On the Real Convergence Rate of the Conjugate Gradient Method

Z. Strakoš
General Computing Center of the
Czechoslovak Academy of Science
Pod vodárenskou věží 2
182 07 Praha 8, Czechoslovakia

Dedicated to Gene Golub, Richard Varga, and David Young

Submitted by Owe Axelsson

ABSTRACT

We present a parametrized class of matrices for which the rate of convergence of the conjugate gradient method varies greatly with the parameter and does not appreciably depend on the algorithm implementation. A small change in the eigenvalue distribution can lead to a large change in the sensitivity of CG to rounding errors. A theorem is proved which gives a necessary and sufficient condition for ordering exact arithmetic CC processes for systems with different spectra according to the energy norm of the error. Theorems 4.1 and 4.2 continue Paige's and Greenbaum's work.

1. INTRODUCTION

Let $Ax = b$ be a system of N linear equations, where $x, b \in R^N$, and A is a symmetric, positive definite matrix of order N. The conjugate gradient method (CG) for the solution of this system can be presented in the form

$$\begin{aligned} &x^0, \qquad r^0 = b - Ax^0, \qquad d^0 = r^0, \\ &\tau_i = \frac{(r^i, r^i)}{(d^i, Ad^i)}, \\ &x^{i+1} = x^i + \tau_i d^i, \\ &r^{i+1} = r^i - \tau_i A d^i, \\ &\varphi_i = \frac{(r^{i+1}, r^{i+1})}{(r^i, r^i)}, \\ &d^{i+1} = r^{i+1} + \varphi_i d^i, \qquad i = 0, 1, 2, \ldots . \end{aligned} \tag{1.1}$$

655 Avenue of the Americas, New York, NY 10010 0024-3795/91/$3.50

The CG method was proposed by Hestenes and Stiefel in 1952 [8]. It gives, in exact arithmetic, the correct solution x within N steps. In practice, however, CG is regarded as an iterative method [16], because a sufficiently accurate approximate solution x^k is often obtained in far fewer than N steps.

The theoretical convergence rate (in the absence of rounding errors) has been studied by many authors. Using Chebyshev polynomials, estimates for the convergence rate were developed ([7], e.g.). It was recognized that the rate of convergence depends strongly on the distribution of eigenvalues of A. The effect of roundoff was discussed in Hestenes and Stiefel's original paper, and again by Reid [16] and Jennings [9]. It was also studied extensively by Wozniakowski [26] and Bollen [4, 5]. Results developed by Paige, Parlett, and Simon for the Lanczos algorithm provided further insight into this problem [12–15, 17–19].

During our experiments we have found a parametrized class of matrices for which the behavior of the CG method is quite surprising. We have studied the change in the number of iterations required for convergence as a function of eigenvalue distribution. For some values of the parameter the CG process is sensitive to rounding errors, which leads to very slow convergence. A small change in the parameter (which implies a small change in the eigenvalue distribution) can lead to a large change in the convergence rate. The fact that the CG rate of convergence depends strongly on the distribution of eigenvalues is well known. However, the known results deal with qualitative characteristics of the spectrum (isolated eigenvalues, clustered eigenvalues, eigenvalue gap). In our experiments, the parameter does not affect the eigenvalue distribution qualitatively.

The purpose of our paper is to report our observations and to give at least a partial explanation of them. For our parametrized eigenvalue distributions CG is sensitive to rounding errors even in the case of a small condition number κ (e.g. $\kappa = 100$) and small N (e.g. $N = 12, 24$). The fact that the CG process is strongly influenced by rounding errors for small N is more surprising than for large N. We have studied the sensitivity of CG to rounding errors as a function of eigenvalue distribution. The small problem size N is convenient for this work. Therefore most of our numerical results illustrated in Figures 1–3 have been obtained for $N = 24$ (we note that they represent a small part of the experimental results obtained during our work for this paper). It can be easily verified that similar results can be obtained for $N \gg 24$.

We do not claim that our particular eigenvalue distribution is the only one for which CG suffers from rounding errors. Nevertheless, we think that studying this distribution may lead to a better understanding of the effect of rounding errors on the CG process.

2. THE REAL CONVERGENCE RATE: DESCRIPTION OF EXPERIMENTS

The aim of our experiments was to study the dependence of the real CG convergence rate on the distribution of eigenvalues of the matrix A (in preconditioned CG methods, the distribution of eigenvalues is that of the preconditioned matrix, which depends on the choice of preconditioning method [3]).

We use the following eigenvalue distributions. For a given N, λ_1, and $\kappa = \lambda_N/\lambda_1$ we generate the inner eigenvalues by the formula

$$\lambda_i = \lambda_1 + \frac{i-1}{N-1}(\lambda_N - \lambda_1)\rho^{N-i}, \qquad i = 2,3,\ldots,N-1. \tag{2.1}$$

Thus for $\rho = 1$ we have a uniformly distributed spectrum, and for $\rho < 1$ the spectrum has the "clusterpoint" λ_1. The parameter ρ describes quantitatively the nonuniformity of the spectrum.

For a given spectrum $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ determined by N, λ_1, κ, and ρ we carried out computations using following five algorithms:

ALGORITHM 2A. The simulation procedure proposed by Hageman and Young [7] generates the Euclidean norm of the kth CG error, $\|x - x^k\|$; the A-norm of the kth CG error, $\|x - x^k\|_A$; and the Euclidean norm of the kth CG residual, $\|r^k\|$. The coordinates η_j of the initial residual r^0 in the basis of normalized eigenvectors of the matrix A are randomly generated with $\eta_j \in \langle 0,1 \rangle$. This simulation procedure is equivalent to the CG algorithm in which all vectors are represented by their coordinates in the basis of A's eigenvectors. The operation $y = Ax$ is thus reduced to $y = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)x \equiv Dx$, and the computation corresponds to the ordinary CG process applied to the diagonal system $Dx = b$.

ALGORITHM 2B. The simulation procedure based on Paige and Saunder's SYMMQL algorithm generates $\|x - x^k\|$, $\|x - x^k\|_A$, and $\|r^k\|$, as well as the Euclidean norms $\|x - x_L^k\|$ of the SYMMQL error, $\|r_L^k\|$ of the SYMMQL residual, and $\|r_M^k\|$ of the MINRES residual; cf. [11].

ALGORITHM 2C. The Rutishauser variant of the CG algorithm [27] is applied to the system $Dx + b = 0$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, $b = -r^0 = -(\eta_1, \eta_2, \ldots, \eta_N)^T$, and the initial guess is $x^0 = 0$.

ALGORITHM 2D. The classical variant of the CG algorithm (referred to by Reid as algorithm version 2; cf. [16]).

ALGORITHM 2E. Jacobi acceleration of the Algorithm 2D; cf. [1].

These last two algorithms were applied to the system $A\bar{x} = \bar{b}$ generated by

$$A = GDG^T, \tag{2.2}$$

where $\bar{b}$ and $\bar{x}$ are related to r^0 and x by

$$\bar{b} = Gr^0, \qquad \bar{x} = Gx, \tag{2.3}$$

and G is a randomly generated orthogonal matrix, $GG^T = I$. The initial guess is $\bar{x}^0 = 0$.

3. THE REAL CONVERGENCE RATE: RESULTS AND DISCUSSION

Our observations and conclusions are summarized in three points.

(1) There exists a critical value of the parameter ρ for which the number of CG iterations required for convergence greatly exceeds N even in the case of a small condition number (e.g. $\kappa = 100$) and small N (e.g. $N = 12, 24$). In its neighborhood a small change in ρ (or a small perturbation of the eigenvalue distribution) may cause a large change in the CG convergence rate.

(2) The critical value of the parameter ρ decreases with increasing precision level. For $\rho \geqslant 0.9$, CG processes are practically ordered in the sense of Theorem 3.1.

(3) The observed results practically do not depend on the variant of the CG algorithm used (that is why the variants are not distinguished in the figures).

A detailed discussion and partial explanation of these results is given in the following subsections.

Point (1)

We have observed that the real CG convergence rate strongly depends on the distribution of eigenvalues of the system matrix. For some eigenvalue

distributions CG is sensitive to roundoff even in the case of a small condition number (e.g. $\kappa = 100$) and small N (e.g. $N = 24$), while for a slightly modified spectrum the sensitivity is remarkably smaller. This is illustrated in Figures 1 and 2. Figure 1 shows the number k of CG iterations required to satisfy the criterion $\|x - x^k\| / \|x - x^0\| < 10^{-L}$, $L = 12$, as a function of the parameter ρ for $\lambda_1 = 0.1$, $\kappa = 10^3$, and $N = 12$, 24, and 48. In Figure 2 the size of the problem, N, is fixed ($N = 24$) and the condition number is set at $\kappa = 10$, 10^2, 10^3, and 10^4. All results have been obtained using IBM double precision (DP) arithmetic (56 bit mantissa). We have found that for a given N, κ, and precision level L (here $L = 12$), there exists a *critical value* ρ^*. As ρ approaches ρ^*, the number of iterations $k = k(\rho, L)$ required to satisfy $\|x - x^k\| < 10^{-L}\|x - x^0\|$ greatly exceeds N, reaching its maximum at $\rho = \rho^*$. In the neighborhood of ρ^* a small change in ρ may cause a large change in $k(\rho, L)$.

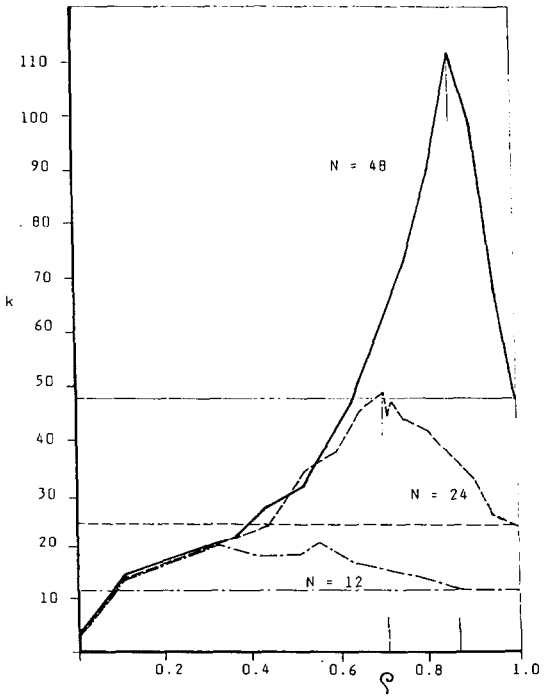

FIG. 1. The number of CG iterations required to reduce the relative CG error measured in the Euclidean norm below 10^{-12}, as a function of the parameter ρ ($\lambda_1 = 0.1$, $\kappa = 10^3$, $N = 12$, 24, and 48).

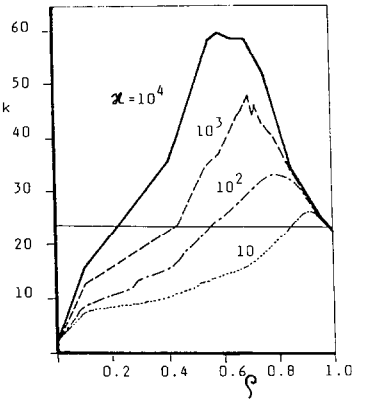

FIG. 2. The number of CG iterations required to reduce the relative CG error measured in the Euclidean norm below 10^{-12}, as a function of the parameter ρ ($\lambda_1 = 0.1$, $N = 24$, $\kappa = 10$, 10^2, 10^3, and 10^4).

The fact that the CG rate of convergence depends strongly on the distribution of eigenvalues is well known. However, the known results deal with qualitative characteristics of the spectrum (isolated eigenvalues, clustered eigenvalues, eigenvalue gap). In our experiments, the parameter ρ does not affect the eigenvalue distribution qualitatively. We wish to call the reader's attention to the fact that a small perturbation of the eigenvalue distribution may cause a large change in the CG convergence rate (or in the sensitivity of CG process to rounding errors).

Point (2)

For a fixed λ_1, N, and κ, the critical value ρ^* decreases with increasing precision level L. This is illustrated in Figure 3, which shows $k = k(\rho, L)$ computed in DP arithmetic for $\lambda_1 = 0.1$, $N = 24$, $\kappa = 10^3$, and precision levels $L = \frac{1}{2}$, 1, 4, 7, and 10. We see a shift of $\rho^*(L)$ from $\cong 0.9$ for $L = 0.5$ to $\cong 0.68$ for $L = 10$. Even more surprising is the fact that $k(\rho^*(1), 1) \gg k(\rho^*(10), 1)$. To reach an accuracy of 10^{-1} for $\rho = \rho^*(1) \cong 0.9$, the CG process needs many more iterations than for $\rho = \rho^*(10) \cong 0.68$. CG rapidly accelerates its rate of convergence in the first case, while in the second it converges slowly with a roughly constant convergence rate. We have not found an explanation.

We see that for any value of L and for any $\rho_1, \rho_2 \to 1$, $\rho_1 > \rho_2$,

$$k(\rho_1, L) \leqslant k(\rho_2, L),$$

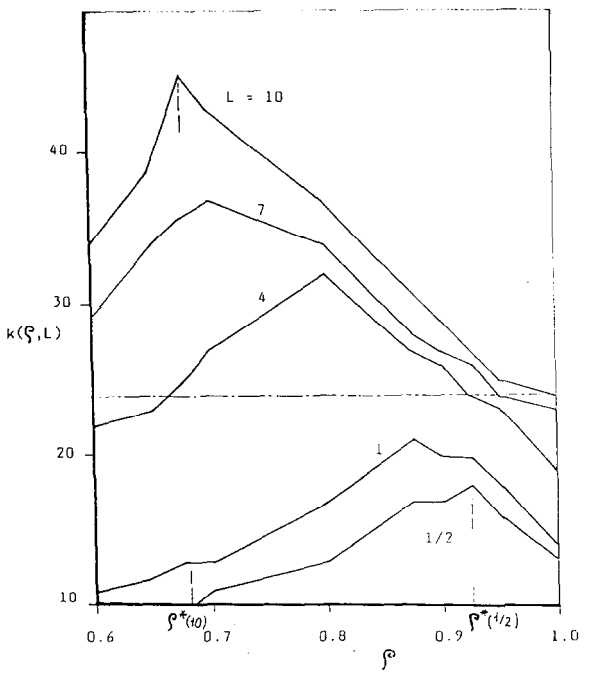

FIG. 3. The function $k = k(\rho, L)$ for $\lambda_1 = 0.1$, $N = 24$, $\kappa = 10^3$, and $L = \frac{1}{2}$, 1, 4, 7, and 10.

i.e., the CG processes computed for different values of ρ, $\rho \to 1$, are in some sense "ordered."

Using Greenbaum's ideas [6], we prove Theorem 3.1, which gives a necessary and sufficient condition for ordering exact arithmetic CG processes for systems with different spectra according to the energy norm of the error.

As above, let x^i be the ith iterate of the exact CG process for $Ax = b$, and r^0 be the initial residual. Corresponding to r^0 there are uniquely determined eigenvalues $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_m$ and orthonormal eigenvectors $z^1, z^2, \ldots, z^m$ of A such that

$$r^0 = \sum_{j=1}^{m} \eta_j z^j, \qquad \eta_j > 0, \quad j = 1, 2, \ldots, m, \tag{3.1}$$

$$\|x - x^0\|_A^2 = \sum_{j=1}^{m} \frac{\eta_j^2}{\lambda_j}. \tag{3.2}$$

Let $By = c$ be a system of linear equations, B a symmetric positive definite matrix. Let y^i be the exact ith CG iterate for this system corresponding to the initial residual $t^0 = c - By^0$, and let t^0 be expressed as

$$t^0 = \sum_{j=1}^{m} \xi_j w^j, \qquad \xi_j > 0, \quad j = 1,2,\ldots,m, \tag{3.3}$$

where $w^1, w^2, \ldots, w^m$ are normalized eigenvectors of B corresponding to its eigenvalues $\mu_1, \mu_2, \ldots, \mu_m$.

DEFINITION 3.1. We call the initial guesses x^0 and y^0 comparable if the corresponding initial residuals are defined by (3.1) and (3.3) and $\eta_j/\sqrt{\lambda_j} = \xi_j/\sqrt{\mu_j}$, $j = 1,2,\ldots,m$.

COMMENT. For the comparable initial guesses x^0 and y^0 we have

$$\|x - x^0\|_A = \|y - y^0\|_B.$$

THEOREM 3.1. *The following three assertions are equivalent:*

(A) *We have*

$$\frac{\lambda_i}{\lambda_j} \leqslant \frac{\mu_i}{\mu_j} \qquad \forall i > j, \quad i,j \in \{1,2,\ldots,m\}. \tag{3.4}$$

(B) *For any comparable initial guesses x^0 and y^0 we have*

$$\|x - x^i\|_A \leqslant \|y - y^i\|_B \qquad \forall i \in \{1,2,\ldots,m\}.$$

(C) *For any comparable initial guesses x^0 and y^0 we have*

$$\|x - x^1\|_A \leqslant \|y - y^1\|_B.$$

Proof. (A) $\Rightarrow$ (B): Using the CG-minimizing property (e.g. [20, 21, 25]) and the expansions (3.1) and (3.3), we can write

$$\|x - x^i\|^2 = \sum_{k=1}^{m} [p_i(\lambda_k)]^2 \zeta_k^2,$$

$$\|y - y^i\|_B^2 = \sum_{k=1}^{m} [q_i(\mu_k)]^2 \zeta_k^2,$$

where p_i, q_i are the minimizing CG polynomials with respect to x^i, y^i, and where $\zeta_k^2 = \eta_k^2/\lambda_k = \xi_k^2/\mu_k$. Greenbaum showed [6, Lemma 1] that there exists an ith degree polynomial s_i, $s_i(0) = 1$, such that $[s_i(\lambda_k)]^2 \leqslant [q_i(\mu_k)]^2$ $\forall k \in \{1, 2, \ldots, m\}$. Applying this, we must have

$$\|x - x^i\|_A^2 = \sum_{k=1}^{m} [p_i(\lambda_k)]^2 \zeta_k^2 \leqslant \sum_{k=1}^{m} [s_i(\lambda_k)]^2 \zeta_k^2$$

$$\leqslant \sum_{k=1}^{m} [q_i(\mu_k)]^2 \zeta_k^2 = \|y - y^i\|_B^2 \qquad \forall i \in \{1, 2, \ldots, m\}.$$

(B) $\Rightarrow$ (C): This implication is trivial.

(C) $\Rightarrow$ (A): Suppose C holds and there are $i, j \in \{1, 2, \ldots, m\}$, $i > j$, such that

$$\frac{\mu_i}{\mu_j} < \frac{\lambda_i}{\lambda_j}.$$

Choosing $\zeta_i = \zeta_j = 1$, $\zeta_k = 0$, $k \in \{\{1, 2, \ldots, m\} \backslash \{i, j\}\}$, we obtain

$$p_1(\lambda) = 1 - \frac{\lambda_i + \lambda_j}{\lambda_i^2 + \lambda_j^2}\lambda, \qquad q_1(\mu) = 1 - \frac{\mu_i + \mu_j}{\mu_i^2 + \mu_j^2}\mu.$$

It is easy to show that

$$\|y - y^1\|_B^2 = \sum_{k=1}^{m} [q_1(\mu_k)]^2 \zeta_k^2 \leqslant \sum_{k=1}^{m} [p_1(\lambda_k)]^2 \zeta_k^2 = \|x - x^1\|_A^2.$$

This contradicts the assumption, and the proof is finished. The ordering

according to the first iteration implies the same ordering for subsequent iterations. ■

Though for $\rho > 0.9$ the CG processes shown in Figure 3 are practically ordered in the sense of Theorem 3.1, this fact can hardly be explained using this theorem. Theorem 3.1 assumes exact arithmetic, while the experimental results in Figure 3 are strongly influenced by rounding errors. Theorem 3.1 can be successfully applied for explanation of experiments, e.g. in the case of CG with full reorthogonalization.

Point (3):

The observed results do not appreciably depend on the variant of the CG algorithm used. Comparable results of Algorithms 2A–2D differ in at most one or two iterations. Jacobi acceleration of CG (Algorithm 2E) does not improve the results shown, e.g., in Figure 3. On the contrary, for $\rho \in \langle 0.5, 0.6 \rangle$ it converges more slowly than ordinary CG. The agreement of real convergence rates obtained using Algorithms 2A, 2C, and 2D shows that the effect of roundoff on the CG process does not depend noticeably on the density of the system matrix (on the accumulation of elemental roundoff). The agreement of the Algorithm 2A and 2B results shows definitely that in our experiments the CG process suffers from rounding errors not due to unstable solution of the implicitly (in Algorithm 2A) or explicitly (in Algorithm 2B) generated tridiagonal system, but due to loss of orthogonality among the residual or normalized residual vectors. This corresponds to Simon's results [17].

4. CONCLUDING REMARKS

In exact arithmetic the convergence of extremal Ritz values implies the acceleration of the CG convergence rate (e.g. [2, 20, 22–24]). On the contrary, the convergence of a computed Ritz pair to an eigenpair is equivalent to the loss of orthogonality caused by rounding errors (e.g. [12]). Jennings observed the influence of large outlying eigenvalues on the CG sensitivity to rounding errors [9]. We present a parameterized eigenvalue distribution with large outlying eigenvalues for which a small change of the parameter can cause a large change in the sensitivity of CG process to rounding errors.

We believe that further progress in our work requires a large number of experiments which relate the real rate of convergence to the loss of orthogonality and the convergence of Ritz pairs.

Fundamental theoretical results concerning the convergence of Ritz pairs and the loss of orthogonality were developed by Paige and Parlett. Greenbaum used Paige's work and proved a very interesting result concerning the real convergence rate of the CG and Lanczos methods. She shows that for a given values of J, finite precision CG applied to a linear system $Ax = b$ converges in steps 1 through J at the same rate as exact CG applied to a certain linear system $\bar{A}_J \bar{x}_J = \bar{b}_J$. The matrix $\bar{A}_J$ has more distinct eigenvalues than A. The eigenvalues of $\bar{A}_J$ all lie within small intervals about the eigenvalues of A [10].

We give a small new contribution to these results. We prove that for any eigenvalue λ_j of A for which the corresponding eigenvector has a nonnegligible component in the initial residual, there is an eigenvalue of $\bar{A}_J$ very near λ_j.

We use the notation similar to [12]. The computed results of the Lanczos process after k steps satisfy

$$AV_k = V_k T_k + \beta_{k+1} v^{k+1} (e^k)^T + \delta V_k. \tag{4.1}$$

Let the exact eigendecomposition of A and T_k be

$$A = U \Lambda U^T, \qquad \Lambda = \operatorname{diag}(\lambda_j), \tag{4.2}$$

$$T_k = Y_k M_k Y_k^T, \qquad M_k = \operatorname{diag}\left(\mu_j^{(k)}\right), \tag{4.3}$$

where the orthonormal matrices U and Y_k have the columns u_j and $y_j^{(k)}$. Then from (4.1)

$$\Lambda U^T V_k Y_k = U^T V_k Y_k M_k + U^T \left[\beta_{k+1} v^{k+1} (e^k)^T + \delta V_k \right] Y_k. \tag{4.4}$$

If we denote

$$W_k = \left(w_{ij}^{(k)} \right)_{i,j} = U^T Z_k = U^T V_k Y_k,$$

$$\Omega_k = \left(\omega_{ij}^{(k)} \right)_{i,j} = U^T \left[\beta_{k+1} v^{k+1} (e^k)^T + \delta V_k \right] Y_k, \tag{4.5}$$

where $Z_k = V_k Y_k$ is the matrix of computed Ritz vectors, then (4.4) can be

written as

$$\Lambda W_k = W_k M_k + \Omega_k \tag{4.6}$$

or

$$\left(\lambda_i - \mu_j^{(k)}\right) w_{ij}^{(k)} = \omega_{ij}^{(k)}. \tag{4.7}$$

We now consider the estimate for $\|\Omega_k\|$. For any vector x

$$\left\|\left[U^T v^{k+1}(e^k)^T Y_k\right]x\right\| = \left\|(e^k)^T (Y_k x) U^T v^{k+1}\right\|$$

$$\leqslant \|x\| \, \|v^{k+1}\| \leqslant \left(1 + \frac{\varepsilon_0}{2}\right)^{1/2} \|x\|,$$

$$\|\Omega_k\| \leqslant \beta_{k+1}\left(1 + \frac{\varepsilon_0}{2}\right)^{1/2} + \sqrt{k}\,\|A\|\varepsilon_1. \tag{4.8}$$

Here $\varepsilon_0, \varepsilon_1$ are small multiples of the relative machine precision ε defined in [12, (2.16)].

We are interested in the value of

$$\min_j |\lambda_i - \mu_j^{(k)}|.$$

The result is obtained in the next theorem.

THEOREM 4.1. *If the eigenvector of A has a nonzero component in the initial vector of the Lanczos process, i.e.*

$$\left|\left(u_i, v^1\right)\right| \neq 0, \tag{4.9}$$

then for some eigenvalue $\mu_j^{(k)}$ *of* T_k

$$|\lambda_i - \mu_j^{(k)}| = \min_t |\lambda_i - \mu_t^{(k)}| \leqslant \frac{k^{1/2}\left(\beta_{k+1} + k^{1/2}\varepsilon_1 \|A\|\right)}{\left|\left(u_i, v^1\right)\right|}\left[1 + o(\varepsilon)\right]. \tag{4.10}$$

Proof. From (4.7)

$$|\lambda_i - \mu_j^{(k)}||w_{it}^{(k)}| \leqslant |\omega_{it}^{(k)}| \leqslant \|\Omega_k\|.$$

Then for any set of real numbers φ_t, $\sum_{t=1}^{k} \varphi_t^2 = 1$, we have

$$|\lambda_i - \mu_j^{(k)}| \left| \sum_{t=1}^{k} \varphi_t w_{it}^{(k)} \right| \leqslant \|\Omega_k\| \sum_{t=1}^{k} |\varphi_t| \leqslant k^{1/2} \|\Omega_k\|, \tag{4.11}$$

$$\sum_{t=1}^{k} \varphi_t w_{it}^{(k)} = u_i^T V_k \left(\sum_{t=1}^{k} \varphi_t y_t^{(k)} \right), \tag{4.12}$$

and we can find coefficients φ_t so that

$$\sum_{t=1}^{k} \varphi_t y_t^{(k)} = e_1. \tag{4.13}$$

Using (4.13), (4.12), (4.11), and (4.8), the proof is completed. ∎

As a consequence we obtain

THEOREM 4.2. *If* (4.9) *holds, then for each eigenvalue* λ_i *of* A *there is some eigenvalue* $\mu_j^{(N+m)}$ *of the Greenbaum matrix* $\bar{A}_J$ *constructed for the* J*th step of the Lanczos process with the property*

$$|\lambda_i - \mu_j^{(N+m)}| \leqslant \frac{(N+m)\xi \|A\|}{|(u_i, v^1)|} [1 + o(\varepsilon)], \tag{4.14}$$

where m *and* ξ *are defined in* [10, *pp.* 22, 36, 46, 50].

Greenbaum warns that Theorem 4.2 does not imply that the finite precision Lanczos process will eventually find all eigenvalues with nonnegligible components in the initial residual. Only if it could be shown that, for some J, the exact Lanczos process applied to $\bar{A}_J$ finds some eigenvalue by step J, would it then follow that the finite precision Lanczos process finds this eigenvalue (A. Greenbaum, personal communication).

I wish to thank Professor A. Greenbaum for suggesting several valuable improvements, and Professor O. Axelsson for his kind support during the preparation of this paper.

REFERENCES

1 O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems — Theory and Computations*, Academic, New York, 1984, 432 pp.
2 O. Axelsson and G. Lindskog, On the rate of convergence of the preconditioned conjugate gradient method, *Numer. Math.* 48:499–523 (1986).
3 O. Axelsson and G. Lindskog, On the eigenvalue distribution of a class of preconditioning methods, *Numer. Math.* 48:479–498 (1986).
4 J. Bollen, Round-off Error Analysis of Descent Methods for Solving Linear Equations, Ph.D. Thesis, Tech. Hogeschool Eindhoven, 1980, 193 pp.
5 J. Bollen, Numerical stability of descent methods for solving linear equations, *Numer. Math.* 43:361–377 (1984).
6 A. Greenbaum, Comparison of splittings used with the conjugate gradient algorithm, *Numer. Math.* 33:181–194 (1979).
7 L. A. Hageman and D. M. Young, *Applied Iterative Methods*, Academic, New York, 1981, 386 pp.
8 M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards* 49:409–436 (1952).
9 A. Jennings, Influence of the eigenvalue spectrum on the convergence rate of the conjugate gradient method, *J. Inst. Math. Appl.* 20:61–72 (1977).
10 A. Greenbaum, Behavior of slightly perturbed Lanczos and conjugate gradient recurrences, *Linear Algebra Appl.* 113:7–63 (1989).
11 C. C. Paige and M. A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM J. Numer. Anal.* 12(4):617–629 (1975).
12 C. C. Paige, Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem, *Linear Algebra Appl.* 34:235–258 (1980).
13 C. C. Paige and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software* 8(1):43–71 (1982).
14 B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, New York, 1980.
15 B. N. Parlett, A new look at the Lanczos algorithm for solving symmetric systems of linear equations, *Linear Algebra Appl.* 29:323–346 (1980).
16 J. K. Reid, On the method of conjugate gradients for solution of large sparse systems of linear equations, in *Large Sparse Sets of Linear Equations* (J. K. Reid, Ed.), Academic, London, 1971, pp. 231–254.
17 H. D. Simon, The Lanczos Algorithm for Solving Symmetric Linear Systems, Ph.D. Thesis, Univ. of California, Berkeley, 1982, 109 pp.
18 H. D. Simon, The Lanczos algorithm with partial reorthogonalization. *Math. Comp.* 42:115–142 (1984).
19 H. D. Simon, Analysis of the symmetric Lanczos algorithm with reorthogonalization methods, *Linear Algebra Appl.* 61:101–131 (1984).

20 A. van der Sluis and H. A. van der Vorst, The rate of convergence of conjugate gradients, preprint 354, Univ. Utrecht, 1984, 29 pp.
21 J. Stoer and R. Freund, On the solution of large indefinite systems of linear equations by conjugate gradient algorithms, preprint 85, Univ. Würzburg, Inst. für Angewandte Mathematik und Statistik, 1981.
22 Z. Strakoš, A remark on the rate of convergence of the conjugate gradient method for the solution of linear systems (in Czech), in *Proceedings of the Conference Algorithms '87*, Slovak Union of Math., Štrbské Pleso, 1987, pp. 287–290.
23 Z. Strakoš, On the rate of convergence of conjugate gradient method for solution of linear systems, in *Proceedings of the European Congress of Simulation*, Prague, 1987, pp. 184–189.
24 Z. Strakoš, The Conjugate Gradient Method for the Solution of Linear Systems (in Czech), Res. Rep. V-279, General Computing Center of the Czechoslovak Academy of Sciences, Prague, 1987, 80 pp.
25 H. A. van der Vorst, Preconditioning by Incomplete Decompositions, Ph.D. Thesis, Univ. Utrecht, 1982, 97 pp.
26 H. Wozniakowski, Roundoff-error analysis of a new class of conjugate-gradient algorithms, *Linear Algebra Appl.* 29:507–529 (1980).
27 J. W. Wilkinson and C. Reinsch, *Handbook for Automatic Computation – Linear Algebra*, Springer, New York, 1971.

Received 5 February 1990; final manuscript accepted 20 November 1990