Numer. Math. 31, 183–208 (1978)

Numerische Mathematik

Adaptive Procedure for Estimating Parameters for the Nonsymmetric Tchebychev Iteration *

Thomas A. Manteuffel

Applied Mathematics Division 8325, Sandia Laboratories, Livermore, CA 94550, USA

Summary. An iteration based upon the Tchebychev polynomials in the complex plane can be used to solve large sparse nonsymmetric linear systems whose eigenvalues lie in the right half plane. The iteration depends upon two parameters which can be chosen from knowledge of the convex hull of the spectrum of the linear operator. This paper deals with a procedure based upon the power method for dynamically estimating the convex hull of the spectrum. The stability of the procedure is discussed in terms of the field of values of the operator. Results show the adaptive procedure to be an effective method of determining parameters. The Tchebychev iteration compares favorably with several competing iterative methods.

Subject Classifications. AMS(MOS): 65F10; CR: 5.14.

Introduction

In the mathematical modeling of physical phenomona one often encounters partial differential equation boundary value problems. A standard technique for solving these problems is to approximate the solution of the differential system at a discrete set of points by the solution of a linear system. In general these systems are large, sparse, and often nonsymmetric.

The Tchebychev iteration, an iterative method based upon the Tchebychev polynomials in the complex plane, can be used to solve nonsymmetric linear systems whose eigenvalues lie in the right (left) half complex plane. Such systems may arise in many ways. For example, if the differential operator being approximated is a positive definite self-adjoint operator perturbed by first order terms, then the associated linear operator will most likely be a positive definite operator perturbed by a skew-symmetric operator. Such an operator will have eigenvalues in the right half plane. In addition, many factorization and splitting

* This work was supported in part by the National Science Foundation under grants NSF GJ-36393 and DCR 74-23679 (NSF)

0029-599X/78/0031/0183/$05.20

techniques applied to symmetric systems yield nonsymmetric systems whose spectrums lie in the right half plane.

It is a standard result, shown in Manteuffel [20], that the Tchebychev iteration for an $N \times N$ real valued linear system

$$A\mathbf{x} = \mathbf{b} \tag{1.1}$$

whose eigenvalues lie in the right half plane can be carried out with two parameters, d and c, as follows:

Given $\mathbf{x}_0$, let

$$\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0,$$
$$\Delta_0 = \frac{1}{d}\mathbf{r}_0,$$
$$\mathbf{x}_1 = \mathbf{x}_0 + \Delta_0,$$

and in general

$$\mathbf{r}_n = \mathbf{b} - A\mathbf{x}_n,$$
$$\Delta_n = \alpha_n \mathbf{r}_n + \beta_n \Delta_{n-1},$$
$$\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta_n,$$

where

$$\alpha_n = \frac{2}{c}\frac{T_n\left(\frac{d}{c}\right)}{T_{n+1}\left(\frac{d}{c}\right)}, \qquad \beta_n = \frac{T_{n-1}\left(\frac{d}{c}\right)}{T_{n+1}\left(\frac{d}{c}\right)}; \tag{1.2}$$

and $T_n(z) = \cosh(n \cosh^{-1}(z))$ is the n^{th} Tchebychev polynomial[1]. The values of α_n, β_n can be generated recursively as follows:

$$\alpha_1 = \frac{2d}{2d^2 - c^2}, \qquad \beta_1 = d\alpha_1 - 1,$$
$$\alpha_n = \left[d - \left(\frac{c}{2}\right)^2 \alpha_{n-1}\right]^{-1}, \qquad \beta_n = d\alpha_n - 1. \tag{1.3}$$

It was also shown in Manteuffel [20] that if the convex hull of the spectrum of A (denoted $H(A)$) were known, then the parameters d and c could be chosen to be optimal in a minimax sense. In this paper a procedure will be discussed for estimating $H(A)$ during iteration from the sequence of residual vectors. Section 2 will provide background and notation. Section 3 will exhibit a variant of the procedure based upon the power method. It will be shown in Section 4 that the power method yields eigenvalue estimates that lie in the field of values of A (denoted $F(A)$) (see Section 4). The relation between $F(A)$ and $H(A)$ will be discussed in terms of a measure of the normality of A. Two variants of the

[1] The branch of $\cosh^{-1}$ with nonnegative real part will be used in the definition of the Tchebychev polynomials

procedure based upon the modified power method will be described in Section 5. Section 6 will describe how the eigenvalue estimates can be put together to form an approximation to $H(A)$. In Section 7 results of tests comparing the method to three other iterative methods for solving nonsymmetric linear systems will be discussed. Section 8 will discuss the acceleration of precondition systems and matrix splittings.

Carre' [2], Reid [23], and Hageman and Kellog [10] among others have done work on dynamic estimation of the optimal S.O.R. parameter. In addition, Hageman [11] discussed a procedure for dynamically estimating a parameter for the Tchebychev iteration for nonsymmetric systems. However, he was restricted to a single parameter representing the eccentricity of a family of ellipses centered at the origin. Wrigley [33] first addressed the problem of dynamic estimation in the two parameter Tchebychev iteration for symmetric as well as nonsymmetric systems. Diamond [4] developed a dynamic procedure for estimation of the Tchebychev parameters for positive definite systems and showed that the procedure was stable. Manteuffel [19, 20] provided a solution to the problem of finding optimal Tchebychev parameters in terms of the spectrum of a nonsymmetric system. This paper deals with a procedure for dynamic estimation of the spectrum of a nonsymmetric system and the stability of the estimates.

2. Background

If we let $\mathbf{e}_n = \mathbf{x} - \mathbf{x}_n$ be the error after n steps of the iteration (1.2) we have (Rutishauser [6], p. 25)

$$\mathbf{e}_n = P_n(A)\,\mathbf{e}_0 \tag{2.1}$$

where

$$P_n(\lambda) = \frac{T_n\left(\frac{d-\lambda}{c}\right)}{T_n\left(\frac{d}{c}\right)} \tag{2.2}$$

is the n^{th} scaled and translated Tchebychev polynomial. The parameters d and c are restricted so that

$$\begin{aligned} &0 < d, \\ &c^2 < d^2. \end{aligned} \tag{2.3}$$

From the definition of cosh we have

$$\begin{aligned} P_n(\lambda) &= \frac{e^{n\cosh^{-1}\left(\frac{d-\lambda}{c}\right)} + e^{-n\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}}{e^{n\cosh^{-1}\left(\frac{d}{c}\right)} + e^{-n\cosh^{-1}\left(\frac{d}{c}\right)}} \\ &= \left(\frac{e^{\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}}{e^{\cosh^{-1}\left(\frac{d}{c}\right)}}\right)^n \cdot \left(\frac{1+e^{-2n\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}}{1+e^{-2n\cosh^{-1}\left(\frac{d}{c}\right)}}\right). \end{aligned} \tag{2.4}$$

Using the identity

$$\cosh^{-1}(z)=\ln(z+(z^2-1)^{1/2}), \tag{2.5}$$

we can write

$$P_n(\lambda)=S(\lambda)^n\,Q_n(\lambda), \tag{2.6}$$

where

$$S(\lambda)=\left(\frac{e^{\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}}{e^{\cosh^{-1}\left(\frac{d}{c}\right)}}\right)=\frac{d-\lambda+((d-\lambda)^2-c^2)^{1/2}}{d+(d^2-c^2)^{1/2}} \tag{2.7}$$

and

$$Q_n(\lambda)=\frac{1+e^{-2n\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}}{1+e^{-2n\cosh^{-1}\left(\frac{d}{c}\right)}}. \tag{2.8}$$

Since the branch of $\cosh^{-1}$ with nonnegative real part is used and since (2.3) insures that $\frac{d}{c}\notin[-1,1]$, then

$$\lim_{n\to\infty} Q_n(\lambda)=1, \quad \text{for } \frac{d-\lambda}{c}\notin[-1,1].$$

If $\delta=e^{-\mathrm{Re}\left(\cosh^{-1}\left(\frac{d}{c}\right)\right)}$, then

$$0\leqq|Q_n(\lambda)|\leqq\frac{2}{1-\delta^n}, \quad \text{for } \frac{d-\lambda}{c}\in[-1,1]. \tag{2.9}$$

Notice that those values of λ for which Q_n does not approach unity quickly, namely those for which $|e^{\cosh^{-1}\left(\frac{d-\lambda}{c}\right)}|$ is near unity, are precisely those for which $S(\lambda)^n$ approaches zero most rapidly. With this in mind, we write

$$P_n(\lambda)\doteq S(\lambda)^n \tag{2.10}$$

for large n.

There exists an operator equivalent to Equation (2.10) (Dunford and Schwartz [5]). Let $S=S(A)$ be the operator such that

$$P_n(A)\doteq S^n \tag{2.11}$$

for large n. In light of (2.1) it is desirable to make the spectral radius of $S=S(A)$ as small as possible. The eigenvalues of S are given by

$$\sigma_i=S(\lambda_i) \tag{2.12}$$

where λ_i is an eigenvalue of A. Thus, d and c should be chosen to satisfy

$$\min_{d,c}\max_{\lambda_i}|S(\lambda_i)|. \tag{2.13}$$

The solution to the mini-max problem (2.13) can be found in terms of the eigenvalues of A that are vertices of $H(A)$ (Manteuffel [20], Sections 3 and 4).

The parameters d and c represent the center and focal length of a family of ellipses in the complex plane, $\mathbf{C}$. Each point $\lambda \in \mathbf{C}$ is associated with an *asymptotic convergene factor,*

$$r(\lambda) = |S(\lambda)| = \left| \frac{(d-\lambda) + ((d-\lambda)^2 - c^2)^{1/2}}{d + (d^2 - c^2)^{1/2}} \right|. \tag{2.14}$$

The ellipses are level lines or $r(\lambda)$ with $r(\lambda)$ increasing monotonically as λ moves outward from the common foci, $d+c$ and $d-c$. The member of this family of ellipses passing through the origin determines a *region of convergence;* that is, if the spectrum of A lies inside this ellipse then $r(\lambda_i) < 1$ for each eigenvalue λ_i; and the iteration (1.2) will converge.

From (2.1) and (2.11) we see that for large n each step of the iteration (1.2) will cause the error in the direction of the eigenvector associated with the eigenvalue λ_i to be multiplied by $r(\lambda_i)$. If A has a complete set of eigenvectors, then after a sufficiently large number of steps of the iteration (1.2), the error will become nearly a linear combination of those eigenvectors associated with the eigenvalues with largest convergence factors; that is, with the eigenvalues on the outermost members of the family of ellipses determined by d and c.

In general, the matrix A may have nonlinear elementary divisors. Suppose we write

$$\mathbf{e}_0 = \sum_{i=1}^{m} \alpha_i \mathbf{w}_i, \tag{2.15}$$

where $\mathbf{w}_i$ is a member of the invariant subspace of A of dimension d_i associated with λ_i. Let

$$\mathbf{w}_i = \sum_{j=1}^{d_i} \beta_i \mathbf{v}_{ij} \tag{2.16}$$

where the $\mathbf{v}_{ij}$'s form the basis of the Jordan form of A; that is,

$$(A-\lambda_i)\, v_{ij} = v_{ij-1},$$
$$(A-\lambda_i)\, v_{il} = 0.$$

From (2.1) we have

$$\mathbf{e}_n = P_n(A)\, \mathbf{e}_0 = \sum_{i=1}^{m} \alpha_i P_n(A)\, \mathbf{w}_i. \tag{2.17}$$

Using the formula for a polynomial of a Jordan block (Varga [29], p. 14) we get

$$P_n(A)\, \mathbf{w}_i = \sum_{j=1}^{d_i} \left(\sum_{k=0}^{d_i - j} \beta_{j+k} \frac{P_n^{(k)}(\lambda_i)}{k!} \right) \mathbf{v}_{ij}. \tag{2.18}$$

It is easy to show that the derivatives of the scaled and translated Tcheb-

ychev polynomials satisfy the relation

$$P_n^{(k)}(\lambda) = n^k S(\lambda)^n B_k(\lambda, n), \tag{2.19}$$

where $B_k(\lambda, n)$ is uniformly bounded for all n. Combining (2.17), (2.18), and (2.19) we have

$$\mathbf{e}_n = \sum_{i=1}^{m} \alpha_i S(\lambda_i)^n \left(\sum_{j=1}^{d_i} \left(\sum_{K=0}^{d_i - j} \beta_{j+k} \frac{n^k B_k(\lambda_i, n)}{k!} \right) \mathbf{v}_{ij} \right). \tag{2.20}$$

From (2.14) and (2.20) we see that for sufficiently large n, the error will be nearly a linear combination of the invariant subspaces associated with the eigenvalues of A having the largest convergence factors, namely those lying on the ellipse furthest from the common foci. It is precisely these eigenvalues that we wish to estimate, for they are the verticies of $H(A)$. In light of the relation

$$\mathbf{r}_n = A\mathbf{e}_n, \tag{2.21}$$

the residual will also become nearly a linear combination of these invariant subspaces. We exploit this property in the next section.

3. Power Method

Given parameters d and c suppose λ_1 and λ_2 are the dominant and subdominant eigenvalues of A; that is,

$$r(\lambda_1) \geqq r(\lambda_2) > r(\lambda_i) \qquad i \neq 1, 2. \tag{3.1}$$

If λ_1 and λ_2 have multiplicity 1, then after n steps of (1.2) the residual can be written

$$\mathbf{r}_n = \beta_1 \mathbf{v}_1 + \bar{\beta}_1 \bar{\mathbf{v}}_1 + \beta_2 \mathbf{v}_2 + \bar{\beta}_2 \bar{\mathbf{v}}_2 + \boldsymbol{\varepsilon}_n, \tag{3.2}$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are the normalized eigenvectors associated with λ_1 and λ_2. If n is large, then $\boldsymbol{\varepsilon}_n$ is small compared to the other terms. Since A and $\mathbf{r}_0$ are real valued, the eigenvalues and eigenvectors will appear in complex conjugate pairs.

The power method (Wilkinson [30], pp. 570–571) is ideally suited to this situation. Consider the Krylov sequence with respect to A and $\mathbf{r}_n$,

$$\mathbf{u}_0 = \mathbf{r}_n, \quad \mathbf{u}_1 = A\mathbf{r}_n, \ldots, \mathbf{u}_j = A^j \mathbf{r}_n, \ldots. \tag{3.3}$$

We have

$$\mathbf{u}_j = \lambda_1^j \beta_1 \mathbf{v}_1 + \bar{\lambda}_1^j \bar{\beta}_1 \bar{\mathbf{v}}_1 + \lambda_2^j \bar{\beta}_2 \bar{\mathbf{v}}_2 + \bar{\lambda}_2^j \bar{\beta}_2 \bar{\mathbf{v}}_2 + A^j \boldsymbol{\varepsilon}_n. \tag{3.4}$$

If we let

$$\begin{aligned} p_4(z) &= z^4 + \rho_3 z^3 + \rho_2 z^2 + \rho_1 z + \rho_0 \\ &= (z - \lambda_1)(z - \bar{\lambda}_1)(z - \lambda_2)(z - \bar{\lambda}_2), \end{aligned} \tag{3.5}$$

then, neglecting the ε_n terms we have[2]

$$\|\mathbf{u}_4+\rho_3\mathbf{u}_3+\rho_2\mathbf{u}_2+\rho_1\mathbf{u}_1+\rho_0\mathbf{u}_0\|\doteq 0. \tag{3.6}$$

Choosing $p_4(z)$ to minimize (3.6) is equivalent to solving the least squares system

$$\begin{pmatrix} | & | & | & | \\ \mathbf{u}_0 & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ | & | & | & | \end{pmatrix}\begin{pmatrix}\rho_0\\ \rho_1\\ \rho_2\\ \rho_3\end{pmatrix}+\begin{pmatrix}|\\ \mathbf{u}_4\\ |\end{pmatrix}=\mathbf{0}. \tag{3.7}$$

Using (3.3), the normal equations of the least squares system (3.7) can be written as

$$\begin{pmatrix} | & | & | & | \\ \mathbf{u}_0 & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ | & | & | & | \end{pmatrix}^T p_4(A)\,\mathbf{u}_0=\mathbf{0}. \tag{3.8}$$

Chosen in this manner, $p_4(z)$ is the orthogonal polynomial of degree 4 with respect to A and $\mathbf{r}_n$ (Householder [14], p. 24). The roots of $p_4(z)$ are approximations to the dominant and subdominant eigenvalues of A.

This technique can be applied with any degree polynomial. Consider the normal equations of (3.7) written as the augmented system

$$\left[\begin{array}{cccc|c} \langle\mathbf{u}_0,\mathbf{u}_0\rangle & \langle\mathbf{u}_0,\mathbf{u}_1\rangle & \langle\mathbf{u}_0,\mathbf{u}_2\rangle & \langle\mathbf{u}_0,\mathbf{u}_3\rangle & \langle\mathbf{u}_0,\mathbf{u}_4\rangle \\ \langle\mathbf{u}_1,\mathbf{u}_0\rangle & \langle\mathbf{u}_1,\mathbf{u}_1\rangle & \langle\mathbf{u}_1,\mathbf{u}_2\rangle & \langle\mathbf{u}_1,\mathbf{u}_3\rangle & \langle\mathbf{u}_1,\mathbf{u}_4\rangle \\ \langle\mathbf{u}_2,\mathbf{u}_0\rangle & \langle\mathbf{u}_2,\mathbf{u}_1\rangle & \langle\mathbf{u}_2,\mathbf{u}_2\rangle & \langle\mathbf{u}_2,\mathbf{u}_3\rangle & \langle\mathbf{u}_2,\mathbf{u}_4\rangle \\ \langle\mathbf{u}_3,\mathbf{u}_0\rangle & \langle\mathbf{u}_3,\mathbf{u}_1\rangle & \langle\mathbf{u}_3,\mathbf{u}_2\rangle & \langle\mathbf{u}_3,\mathbf{u}_3\rangle & \langle\mathbf{u}_3,\mathbf{u}_4\rangle \end{array}\right]. \tag{3.9}$$

The normal system corresponding to any smaller degree polynomial can be found by taking the appropriate submatrix from the upper left corner of (3.9). If the vector $\mathbf{r}_n$ is nearly a linear combination of fewer than four eigenvectors, the system (3.9) will be nearly rank deficient. After scaling the system by dividing each element by $\langle u_0, u_0\rangle$ one can choose the proper degree by performing Gaussian elimination without pivoting and using the largest principle subsystem with acceptable determinant[3].

If the vector $\mathbf{r}_n$ is a linear combination of more than four eigenvectors or if the dominant eigenvalues have nonlinear elementary divisors, the eigenvalue estimates will not be accurate but will still be useful. For our purposes we would like the eigenvalue estimates to lie inside $H(A)$. In Section 4 we will see that, in fact, the estimates will lie in $F(A)$, the field of values of A. In Section 7 we will see that iteration parameters choosen in terms of estimates of $F(A)$ rather than $H(A)$ still yield good numerical results.

---

[2] Here $\|\cdot\|$ represents the l_2-norm of the vector

[3] In order to retain as much accuracy as possible, all calculations involved in the construction and solution of (3.9) should be performed in double presision (Stewart [23], pp. 217–226)

4. Stability of the Eigenvalue Estimates

Definition. The *field of values* of the matrix A is the set

$$F(A)=\left\{\lambda\in\mathbf{C}/\lambda=\frac{\langle A\mathbf{x},\mathbf{x}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle} \text{ for some possibly complex } \mathbf{x}\right\}.$$

The following useful results can be found in Householder [14], pp. 74–80.

Theorem 4.1. The field of values of the matrix A is a convex set in $\mathbf{C}$.

Theorem 4.2. If A is real, then $F(A)$ is symmetric with respect to the real axis.

Theorem 4.3. In general, we have

$$H(A)\subseteq F(A), \tag{4.1}$$

and if A is normal, then

$$H(A)=F(A). \tag{4.2}$$

Consider the Krylov sequence

$$\mathbf{u}_0, \mathbf{u}_1=A\mathbf{u}_0, \ldots, \mathbf{u}_j=A^j\mathbf{u}_0, \ldots. \tag{4.3}$$

If we extract a sequence of orthogonal vectors from (4.3),

$$\begin{aligned}
&\mathbf{p}_0=\mathbf{u}_0\\
&\mathbf{p}_1=\mathbf{u}_1-\alpha_{10}\mathbf{p}_0\\
&\mathbf{p}_2=\mathbf{u}_2-\alpha_{21}\mathbf{p}_1-\alpha_{20}\mathbf{p}_0\\
&\vdots\\
&\mathbf{p}_j=\mathbf{u}_j-\alpha_{jj-1}\mathbf{p}_{j-1}\ldots-\alpha_{j0}\mathbf{p}_0,
\end{aligned} \tag{4.4}$$

then

$$\mathbf{p}_j=p_j(A)\,\mathbf{u}_0 \tag{4.5}$$

where $p_j(\lambda)$ is the j^{th} orthogonal polynomial with respect to A and $\mathbf{u}_0$.

Theorem 4.4. Let $\mathbf{q}_i=\dfrac{\mathbf{p}_i}{\|\mathbf{p}_i\|}$, $i=1,\ldots,j-1$, and let

$$Q=\begin{pmatrix} | & | & & | \\ \mathbf{q}_0 & \mathbf{q}_1 & \cdots & \mathbf{q}_{j-1} \\ | & | & & | \end{pmatrix}$$

then, Q is orthogonal and the eigenvalues of the orthogonal section of A given by

$$A'=Q^*AQ$$

are the roots of $P_j(\lambda)$.

Proof. The proof follows from the discussion on pages 24–26 of Householder [14].

Theorem 4.5. If A' is an orthogonal section of A, then

$$F(A') \subseteq F(A).$$

Proof. Suppose $\lambda \in F(A')$; then, there exists a vector $\mathbf{z}$ such that

$$\lambda = \frac{\langle A'\mathbf{z}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle}.$$

Since A' is an orthogonal section of A, there is some orthogonal Q such that

$$Q^* A Q = A'.$$

Thus,

$$\lambda = \frac{\langle Q^* A Q\mathbf{z}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle} = \frac{\langle AQ\mathbf{z}, Q\mathbf{z}\rangle}{\langle Q\mathbf{z}, Q\mathbf{z}\rangle} \in F(A).$$

Theorems 4.3, 4.4, 4.5 yield the following result.

Corollary 4.6. If λ is a root of $p_j(\lambda)$, the j^{th} orthogonal polynomial with respect to A and $\mathbf{u}_0$, then $\lambda \in F(A)$.

For our purposes, $\mathbf{u}_0$ is nearly a linear combination of just a few eigenvectors. If $\mathbf{u}_0$ is a linear combination of a set of eigenvectors that are *mutually orthogonal*, then the power method will yield estimates that are in $H(A)$.

Theorem 4.7. Suppose $\mathbf{u}_0$ in (4.3) is in the span of the mutually orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Let Q be as in Theorem 4.4 with $j \leqq k$ and let

$$V = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{pmatrix};$$

then,

$$F(Q^* A Q) \subseteq F(V^* A V) = H(V^* A V) \subseteq H(A).$$

Proof. We have

$$V^* A V = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ & 0 & & \lambda_k \end{pmatrix}.$$

Thus,

$$F(V^* A V) = H(V^* A V) \subseteq H(A).$$

Since $\mathbf{u}_0$ is in the span of $\{\mathbf{v}_1 \ldots \mathbf{v}_k\}$, each $\mathbf{q}_i$ in Theorem 4.4 is also. There exists a

$k \times j$ matrix U such that

$$Q = VU.$$

Since Q and V are orthogonal, U is also, we have

$$Q^* AQ = U^*(V^* AV)\, U$$

and from Theorem 4.5

$$F(Q^* AQ) \subseteq F(V^* AV).$$

A relation between the size of $H(A)$ and the size of $F(A)$, due to Henrici [12], can be expressed in terms of a measure of the normality of A. If A is any matrix, then a classical result due to Schur (Birkhoff and Mclane [1]) states that there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*. \tag{4.6}$$

In general, T is not unique. Let

$$T = D + E \tag{4.7}$$

where D is the diagonal and E is the nilpotent part.

Definition. The departure from normality of A is[4]

$$\Delta(A) = \inf \|E\| \tag{4.8}$$

where inf is taken with respect to all E that can appear in a Schur Triangular form of A.

Theorem 4.8. If $\mu \in F(A)$, then there exists $\lambda \in H(A)$ such that

$$|\mu - \lambda| \leqq \Delta(A). \tag{4.9}$$

Proof. The proof follows with minor changes from Henrici ([12], Theorem 7, p. 36).

The field of values of A shrinks down upon the convex hull of the spectrum of A as the norm of E decreases to zero. In general, though, the field of values can be much larger than the convex hull of the spectrum as will be seen in the last section.

Simple bounds on $F(A)$ can be found in terms of the symmetric and skew-symmetric parts of A. Let

$$M = 1/2(A + A^T), \tag{4.10}$$

$$N = 1/2(A - A^T). \tag{4.11}$$

Since M and N are normal, we have from Theorem 4.3 that there exist real

[4] Here $\|E\|$ is the lub norm of the matrix associated with the Euclidian vector norm

numbers a, b, c such that

$$F(M)=[a, b], \tag{4.12}$$

$$F(N)=[-ic, ic]. \tag{4.13}$$

We have the following result (Householder [14], p. 79).

Theorem 4.9. If M and N are as in (4.10) and (4.11), then

$$F(A)\subseteq[a, b]\times[-ic, ic]. \tag{4.14}$$

This is the smallest such rectangle to contain $F(A)$.

Corollary 4.10. If M is positive definite $F(A)$ lies in the right half plane.

5. Modified Power Method

Notice that the construction of the Krylov Sequence (4.3) required four matrix vector multiplications and the storage of 4 additional vectors. The matrix vector multiplications can be avoided. In light of (2.11) and (2.21) we can write

$$\mathbf{r}_n \doteq S^n \mathbf{r}_0. \tag{5.1}$$

The residuals provide a Krylov sequence for the operator $S=S(A)$. If we let

$$\mathbf{u}_0=\mathbf{r}_n, \quad \mathbf{u}_1=\mathbf{r}_{n+1}, \ldots, \mathbf{u}_4=\mathbf{r}_{n+4}, \tag{5.2}$$

then the solution of (3.9) will yield estimates of $\sigma_1=S(\lambda_1)$ and $\sigma_2=S(\lambda_2)$, where again λ_1 and λ_2 are the dominant and subdominant eigenvalues of A as in (3.1).

This method is similar to the modified power method (Wilkinson [32]). Let us examine the relation between σ_i and λ_i. From (2.7) and (2.12) we have that if

$$g=d+(d^2-c^2)^{1/2},$$

then

$$\lambda_i=d-\frac{1}{2}\left(g\sigma_i+\frac{c^2}{g\sigma_i}\right). \tag{5.3}$$

From the exponential form of $S(\lambda)$ in (2.7) and the fact that the branch of $\cosh^{-1}$ with positive real part is used we have

$$\sigma_i \geqq \left|\frac{c}{g}\right|. \tag{5.4}$$

The relation (5.3) maps the values $\sigma \leqq \left|\frac{c}{g}\right|$ onto the entire λ plane. Thus, if poor separation of the eigenvalues of S or the presence of non-linear elementary divisors causes the estimate of σ_i to lie near the center of $H(S)$ in such a way as to violate (5.4), it must be discarded.

The error that results from the power method (Method 1) will be due entirely

to poor separation of the eigenvalues and the presence of nonlinear elementary divisors. The modified power method (Method 2), on the other hand, also introduces error from the relation $P_n(A) \doteq S^n$. The transformation (5.3) may warp and magnify this error. We can eliminate the need for this transformation at the expense of storing an additional vector. Consider the operator equivalent of (5.3),

$$A \doteq dI - \frac{1}{2}\left(gS - \frac{c^2}{g} S^{-1}\right). \tag{5.5}$$

Equation (5.5) holds in the sense that for sufficiently large n the error in the direction of the invariant subspaces for which (5.5) is poor has been greatly suppressed. Consider the operator

$$\hat{A} = 2g(dI - A). \tag{5.6}$$

We can write

$$\begin{aligned}
\hat{A} &= (g^2 S + c^2 S^{-1}),\\
\hat{A}^2 &= (g S^2 + 2g^2 c^2 I + c^4 S^{-2}),\\
\hat{A}^3 &= (g^6 S^3 + 3g^4 c^2 S + 3g^2 c^4 S^{-1} - c^6 S^{-3}),\\
\hat{A}^4 &= (g^8 S^4 + 4g^6 c^2 S^2 + 6g^4 c^4 I + 4g^2 c^6 S^{-2} + c^8 S^{-4}).
\end{aligned} \tag{5.7}$$

If we let

$$\mathbf{u}_0 = \mathbf{r}_n, \tag{5.8}$$

then with the use of (5.1) we can form the Krylov sequence for $\hat{A}$ and $\mathbf{u}_0$,

$$\begin{aligned}
\mathbf{u}_1 &= g^2 \mathbf{r}_{n+1} + c^2 \mathbf{r}_{n-1},\\
\mathbf{u}_2 &= g^4 \mathbf{r}_{n+2} + 2g^2 c^2 \mathbf{r}_n + c^4 \mathbf{r}_{n-2},\\
\mathbf{u}_3 &= g^6 \mathbf{r}_{n+3} + 3g^4 c^2 \mathbf{r}_{n+1} + 3g^2 c^4 \mathbf{r}_{n-1} + c^6 \mathbf{r}_{n-3},\\
\mathbf{u}_4 &= g^8 \mathbf{r}_{n+4} + 4g^6 c^2 \mathbf{r}_{n+2} + 6g^4 c^4 \mathbf{r}_n + 4g^2 c^6 \mathbf{r}_{n-2} + c^8 g \mathbf{r}_{n-4}.
\end{aligned} \tag{5.9}$$

Again using (3.9) we can get approximations to the appropriate eigenvalues of $\hat{A}$ which can then be translated to eigenvalues of A with the use of (5.6).

Let us examine the relative strengths of the three methods. The extra storage and work requirements of the three methods are displayed in Table 1. Here MM represents the work required to perform a matrix vector multiplication, IP represents the work required to perform one vector inner product, and N is the dimension of the system.

Table 1

	Work	Storage
Method 1	$4MM + 14IP$	$4N$
Method 2	$14IP$	$4N$
Method 3	$14IP$	$5N$

The first method tends to favor eigenvalues of large modulus. If $\mathbf{u}_0$ in (3.3) contains components of equal weight in the direction of eigenvectors associated with both very small eigenvalues and very large eigenvalues, then the vectors $\mathbf{u}_j$, $j=1, \ldots, 4$ will be biased toward the large eigenvalues. This bias does not appear in either of the other two methods. The desired eigenvalues of A correspond exactly with the eigenvalues of S with largest modulous. Similarly, the translation from A to $\hat{A}$ in (3.15) gives equal weight to the small eigenvalues. However, a small relative error in computing an eigenvalue estimate of S or $\hat{A}$ may become a large relative error if the estimate corresponds to a small eigenvalue of A.

As was shown in Section 4, the first method is guaranteed to yield estimates that lie in $F(A)$. The other two methods depend upon the relation (2.11), $P_n(A) \doteq S^n$. The validity of this relation depends upon $H(A)$ as well as n. If $H(A)$ is long and thin, then the ellipses enclosing $H(A)$ will be close to the degenerate ellipse. In this case n must be large for (2.11) to hold.

If n is choosen large enough so that (2.11) holds, then Method 2 will yield eigenvalue estimates that lie in $F(S)$ and Method 3 will yield estimates in $F(\hat{A})$. Although the relations among the eigenvalues of A, $\hat{A}$, and S are well defined, there is no exact relation between $F(A)$ and $F(S)$. Thus, if an eigenvalue estimate σ_i lies in $F(S)$, there is no guarantee that the corresponding λ_i from (5.3) will lie in $F(A)$. However, because it is linear, (5.6) defines a relation between $F(A)$ and $F(\hat{A})$. Thus, if A is know to be nearly symmetric (M is large compared to N in 4.10 and 4.11) or very nonsymmetric (N is large compared to M), then Method 1 is preferable for stability of the eigenvalue estimates. Otherwise, Method 3 is preferable because it requires less work. In a series of test problems described in Section 7 the three methods were nearly indistinguishable.

6. Adaptive Procedure

The method of choosing the parameters d and c outlined in Manteuffel [20] can be combined with one of the three methods for finding eigenvalue estimates. Since it is $H(A)$ that we wish to approximate, a sequence of eigenvalue estimates, each of which lies inside $H(A)$ will yield a sequence of approximations to $H(A)$. Consider the following procedure: First, choose K_0, a set of points in $\mathbf{C}$, from some prior knowledge of $H(A)$ such that $H(K_0)$ is an approximation to $H(A)$ and $H(K_0) \subseteq H(A)$. Choose d and c optimal for K_0 and perform the iteration (1.2) for a predetermined number of steps (20 or 30 have been sufficient in practice). One of the methods outlined above will yield a set of eigenvalue approximations denoted by L_1. Let K_1 be the smallest set of points such that $H(K_1) = H(L_1 \cup K_0)$ and choose new parameters d and c optimal for K_1. Again iterate for a predetermined number of steps before getting a new set of eigenvalue estimates L_2. Let K_2 be the smallest set such that $H(K_2) = H(L_2 \cup K_1)$. Repeating this cycle, a sequence of sets, $K_0, K_1, \ldots, K_j$ will be formed. If all of the eigenvalue estimates lie in $H(A)$, then

$$H(K_0) \subseteq H(K_1) \subseteq \cdots H(K_j) \subseteq H(A), \tag{6.1}$$

and each K_j gives at least as good an approximation to $H(A)$ as its predecessor.

Unfortunately, as was shown in Section 4, we can only guarantee that each of the eigenvalue estimates will lie in $F(A)$. Equation (5.1) must be re-written as

$$H(K_0) \subseteq H(K_1) \subseteq \cdots H(K_j) \subseteq F(A). \tag{6.2}$$

If A is nearly normal – that is, if $\Delta(A)$ in (4.8) is small – then $F(A)$ will differ only slightly from $H(A)$ and $H(K_j)$ will provide a good approximation to $H(A)$. Even if $\Delta(A)$ is not small, $H(K_j)$ may still provide a good choice of parameters as will be exhibited in Section 7. This may be explained, in part, by the fact that in the test problems $F(A)$ somewhat symmetrically surrounds $H(A)$. The best family of ellipses to fit $F(A)$ may closely approximate the best family of ellipses to fit $H(A)$. In general, however, iteration parameters based upon $F(A)$ will not perform as well as iteration parameters based upon $H(A)$.

If some K_j yields a choice of d and c and corresponding convergence region that does not include $H(A)$ in its interior, then some eigenvalue, λ_i, will have convergence factor $r(\lambda_i) > 1$. The error in the direction of this eigenvector will be magnified at each step. Since the error in the direction of most of the other eigenvectors will be suppressed, the power method will yield a very accurate estimate of λ_i. Subsequent choices of parameters will take this into account and will produce parameters d and c with associated convergence region that will contain $H(A)$. If a provision is made to return to the solution vector at the beginning of the cycle, then a poor choice of K_j will not move the iteration toward convergence, but will give rise to new iteration parameters that will produce convergence.

The Tchebychev iteration is restricted to systems whose eigenvalues lie in the right half plane. More specifically, the iteration is restricted to systems whose spectrums can be enclosed in an ellipse that does not contain the origin (Manteuffel [20], Section 2). Likewise, each of the sets $H(K_j)$ must not include the origin. From Corollary 4.10 we have that if M in (4.10) is positive definite, then $F(A)$ – and thus each $H(K_j)$ – lies in the right half plane. We can say that if A has eigenvalues in the open right half plane, if $M = 1/2(A + A^T)$ is positive definite, and if K_0 is chosen such that $H(K_0) \subseteq H(A)$, then the procedure described above will yield iteration parameters for which (1.2) will converge.

It is possible to have a system A with spectrum in the right half plane whose symmetric part, M, is not positive definite. In this case, $F(A)$ includes the origin. It may be possible to determine a region $\hat{H}$ in the right half plane known to contain $H(A)$. If any eigenvalue estimate outside $\hat{H}$ is discarded, then we would have $H(K_j) \subseteq F(A) \cap \hat{H}$ for each K_j. With this provision, the procedure described above will yield iteration parameters for which (1.2) will converge.

7. Results

The adaptive Tchebychev algorithm (TCHEB) was compared with three other iterative methods for solving nonsymmetric linear systems on a series of test problems. The other methods are the bidiagonalization method (BIDIAG)

Table 2

Method	Work	Storage
TCHEB	$MM+2IP+0$	$9N+S$
BIDIAG	$2MM+7IP$	$4N+S$
CGonA^TA	$2MM+5IP$	$4N+S$
CGNS	$2MM+5IP$	$4N+S$

(Golub and Kahan [9], Paige [22]), the method of conjugate gradients applied to the normalized system $A^TA\mathbf{x}+A^T\mathbf{b}$(CGon A^TA) (Hestenes and Stiefel [13]), and a varient of the conjugate gradient method for nonsymmetric systems described by D. Kershaw [15] (CGNS).

The other methods require more work per iterative step than the Tchebychev iteration, but the adaptive procedure requires extra storage. Table 2 shows a comparison of the work per iterative step and storage requirements of the four methods. As in Section 5, MM represents the work required to perform a matrix vector multiplication and IP represents the work required to perform an inner product. The Symbol 0 is the overhead created by the adaptive procedure. This depends upon the number of steps of (1.2) taken between eigenvalue estimates and the variant of the adaptive procedure used (see Table 1). The symbol N is the dimension of the system and S is the storage required for the matrix A. If A is a 5-point difference matrix, then $S=5N$. Finite Element matrices are more dense and require much more storage. As the density of nonzeros in A increases the extra storage required by TCHEB becomes less of a factor, while the work required by the other methods remains twice as much as that required by TCHEB.

Each of the three competing methods can be considered to be a polynomial method on A^TA, while TCHEB is a polynomial method on A (Rutishauser [6], p. 25). Because of this advantage and the advantage of less work per iterative step, TCHEB was considerably faster than the other methods on a series of test problems. In each of the tests, CG on A^TA performed identically with BIDIAG[5]. For that reason results are not shown for CG on A^TA.

The convergence properties of the methods above depend only upon the spectral properties of the matrix A and not upon any special zero structure. It is desirable, however, to test the algorithms on easily constructable systems whose spectrums have known properties. The methods described by Concus and Golub [3] and Widland [31] are better suited to some of the test problems below. However, these problems demonstrate the power of the Tchebychev algorithm and the Tchebychev algorithm can be applied to more complex problems. Consider the differential operator

$$-\Delta+\beta\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right), \quad \beta\geqq 0 \tag{7.1}$$

[5] BIDIAG and CG on A^TA are theoretically equivalent if CG on A^TA has initial vector $\mathbf{x}=\mathbf{0}$

on a square domain $[0, L] \times [0, L]$ with Dirichlet boundary conditions. The discrete analog using central differences on a square mesh of mesh size $h = \frac{L}{n+1}$ yields an $n^2 \times n^2$ matrix A such that

$$A = \frac{1}{h^2} M + \frac{\beta}{2h} N. \tag{7.2}$$

The positive definite M can be written in block tridiagonal form as

$$M = \left[\begin{array}{cccc|cccc|c}
4 & -1 & & & -1 & & & & \\
-1 & \ddots & \ddots & & & \ddots & & & \\
 & \ddots & \ddots & -1 & & & \ddots & & \\
 & & -1 & 4 & & & & -1 & \\
\hline
-1 & & & & 4 & -1 & & & \\
 & \ddots & & & -1 & \ddots & \ddots & & \\
 & & \ddots & & & \ddots & \ddots & -1 & \\
 & & & -1 & & & -1 & 4 & \\
\hline
 & & \mathbf{0} & & & & & & \ddots
\end{array}\right]$$

where each block is $n \times n$ and the skew-symmetric N can be written in block tridiagonal form as

$$N = \left[\begin{array}{cccc|cccc|c}
0 & 1 & & & 1 & & & & \\
-1 & \ddots & \ddots & & & \ddots & & & \\
 & \ddots & \ddots & 1 & & & \ddots & & \\
 & & -1 & 0 & & & & 1 & \\
\hline
-1 & & & & 0 & 1 & & & \\
 & \ddots & & & -1 & \ddots & \ddots & & \\
 & & \ddots & & & \ddots & \ddots & 1 & \\
 & & & -1 & & & -1 & 0 & \\
\hline
 & & \mathbf{0} & & & & & & \ddots
\end{array}\right].$$

In particular, let $h = 1.0$ and $L = 41$. Then,

$$A = M + \frac{\beta}{2} N$$

is of dimension $n^2 = 1600$ with eigenvalues given by

$$\lambda_{jk}=2\left[2-\sqrt{1-\left(\frac{\beta}{2}\right)^2}\left(\cos\left(\frac{j\pi}{41}\right)+\cos\left(\frac{k\pi}{41}\right)\right)\right] \quad \begin{matrix} j=1,\ldots,40 \\ k=1,\ldots,40. \end{matrix} \tag{7.3}$$

Notice that for $\beta \leqq 2$ all of the eigenvalues are real and for $\beta > 2$ they are all complex with real part 4.0.

The eigenvalues of M are given by

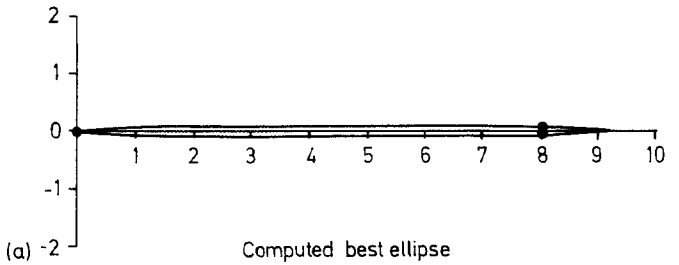
$$\mu_{jk}=2\left[2-\left(\cos\left(\frac{j\pi}{41}\right)+\cos\left(\frac{k\pi}{41}\right)\right)\right] \quad \begin{matrix} j=1,\ldots,40 \\ k=1,\ldots,40, \end{matrix} \tag{7.4}$$

and the eigenvalues of N are given by

$$\eta_{jk}=2\left[\cos\left(\frac{j\pi}{41}\right)+\cos\left(\frac{k\pi}{41}\right)\right] i \quad \begin{matrix} j=1,\ldots,40 \\ k=1,\ldots,40. \end{matrix} \tag{7.5}$$

From Theorem 4.9 we have that

$$F(A) \subseteq [a, b] \times [-ic, ic], \tag{7.6}$$



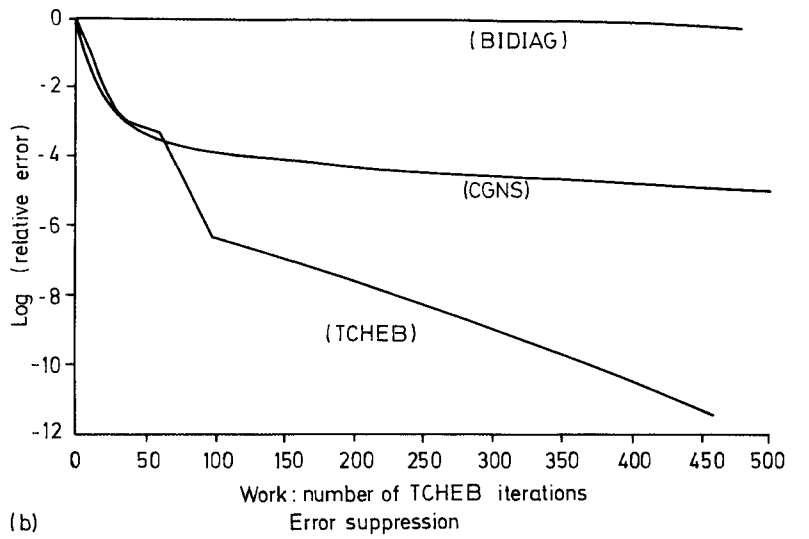
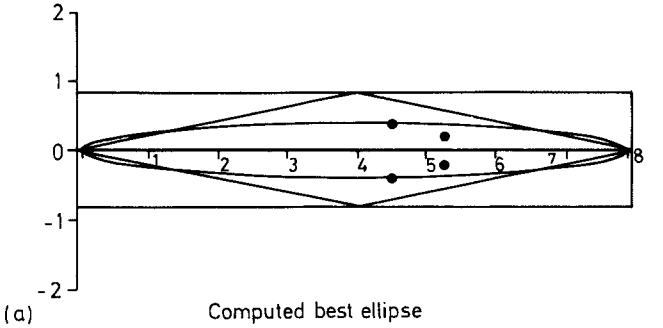
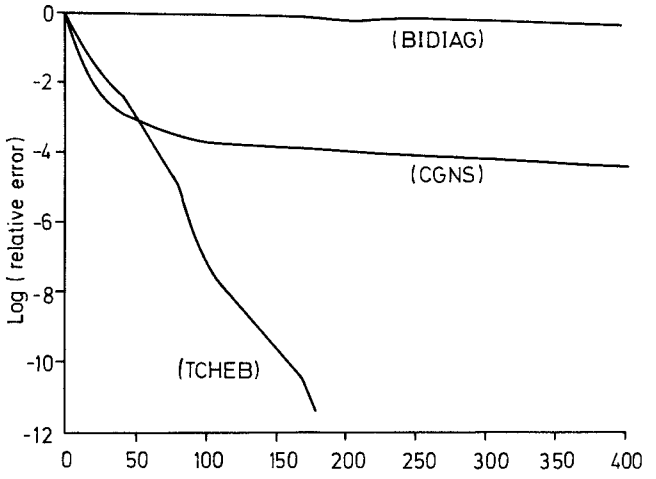


Fig. 1a and b. $\beta = 0.1$



(a) Computed best ellipse



(b) Error suppression

Fig. 2a and b. $\beta = 0.4$

where

$$a = 4\left[1 - \cos\left(\frac{\pi}{41}\right)\right],$$

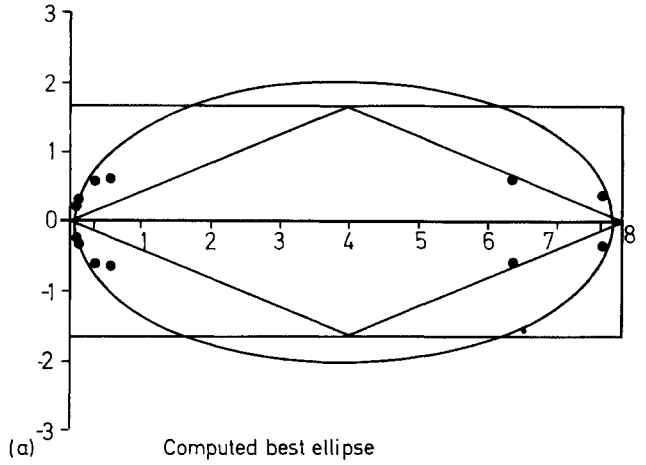
$$b = 4\left[1 + \cos\left(\frac{\pi}{41}\right)\right],$$

$$c = 2\beta \cos\left(\frac{\pi}{41}\right). \tag{7.7}$$

It can be shown that

$$a, b, 4 \pm ic \in F(A). \tag{7.8}$$

Since $F(A)$ is convex, (7.8) yields a rhombus that lies inside $F(A)$. Figures 2a–7a



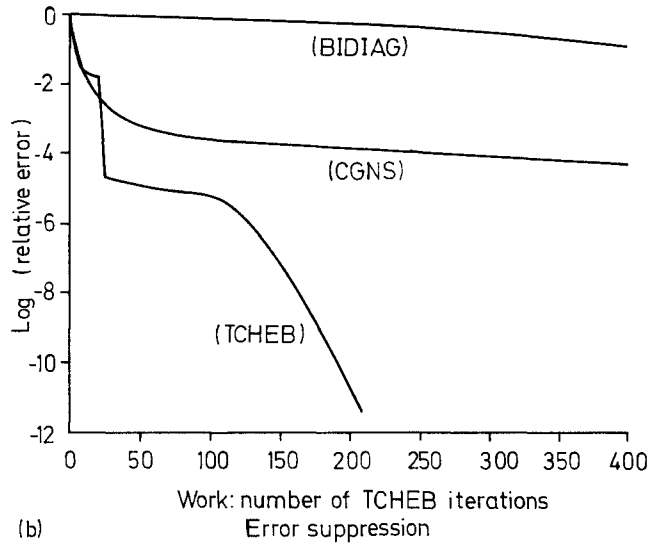
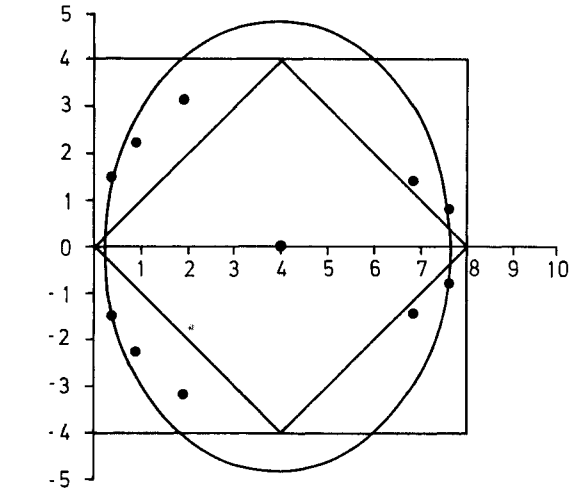


Fig. 3a and b. $\beta = 0.8$

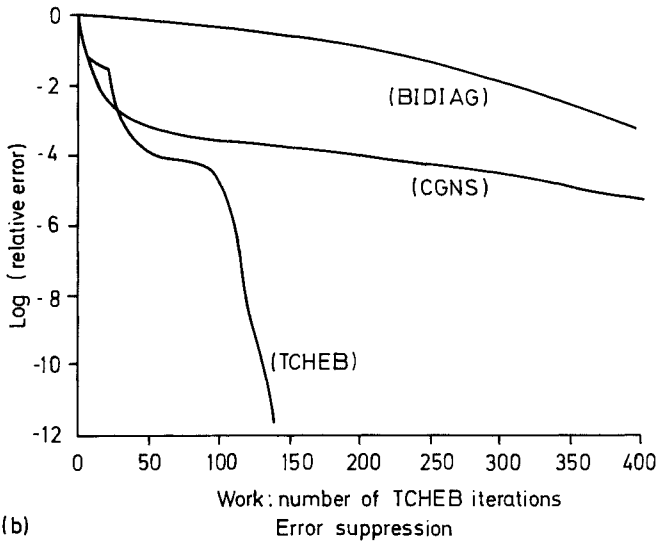
show the rectangle known to contain $F(A)$ and the rhombus inside $F(A)$ for the indicated values of β. Also, $H(A)$ is indicated by a heavy line.

TCHEB was tested against the other methods for values of β ranging from 0.1 to 40. In each case 20 steps of (1.2) were taken between each adaptive procedure. Figures 1a–9a show the hull of the approximate spectrum computed by TCHEB (using Method 2 for the adaptive procedure) as well as the best ellipse enclosing the approximate spectrum.

Figures 1b–9b show a comparison of the error suppression of BIDIAG, CGNS, and TCHEB (using Method 2 for the adaptive procedure), versus the



(a) Computed best ellipse



(b) Error suppression

Fig. 4a and b. $\beta = 2$

work required. Error suppression was measured by the log of the relative error; that is, by $\log(\|e_n\|/\|e_0\|)$ where $\|e_n\|$ is the l_2-norm of the error vector at the n^{th} step. Work was measured in terms of the number of Tchebychev iterations.

For A nearly symmetric, the condition of A^TA is significantly worse than the condition of A. Since the other methods are sensitive to the condition of A^TA, they did rather poorly for nearly symmetric A (see Figs. 1b–3b). For large β the two methods were more comparable, but the Tchebychev method still held a significant advantage (see Fig. 9b).

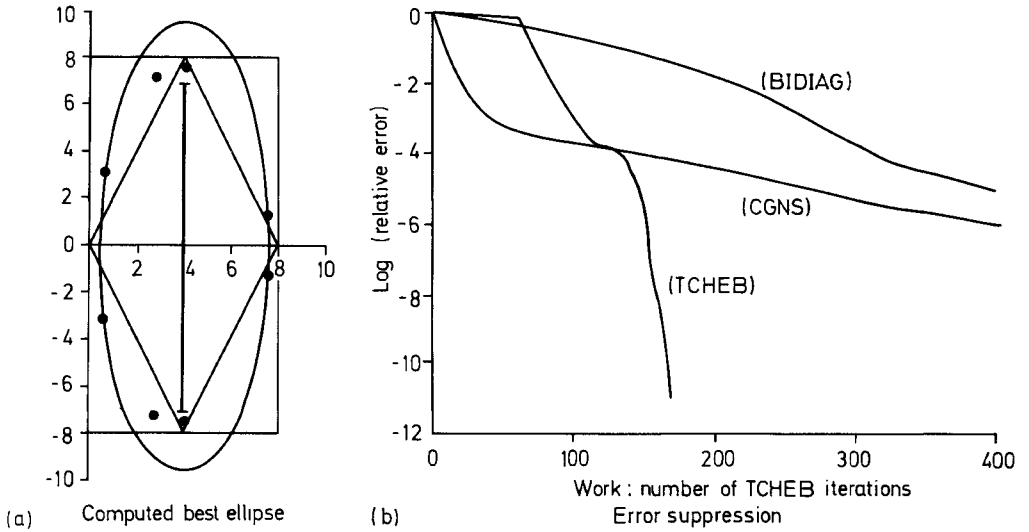


Fig. 5a and b. $\beta = 4$

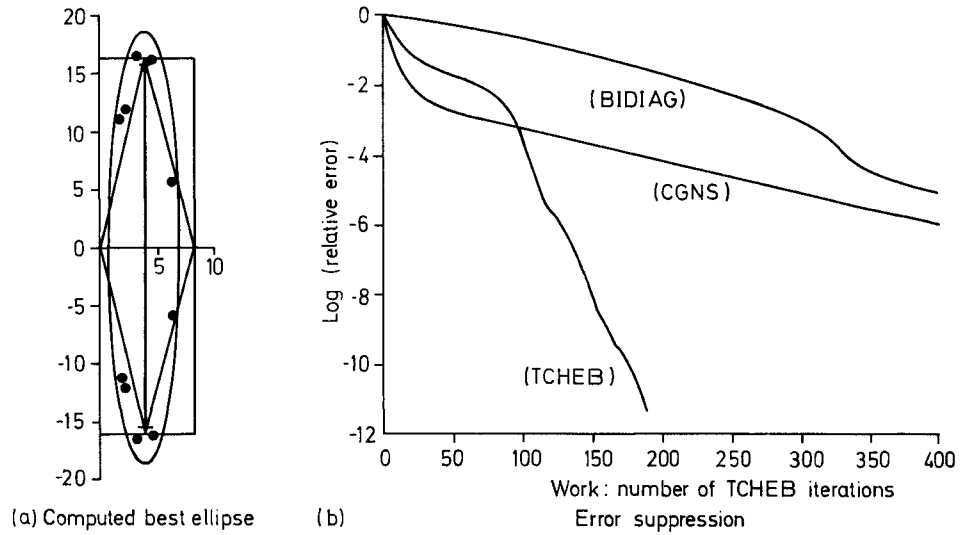


Fig. 6a and b. $\beta = 8$

The initial choice of parameters d and c was based on the rectangle known to contain the spectrum of A (see Table 3). In some cases a rather poor choice was used to show that the adaptive procedure would still work. For $\beta = 4$, and $\beta = 20$, the convergence region associated with the initial parameters did not include the entire spectrum. The error was allowed to grow until the adaptive procedure extracted eigenvalue estimates. The solution vector was set back to its

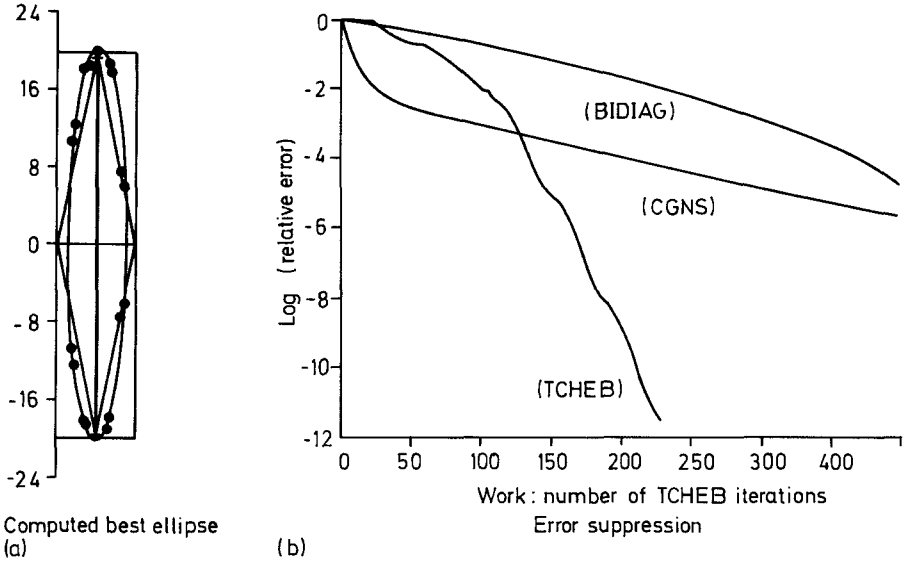


Fig. 7a and b. $\beta = 10$

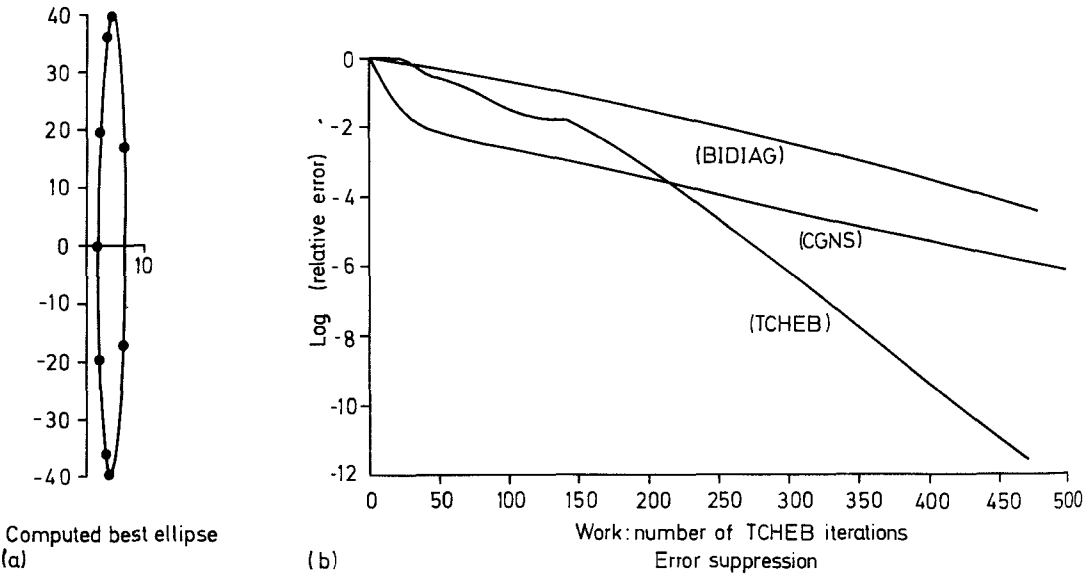


Fig. 8a and b. $\beta = 20$

original value and the new parameters led to convergence (see Figs. 5b and 7b).

Each variant of the adaptive procedure was used on each test problem. They performed almost identically. TCHEB was also run on each test problem with fixed parameters computed from the known spectrum (7.3). The adaptive procedures found parameters for which convergence was not significantly slower

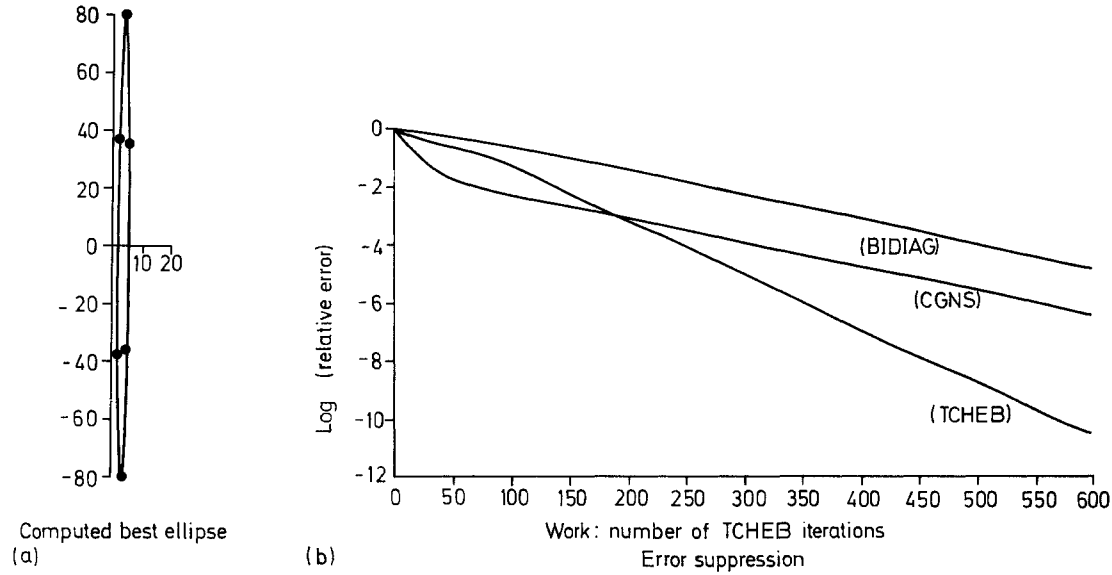


Fig. 9a and b. $\beta = 40$

Table 3

β	Initial		Method 1		Method 2		Method 3		Exact	
	d	c	d	c	d	c	d	c	d	c
0.1	4.0	3.872	4.01	3.984	4.00	3.980	4.00	3.983	4.0	3.983
0.4	4.0	3.872	4.01	3.911	4.03	3.922	4.01	3.906	4.0	3.908
0.8	4.0	0	3.96	3.429	3.95	3.317	3.95	3.386	4.0	3.655
2	4.0	0	4.00	$5.20i$	3.94	$3.10i$	3.92	$3.18i$	4.0	0
4	4.0	0	4.09	$9.35i$	4.11	$9.38i$	4.06	$9.26i$	4.0	$6.91i$
8	4.0	$15.00i$	4.05	$20.48i$	3.86	$18.45i$	3.88	$19.42i$	4.0	$15.45i$
10	4.0	$14.14i$	3.94	$22.84i$	4.06	$20.19i$	4.05	$19.45i$	4.0	$19.54i$
20	4.0	$31.62i$	3.99	$39.81i$	4.25	$40.39i$	3.99	$39.85i$	4.0	$39.68i$
40	4.0	$75.00i$	3.92	$79.88i$	4.10	$79.94i$	3.93	$79.82i$	4.0	$79.67i$

than for the exact parameters. Table 3 shows the initial parameters, the parameters found by each method and the exact parameters. Table 4 shows the number of steps required by each method to reduce the error by the factor shown.

Figures 1a–9a show that the eigenvalue estimates tend to approximate $F(A)$ when $F(A)$ is considerably larger than $H(A)$. Notice that for $\beta = 2$ all of the eigenvalues (7.3) have the value 4. In fact $p(\lambda) = (\lambda - 4)^{79}$ is the minimum polynomial for the matrix $A = M + N$. The exact parameters $d = 4$, $c = 0$ which correspond to circles centered at 4 yield a solution in 79 steps. In each case, the parameters based on the eigenvalue estimates were close to the exact parameters and convergence was only slightly slower. This may be explained, in part, by the

Table 4

β	Method 1	Method 2	Method 3	Exact	Factor
0.1	238	255	256	268	10^{-10}
0.4	151	154	152	150	10^{-10}
0.8	177	195	181	112	10^{-10}
2	121	131	135	79	10^{-10}
4	162	164	165	106	10^{-10}
8	181	175	184	137	10^{-10}
10	225	211	207	153	10^{-10}
20	324	411	348	264	10^{-10}
40	572	571	523	402	10^{-8}

Table 5

β	α	Steps	β	α	Steps
0.4	0.3	49	4.0	0.1	48
	0.5	43		0.3	46
	0.7	34		0.5	43
	0.8	40		0.7	56
	1.0	53		1.0	155

fact that for these problems $F(A)$ somewhat symmetrically surrounds $H(A)$. The best family of ellipses to fit $F(A)$ may closely approximate the best family of ellipses to fit $H(A)$.

8. Acceleration of Matrix Splittings

The greatest advantage of this method lies in the ability to accelerate preconditioned systems with the Tchebychev iteration. For example, the matrix A may be split into

$$A = M - N \tag{8.1}$$

where M is easily invertable. The preconditioned system

$$M^{-1}A\mathbf{x} + M^{-1}\mathbf{b} \tag{8.2}$$

may have better condition than (1.1). If N is "small", then

$$M^{-1}A = I - M^{-1}N$$

will have spectrum close to 1.0, and thus in the right half plane. The Tchebychev iteration can be applied to (8.2). The SIP splitting (Stone [26]), the incomplete Cholesky decomposition (Meijerink and van der Vorst [21]), and preconditioning by use of fast Poisson solvers (Concus and Golub [3]) are examples of such splittings. The dynamic procedure outlined in this paper can be used to estimate the spectrum of $M^{-1}A$ and thus good iteration parameters can be found.

The SIP splitting, which depends upon a parameter α (see Stone [26]), was applied to the test problem with $\beta=0.4$ and $\beta=4.0$ for several values of α and then accelerated by the Tchebychev iteration. Table 5 shows the number of iterations required to reduce the relative error by a factor of 10^{-10}. Each iteration required approximately twice as much work as TCHEB alone. In each case the iteration parameters were initially set at $d=1.0$, $c=0.0$. The results indicate that the SIP splitting is very sensitive to the choice of parameter α. A comparison with Table 4 shows a considerable savings in the work required. More detailed numerical results of the acceleration of the SIP splitting appear in Manteuffel [19, p. 133].

Acknowledgments. I would like to thank Gene Golub for introducing me to the literature about the field of values of a matrix.

References

1. Birkhoff, G., Maclane, S.: A survey of modern algebra. New York: MacMillan 1953
2. Carre', B.A.: The determination of the optimum acceleration factor for successive overrelaxation. Comput. J. **4**, 73–78 (1961)
3. Concus, P., Golub, G.H.: A generalized conjugate gradient method for non-symmetric systems of linear equations. Proc. Second International Symposium on Computing Methods in Applied Sciences and Engineering. Versailles, France, December 1975
4. Diamond, M.A.: An economical algorithm for the solution of elliptic difference equations independent of user-supplied parameters. Ph.D. Dissertation, Department of Computer Science, University of Ill., 1972
5. Dunford, N., Schwartz, J.L.: Linear operators. New York: Interscience 1958
6. Engeli, M., Ginsburg, T.H., Rutishauser, H., Stiefel, E.L.: Refined iterative methods for computation of the solution and eigenvalues of self-adjoint boundary value problems. Mitteilungen aus dem Institut für Angewandte Mathematik, No. **8**, pp. 1–78, 1959
7. Faddeev, D.K., Faddeeva, U.N.: Computational methods of linear algebra. San Francisco: Freeman 1963
8. Golub, G.H., Varga, R.S.: Chebyshev semi-iterative methods, successive over relaxation iterative methods and second order Richardson iterative methods. Numer. Math. **3**, 147 (1961)
9. Golub, G., Kahan, W.: Calculating the singular values and pseudo-inverse of a matrix. SIAM J. Numer. Anal. **2**, 205–224 (1965)
10. Hageman, L.A., Kellogg, R.B.: Estimating optimum overrelaxation parameters. Math Comput. **22**, 60–68 (1968)
11. Hageman, L.A.: The estimation of acceleration parameters for the Chebyshev polynomial and the successive over relaxation iteration methods. AEC Research and Development Report WAPD-TM-1038, June 1972
12. Henrici, P.: Bounds for iterates, inverses, spectral varation and fields of values of non-normal matricies. Numer. Math. **4**, 24–40 (1962)
13. Hestenes, M.R., Stiefel, E.L.: Methods of conjugate gradients for solving linear systems. N.B.S.J. Res. **49**, 409–436 (1952)
14. Householder, A.S.: The theory of Matricies in numerical analysis, pp. 37–57. New York: Blaisdell 1964
15. Kershaw, D.S.: The incomplete Cholesky-conjugate gradient method for the iterative solution of systems of linear equations. Lawrence Livermore Lab Report, UCRL-78333, Livermore, California, 1976
16. Kincaid, D.R.: On complex second-degree iterative methods. SIAM J. Numer. Anal. **II**, No. 2, 211–218 (1974)
17. Kincaid, D.R.: Numerical results of the application of complex second-degree and semi-iterative methods. Center for Numerical Analysis Report, CNA-90, Oct. 1974

18. Kjellberg, G.: On the convergence of successive over relaxation applied to a class of linear systems of equations with complex eigenvalues. Ericsson Technics **2**, 245–258 (1958)
19. Manteuffel, T.A.: An iterative method for solving nonsymmetric linear systems with dynamic estimation of parameters. Digital Computer Laboratory Reports, Rep. UIUCDS-R-75-758, University of Ill., Oct. 1975
20. Manteuffel, T.A.: The Tchebychev iteration for nonsymmetric linear systems. Numer. Math. **28**, 307–327 (1977)
21. Meijerink, J.A., van der Vorst, H.A.: An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Math. Comput. **31**, 148–162 (1977)
22. Paige, C.C.: Bidiagonalization of matrices and solution of linear equations. SIAM J. Numer. Anal. **11**, 197 (1974)
23. Reid, J.K.: A method for finding the optimum successive overrelaxation parameter. Comput. J. **9**, 200–204 (1966)
24. Stewart, G.W.: Introduction to matrix computation. New York: Academic Press 1973
25. Stiefel, E.L.: Kernel polynomials in linear algebra and their applications. U.S. N.B.S. Applied Math Series **49**, 1–22 (1958)
26. Stone H.L.: Iterative solutions of implicit approximations of multidimensional partial differential equations. SIAM J. Numer. Anal. **5**, 530 (1968)
27. Taussky, O.: Some topics concerning bounds for eigenvalues of finite matrices. In: Survey of numerical analysis (J. Todd, ed.), Ch. 8, pp. 279–297. New York: McGraw Hill 1962
28. Varga, R.S.: A comparison of successive over relaxation and semi-iterative methods using Chebyshev polynomials. SIAM J. Numer. Anal. **5**, 39–46 (1957)
29. Varga, R.S.: Matrix iterative analysis. Englewood Cliffs, N.J.: Prentice-Hall 1962
30. Wachspress, E.L.: Iterative solution of elliptic systems, pp. 157–158. Englewood Cliffs, N.J.: Prentice-Hall 1962
31. Widland, O.: A Lanczos method for a class of non-hermitian systems of linear equations. Tech. Rep., Courant Institute, New York (to appear)
32. Wilkinson, J.H.: The algebraic eigenvalue problem. Oxford: Clarendon Press 1965
33. Wrigley, H.E.: Accelerating the Jacobi method for solving simultaneous equations by Chebyshev extrapolation when the eigenvalues of the iteration matrix are complex. **6**, 169–176 (1963)
34. Young, D.M., Edison, H.D.: On the determination of the optimum relaxation factor for the SOR method when the eigenvalues of the Jacobi method are complex. Center for Numerical Analysis Report, CNA-1, September 1970
35. Young, D.: Iterative solution of large linear systems, pp. 191–200. New York-London: Academic Press 1971

Received April 28, 1978