SLAC-PUB-774
CS 141
October 1969
(MISC)

BOUNDS FOR THE ERROR OF LINEAR SYSTEMS OF EQUATIONS USING THE THEORY OF MOMENTS

Germund Dahlquist[+]

Stanley C. Eisenstat

Gene H. Golub[*]

+ Royal Institute of Technology, 10044 Stockholm 70, Sweden.

* The work of this author was in part supported by the Atomic Energy Commission and by the National Science Foundation.

(To be submitted for publication)

Abstract

Consider the system of linear equations $A\underset{\sim}{x} = \underset{\sim}{b}$ where A is an $n \times n$ real symmetric, positive definite matrix and $\underset{\sim}{b}$ is a known vector. Suppose we are given an approximation to $\underset{\sim}{x}$, $\underset{\sim}{\xi}$, and we wish to determine upper and lower bounds for $\|\underset{\sim}{x} - \underset{\sim}{\xi}\|$ where $\|\ldots\|$ indicates the euclidean norm. Given the sequence of vectors $\{\underset{\sim}{r}_i\}_{i=0}^{k}$ where $\underset{\sim}{r}_i = A\underset{\sim}{r}_{i-1}$ and $\underset{\sim}{r}_o = \underset{\sim}{b} - A\underset{\sim}{\xi}$, it is shown how to construct a sequence of upper and lower bounds for $\|\underset{\sim}{x} - \underset{\sim}{\xi}\|$ using the theory of moments.

In addition, consider the Jacobi algorithm for solving the system $\underset{\sim}{x} = M\underset{\sim}{x} + \underset{\sim}{b}$ *viz.* $\underset{\sim}{x}_{i+1} = M\underset{\sim}{x}_i + \underset{\sim}{b}$. It is shown that by examining $\underset{\sim}{\delta}_i = \underset{\sim}{x}_{i+1} - \underset{\sim}{x}_i$, it is possible to construct upper and lower bounds for $\|\underset{\sim}{x}_i - \underset{\sim}{x}\|$.

1. Introduction

Consider the system of linear algebraic equations

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b} \tag{1.1}$$

where A is an $n \times n$ real symmetric, positive definite matrix and $\underset{\sim}{b}$ is a given vector. Assume we have an approximation to $\underset{\sim}{x}$ so that

$$\underset{\sim}{x} = \underset{\sim}{\xi} + \underset{\sim}{e} \tag{1.2}$$

where $\underset{\sim}{\xi}$ is the approximation vector and $\underset{\sim}{e}$ is the error vector. We are concerned with determining upper and lower bounds for $\|\underset{\sim}{e}\|$ where $\|\ldots\|$ indicates the euclidean norm of the vector.

In order to compute bounds for the norm of the error vector, it is natural to compute the residual vector,

$$\underset{\sim}{r}_o = \underset{\sim}{b} - A\underset{\sim}{\xi} \quad . \tag{1.3}$$

Thus since $\underset{\sim}{r}_o = A\underset{\sim}{e}$,

$$\frac{\|\underset{\sim}{r}_o\|}{\|A\|} \leq \|\underset{\sim}{e}\| \leq \|A^{-1}\| \, \|\underset{\sim}{r}_o\| \quad . \tag{1.4}$$

Here $\|A\|$ indicates the spectral norm of the matrix A . Assuming that $\|A\| = 1$ (this can be accomplished via a simple scaling of (1.1)), we see that even though $\|\underset{\sim}{r}_o\|$ is "small", the bound for $\|\underset{\sim}{e}\|$ can be quite large when $\|A^{-1}\|$ is very large.

By computing additional information, it is possible to obtain more precise upper and lower bounds on the euclidean length of the error vector. In Section 2, we give an algorithm which depends upon computing an auxilliary sequence of vectors and an explicit knowledge of all the eigenvalues of the matrix A . The bounds are actually obtained as a solution to a linear programming problem.

In Section 3, we use the same sequence of vectors as described in Section 2 but we assume that the only information that the investigator has is an upper bound on the largest eigenvalue of A and a non-trivial lower bound for the smallest eigenvalue. Using the theory of moments, an algorithm is given for determining upper and lower bounds. Then in Section 4, we consider the Jacobi iterative method for solving the system of equations (1.1), and we show it is possible to establish bounds for the error by examining the difference of successive iterates. Finally, a numerical example is given in Section 5. In a future report, we shall give methods for improving the approximate solution using the techniques described in this paper.

2. Bounds using linear programming

Consider the Krylov sequence [6],

$$\underset{\sim}{r}_{i+1} = A\underset{\sim}{r}_i \qquad (i = 0,1,\ldots,k-1 \leq n)$$

where $\underset{\sim}{r}_o$ is defined by (1.3). Thus

$$\underset{\sim}{r}_i = A^i\underset{\sim}{r}_o \qquad (i = 0,1,\ldots,k) \quad .$$

We define

$$(x,y) = \sum_{i=1}^{n} x_i y_i$$

so that

$$\begin{aligned}(r_p,r_q) &= (A^p r_o, A^q r_o)\\ &= (A^{p+q} r_o, r_o)\\ &\equiv \mu_{p+q} \qquad (p,q = 0,1,\ldots,k) \quad .\end{aligned}$$

Since A is symmetric and positive definite, we have

$$A\underset{\sim}{u}_i = \lambda_i \underset{\sim}{u}_i \qquad (i = 1,2,\ldots,n)$$

with

$$(\underset{\sim}{u}_i,\underset{\sim}{u}_j) = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

and

$$0 < a \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq b \quad .$$

Now writing

$$\underset{\sim}{r}_o = \sum_{i=1}^{n} \alpha_i \underset{\sim}{u}_i \quad ,$$

we have

$$\mu_m = (A^m r_o, r_o) = \sum_{i=1}^{n} \alpha_i^2 \lambda_i^m \qquad (m = 0,1,\ldots,2k) \quad . \tag{2.1}$$

Since $\underset{\sim}{e} = A^{-1}\underset{\sim}{r}_o$,

$$\|\underset{\sim}{e}\|^2 = (A^{-2}\underset{\sim}{r}_o, \underset{\sim}{r}_o) = \sum_{i=1}^{m} \alpha_i^2 \lambda_i^{-2} = \mu_{-2} \quad . \tag{2.2}$$

Equations (2.1) and (2.2) are equivalent to

$$\mu_m = \int_a^b \lambda^m d\alpha(\lambda) \qquad (m = -2,0,1,\ldots,2k) \tag{2.3}$$

where the weight function of the Stieltjes integral is determined as follows:

$$\alpha(\lambda) = 0 \qquad \text{for } a \leq \lambda \leq \lambda_1 \ ,$$

$$\alpha(\lambda) = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_t^2 \qquad \lambda_t < \lambda \leq \lambda_{t+1} \quad (t = 1,2,\ldots,n-1) \ ,$$

$$\alpha(\lambda) = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2 \qquad \lambda_n < \lambda \leq b \quad .$$

The problem of determining an upper and lower bound for $\|\underset{\sim}{e}\|$ is equivalent to the following:

Given the $(2k+1)$ moments μ_i , determine upper and lower bounds for μ_{-2} .

The solution to this classical problem (cf. [7]) is dependent upon the amount of information available.

Suppose we know the eigenvalues of the matrix A . An example of this is the usual five point approximation to Poisson's equation with Dirichlet boundary conditions in a rectangular region. Thus to determine an upper bound for $\|\underset{\sim}{e}\|^2$, we wish to maximize

$$\sum_{i=1}^{n} \gamma_i \lambda_i^{-2}$$

subject to the constraints

$$\left.\begin{aligned} \sum_{i=1}^{n} \gamma_i \lambda_i^m &= \mu_m \qquad (m = 0,1,\ldots,2k) \\ \gamma_i &\geq 0 \qquad (i = 1,\ldots,n) \quad . \end{aligned}\right\} \tag{2.4}$$

The numerical solution to this problem can be obtained by the simplex algorithm of G. Dantzig [3]. Special techniques may be used to take advantage of the fact that a Vandermonde system is solved at each iteration of the simplex algorithm. A lower bound for $\|\underset{\sim}{e}\|^2$ may be obtained by determining the minimum of $\sum_{i=1}^{n} \gamma_i \lambda_i^{-2}$ subject to the constraints (2.4) by the simplex algorithm.

3. Error bounds using the theory of moments

In the more usual situation, one has the information that

$$0 < a \leq \lambda_i \leq b \quad \text{for} \quad i = 1,\ldots,n \ .$$

This is a problem in the classical theory of moments which has been solved by A. A. Markov. In order to give a numerical algorithm for determining bounds for $\|\underset{\sim}{e}\|$, we review some facts from the theory of Gaussian quadrature.

Suppose we are given $\{\mu_i\}_{i=0}^{2k}$, and a function $\varphi(\lambda)$ $(a \leq \lambda \leq b)$, and we wish to determine (L,U) so that

$$L \leq \int_a^b \varphi(\lambda)d\alpha(\lambda) \leq U \ .$$

We can determine a quadrature rule such that

$$\mu_r = \int_a^b \lambda^r d\alpha(\lambda) = \sum_{i=1}^{k} A_i t_i^r + \sum_{j=1}^{m} B_j z_j^r \quad \text{for} \quad r = 0,1,\ldots,2k+m-1$$

where $\{A_i, t_i\}_{i=1}^{k}$ and $\{B_j\}_{j=1}^{m}$ are unknown and $\{z_j\}_{j=1}^{m}$ is specified. Then

$$\int_a^b \varphi(\lambda)d\alpha(\lambda) = \sum_{i=1}^{k} A_i\varphi(t_i) + \sum_{j=1}^{m} B_j\varphi(z_j) + R[\varphi]$$

where

$$R[\varphi] = \frac{\varphi^{(2k+m)}(\eta)}{(2k+m)!} \int_a^b \prod_{j=1}^{m}(\lambda-z_j)\Big[\prod_{i=1}^{k}(\lambda-t_i)\Big]^2 d\alpha(\lambda) \tag{3.1}$$

$$a < \eta < b \ .$$

Thus if $\varphi(\lambda) = \lambda^{-2}$ and $m = 1$,

$$R[\lambda^{-2}] = -2(k+1)\eta^{-(2k+3)} \int_a^b (\lambda-z_1)\Big[\prod_{i=1}^{k}(\lambda-t_i)\Big]^2 d\alpha(\lambda) \ .$$

Hence for $z_1 = a$, the Gauss-Radau type quadrature rule yields an upper bound for $\int_a^b \lambda^{-2} d\alpha(\lambda)$ and if $z_1 = b$, we have a lower bound. It can be shown (cf. [7, pg. 80]) that these bounds are attainable.

It is not necessary to solve the equations for the quadrature rule. Let us note that

$$\mu_r = \sum_{i=1}^{k} A_i t_i^r + B_1 z_1^r \qquad (r = 0,1,\ldots,2k)$$

where z_1 may be a or b. Let us write

$$\bar{\mu}_r = \sum_{i=1}^{k} \bar{A}_i \bar{t}_i^r + \bar{B}_1 a^r \qquad \text{for all } r \text{ and for } z_1 = a \qquad (3.2)$$

so that

$$\bar{\mu}_{-2} \geq \mu_{-2} \quad .$$

From (3.2), we see that $\bar{\mu}_r$ satisfies a $(k+1)^{\text{th}}$ order difference equation

$$g_o \bar{\mu}_r + \bar{g}_1 \bar{\mu}_{r-1} + \ldots + \bar{g}_k \bar{\mu}_{r-k} - \bar{\mu}_{r-(k+1)} = 0 \qquad (3.3)$$

and $\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_k$, and a are the roots of the characteristic polynomial

$$p(\xi) = \bar{g}_o \xi^{k+1} + \bar{g}_1 \xi^k + \ldots + \bar{g}_k \xi - 1 \quad .$$

Since $\bar{p}(a) = 0$, we must have

$$\bar{g}_o a^{k+1} + \bar{g}_1 a^k + \ldots + \bar{g}_k a - 1 = 0 \quad . \qquad (3.4)$$

Thus using (3.3) and (3.4) and the fact that $\mu_r = \bar{\mu}_r$ for $r = 0,1,\ldots,2k$, we have enough equations to determine $\{\bar{g}_i\}_{i=0}^k$. Having determined $\{\bar{g}_i\}_{i=0}^k$, one can solve for $\bar{\mu}_{-2}$ by recurring twice backwards with equation (3.3).

To determine a lower bound for the error viz. μ_{-2}, it is necessary to solve for $\{g_i\}_{i=0}^k$ from equations similar to (3.3) and the additional equation

$$\underline{g}_0 b^{k+1} + \underline{g}_1 b^k + \ldots - \underline{g}_k b - 1 = 0 \quad .$$

Note to solve for $\{\underline{g}_i\}_{i=0}^{k}$, it is necessary to change only one row in the matrix and one can use the devices given in [2] for solving such a modified system efficiently.

For large k , the system of linear equations which one solves for the coefficients of the difference equation may be quite ill-conditioned. For that reason it is sometimes preferable to solve explicitly for the quadrature rule. As is well known, the nodes of the quadrature rule are the roots of orthogonal polynomials. Now the orthogonal polynomials satisfy a three term recurrence relationship viz.

$$p_{j+1}(\lambda) = (\xi_{j+1} - \lambda)p_j(\lambda) - \eta_j^2 p_{j-1}(\lambda)$$

$$p_{-1}(\lambda) = 0 \quad , \quad p_0(\lambda) = 1 \quad .$$

The coefficients $\{\xi_j\}_{j=1}^{k}$, $\{\eta_j\}_{j=1}^{k-1}$ can be computed directly using the Lanczos algorithm [8].

Again, let

$$\underset{\sim}{r}_0 = \underset{\sim}{b} - A\underset{\sim}{\xi} \quad .$$

We generate a sequence of vectors $\{\underset{\sim}{z}_i\}_{i=0}^{k+1}$ such that

$$(\underset{\sim}{z}_i, \underset{\sim}{z}_j) = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases} \quad .$$

Let $\underset{\sim}{z}_0 = \underset{\sim}{r}_0 \times (\|\underset{\sim}{r}_0\|)^{-1}$.

Then for $j = 0,1,\ldots,k$,

$$\xi_{j+1} = (\underset{\sim}{z}_j, A\underset{\sim}{z}_j) \quad ,$$

$$\underset{\sim}{w}_{j+1} = A\underset{\sim}{z}_j - \xi_{j+1}\,\underset{\sim}{z}_j - \eta_j\,\underset{\sim}{z}_{j-1} \quad , \qquad (\eta_o = 0)$$

$$\eta_{j+1} = \|\underset{\sim}{w}_{j+1}\| \quad ,$$

$$\underset{\sim}{z}_{j+1} = \eta_{j+1}^{-1} \times \underset{\sim}{w}_{j+1} \quad .$$

For numerical stability, one must re-orthogonalize $\underset{\sim}{z}_{j+1}$ with respect to all the previous $\underset{\sim}{z}_j$'s . Let

$$J = \begin{bmatrix} \xi_1 & \eta_1 & & & & \\ \eta_1 & \cdot & \cdot & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \eta_k \\ & & & & \eta_k & \xi_{k+1} \end{bmatrix} \quad .$$

It is well known (cf. [5]) that the eigenvalues of J are the roots of the polynomial $p_{k+1}(\lambda)$. In order to compute the upper bound for $\|\underset{\sim}{e}\|$, we need to compute the Gauss-Radau quadrature rule with the fixed node a . This can be accomplished by the following algorithm suggested by Mr. David Galant [4]. Let

$$C = J - aI$$

so C is a real symmetric positive definite or semi-definite matrix. Let

$$C = FF^T \quad ,$$

where F is the lower triangular Choleski factor of C . Now let

$$\bar{C} = F^T F = \begin{bmatrix} \bar{\xi}_1 & \bar{\eta}_1 & & & \\ \bar{\eta}_1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \bar{\eta}_k \\ & & & \bar{\eta}_k & \bar{\xi}_{k+1} \end{bmatrix}_{(k+1)\times(k+1)}$$

and

$$\tilde{C} = \begin{bmatrix} \bar{\xi}_1 & \bar{\eta}_1 & & & \\ \bar{\eta}_1 & \ddots & \ddots & & \\ & \ddots & \ddots & \bar{\eta}_{k-1} & \\ & & \bar{\eta}_{k-1} & \bar{\xi}_k & 0 \\ & & & 0 & 0 \end{bmatrix}_{(k+1)\times(k+1)}$$

Then the eigenvalues of $(\tilde{C} + aI)$ yield the nodes of Gauss-Radau rule. By using a variant of the algorithm described in [5], it is possible to compute the quadrature rule. In order to compute the Gauss-Radau rule with the fixed node b, one performs similar operations to those described above on the matrix $C = bI - J$.

4. Error bounds for the Jacobi method

Consider the system of equations,

$$\underset{\sim}{C}\underset{\sim}{y} = \underset{\sim}{f} \tag{4.1}$$

where C is a real, symmetric positive definite matrix of order n. Let $D = \text{diag}[(c_{11})^{-1/2},\ldots,(c_{nn})^{-1/2}]$. We may write (4.1) in the form

$$DCDD^{-1}\underset{\sim}{y} = D\underset{\sim}{f} \tag{4.2}$$

or equivalently,

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b} \quad . \tag{4.3}$$

Note the diagonal elements of the matrix A are all equal to one. Hence

$$A = I - M$$

where the diagonal elements of M are zero. We shall assume that M is convergent viz. $\max_{1\leq i\leq n} |\lambda_i(M)| < 1$. The Jacobi method viz.

$$\underset{\sim}{x}_{i+1} = M\underset{\sim}{x}_i + \underset{\sim}{b} \qquad (i = 0,1,\ldots)$$

is frequently used to solve (4.3).

Let

$$\underset{\sim}{e}_i = \underset{\sim}{x} - \underset{\sim}{x}_i = M^i \underset{\sim}{e}_o$$

and

$$\underset{\sim}{\delta}_i = \underset{\sim}{x}_{i+1} - \underset{\sim}{x}_i = M^i \underset{\sim}{\delta}_o \quad :$$

The vector $\underset{\sim}{\delta}_i$ is the difference vector. Since $\underset{\sim}{\delta}_i = \underset{\sim}{x}_{i+1} - \underset{\sim}{x}_i = M\underset{\sim}{x}_i + \underset{\sim}{b} - \underset{\sim}{x}_i = \underset{\sim}{b} - A\underset{\sim}{x}_i$, the difference vector is the residual vector associated with $\underset{\sim}{x}_i$. Note

$$\underset{\sim}{e}_i = (I-M)^{-1}\,\underset{\sim}{\delta}_i = (I-M)^{-1}\,M^i\,\underset{\sim}{\delta}_o \quad .$$

Given $\{\underset{\sim}{\delta}_i\}_{i=0}^{k}$, we compute

$$(\underset{\sim}{\delta}_p, \underset{\sim}{\delta}_q) = (\underset{\sim}{\delta}_o, M^{p+q}\,\delta_o) \equiv \nu_{p+q} \quad , \quad (p+q = 0,\ldots,2k).$$

Thus

$$\|\underset{\sim}{e}_{k+1}\|^2 = (\underset{\sim}{e}_{k+1}, \underset{\sim}{e}_{k+1}) = ((I-M)^{-1}\,M^{k+1}\,\underset{\sim}{\delta}_o, (I-M)^{-1}\,M^{k+1}\,\underset{\sim}{\delta}_o) \quad .$$

Since M is symmetric, we have

$$M\underset{\sim}{u}_i = \xi_i\,\underset{\sim}{u}_i \quad , \quad (i = 1,2,\ldots,n)$$

$$(\underset{\sim}{u}_i, \underset{\sim}{u}_j) = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

and we assume

$$-1 < c \leq \xi_1 \leq \xi_2 \leq \cdots \leq \xi_n \leq d < 1 \quad .$$

Thus

$$\nu_m = \int_c^d \xi^m d\beta(\xi) \qquad (m = 0,1,\ldots,2k)$$

and

$$(\underset{\sim}{e}_{k+1}, \underset{\sim}{e}_{k+1}) = \int_c^d \frac{\xi^{2k+2}}{(1-\xi)^2}\, d\beta(\xi) \quad .$$

We wish to determine upper and lower bounds for $\|\underset{\sim}{e}_{k+1}\|$. This problem was first discussed by H. Weinberger [9] for $k = 1$.

Again, if the eigenvalues of M are known then one can use linear programming for determining upper and lower bounds of $\|e_{k+1}\|^2$. Thus to determine an upper bound for $\|e_{k+1}\|^2$, we wish to maximize

$$\sum_{i=1}^{n} \omega_i \frac{\xi_i^{2k+2}}{(1-\xi_i)^2}$$

subject to the constraints

$$\sum_{i=1}^{n} \omega_i \xi_i^m = \nu_m \qquad (m = 0,1,\ldots,2k)$$

$$\omega_i \geq 0 \qquad (i = 1,2,\ldots,n) \quad .$$

If the eigenvalues are unknown, then we are unable to use the arguments associated with the residual vector since the $(2k+1)\underline{st}$ derivative of $\varphi(\xi) = \xi^{2k+2}/(1-\xi)^2$ is not of constant sign in the interval (c,d).

Now, if we can determine a polynomial $p_{2k}(\xi)$ such that

$$p_{2k}(\xi) = c_0 + c_1\xi + \ldots + c_{2k}\xi^{2k} \geq \frac{\xi^{2k+2}}{(1-\xi)^2}$$

for $c \leq \xi \leq d$

then this will determine an upper bound for $\|e_{k+1}\|^2$ since

$$\int_c^d p_{2k}(\xi)d\beta(\xi) = c_0\nu_0 + \ldots + c_{2k}\nu_{2k} \geq \int_c^d \frac{\xi^{2k+2}}{(1-\xi)^2} d\beta(\xi) \quad .$$

The polynomial $p_{2k}(\xi)$ is not unique and consequently we desire that polynomial for which

$$c_o\nu_o + \ldots + c_{2k}\nu_{2k} = \min.$$

Unfortunately, there does not seem to be any numerical algorithms which will satisfactorily solve this problem in general.

Let $\lambda = (1-\xi)$ so that

$$\int_c^d \frac{\xi^{2k+2}}{(1-\xi)^2}\, d\beta(\xi) = \int_a^b \frac{(1-\lambda)^{2k+2}}{\lambda^2}\, d\alpha(\lambda)$$

$$= \mu_{-2} - 2(k+1)\mu_{-1} + \sum_{s=0}^{2k} (-1)^s \binom{2k+2}{s+2} \mu_s$$

where

$$a = 1-d \quad , \quad b = 1-c$$

$$\mu_s = \int_a^b \lambda^s d\alpha(\lambda) \qquad (s = -2,-1,\ldots,2k) \quad .$$

It is easy to verify that

$$\mu_s = (-1)^s \Delta^s \nu_o$$

where

$$\Delta\nu_o = \nu_1 - \nu_o$$

$$\Delta^s\nu_o = \Delta(\Delta^{s-1}\nu_o)$$

and hence $\mu_s = (\underset{\sim}{\delta}_o, (I-M)^s \underset{\sim}{\delta}_o)$. Our problem now is to determine upper and lower bounds for

$$\mu_{-2} - 2(k+1)\mu_{-1} \quad .$$

In order that there exist a distribution function $\alpha(\lambda)$ in the interval (a,b) associated with $\{\mu_s\}_{s=-2}^{2k}$, it is necessary and sufficient that

$$M = \begin{bmatrix} \mu_{-2} & \mu_{-1} & \cdot & \cdot & \cdot & \mu_{k-1} \\ \mu_{-1} & \mu_{0} & \cdot & \cdot & \cdot & \mu_{k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu_{k-1} & \mu_{k} & \cdot & \cdot & \cdot & \mu_{2k} \end{bmatrix}_{(k+2)\times(k+2)}$$

and

$$G = \begin{bmatrix} \gamma_{-2} & \gamma_{-2} & \cdot & \cdot & \cdot & \gamma_{k-1} \\ \gamma_{-1} & \gamma_{0} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{k-2} & \cdot & \cdot & \cdot & \cdot & \gamma_{2k-2} \end{bmatrix}_{(k+1)\times(k+1)}$$

where

$$\gamma_j = -[ab\,\mu_j - (a+b)\mu_{j+1} + \mu_{j+2}] \tag{4.4}$$

be positive semi-definite (cf. [1]). It is easy to see why G must be positive semi-definite. Note from (4.4)

$$\gamma_j = -\int_a^b (ab\lambda^j - (a+b)\lambda^{j+1} + \lambda^{j+2})d\alpha(\lambda)$$

$$= \int_a^b \lambda^j(\lambda-a)(b-\lambda)d\alpha(\lambda) \quad .$$

Hence,

$$\underset{\sim}{z}^T G \underset{\sim}{z} = \sum_{i=-1}^{k-1} \sum_{j=-1}^{k-1} \gamma_{i+j} z_i z_j$$

$$= \int_a^b \left(\sum_{i=-1}^{k-1} z_i \lambda^i \right)^2 (\lambda - a)(b-a) d\alpha(\lambda) \geq 0 \quad .$$

A similar argument shows that M must be positive semi-definite. Observe that there are two elements which are unknown in G and two elements which are unknown in M and they occur in either the first row or column of the matrix.

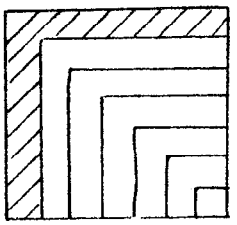

Figure I

Since M and G are positive semi-definite, it is necessary and sufficient that $\det(M) \geq 0$ and $\det(G) \geq 0$ in order for the values μ_{-2}, μ_{-1} be consistent with some distribution $\alpha(\lambda)$ with moments $\mu_0, \mu_1, \ldots, \mu_{2k}$. The positive semi-definite property of M and G is equivalent to the non-negativity of the sub-determinants indicated in Figure I.

We partition the Hankel matrix M as follows:

$$M = \left[\begin{array}{cc|c} \mu_{-2} & \mu_{-1} & \mu_0, \ldots, \mu_{k-1} \\ \mu_{-1} & \mu_0 & \mu_1, \ldots, \mu_k \\ \hline \mu_0 & \mu_1 & \mu_2, \ldots, \mu_{k+1} \\ \vdots & \vdots & \vdots \\ \mu_{k-1} & \mu_k & \mu_{k+1}, \ldots, \mu_{2k} \end{array}\right] \equiv \left[\begin{array}{c|c} A & B \\ \hline B^T & C \end{array}\right] \quad .$$

Since M is positive semi-definite

$$\det(M) = \det(C)\,\det(A - BC^{-1}B^T) \geq 0$$

so that

$$\det\begin{bmatrix} \mu_{-2} + r_1 & \mu_{-1} + r_2 \\ \mu_{-1} + r_2 & \mu_0 + r_3 \end{bmatrix} \geq 0 \qquad (4.5)$$

where

$$\begin{bmatrix} r_1 & r_2 \\ r_2 & r_3 \end{bmatrix} = -BC^{-1}B^T \quad .$$

The matrix $-BC^{-1}B^T$ can easily be computed by applying the Choleski algorithm to the matrix

$$\left[\begin{array}{c|c} 0 & B \\ \hline B^T & C \end{array}\right] \quad .$$

One must begin the pivoting operation, however, from the bottom diagonal element and after k eliminations, the upper 2×2 matrix will contain $-BC^{-1}B^T$. In a similar fashion,

$$\det\begin{bmatrix} -ab\mu_{-2} + (a+b)\mu_{-1} - \mu_0 + s_1 & , & -ab\mu_{-1} + (a+b)\mu_0 - \mu_1 + s_2 \\ -ab\mu_{-1} + (a+b)\mu_0 - \mu_1 + s_2 & , & -ab\mu_0 + (a+b)\mu_1 - \mu_2 + s_3 \end{bmatrix} \geq 0 . \qquad (4.6)$$

From equation (4.5), we see that

$$(\mu_{-2} + r_1)(\mu_0 + r_3) - (\mu_{-1} + r_2)^2 \geq 0$$

and hence since $\mu_0 + r_3 \geq 0$ by the positive semi-definiteness of M

$$\mu_{-2} \geq \frac{(\mu_{-1} + r_2)^2}{\mu_0 + r_3} - r_1 \equiv \frac{(\mu_{-1} + R_2)^2}{\mu_0 + R_3} - R_1 \quad .$$

From (4.6), we have

$$\det \begin{bmatrix} \mu_{-2} + S_1 & \mu_{-1} + S_2 \\ \mu_{-1} + S_2 & S_3 \end{bmatrix} \geq 0$$

where

$$S_3 = \frac{ab\mu_0 - (a+b)\mu_1 + \mu_2 - s_3}{ab} < 0$$

since $ab > 0$,

and hence

$$\mu_{-2} \leq \frac{(\mu_{-1} + S_2)^2}{S_3} - S_1 \quad .$$

Therefore

$$\frac{(\mu_{-1} + R_2)^2}{R_3} - R_1 \leq \mu_{-2} \leq \frac{(\mu_{-1} + S_2)^2}{S_3} - S_1 \quad . \qquad (4.7)$$

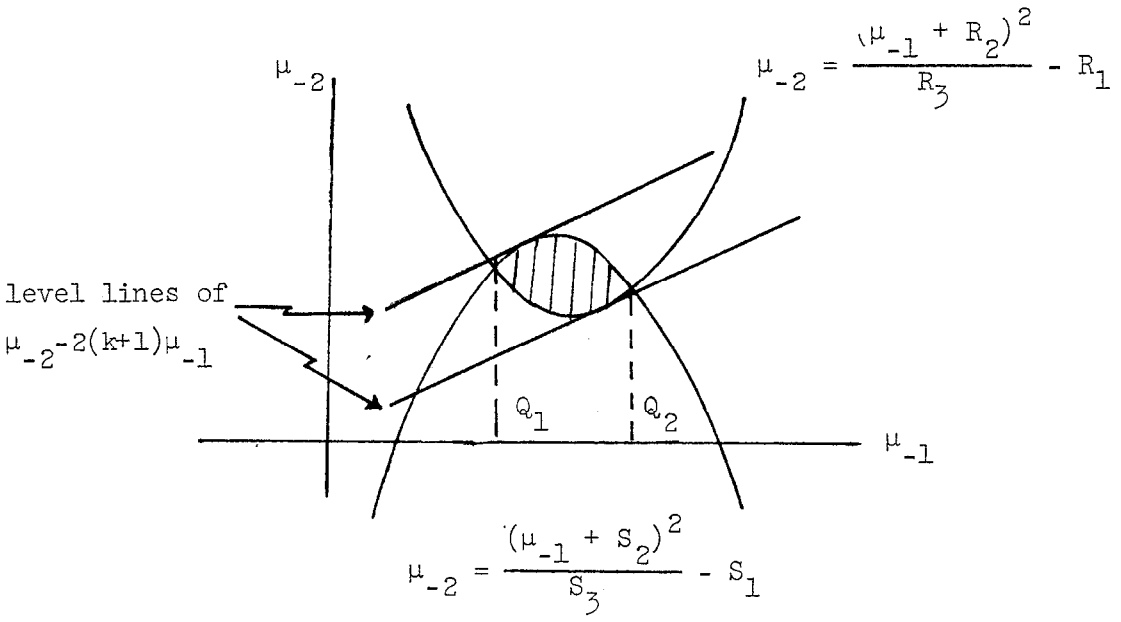

Figure II

Thus to determine the maximum of

$$\mu_{-2} - 2(k+1)\mu_{-1} \quad ,$$

it is simply necessary to examine the boundary of the shaded region in Figure II. A short calculation yields (Q_1, Q_2) for which

$$Q_1 \leq \mu_{-1} \leq Q_2 \quad .$$

Then $\mu_{-2} - 2(k+1)\mu_{-1}$ = maximum subject to (4.7) if

$$\frac{d}{d\mu_{-1}} \frac{(\mu_{-1} + S_2)^2}{S_3} - S_1 = \frac{d}{d\mu_{-1}} 2(k+1)\mu_{-1}$$

$$\text{with} \quad \mu_{-1}^{U} = -S_2 + (k+1)S_3$$

and

$$Q_1 \leq \mu_{-1}^U \leq Q_2 \quad ;$$

otherwise the maximum occurs at $\mu_{-1}^U = Q_1$ or $\mu_{-1}^U = Q_2$ according to

$$\mu_{-1}^U = \max\{Q_1, \min\{-S_2 + (k+1)S_3, Q_2\}\} \quad .$$

Similarly, the minimum occurs at

$$\mu_{-1}^L = \max\{Q_1, \min\{-R_2 + (k+1)R_3, Q_2\}\} \quad .$$

Thus, it is possible to determine upper and lower bounds for $\|e_{k+1}\|$, and these bounds are attainable.

5. A numerical example

Consider the system of equations

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$$

where A is a tri-diagonal matrix with elements $(-1, 2, -1)$ and $\underset{\sim}{b} = \underset{\sim}{\theta}$, the null vector. It is well known that

$$\lambda_j(A) = 2 + 2\cos\frac{j\pi}{n+1} \quad , \quad j = 1, 2, \ldots, n \quad .$$

The Jacobi matrix M is also tri-diagonal and has elements $(1/2,\ 0,\ 1/2)$. Here

$$\lambda_j(M) = \cos\frac{j\pi}{n+1} \quad , \quad j = 1, 2, \ldots, n \quad .$$

The Jacobi method was used for solving the system for $n = 20$ and $\underset{\sim}{x}_0^T = (1, 1, \ldots, 1)$. In Tables I, II, and III, we give the error bounds associated with the error vector of $\underset{\sim}{x}_{10}$. To use the methods of Section 3, we must compute in addition $\{\underset{\sim}{r}_p\}$ for $p = 0, 1, \ldots, k$. In Tables II and III, we give bounds for the error using the difference vectors $\{\underset{\sim}{\delta}_{9-p}\}$ for $p = 0, 1, \ldots, k$. Note that the bound using the residual vector is slightly better than those computed using the difference vectors but it requires additional work to compute $\{\underset{\sim}{r}_p\}_{p=0}^{k}$ whereas the difference vectors are computed in the natural sequence of events. In addition, note that the lower bounds are less influenced by the interval of the eigenvalues than are the upper bounds. Furthermore, we see that in this case that a knowledge of all the eigenvalues does <u>not</u> provide much smaller intervals for the error.

The authors are very pleased to acknowledge the stimulating comments of Professor H. Weinberger of the University of Minnesota and Mr. David Galant of the Ames Research Center.

Error bounds after 10 iterations

$$\| \underset{\sim}{e} \|^2 = 2.700138$$

Table I

Error bounds computed from residual vector using Gauss-Radau quadrature rule

k	Lower bounds		Upper bounds	
	$a=1.116917_{10}-2$ $b=1.988831$	$a=10^{-2}$ $b=1.99$	$a=1.116917_{10}-2$ $b=1.988831$	$a=10^{-2}$ $b=1.99$
1	1.35	1.35	$3.40_{10}1$	$4.21_{10}1$
2	1.55	1.54	$2.05_{10}1$	$2.52_{10}1$
3	1.66	1.66	9.40	$1.12_{10}1$
4	1.81	1.81	6.89	8.08
5	1.89	1.89	4.84	5.50

Table II

Error bounds computed from difference vectors using determinantal inequalities

k	Lower bounds		Upper bounds	
	$a = 1.116917_{10}-2$ $b = 1.988831$	$a = 10^{-2}$ $b = 1.99$	$a = 1.116917_{10}-2$ $b = 1.988831$	$a = 10^{-2}$ $b = 1.99$
1	1.35	1.35	$5.29_{10}1$	$6.59_{10}1$
2	1.43	1.43	$3.88_{10}1$	$4.82_{10}1$
3	1.48	1.48	$2.05_{10}1$	$2.51_{10}1$
4	1.56	1.56	$1.70_{10}1$	$2.08_{10}1$
5	1.59	1.59	$1.29_{10}1$	$1.57_{10}1$

Table III

Error bounds computed from difference vectors using linear programming

k	Lower bounds	Upper bounds
1	1.35	$5.04_{10}1$
2	1.45	$3.86_{10}1$
3	1.55	$1.73_{10}1$
4	1.61	$1.50_{10}1$
5	1.62	$1.04_{10}1$

References

[1] N. I. Aheizer and M. Krein, "Some Questions in the Theory of Moments," American Mathematical Society, Providence, Rhode Island, 1962.

[2] R. Bartels and G. H. Golub, "Stable numerical methods for obtaining the Chebyshev solution to an overdetermined system of equations," Comm. A.C.M., 11 (1968), 401-406.

[3] G. B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

[4] D. Galant, personal communication.

[5] G. H. Golub and J. Welsch, "Calculation of Gauss quadrature rules," M.O.C., 23 (1969), 221-230.

[6] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Co., New York, 1964.

[7] S. Karlin and W. J. Studden, Tchebycheff Systems: With Application in Analysis and Statistics, Interscience Publishers, New York, 1966.

[8] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," J. Res. Nat. Bur. Standards, 45 (1950), 255-282.

[9] H. Weinberger, "A posteriori error bounds in iterative matrix inversion," in Symposium on the Numerical Solution of Partial Differential Equations, James H. Bramble, ed., Academic Press, New York, 1966, 153-163.

Key words

linear algebra

theory of moments

quadrature rules