A Two-Step Iterative Block Lanczos Algorithm for a Dominant Eigenspan

H. G. MILLER AND M. C. STEENKAMP,

Theoretical Physics Division, National Research Institute for Mathematical Sciences of the CSIR, P.O. Box 395. Pretoria 0001. Republic of South Africa

AND

C. J. WRIGHT

Department of Applied Mathematics, University of the Witwatersand, 1 Jan Smuts Avenue, Johannesburg 2001, Republic of South Africa

A two-step iterative block Lanczos algorithm for determining a dominant span of eigenvectors. which under certain conditions may be alternative to the full Lanczos algorithm and the block Lanczos algorithm, is presented. The present method is an extension of the 2×2 Lanczos method for iteratively determining a dominant eigenvector and exhibits better convergence properties than the simple power method. The properties of the method are discussed and a numerical example is presented. © 1987 Academic Press Inc

Iterative methods for determining dominant (or extremal) eigenstates of a symmetric (or Hermitian) matrix are, in some cases, a viable alternative to the full Lanczos algorithm [1–11], particularly if a full reorthogonalization [10, 12] or a selective orthogonalization [13, 15] of the Lanczos vectors is deemed necessary. The 2×2 Lanczos algorithm [12], for example, represents a simple convergent means of obtaining the dominant eigenstate of a Hermitian matrix, A, which avoids the aforementioned problems by confining all iterations to a two-dimensional Krylov subspace [15], $K_n^{(2)}$ spanned by the linearly independent, but non-orthogonal vectors $|n\rangle$ and $A |n\rangle$. Choice of $|n\rangle$ as the eigenstate, with the smallest (largest) eigenvalue obtained from the Rayleigh–Ritz (RR) procedure applied to the projection of A onto $K_{n-1}^{(2)}$, yields an algorithm that ultimately will converge to the exact eigenstate of A with the smallest (largest) eigenvalue, provided that this eigenstate is present in the original start vector $|1\rangle$. In practice, the effect of roundoff errors always ensures that some component of the dominant eigenstate enters the iterated vectors even if it is absent from the start vector. The 2×2 Lanczos algorithm combines the programming simplicity of subspace iteration methods with the better convergence properties of the Lanczos algorithm.

383

Since only two vectors are involved in each iteration step, less computer storage is required and no numerical problems should arise from the orthogonalization of the two vectors in $K^{(2)}$ required by the RR procedure. The 2×2 Lanczos algorithm, although convergent, converges at a rate slower than the full Lanczos algorithm, but faster than the power method and, as previously indicated, yields only one eigenvector.

It is not always clear which of the variations on the Lanczos algorithm is best suited to a particular eigenproblem and computer environment. We show the usefulness of a two-step iterative block Lanczos variant for determining the dominant eigenspan of a symmetric matrix, particularly when computer memory is limited. This algorithm is a simple extension of the 2×2 Lanczos algorithm. The idea is as follows:
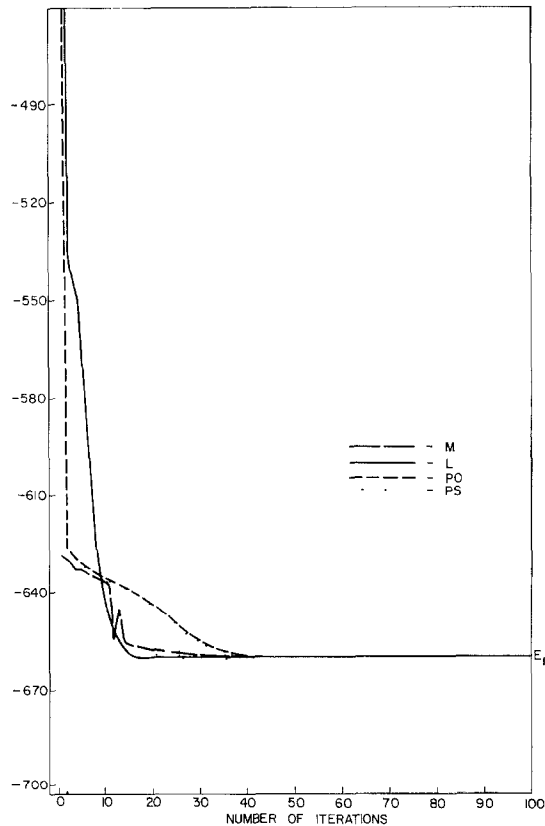

FIGURE 1

FIG. 1–5. The convergence rate of the five lowest-lying eigenvalues in a random 100×100 symmetric matrix for: L = Lanczos method; M = present method; PO = power method without acceleration; PS = power method with RR acceleration. The five startvectors have all been chosen as $10^{-6} \times$ (sum of first ten exact eigenvectors) $+ 0.1 \times$ (sum of next seven exact eigenvectors) before normalization.

From the span of vectors $Q_N^{(n)} = \{|n_1\rangle, |n_2\rangle \cdots |n_N\rangle\}$ construct the span of vectors $AQ_N^{(n)} = \{A|n_1\rangle, A|n_2\rangle \cdots A|n_N\rangle\}$. Orthogonalize the $2N$ vectors contained in $Q_N^{(n)} + AQ_N^{(n)}$ and diagonalize A in this subspace, (that is, perform a RR step in the subspace $Q_N^{(n)} + AQ_N^{(n)}$). Choose the N eigenvectors with the smallest (largest) eigenvalues as the span $Q_N^{(n+1)}$. Such iterations will yield the dominant span of the exact eigenvector of A, that is, those with the smallest (largest) eigenvalues, provided, of course, that these eigenvectors are present in the initial span of vectors. As before, roundoff effects ensure that this proviso is satisfied in practice.

The present algorithm is a sub-algorithm of an algorithm originally proposed by Karush [16] and which was later extended to become the block Lanczos algorithm [17–20]. Here, as in the present algorithm, one chooses a set of n mutually orthogonal vectors which are taken to be columns of a matrix Q_1. Associated with Q_1 is a big Krylov subspace

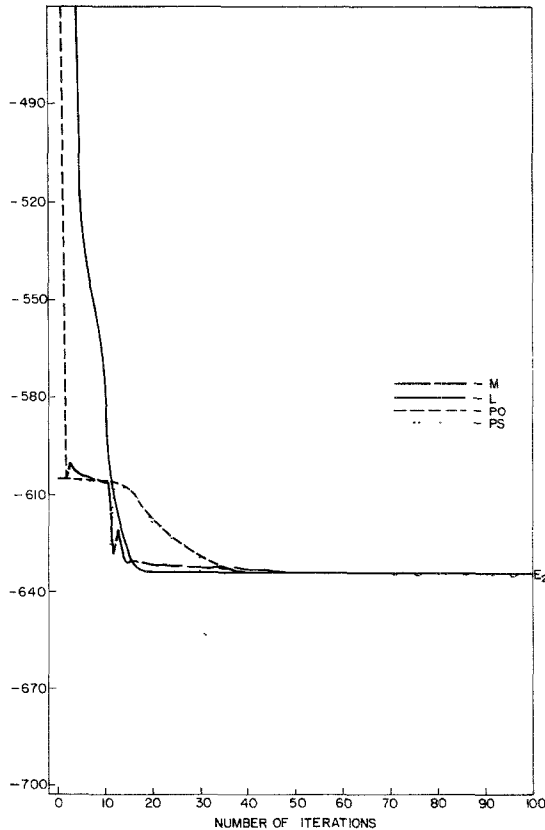$$K^J(Q_1) = \text{span}(Q_1, AQ_1 \cdots A^{J-1}Q_1).$$

FIGURE 2

Reorthogonalizing sequentially the vectors in each member of the span with the vectors contained in the preceding elements of the span yields an orthonormal basis

$$\hat{Q}_1 = (Q_1, Q_2,..., Q_j).$$

In this basis, the projection of A is a block tridiagonal matrix.

In the block Lanczos as in the present algorithm, multiple eigenvalues with multiplicities up to the block size can be computed. However, in the block Lanczos algorithm the first n eigenpairs which converge need not necessarily be the eigenpairs with the largest (or smallest) eigenvalues. In general the n eigenpairs which converge the fastest are a mixture of the eigenvectors with highest and lowest eigenvalues. This may lead to some problems in choosing the block size. This problem is avoided in the present algorithm.

The present algorithm is related to the simple power method for obtaining the dominant span of eigenvectors [15]. The matrix A must possess either a positive or a negative definite spectrum. This presents no problem, since the spectrum may be
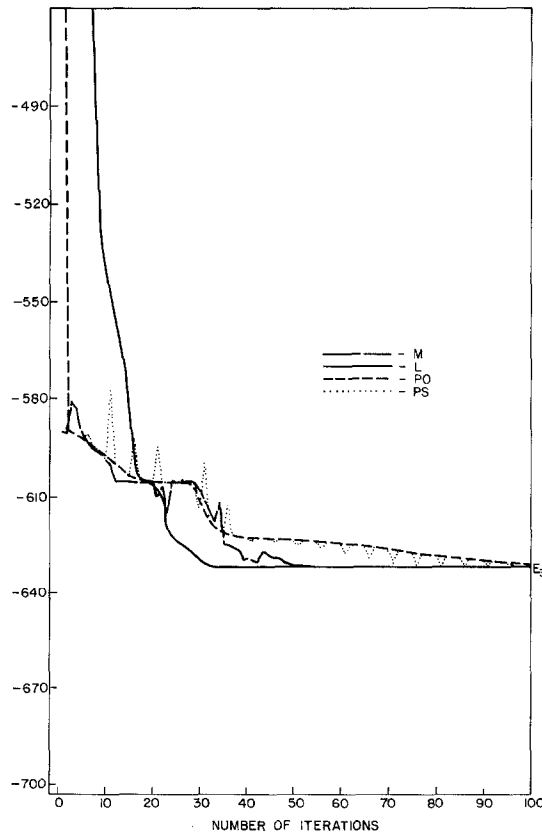

FIGURE 3

shifted to ensure that either of these conditions is fulfilled. The present algorithm improves the convergence rate of the power method.

It is well known that the convergence of the power method can, in principle, be improved by allowing for a spectral shift of the matrix A during the iterations. The iterated span of N vectors may be given by

$$\tilde{Q}_N^{(k+1)} = N^{k+1} \prod_{n=1}^{k+1} (A - \Lambda_n)\, \tilde{Q}_N^{(1)} = N^{k+1}(A - \Lambda_{k+1})\, \tilde{Q}_N^{(k)}, \qquad (1)$$

where Λ is a diagonal matrix containing the spectral shifts, N is an orthonormalization matrix, and the vectors in Q_N have been orthonormalized. To obtain the dominant span of eigenvectors, one may determine the spectral shifts in an optimal manner by requiring that the trace of the matrix

$$\Omega_N^{(k+1)} = (\tilde{Q}_N^{(k)})^\dagger\, (A - \Lambda_N^{k+1})(N^{k+1})^\dagger\, A N^{k+1}(A - \Lambda_N^{(k+1)})\, \tilde{Q}_N^{(k)} \qquad (2)$$

be an extremum. This implies that the trace of the eigenvalues of the matrix A projected onto the subspace spanned by the vectors in $\tilde{Q}_N^{(k+1)}$ must be an extremum.
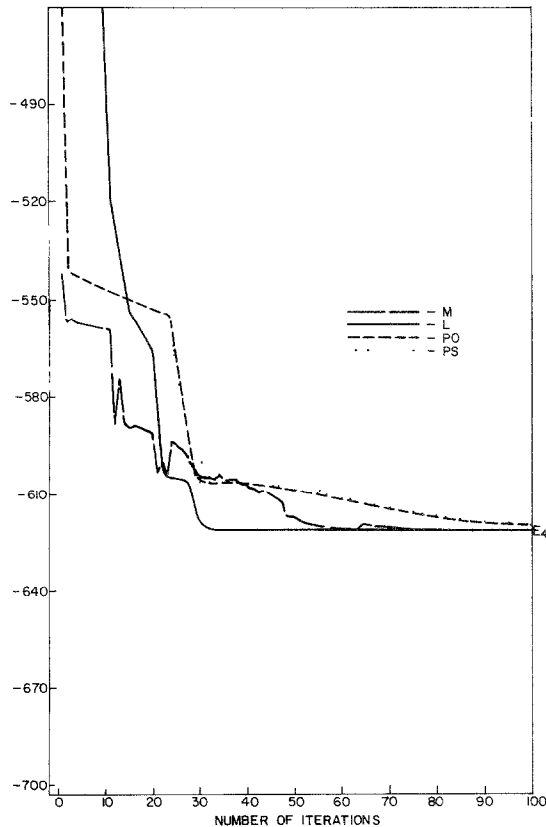

FIGURE 4

Furthermore, the span of vectors in $\tilde{Q}_N^{(k+1)}$ may always be expressed in the following manner

$$\tilde{Q}_N^{(k+1)} = C\tilde{Q}_N^{(k)} + (A - \Omega_N^{(k)})\,\tilde{Q}_N^{(k)}, \tag{3}$$

where

$$\Omega_N^k = (Q_N^{(k)})^\dagger\, AQ_N^{(k)} \tag{4}$$

and C is a suitably constructed matrix of constants. Since the trace of the eigenvalues of the dominant N-dimensional span of the eigenvectors obtained from the diagonalization of A projected onto the subspace spanned by the $2N$ vector contained in $\tilde{Q}_N^{(k)}$ and $(A - \Omega_N^k)\,\tilde{Q}_N^{(k)}$ is an extremum, choosing this set of eigenvectors for $Q_N^{(k+1)}$ corresponds to making an optimal choice for the corresponding set of spectral shifts. Therefore, the requirement that spectral shifts be determined in an optimal manner in each iteration step is satisfied in the present algorithm and, hence, the convergence rate of the dominant span of vectors will be improved.
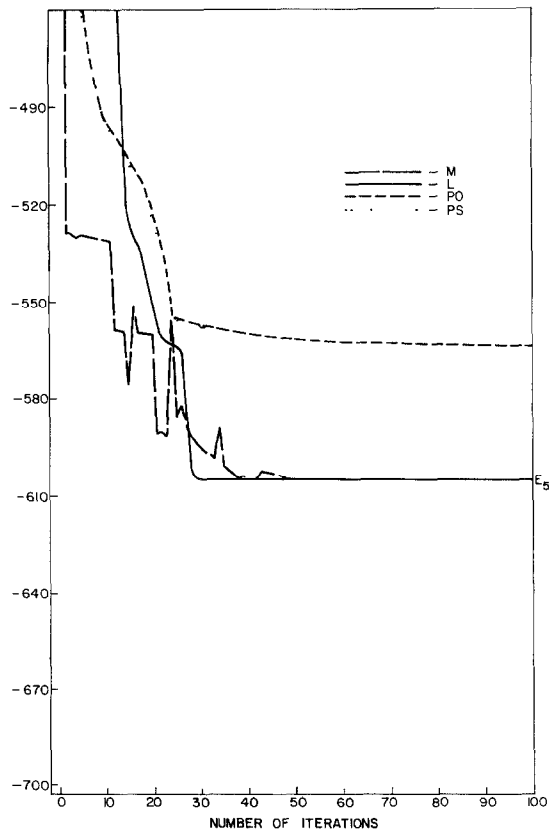

FIGURE 5

To illustrate the convergence of the present method as compared with other methods, the following calculations were performed. A 100×100 symmetric random matrix was generated and shifted by an amount somewhat greater than the largest eigenvalue of the matrix. This ensures that the matrix has a negative definite spectrum.

A dominant span of eigenvectors, containing the first five eigenvectors with the smallest eigenvalues, has been calculated for various choices of the initial span of start vectors by use of the present method, the power method, and, for comparison, the full Lanczos algorithm. For the various cases, the start vector for the Lanczos algorithm has been taken to be the first vector in the initial span.

Numerical results are presented in Figs. 1–7. As might be expected [12] the convergence rate of the present method is faster than that of the power method but, in most cases, slower than the Lanczos method for each member of the span (see Figs. 1–5). This is independent of the choice of the initial span of start vectors, as
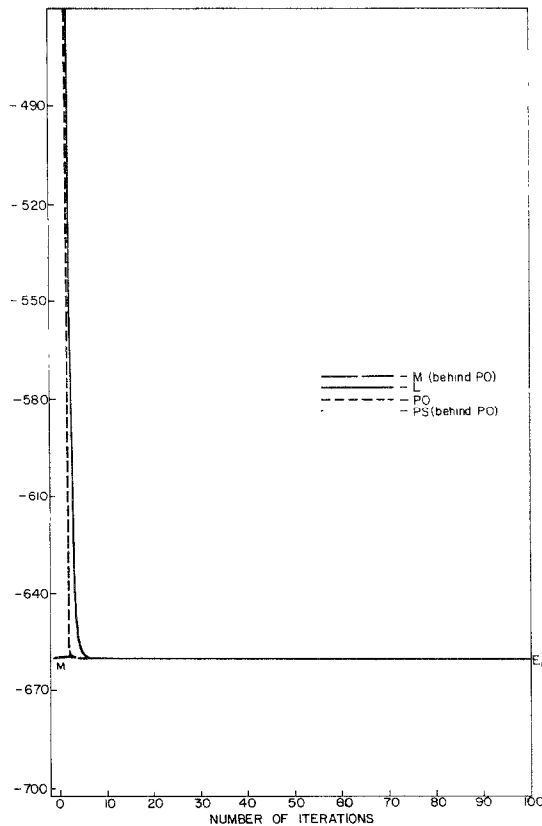

FIG. 6. The lowest lying eigenvalue of the first five eigenvalues that have been calculated in the same matrix by using a second span of start vectors, namely, $0.1 \times$ (ith eigenvector) $+ 10^{-6} \times$ (sum of other 16 out of first 17) eigenvectors for the ith start vector, $i = 1,...,5$ before normalization. L, M, PO, and PS as in Figs. 1–5.

may be seen in Figs. 1, 6, and 7 in which the convergent rate of the smallest eigen-value for different choices of the initial span of vectors is given. The question of block size is relevant. We have chosen a block size equal to the size of the span being sought.

In the numerical calculations using the power method, the reorthogonalization of the vectors in the span in each iteration step is accomplished by a diagonalization of the corresponding norm matrix. This method of reorthogonalization yields a somewhat better convergence rate for all vectors in the span than the Gram–Schmidt (GS) method.

A slight improvement in the convergence rate of the power method has also been obtained by performing an RR diagonalization of A in the subspace of the span of vectors in every 5th iteration step and by using the eigenvectors obtained in the next iteration step of the power method (see Figs. 1–7). Implementation of an RR diagonalization in every iteration step, however, is computationally costly and need not necessarily accelerate the rate of convergence [15].
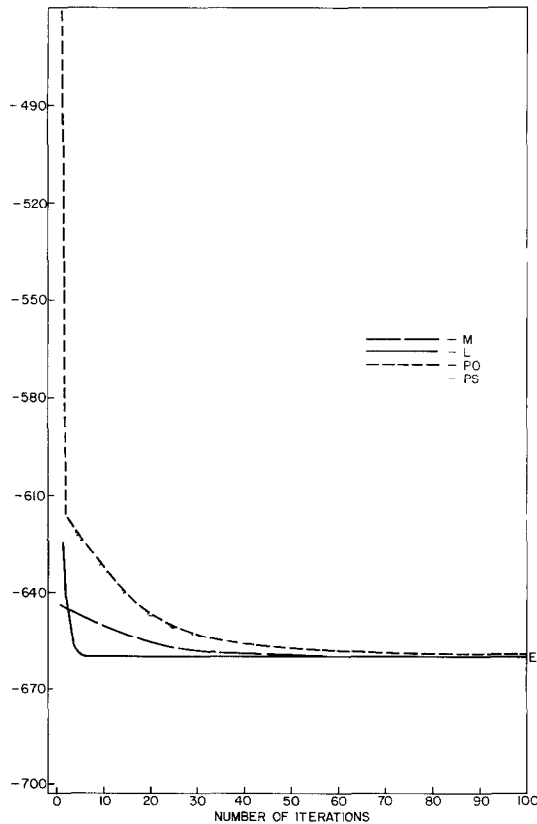

FIG. 7.   The lowest lying eigenvalue for a third span of start vectors, namely, $0.1 \times$ (sum of first ten exact eigenvectors) $+ 10^{-6} \times$ (sum of next seven exact eigenvectors) before normalization. L, M, PO, and PS as in Fig. 5.

To make it clear that the convergence of all the methods depends very strongly on the choice of the span of start vectors, we present in Table I the approximate values of the dominant eigenvalue after twenty iterations. The Lanczos method is the least sensitive to the initial choice of the start vector.

The amount of computational work in the present algorithm is only slightly more than that in the power method. An additional $2N \times 2N$ diagonalization is required to determine the start span of N vectors for the next iteration step. However, the present method is superior to the simple power method because it corresponds to the power method with a spectral shift in which the spectral shifts are determined optimally in each iteration step. Bear in mind that each iteration step in the present method as well as the power method is computationally more expensive than an iteration step of the Lanczos vectors, since in each iteration step the matrix A must be multiplied by each member of the span.

Although the present method yields the dominant span of eigenvectors with a convergence rate faster than that of the power method, its convergence rate is generally somewhat slower than that of the full Lanczos algorithm. Two steps of the present algorithm require four operations of the form "operator times a block" while three steps of the block Lanczos algorithm require three operations of this form. The three-step block Lanczos algorithm obtains an approximation from a $3p$-dimensional Krylov subspace while this algorithm obtains its approximation from a $2p$-dimensional restricted subspace of the same $3p$-dimensional Krylov space. Hence if space permits a multi-step algorithm is preferable.

In spite of the fact that the full Lanczos algorithm or the block Lanczos algorithm may be used to obtain a dominant span of eigenvectors, there are cir-

TABLE I

The Approximate Value of the Dominant Eigenvalue after Twenty Iterations

	Dominant eigenvalue $= -660.0569$		
	a	b	c
Lanczos	-660.0569	-660.0569	-660.0480
Present method	-660.0359	-655.4742	-657.7790
Power with diagonalization of norm matrix and RR every fifth step	-660.0352	-646.5317	-644.7444
Power with GS and RR every fifth step	-660.0352	-646.5317	-643.6314
Power with diagonalization of norm matrix, no RR	-660.0323	-645.8453	-643.9544
Power with GS; no RR	-659.9911	-645.8429	-585.9765

Note. The span of start vectors for column a is the same as in Fig. 6; those for column b as in Fig. 7; and those for column c as in Fig. 1–5.

cumstances in which the use of the present method is warranted and may even be advantageous. In particular, if limited storage is available and fast store can only accomodate a few of the Lanczos vectors, reorthogonalization or selective reorthogonalization of the Lanczos vectors becomes a prohibitive task. In this case the use of the present iteration method is really a viable alternative. Second, if the dominant span of eigenvectors has evenly spaced eigenvalues which are well separated from those of the undesired eigenvectors, the convergence rate of the present method as well as the power method will be very fast. In this case the advantages of the Lanczos algorithm will not be needed. Last, relaxation techniques [20] used to accelerate the convergence rate of the 2×2 Lanczos algorithm may also be generalized and should lead to improvements in the convergence rate of the present method.

ACKNOWLEDGMENTS

REFERENCES

1. C. J. LANCZOS, *J. Res. Nat. Bur. Stand.* **45**, 255 (1950).
2. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *Phys. Lett. B* **45**, 429 (1973).
3. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *Phys. Lett. B* **49**, 133 (1974a).
4. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. A: Math. Nucl. Gen.* **7**, 1374 (1974).
5. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. A: Math. Nucl. Gen.* **7**, 1399 (1974).
6. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. G: Math. Nucl. Gen.* **1**, 17 (1975).
7. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. G: Math. Nucl. Gen.* **1**, 213 (1974).
8. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. G: Math. Nucl. Gen.* **1**, 303 (1975).
9. B. J. COLE, A. WATT, AND R. R. WHITEHEAD, *J. Phys. G: Math. Nucl. Gen.* **1**, 935 (1975).
10. R. R. WHITEHEAD, *Nucl. Phys. A* **182**, 290 (1972).
11. T. SEBE AND J. NACHAMKIN, *Ann. Phys. (N.Y.)* **51**, 100 (1969).
12. W. A. BERGER, H. G. MILLER, K. G. KREUZER, AND R. M. DREIZLER, *J. Phys. A: Math. Gen.* **7**, 1089 (1977).
13. C. C. PAIGE, *J. Inst. Math. Its. Appl.* **10**, 373 (1972).
14. C. C. PAIGE, *J. Inst. Math. Its. Appl.* **18**, 341 (1976).
15. B. N. PARLETT, "The Symmetric Eigenvalue Problem" (Prentice–Hall, Englewood Cliffs, N. J., 1980).
16. W. KARUSH, *Pac. J. Math.* **1**, 232 (1951).
17. W. KAHAN AND B. N. PARLETT, in *Sparse Matrix Computations*, edited by J. Bunch and D. Rose (Academic Press, New York), p. 131.
18. J CULLUM AND W. E. DONATH, Report RC4845 IBM T. J. Watson Research, T. J. Watson Research Labs, Yorktown Heights, New York, 1974 (unpublished).
19. R. R. UNDERWOOD, Ph.D. thesis, Stanford University, Stanford California, 1975 (unpublished).
20. R. M. QUICK, M. C. STEENKAMP, AND H. G. MILLER, *Comput. Math. Appl. Math. Eng.* **48**, 301 (1985).