11}$$

$$\delta V_k = [\delta v_1, \ldots, \delta v_k], \qquad I_k = [e_1, \ldots, e_k],$$

$$|v_j^T v_j - 1| \leqslant \varepsilon_0/2, \qquad j = 1, 2, \ldots, k+1, \tag{2.12}$$

while for $j = 1, 2, \ldots, k$ with $\beta_1 \triangleq 0$,

$$\|\delta v_j\| \leqslant \sigma\varepsilon_1, \tag{2.13}$$

$$\beta_{j+1} |v_j^T v_{j+1}| \leqslant 2\sigma\varepsilon_0, \tag{2.14}$$

$$|\beta_j^2 + \alpha_j^2 + \beta_{j+1}^2 - \|Av_j\|^2| \leqslant 4j(3\varepsilon_0 + \varepsilon_1)\sigma^2. \tag{2.15}$$

It is assumed that

$$\beta_{j+1} \neq 0, \quad j = 1, 2, \ldots, k; \qquad k(3\varepsilon_0 + \varepsilon_1) \leqslant 1;$$

$$\varepsilon_0 \triangleq 2(n+4)\varepsilon < \tfrac{1}{12}; \qquad \varepsilon_1 \triangleq 2(7 + m\alpha)\varepsilon. \tag{2.16}$$

Here ε_0 and ε_1 are twice the values given in [4], and by increasing the restriction on k here we could similarly decrease our error bounds. The existence of the bounds is far more important than the exact size of

the multiplicative constant. Note that n in (2.16) comes from vector inner products, while $m\alpha$ comes from matrix-vector multiplications.

Perhaps the key to understanding the computational behavior of the Lanczos algorithm is an expression for the loss of orthogonality. It was shown in [4] that if R_k is the strictly upper triangular matrix such that

$$V_k^T V_k = R_k^T + \operatorname{diag}(v_i^T v_i) + R_k, \tag{2.17}$$

then

$$T_k R_k - R_k T_k = \beta_{k+1} V_k^T v_{k+1} e_k^T + \delta R_k, \tag{2.18}$$

where δR_k is upper triangular, and it can be shown that

$$\|\delta R_k\|_F^2 \leqslant 2\sigma^2 \left[2(5k-4)\varepsilon_0^2 + 4(k-1)\varepsilon_0\varepsilon_1 + k(k-1)\varepsilon_1^2\right], \tag{2.19}$$

subscript F representing the square root of the sum of squared elements. Expressions of this form are messy, and in order to make the work a little more tractable and the results understandable, we will regularly simplify and thus weaken bounds. The important result is that the bounds exist, and since in many cases even tight bounds would not be approached in realistic examples, indications will be given where possible of the likely errors. If we define

$$\varepsilon_2 \triangleq \sqrt{2}\ \max(6\varepsilon_0, \varepsilon_1), \tag{2.20}$$

then (2.19) gives

$$\|\delta R_k\|_F \leqslant k\sigma\varepsilon_2, \tag{2.21}$$

where we realize that for very large k the bound is effectively $\sqrt{2}\ k\sigma\varepsilon_1$.

This section has described one variant of the Lanczos algorithm and its rounding errors. The computed V_k and T_k can be used in solving many problems involving A, and the results here can be used to analyse the methods used, but here we will restrict ourselves to the eigenproblem of A. We will assume that no further rounding errors occur; for example, we will assume the eigendecomposition of T_k is known exactly. The analysis will thus show the potential of the Lanczos algorithm for the eigenproblem. However, since we know the remaining computations are very well behaved, the analysis will in fact give us all the understanding we need. Chapter 9 of [2] also considers rounding errors in the remaining computations.

3. ACCURACY OF THE COMPUTED EIGENVALUES

Here we will examine how accurately the eigenvalues of T_k will approximate those of A, but to do this we first have to give a long series of definitions and results. Let the eigendecomposition of T_k be

$$T_k Y^{(k)} = Y^{(k)} \mathrm{diag}(\mu_j^{(k)}), \tag{3.1}$$

where the orthonormal matrix $Y^{(k)}$ has jth column $y_j^{(k)}$ and (i,j)th element $\eta_{ij}^{(k)}$, and the eigenvalues are ordered

$$\mu_1^{(k)} > \mu_2^{(k)} > \cdots > \mu_k^{(k)}. \tag{3.2}$$

If $\mu_j^{(k)}$ is an approximation to an eigenvalue λ_i of A, then the corresponding approximate eigenvector is $z_j^{(k)}$, the jth column of

$$Z^{(k)} \triangleq V_k Y^{(k)}. \tag{3.3}$$

We will need some properties of T_k. If $\nu_j^{(k)}$, $j=1,\ldots,k-1$, are the eigenvalues of the matrix obtained by deleting the $(t+1)$st row and column of T_k, and ordered so that

$$\mu_1^{(k)} \geqslant \nu_1^{(k)} \geqslant \mu_2^{(k)} \geqslant \cdots \geqslant \nu_{k-1}^{(k)} \geqslant \mu_k^{(k)}, \tag{3.4}$$

then it was shown in [8] that

$$(\eta_{t+1,j}^{(k)})^2 = \prod_{\substack{i=1 \\ i \neq j}}^{k} \delta_i(t+1, j, k), \tag{3.5}$$

$$\delta_i(t+1, j, k) \triangleq \begin{cases} \dfrac{\mu_j^{(k)} - \nu_i^{(k)}}{\mu_j^{(k)} - \mu_i^{(k)}}, & i = 1,2,\ldots, j-1, \\ \dfrac{\mu_j^{(k)} - \nu_{i-1}^{(k)}}{\mu_j^{(k)} - \mu_i^{(k)}}, & i = j+1,\ldots,k, \end{cases} \tag{3.6}$$

$$0 \leqslant \delta_i(t+1, j, k) \leqslant 1, \qquad i = 1,\ldots, j-1, j+1,\ldots,k. \tag{3.7}$$

Next if we apply T_k to an eigenvector of T_t, $t<k$,

$$T_k\begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_r^{(t)} y_r^{(t)} \\ \beta_{t+1}\eta_{tr}^{(t)} e_1 \end{bmatrix}, \tag{3.8}$$

then from [9, p. 171]

$$\delta_{tr} \triangleq \beta_{t+1}|\eta_{tr}^{(t)}| \geqslant \min_i |\mu_i^{(k)} - \mu_r^{(t)}|, \qquad t<k. \tag{3.9}$$

DEFINITION 1. We will say an eigenvalue $\mu_r^{(t)}$ of the $t\times t$ symmetric tridiagonal matrix T_t has *stabilized to within* δ_{tr} if for every $k>t$ we know there is an eigenvalue of T_k within δ_{tr} of $\mu_r^{(t)}$. We will say $\mu_r^{(t)}$ has *stabilized* when we know it has stabilized to within $\gamma k^{\theta}\sigma\varepsilon_2$, where γ and θ are small positive constants.

We see from (3.9) that after t steps $\mu_r^{(t)}$ has necessarily stabilized to within δ_{tr}, and this quantity will be much used in the remaining analysis. Another useful result is obtained from (3.8) by multiplying it by $y_i^{(k)T}$ to give

$$\left(\mu_i^{(k)} - \mu_r^{(t)}\right) y_i^{(k)T}\begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} = \beta_{t+1}\eta_{t+1,\,i}^{(k)}\eta_{tr}^{(t)}. \tag{3.10}$$

A basic result is given by applying eigenvectors of T_k to each side of (2.18):

$$\left(\mu_i^{(k)} - \mu_j^{(k)}\right) y_i^{(k)T} R_k y_j^{(k)} = \beta_{k+1} z_i^{(k)T} v_{k+1}\eta_{kj}^{(k)} + \varepsilon_{ij}^{(k)}, \tag{3.11}$$

$$\varepsilon_{ij}^{(k)} \triangleq y_i^{(k)T}\,\delta R_k\, y_j^{(k)},$$

$$|\varepsilon_{ij}^{(k)}| \leqslant k\sigma\varepsilon_2, \tag{3.12}$$

from (2.21). If we take $j=i$, then

$$z_j^{(k)T} v_{k+1} = -\frac{\varepsilon_{jj}^{(k)}}{\left(\beta_{k+1}\eta_{kj}^{(k)}\right)}, \tag{3.13}$$

so that with (3.9), $z_j^{(k)}$ is almost orthogonal to v_{k+1} if we have not yet

obtained a small eigenvalue interval about $\mu_j^{(k)}$, and the eigenvector approximation $z_j^{(k)}$ does not have a small norm. The possibility of this eigenvector approximation having a small norm also causes difficulty in showing that when $\mu_j^{(k)}$ has stabilized then $z_j^{(k)}$ is effectively an eigenvector of A.

DEFINITION 2. We will say an eigenpair (μ, z) *represents an eigenpair of A to within δ* if we know that

$$\frac{\|Az-\mu z\|}{\|z\|} \leqslant \delta.$$

It follows that if (μ, z) represents an eigenpair of A to within δ, then (μ, z) is an exact eigenpair of A perturbed by a matrix whose 2-norm is no greater than δ, and if μ is the Rayleigh quotient of A with z, then the perturbation will be taken symmetric [9, p. 175].

We see from (2.11) that

$$Az_j^{(k)}-\mu_j^{(k)}z_j^{(k)}=\beta_{k+1}\eta_{kj}^{(k)}v_{k+1}+\delta V_k y_j^{(k)}, \tag{3.14}$$

so from (2.12), (2.13), (3.9), and [9,p. 171], if λ_i are the eigenvalues of A, then

$$\min_i|\lambda_i-\mu_j^{(k)}| \leqslant \frac{\|Az_j^{(k)}-\mu_j^{(k)}z_j^{(k)}\|}{\|z_j^{(k)}\|}$$

$$\leqslant \frac{\delta_{kj}(1+\varepsilon_0)+k^{1/2}\sigma\varepsilon_1}{\|z_j^{(k)}\|}, \tag{3.15}$$

and if

$$\|z_j^{(k)}\| \doteq 1, \tag{3.16}$$

then $(\mu_j^{(k)}, z_j^{(k)})$ represents an eigenpair of A to within about δ_{kj}. Unfortunately computations carried out by the author indicate that (3.16) need not hold; in fact norms of 10^{-6} have been observed on a computer with $\varepsilon \doteq 10^{-11}$.

We see from (2.12) and (2.17) that

$$\|z_j^{(k)}\|^2-1=2y_j^{(k)T}R_k y_j^{(k)}+y_j^{(k)T}\mathrm{diag}\left(v_i^Tv_i-1\right)y_j^{(k)}, \tag{3.17}$$

the last term on the right hand side having magnitude no greater than $\varepsilon_0/2$.

We use (3.13) to obtain the significant part of the $(t+1)$st column of R_K:

$$V_t^T v_{t+1} = Y^{(t)} b_t, \qquad e_r^T b_t \triangleq -\frac{\varepsilon_{rr}^{(t)}}{\beta_{t+1}\eta_{tr}^{(t)}}, \tag{3.18}$$

which with (3.10) and (3.5) gives

$$y_j^{(k)T} R_k y_j^{(k)} = -\sum_{t=1}^{k-1} \eta_{t+1,j}^{(k)} \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\beta_{t+1}\eta_{tr}^{(t)}} y_j^{(k)T} \begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} \tag{3.19}$$

$$= -\sum_{t=1}^{k-1} \left(\eta_{t+1,j}^{(k)}\right)^2 \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\mu_j^{(k)} - \mu_r^{(t)}} \tag{3.20}$$

$$= -\sum_{t=1}^{k-1} \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\mu_j^{(k)} - \mu_{s(r)}^{(k)}} \prod_{\substack{i=1 \\ i\neq j \\ i\neq s(r)}}^{k} \delta_i(t+1, j, k). \tag{3.21}$$

The last equation has this form because t of the $\nu_i^{(k)}$ in (3.4) are the eigenvalues $\mu_r^{(t)}$. The index $s(r)$ indicates that the numerator of $\delta_{s(r)}(t+1, j, k)$ cancels with $1/(\mu_j^{(k)} - \mu_r^{(t)})$ in (3.20), and we know $s(r) \neq j$. These three equations give some useful insights. From (3.17), $\| z_j^{(k)} \|$ will be significantly different from unity only if the right hand sides of these last three equations are large. In this case (3.19) shows there must be a small $\delta_{tr} = \beta_{t+1} |\eta_{tr}^{(t)}|$, and some $\mu_r^{(t)}$ has therefore stabilized. Equation (3.20) shows that some $\mu_r^{(t)}$ must be close to $\mu_j^{(k)}$, and combining this with (3.19) we will show that at least one such $\mu_r^{(t)}$ has stabilized. Finally from (3.21) we see that there is at least one $\mu_s^{(k)}$ close to $\mu_j^{(k)}$, so that $\mu_j^{(k)}$ cannot be a well-separated eigenvalue of T_k. Conversely if $\mu_j^{(k)}$ is a well-separated eigenvalue of T_k, then (3.16) holds, and if $\mu_j^{(k)}$ has stabilized, then it and $z_j^{(k)}$ are a satisfactory approximation to an eigenvalue-eigenvector pair of A. We will now quantify these results.

We note from (3.12) and (2.21)

$$\sum_{r=1}^{t} \left(\varepsilon_{rr}^{(t)}\right)^2 \leqslant \sum_{r=1}^{t} \sum_{s=1}^{t} \left(\varepsilon_{rs}^{(t)}\right)^2 = \| \delta R_t \|_F^2 \leqslant t^2 \sigma^2 \varepsilon_2^2, \tag{3.22}$$

and using the Cauchy-Schwarz inequality,

$$\left(\sum_{r=1}^{t} |\varepsilon_{rr}^{(t)}| \right)^2 \leqslant \sum_{r=1}^{t} \left(\varepsilon_{rr}^{(t)}\right)^2 \sum_{r=1}^{t} 1 \leqslant t^3 \sigma^2 \varepsilon_2^2. \tag{3.23}$$

In a similar manner (3.21) with (3.7) gives

$$|y_j^{(k)T} R_k y_j^{(k)}| \leqslant \frac{k^{5/2}\sigma\varepsilon_2}{6^{1/2}\min_{i\neq j}|\mu_j^{(k)}-\mu_i^{(k)}|}, \tag{3.24}$$

which is a weak bound, but shows that if

$$\min_{i\neq j}|\mu_j^{(k)}-\mu_i^{(k)}| \geqslant k^{5/2}\sigma\varepsilon_2, \tag{3.25}$$

then

$$0.42 < \|z_j^{(k)}\| < 1.4, \tag{3.26}$$

with obvious implications for (3.15). Just as in (2.21), we can effectively replace ε_2 by $\sqrt{2}\,\varepsilon_1$ when k is very large, but it will still give a very conservative result. Note from (3.3) and (2.12) that

$$\left|\sum_{j=1}^{k}\|z_j^{(k)}\|^2 - k\right| \leqslant \frac{k\varepsilon_0}{2}. \tag{3.27}$$

It was also proven in [2, pp. 122–126] that if $\mu_j^{(k)},\ldots,\mu_{j+s}^{(k)}$ are $s+1$ eigenvalues of T_k which are close to each other but separated from the rest, then

$$\sum_{i=j}^{j+s}\|z_i^{(k)}\|^2 \doteq s+1. \tag{3.28}$$

This result was proven for a different variant of the Lanczos algorithm, but will also hold here with "separated" interpreted similarly to (3.25). The proof is too lengthy and awkward to reproduce here, but the result has an interesting possible explanation. It is possible to have several close eigenvalues of T_k corresponding to one simple eigenvalue of A; if this is the case here, then the columns of

$$Z_s \triangleq \left[z_j^{(k)},\ldots,z_{j+s}^{(k)}\right] \tag{3.29}$$

will all correspond in some sense to one eigenvector z of A having $z^Tz=1$. Ideally $Z_s = ze^T$ where e is the vector of ones, but because of the indeterminacy of eigenvectors corresponding to equal eigenvalues we apparently

obtain $Z_s \doteq ze^TQ$ for an "arbitrary" orthogonal matrix Q. This still gives (3.28), yet any particular column of Z_s can have arbitrarily small norm.

So far the results here have been parallels of results in [2] and [4], but we now give a new one.

LEMMA 3.1. *Let T_k and V_k be the result of k steps of the Lanczos algorithm in Sec. 2 with (2.16) and (2.20), and let R_k be the strictly upper triangular matrix defined in (2.17). Then for each eigenpair $(\mu_j^{(k)}, y_j^{(k)})$ of T_k, using the notation of (3.1) and following, there exists a pair of integers (r, t) with $1 \leqslant r \leqslant t < k$ such that*

$$\beta_{t+1}|\eta_{tr}^{(t)}| \leqslant \psi_{jk}, \qquad |\mu_j^{(k)} - \mu_r^{(t)}| \leqslant \psi_{jk},$$

$$\psi_{jk} \triangleq \frac{3^{-1/2}k^2\sigma\varepsilon_2}{|y_j^{(k)T}R_k y_j^{(k)}|}.$$

Proof. For $r \leqslant t < k$ write, using (3.10),

$$\gamma_{rt} \triangleq \left(\beta_{t+1}\eta_{tr}^{(t)}\right)^{-1} y_j^{(k)T}\begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} = \frac{\eta_{t+1,j}^{(k)}}{\mu_j^{(k)} - \mu_r^{(t)}}, \tag{3.30}$$

so with (3.19) and (3.20)

$$y_j^{(k)T}R_k y_j^{(k)} = -\sum_{t=1}^{k-1} \eta_{t+1,j}^{(k)} \sum_{r=1}^{t} \gamma_{rt}\varepsilon_{rr}^{(t)}, \tag{3.31}$$

$$\triangleq -e^TC\bar{y},$$

where e is the vector with every element unity, C is upper triangular with (r, t) element $\gamma_{rt}\varepsilon_{rr}^{(t)}$, and $\bar{y}$ contains the last $k-1$ elements of $y_j^{(k)}$. Letting E be the $(k-1)$-square matrix with (r, t) element $\varepsilon_{rr}^{(t)}$ and combining this with (3.22) gives

$$|y_j^{(k)T}R_k y_j^{(k)}| \leqslant \|C^Te\| \leqslant k^{1/2}\|C\|_F \tag{3.32}$$

$$\leqslant k^{1/2} \max_{r \leqslant t < k} |\gamma_{rt}| \|E\|_F$$

$$\leqslant \frac{k^2\sigma\varepsilon_2|\gamma_{rt}|}{\sqrt{3}},$$

where we take the particular r and t giving the maximum. For this r and t, (3.30) then gives the desired results

$$\delta_{tr}=\beta_{t+1}|\eta_{tr}^{(t)}|\leqslant\frac{k^2\sigma\varepsilon_2}{|\sqrt{3}\ y_j^{(k)T}R_k y_j^{(k)}|}, \tag{3.33}$$

$$|\mu_j^{(k)}-\mu_r^{(t)}|\leqslant\frac{k^2\sigma\varepsilon_2}{|\sqrt{3}\ y_j^{(k)T}R_k y_j^{(k)}|}. \tag{3.34}$$

■

COMMENTS. Bounds proportional to $k^{3/2}$ have been found for each of these expressions individually, but the author has been unable to show that one (r,t) pair must satisfy both these tighter bounds simultaneously. The present expressions show that if $\|z_j^{(k)}\|$ is significantly different from unity, then for some $t<k$ there is an eigenvalue of T_t which has stabilized and is close to $\mu_j^{(k)}$. This will be used to prove that stabilized eigenvalues of T_k are always close to eigenvalues of A, and so the Lanczos process does not produce any "spurious" eigenvalues.

THEOREM 3.1. *If, with the conditions and notation of Lemma* 3.1, *an eigenvalue* $\mu_j^{(k)}$ *of* T_k *is stabilized so that*

$$\delta_{kj}\triangleq\beta_{k+1}|\eta_{kj}^{(k)}|\leqslant\sqrt{3}\ k^2\sigma\varepsilon_2, \tag{3.35}$$

then for some eigenvalue λ_s *of* A

$$|\lambda_s-\mu_j^{(k)}|\leqslant(k+1)^3\sigma\varepsilon_2. \tag{3.36}$$

Proof. (i) Suppose (3.35) holds. If

$$|y_j^{(k)T}R_k y_j^{(k)}|<\frac{3}{8}-\frac{\varepsilon_0}{2}, \tag{3.37}$$

then from (3.17)

$$\|z_j^{(k)}\|\geqslant\tfrac{1}{2}, \tag{3.38}$$

and using (3.15), (2.20), and (2.16), it follows that (3.36) holds. The other

possibility is that (3.37) is false, and then we take $i=1$ and write

$$t_1=k, \qquad r_1=j. \tag{3.39}$$

(ii) In this case we know from (3.33) and (3.34) in Lemma 3.1 that there exist positive integers r_{i+1} and t_{i+1} with

$$r_{i+1} \leqslant t_{i+1} < t_i \tag{3.40}$$

such that

$$\delta_{t_{i+1}r_{i+1}}\left|\mu_{r_i}^{(t_i)}-\mu_{r_{i+1}}^{(t_{i+1})}\right| \leqslant \frac{t_i^2\sigma\varepsilon_2}{\sqrt{3}\left(\frac{3}{8}-\varepsilon_0/2\right)}$$

$$\leqslant \sqrt{3}\, t_i^2\sigma\varepsilon_2. \tag{3.41}$$

If the equivalent of (3.37), and so (3.38), holds for (r_{i+1},t_{i+1}), then for some eigenvalue λ_s of A

$$\left|\lambda_s-\mu_{r_{i+1}}^{(t_{i+1})}\right| \leqslant \sigma\varepsilon_2\left[2\sqrt{3}\, t_i^2(1+\varepsilon_0)+(2t_{i+1})^{1/2}\right], \tag{3.42}$$

which gives

$$\left|\lambda_s-\mu_j^{(k)}\right| \leqslant \left|\lambda_s-\mu_{r_{i+1}}^{(t_{i+1})}\right| + \sum_{p=1}^{i}\left|\mu_{r_p}^{(t_p)}-\mu_{r_{p+1}}^{(t_{p+1})}\right|$$

$$\leqslant \sigma\varepsilon_2\left[\frac{13\sqrt{3}\, t_i^2}{6}+(2t_{i+1})^{1/2}+\sqrt{3}\sum_{p=1}^{i} t_p^2\right]$$

$$\leqslant (k+1)^3\sigma\varepsilon_2, \tag{3.43}$$

as required by (3.36), this last following from (3.39) and (3.40). If the equivalent of (3.37) does not hold, then we replace i by $i+1$ and return to (ii).

We see that $T_1=\alpha_1$, $y_1^{(1)}=1$, so that $z_1^{(1)}=v_1$ certainly satisfies (3.38), proving that we must encounter an (r_{i+1},t_{i+1}) pair satisfying (3.38), which completes the proof. ■

COMMENT. It is clear from the derivation that (3.36) is not at all tight for realistic problems and is no indicator of the obtainable accuracy using the

Lanczos algorithm. In practice well-separated eigenvalues of A are usually given with an error no worse than $k\sigma\varepsilon$ times a small constant. The following shows we have an eigenvalue with a superior error bound to (3.43), and that we also have a good eigenvector approximation.

COROLLARY 3.1. *If* (3.35) *holds, then for the final* (r,t) *pair in Theorem* 3.1, $(\mu_r^{(t)}, V_t y_r^{(t)})$ *is an exact eigenpair for a matrix within* $5t^2\sigma\varepsilon_2$ *of* A.

Proof. Theorem 3.1 shows that if there is a j, $1 \leqslant j \leqslant k$, such that (3.35) holds, then there exist r and t, $1 \leqslant r \leqslant t \leqslant k$, such that

$$\delta_{tr} \leqslant \sqrt{3}\, t^2\sigma\varepsilon_2, \|z_r^{(t)}\| \geqslant \tfrac{1}{2}, \tag{3.44}$$

and both $\mu_r^{(t)}$ and $\mu_j^{(k)}$ are close to the same eigenvalue of A. It follows from (3.14) that

$$\left(A+\delta A_r^{(t)}\right)z_r^{(t)} = \mu_r^{(t)} z_r^{(t)}, \tag{3.45}$$

$$\delta A_r^{(t)} \triangleq -\left(\beta_{t+1}\eta_{tr}^{(t)} v_{t+1} + \delta V_t y_r^{(t)}\right)\frac{z_r^{(t)T}}{\|z_r^{(t)}\|^2},$$

$$\|\delta A_r^{(t)}\| \leqslant 5t^2\sigma\varepsilon_2,$$

so $z_r^{(t)}$, which lies in the range of V_r, is an exact eigenvector of a matrix very close to A, and $\mu_r^{(t)}$ is the corresponding exact eigenvalue. ■

COMMENT. This is clearly also the result we obtain for a stabilized, well-separated eigenvalue of T_k, where the exact meaning of these terms is given in (3.35) and (3.25) respectively.

We now consider the accuracy of the $\mu_j^{(k)}$ as Rayleigh quotients. Without rounding errors $\mu_j^{(k)}$ is the Rayleigh quotient of A with $z_j^{(k)}$, and this gives the best bound from (3.14) and (3.15) with $\varepsilon = 0$. Here (3.13) and (3.14) combine to give

$$z_j^{(k)T} A z_j^{(k)} - \mu_j^{(k)} z_j^{(k)T} z_j^{(k)} = -\varepsilon_{jj}^{(k)} + z_j^{(k)T}\delta V_K y_j^{(k)}, \tag{3.46}$$

so that if (3.16) holds, then $\mu_j^{(k)}$ is remarkably close to the Rayleigh quotient $\rho_j^{(k)}$. In fact, if (3.25) holds, then

$$|\rho_j^{(k)} - \mu_j^{(k)}| \leqslant 8k\sigma\varepsilon_2. \tag{3.47}$$

If $\|z_j^{(k)}\|$ is small, then it is unlikely that $\mu_j^{(k)}$ will be very close to $\rho_j^{(k)}$, but this is because a small $z_j^{(k)}$ will probably have a relatively large rounding error component and so $\rho_j^{(k)}$ will probably be inaccurate. On the other hand, (3.28) suggests that at least one of a group of close eigenvalues will have $\|z_j^{(k)}\| \gtrsim 1$, and for this (3.47) follows. In fact (3.20), (3.23), and an argument like that used in Theorem 3.1 show that every $\mu_j^{(k)}$ lies within $k^{5/2}\sigma\varepsilon_2$ of a Rayleigh quotient of A, and so with (2.16) and (2.20), all the $\mu_j^{(k)}$ lie in the interval

$$\lambda_{\min}(A) - k^{5/2}\sigma\varepsilon_2 \leqslant \mu_j^{(k)} \leqslant \lambda_{\max}(A) + k^{5/2}\sigma\varepsilon_2. \tag{3.48}$$

This is different from (3.36) in that here we do not require $\mu_j^{(k)}$ to have stabilized.

To complete this section we emphasize that whatever the size of δ_{kj}, the eigenvalue $\mu_j^{(k)}$ of T_k with eigenvector $y_j^{(k)}$ has necessarily stabilized to within $\delta_{kj} \triangleq \beta_{k+1}|e_k^T y_j^{(k)}|$. If $\mu_j^{(k)}$ is a separated eigenvalue of T_k, so that (3.25) holds, then it follows from (3.26), (3.14), and (3.15) that the eigenpair

$$\left(\mu_j^{(k)}, V_k y_j^{(k)}\right) \tag{3.49}$$

represents an eigenpair of A to within

$$2.5\left(\delta_{kj} + k^{1/2}\sigma\varepsilon_1\right), \tag{3.50}$$

and apart from reducing the constant to about unity, this is about all we can say. But if $\mu_j^{(k)}$ is one of a close group of eigenvalues of T_k, so that (3.25) does not hold, then we have found a good approximation to an eigenvalue of A. For in this case either (3.37) holds, in which case (3.38), (3.14), and (3.15) show that (3.49) represents an eigenpair of A to within (3.50), or there exists $1 \leqslant r \leqslant t < k$ such that

$$\delta_{tr}, \left|\mu_j^{(k)} - \mu_r^{(t)}\right| \leqslant \sqrt{3}\, k^2 \sigma\varepsilon_2,$$

as can be seen from Lemma 3.1. It then follows from Theorem 3.1 that $\mu_j^{(k)}$ is within $[(k+1)^3 + \sqrt{3}\, k^2]\sigma\varepsilon_2$ of an eigenvalue of A. As noted earlier, bounds like this last one are not at all tight. The δ_{kj} and $\mu_j^{(k)}$ can be computed from T_k quite quickly (see [3]), and these results show how we can obtain intervals from them which are known to contain eigenvalues of A, whether δ_{kj} is large or small.

4. CONVERGENCE OF EIGENVALUES

Theorem 3.1 showed if an eigenvalue of T_k has stabilized to within $\sqrt{3}\ k^2\sigma\varepsilon_2$, then it is within $(k+1)^3\sigma\varepsilon_2$ of an eigenvalue of A, no matter how many other eigenvalues of T_k are close, and Corollary 3.1 showed we had an eigenpair of a matrix within $5k^2\sigma\varepsilon_2$ of A. We will now show that eigenvalues do stabilize to this accuracy and give an indication of how quickly this occurs.

It was shown in [2] that at least one eigenvalue of T_k must have stabilized by $k=n$. The argument is based on (3.13), which indicates that significant loss of orthogonality implies stabilization of at least one eigenvalue. In fact, if at step k

$$\delta_{ij} \triangleq \beta_{i+1}|\eta_{ij}^{(i)}| \geqslant \sqrt{3}\ k^2\sigma\varepsilon_2, \qquad 1 \leqslant j \leqslant i < k, \tag{4.1}$$

then we have with (3.18) and the bound (3.12)

$$\|R_k\|_F^2 \leqslant \sum_{t=1}^{k-1} \frac{t^3}{3k^4} < \frac{1}{12}, \tag{4.2}$$

and if $\sigma_1 \geqslant \cdots \geqslant \sigma_k$ are the singular values of V_k, then (2.17) and (2.12) give

$$0.41 < \sigma_k \leqslant \sigma_1 < 1.6. \tag{4.3}$$

Note that if (4.1) does not hold, then we already have an eigenpair of a matrix close to A, and many eigenvalues of A are usually given accurately by this time. If we now consider the $y_j^{(k)}$ giving smallest δ_{kj} for T_k, we see from (3.18) and (3.12) that

$$\|\beta_{k+1}\eta_{kj}^{(k)}V_k^T v_{k+1}\| \leqslant k^{3/2}\sigma\varepsilon_2. \tag{4.4}$$

THEOREM 4.1. *For the Lanczos algorithm in Sec. 2, if* $n(3\varepsilon_0+\varepsilon_1) \leqslant 1$, *then at least one eigenvalue of* T_n *must be within* $(n+1)^3\sigma\varepsilon_2$ *of an eigenvalue of the* $n \times n$ *matrix* A, *and there exist* $r \leqslant t \leqslant n$ *such that* $(\mu_r^{(t)}, z_r^{(t)})$ *(see* (3.1) *to* (3.3)) *is an exact eigenpair of a matrix within* $5t^2\sigma\varepsilon_2$ *of* A.

Proof. If (4.1) does not hold for $k=n$, then an eigenvalue has stabilized to that accuracy before $k=n$. Otherwise (4.1) holds for $k=n$, so from (4.3)

V_n is nonsingular, and then (4.4) shows that for the smallest δ_{nj} of T_n

$$\delta_{nj} \leqslant \frac{n^{3/2}\sigma\varepsilon_2}{0.4} \leqslant \sqrt{3}\, n^2\sigma\varepsilon_2 \tag{4.5}$$

if $n \geqslant 2$, since $\| V_k^T v_{k+1} \| \geqslant \sigma_k \| v_{k+1} \| > 0.4$ from (4.3) and (2.12). The case of $n=1$ is trivial, so at least one eigenvalue must have stabilized to within $\sqrt{3}\ k^2\sigma\varepsilon_2$ by $k=n$, and from Theorem 3.1 this eigenvalue must be within $(k+1)^3\sigma\varepsilon_2$ of an eigenvalue of A. In fact Corollary 3.1 shows that there is an exact eigenpair $(\mu_r^{(t)}, z_r^{(t)})$, $r \leqslant t \leqslant n$, of a matrix within $5t^2\sigma\varepsilon_2$ of A. ■

COMMENT. In [10] Scott shows how to arrange a problem so that even without rounding errors no eigenvalue of A is given accurately until step $k=n$, and the above result seems to be all that can be said in the most general case.

Having shown that the Lanczos algorithm gives at least one eigenvalue of A to high accuracy by $k=n$, we now show just how quickly we can expect to find eigenvalues and eigenvectors of A in practice. To do this we will first consider the Krylov sequence on which the Lanczos algorithm and several other methods are based (see [9]). For symmetric A one way of using k steps of the Krylov sequence is to form an $n \times k$ matrix V whose columns span the range of

$$\left[v_1, Av_1, \ldots, A^{k-1}v_1 \right] \tag{4.6}$$

and use the eigenvalues of

$$V^T A V y = \mu V^T V y \tag{4.7}$$

as approximations to some of the eigenvalues of A. Clearly the Lanczos algorithm does just this in a very efficient way when there are no rounding errors. As k increases these approximations improve, and for the error-free case Kaniel [11] obtained *a priori* bounds showing just how good these approximations can be: for many distributions of eigenvalues they can be very impressive. Kaniel's results were essentially correct, but as the proofs were a little faulty they were reworked in [2].

Kaniel's results therefore apply to the theoretical Lanczos algorithm. Here we will describe how in the presence of rounding errors the Lanczos method with full reorthogonalization closely approximates the theoretical method, and in what way the Lanczos algorithm without reorthogonalization does the same.

In the presence of rounding errors, k steps of a correctly programmed Lanczos algorithm with full reorthogonalization form $k\times k$ T and $n\times k$ V so that the columns of V span the exact Krylov subspace of $A+\delta A$ starting with v_1. A fairly sloppy analysis in [12] shows that for the algorithm there under mild restrictions on problem size, and taking $k \geqslant 100$ to give simpler bounds,

$$\|\delta A\| \leqslant (67+1.6m\alpha)k^{3/2}\sigma\varepsilon, \tag{4.8}$$

$$\|V^TAV - V^TVT\| \leqslant 1.02\|\delta A\| \tag{4.9}$$

$$\|V^TV - I\| \leqslant 15k\varepsilon. \tag{4.10}$$

As a result the Lanczos algorithm with full reorthogonalization forms Krylov subspaces for a matrix very close to A, and the eigenvalues of T are very close to those of (4.7). It thus approximates the error-free case nicely.

We will now show in what way the ordinary Lanczos algorithm parallels these results.

THEOREM 4.2. *For k steps of the Lanczos algorithm in Sec. 2 with* (2.16) *and* (2.20), *and k such that* (4.1) *holds, the Lanczos vectors span Krylov spaces of a matrix within $(3k)^{1/2}\sigma\varepsilon_2$ of A.*

Proof. We have from (2.11)

$$AV_k = V_{k+1}T_{k+1,k} + \delta V_k, \tag{4.11}$$

where $T_{k+1,k}$ is the matrix of the first k columns of T_{k+1}, so that with (4.3), (2.13), and (2.20),

$$(A+\delta A_k)V_k = V_{k+1}T_{k+1,k}, \qquad \delta A_k \triangleq -\delta V_k\left(V_k^TV_k\right)^{-1}V_k^T,$$

$$\|\delta A_k\|_F = \operatorname{trace}\left(\delta A_k\delta A_k^T\right)^{1/2} \leqslant (3k)^{1/2}\sigma\varepsilon_2. \tag{4.12}$$

Since T_{k+1} is tridiagonal with nonzero next to diagonal elements, we see that for $j \leqslant k+1$, $v_1,\ldots,v_j$ span the same space as the first j Krylov vectors for $A+\delta A_k$ starting with v_1. ∎

COMMENT. This important result says that until an eigenvalue of T_{k-1} has stabilized, i.e. while (4.1) holds, the vectors $v_1,\ldots,v_{k+1}$ we compute correspond to an exact Krylov sequence for a matrix very close to A. We

cannot in general find an equivalent symmetric perturbation δA_k with a small bound, but this does not worry us too much, as even with complete reorthogonalization we can say no better, at least before an eigenvalue has stabilized — and the bound (4.8) is actually worse than that for (4.12). In fact the result (4.12) is better than we could hope for if we computed the Krylov sequence directly. As a result of this and (3.45), the Lanczos algorithm can be thought of as a numerically stable way (in the backward error analysis sense [9]) of computing a Krylov sequence, at least until the corresponding Krylov subspace contains an exact eigenvector of a matrix within $5k^2\sigma\varepsilon_2$ of A. However, it can be shown that a Krylov subspace can be very sensitive to small perturbations in A.

In the particular case here where T_k and V_k are used to solve the eigenproblem of A, if we follow (4.6) and (4.7) we would like the eigenvalues and vectors of T_k to be close to those of

$$V_k^T A V_k y = \mu V_k^T V_k y, \qquad y^T y = 1, \tag{4.13}$$

as we showed would be the case with full reorthogonalization. In fact with (4.1) holding, the range of V_k is very much what we would expect using full reorthogonalization, so the eigensolutions of (4.13) correspond closely to those we could obtain in the Lanczos method with full reorthogonalization, and would thus approximate the error-free case nicely.

THEOREM 4.3. *If V_k comes from the Lanczos algorithm in Sec. 2 with (2.16) and (2.20), and (4.1) holds, then for μ and y in (4.13), $(\mu, V_k y)$ is an exact eigenpair for a matrix within $2.5\delta + 2k^{1/2}\sigma\varepsilon_2$ of A, where we have defined*

$$\eta \triangleq e_k^T y, \qquad \delta \triangleq \beta_{k+1}|\eta|. \tag{4.14}$$

Proof. Define

$$\begin{aligned} r &\triangleq A V_k y - \mu V_k y \qquad (4.15)\\ &= V_k(T_k - \mu I)y + \beta_{k+1}\eta v_{k+1} + \delta V_k y, \end{aligned}$$

where we make use of (2.11). Since from (4.13) $V_k^T r = 0$,

$$(T_k - \mu I)y = -\left(V_k^T V_k\right)^{-1} V_k^T(\beta_{k+1}\eta v_{k+1} + \delta V_k y), \tag{4.16}$$

$$r = P_k(\beta_{k+1}\eta v_{k+1} + \delta V_k y), \tag{4.17}$$

where P_k is the projector orthogonal to the range of V_k. But (4.4) gives with (2.12), (2.13), and (4.3)

$$\| V_k^T v_{k+1} \| \leqslant (3k)^{-1/2} \tag{4.18}$$

$$\| P_k v_{k+1} \|^2 = v_{k+1}^T P_k v_{k+1} = v_{k+1}^T v_{k+1} - v_{k+1}^T V_k \left(V_k^T V_k \right)^{-1} V_k^T v_{k+1}, \tag{4.19}$$

$$1 - \varepsilon_0 - 2/k \leqslant \| P_k v_{k+1} \|^2 \leqslant 1 + \varepsilon_0, \tag{4.20}$$

$$\| r \| \leqslant \delta(1+\varepsilon_0) + (k/2)^{1/2} \sigma \varepsilon_2, \tag{4.21}$$

$$0.41 \leqslant \| V_k y \| \leqslant 1.6. \tag{4.22}$$

We then have from (4.15)

$$(A - \delta A) V_k y = \mu V_k y, \qquad \delta A \triangleq \frac{r y^T V_k^T}{\| V_k y \|^2}$$

$$\| \delta A \|_F = \frac{\| r \|}{\| V_k y \|} \leqslant 2.5\delta + 2k^{1/2} \sigma \varepsilon_2, \tag{4.23}$$

which completes the proof. ■

If we now order the eigenvalues $\mu_i^{(k)}$ of T_k so that

$$\delta_{k1} \geqslant \delta_{k2} \geqslant \cdots \geqslant \delta_{kk}, \tag{4.24}$$

and in the remainder of Sec. 4 assume (4.1) is also true for $i = k$, then for any eigenpair of (4.13), Eq. (4.16) gives

$$\| T_k y - \mu y \| \leqslant \left(7 + \frac{6k\delta}{\delta_{kk}} \right) k^{1/2} \sigma \varepsilon_{22}, \tag{4.25}$$

$$| \mu_m^{(k)} - \mu | \triangleq \min_i | \mu_i^{(k)} - \mu | \leqslant \left(7 + \frac{6k\delta}{\delta_{kk}} \right) k^{1/2} \sigma \varepsilon_2$$

$$\leqslant 7k^{1/2} \sigma \varepsilon_2 + \frac{6\delta}{(3k)^{1/2}}. \tag{4.26}$$

In fact, for any $t>k$

$$T_t\begin{bmatrix} y \\ 0 \end{bmatrix}=\begin{bmatrix} T_k y \\ \beta_{k+1}\eta e_1 \end{bmatrix},$$

and combining this with (4.25),

$$\min_i |\mu_i^{(t)}-\mu| \leqslant 7k^{1/2}\sigma\varepsilon_2+\delta\left[1+k^3\left(\frac{6\sigma\varepsilon_2}{\delta_{kk}}\right)^2\right]^{1/2}$$

$$\leqslant 7k^{1/2}\sigma\varepsilon_2+\delta\left(1+\frac{12}{k}\right)^{1/2}. \tag{4.27}$$

We can combine (4.26) and (4.27) to give

$$\min_i |\mu_i^{(t)}-\mu_m^{(k)}| \leqslant 14k^{1/2}\sigma\varepsilon_2+2.5\delta \qquad \text{if} \quad k\geqslant 6, \tag{4.28}$$

so that an eigenvalue of T_k close to μ has certainly stabilized to about 2.5δ, where from (4.23) μ is within about 2.5δ of an eigenvalue of A.

These results go a long way towards explaining the excellent behavior of the Lanczos algorithm. When (4.1) holds with $i\leqslant k$, all the pairs $(\mu, V_k y)$ in (4.13) could, apart from normalization, have come from k steps of the Lanczos algorithm with reorthogonalization. From (4.23) we know that any pair $(\mu, V_k y)$ represents an eigenpair of A to within about 2.5δ, where δ is defined in (4.14). But (4.26) to (4.28) show that there is an eigenvalue $\mu_m^{(k)}$ of T_k which is within about δ of μ for $k\geqslant 12$, and has stabilized to about 2.5δ, and we know then that $\mu_m^{(k)}$ is within about 3.5δ of an eigenvalue of A. Note that for small k (4.26) does not say μ and $\mu_m^{(k)}$ will be very close unless δ is as small as δ_{kk}.

It can also be shown that for each $\mu_j^{(k)}$ of T_k

$$\min_{\mu \text{ in } (4.13)} |\mu-\mu_j^{(k)}| \leqslant \left(7+\frac{6k\delta_{kj}}{\delta_{kk}}\right)k^{1/2}\sigma\varepsilon_2$$

$$\leqslant 7k^{1/2}\sigma\varepsilon_2+\frac{6\delta_{kj}}{(3k)^{1/2}}, \tag{4.29}$$

so that when we know $(\mu_j^{(k)}, V_k y_j^{(k)})$ represents an eigenpair of A to within about δ_{kj}, we know there is a μ of (4.13) within about δ_{kj} of $\mu_j^{(k)}$, at least for $k\geqslant 12$.

These results say that until an eigenvalue has stabilized, the Lanczos algorithm behaves very much like the error-free process, or the algorithm with reorthogonalization. It may not have all its eigenvalues close to those in the latter processes, but the resulting eigenpairs will still represent those of A to within about the same amount, and (4.29) shows that the most stabilized eigenvalue of T_k is very close to an eigenvalue in these latter processes.

These results help us to understand why in practice the algorithm will give several eigenvalues to machine precision in the same number of steps that the algorithm with full reorthogonalization takes. For example, using FORTRAN double precision on the IBM 360 at McGill University in 1976, Nabil Rafla took

$$A_{100\times 100} = \begin{bmatrix} 5 & -4 & 1 & & & & \\ -4 & 6 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \cdot & \cdot & \cdot & \cdot & \cdot & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 6 & -4 \\ & & & & 1 & -4 & 5 \end{bmatrix}$$

and applied the algorithm with and without reorthogonalization. For $v_1 = e_1$ neither algorithm gave any eigenvalues of A to machine accuracy until step 78, but by step 85 both had given 42 eigenvalues of A to machine accuracy. Then the algorithms diverged, the one without reorthogonalization only finding 62 eigenvalues of A in 100 steps, and taking 350 steps to find all the eigenvalues of A. This is perhaps a biased example in that high precision was used and it took many steps before the first eigenvalue stabilized. With v_1 the vector of ones, normalized, the first eigenvalues stabilized at $k=40$ and the algorithms diverged at $k=50$ when both had found 32 eigenvalues to machine accuracy. At $k=100$ the basic algorithm had only found 46, and the first repeated eigenvalue appeared at $k=110$. All eigenvalues of A were given by $k=400$.

It is hard to imagine there being a better algorithm than the basic Lanczos algorithm when only a few eigenvalues of a large sparse symmetric matrix are needed, and these eigenvalues are among the early ones to be found, as will often be the case. Even when all the eigenvalues are required the algorithm can be surprisingly effective; see for example [13].

It should be a straightforward exercise to show that the behavior described here also holds for the skew-symmetric case, and that much of the good behavior carries over to other more general problems (see for example [15]). Section 4 of [16] indicates the relationship in finite precision between the Lanczos algorithm and the method of conjugate gradients, and the

results here could help in analyzing that method, though this may not be an easy task. The author has found the variant of the conjugate gradients method in Sec. 4 of [16] to be very reliable.

Finally the author would like to point out once more (see [2] and [4]) that he is convinced that the expression (2.18) with the bound (2.21) will play an important role in the realistic analysis of any algorithm based on the Lanczos method, and this of course includes the conjugate gradient methods. Here this expression and bound led to Lemma 3.1, and so to all the subsequent results.

Talks with John Lewis, Beresford Parlett, and Henk van der Vorst were very helpful to the author. Nabil Rafla carried out many computer experiments that provided useful insights. The work was supported by Natural Sciences and Engineering Research Council of Canada Grant A8652, and was partly carried out while the author was on sabbatical from the Department of Mathematics, Imperial College, London, England, enjoying the hospitality of Mike Bernal and others in the numerical group.

REFERENCES

1 C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Standards* 45: 255–282 (1950).
2 C. C. Paige, The computation of eigenvalues and eigenvectors of very large sparse matrices, Ph. D. Thesis, Univ. of London, London, 1971.
3 C. C. Paige, Computational variants of the Lanczos method for the eigenproblem, *J. Inst. Math. Appl.* 10: 373–381 (1972).
4 C. C. Paige, Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix, *J. Inst. Math. Appl.* 18: 341–349 (1976).
5 J. G. Lewis, Algorithms for sparse matrix eigenvalue problems, Ph.D. Thesis, Report STAN-CS-77-595, Computer Science Dept., Stanford Univ., Stanford, Calif., 1977.
6 B. N. Parlett and D. S. Scott, The Lanczos algorithm with selective orthogonalization, *Math. Comp.* 33: 217–238 (1979).
7 R. R. Underwood, An iterative block Lanczos method for the solution of large, sparse symmetric eigenproblems, Ph.D. Thesis, Report STAN-CS-75-496, Computer Science Dept., Stanford Univ., Stanford, Calif., 1975.
8 R. C. Thompson and P. McEnteggert, Principal submatrices II: The upper and lower quadratic inequalities, *Linear Algebra and Appl.* 1: 211–243 (1968).
9 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.
10 D. S. Scott, How to make the Lanczos algorithm converge slowly, *Math. Comp.* 33: 239–247 (1979).
11 S. Kaniel, Estimates for some computational techniques in linear algebra, *Math. Comp.* 20: 369–378 (1966).

12 C. C. Paige, Practical use of the symmetric Lanczos process with re-orthogonalization, *Nordisk Tidskr. Informationsbehandling* (*Bit*) 10: 183–195 (1970).
13 J. T. Edwards, D. C. Licciardello, and D. J. Thouless, Use of the Lanczos method for finding complete sets of eigenvalues of large sparse matrices, *J. Inst. Math. Appl.* 23: 277–283 (1979).
14 W. Kahan, personal communication.
15 J. M. van Kats and H. A. van der Vorst, Automatic monitoring of Lanczos-schemes for symmetric or skew-symmetric generalized eigenvalue problems, Report TR-7, Academisch Computer Centrum, Utrecht, The Netherlands, 1977.
16 C. C. Paige and M. A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM J. Numer. Anal.* 12: 617–629 (1975).

Received 25 May 1979