The Sign Matrix and the Separation of Matrix Eigenvalues*

James Lucien Howland
Department of Mathematics
University of Ottawa
Ottawa, Ontario, Canada K1N 9B4

Submitted by David H. Carlson

ABSTRACT

The sign matrices uniquely associated with the matrices $(M-\zeta_j I)^2$, where ζ_j are the corners of a rectangle oriented at $\pi/4$ to the axes of a Cartesian coordinate system, may be used to compute the number of eigenvalues of the arbitrarily chosen matrix M which lie within the rectangle, and to determine the left and right invariant subspaces of M associated with these eigenvalues. This paper is concerned with the proof of this statement, and with the details of the computation of the required sign matrices.

I. INTRODUCTION

A unique matrix, the sign matrix M_∞, may be associated with any real or complex square matrix $M=M_0$, none of whose eigenvalues is pure imaginary, as the limit of the convergent matrix sequence $\{M_j\}$ generated by the recursion

$$M_{j+1}=\frac{M_j+M_j^{-1}}{2},\qquad M_0=M. \tag{1}$$

The definition, iterative construction and application of the sign matrix to the solution of matrix Riccati and Liapounov equations have been given by J. D. Roberts [10]; further applications have been developed by A. N. Beavers, Jr. and E. D. Denman [3–5] and by A. Halbersberg and Y. Bar-ness [7]. Consideration of the Jordan form of M and of the properties of the Newton iteration $x_{j+1}=(x_j+x_j^{-1})/2$ for determining the square roots of $+1$ shows that the eigenvalues of M are $+1$ and -1; each eigenvalue $+1$ corresponding to an eigenvalue of M with positive real part, and each eigenvalue -1

*Presented at the Matrix Theory Conference held at Auburn University, 19–22 March 1980.

© Elsevier Science Publishing Co., Inc., 1983
52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/010221-12$3.00

corresponding to an eigenvalue of M with negative real part. Further, regardless of the complexity of the Jordan structure of M, M_∞ is always diagonable.

In the present context, the most important property of M_∞ is expressed by the formula

$$\mathrm{Tr}(M_\infty) = p - q,$$

where p, q are the numbers of eigenvalues of M with positive and negative real parts, respectively, and $p + q = \mathrm{order}(M)$. $\mathrm{Tr}(M_\infty)$ is thus analogous to the signature of a quadratic form, such as that formulated by Hermite [8, §II] to count the numbers of zeros of a complex polynomial with positive and negative imaginary parts.

Adapting Hermite's method [8, §IV], it may be noted that the eigenvalues of $(M - \zeta I)^2$, with $\zeta = a + ib$, are $(x - a)^2 - (y - b)^2 + i2(x - a)(y - b)$, where $\lambda = x + iy$ is a typical eigenvalue of M. An eigenvalue of $(M - \zeta I)^2$ has positive real part whenever

$$(x-a)^2 - (y-b)^2 = -[y - b - (x-a)][y - b + (x-a)] > 0,$$

that is, whenever λ lies in one of the sectors defined by the inequality

$$[(y-b) - (x-a)][(y-b) + (x-a)] < 0.$$

Thus, $N(\zeta) = \mathrm{Tr}\left(\left((M - \zeta I)^2\right)_\infty\right)$ counts the number of such eigenvalues with respect to any point ζ. Selecting four points n, s, e, w at the corners of a rectangle whose sides are formed from the lines

$$[(y-b) - (x-a)][(y-b) + (x-a)] = 0 \tag{2}$$

at each point, as illustrated in the diagram Figure 1, the number S of eigenvalues within the rectangle is just

$$S = \frac{N(w) + N(e) - N(n) - N(s)}{4}. \tag{3}$$

When such a rectangle has been shown to contain eigenvalues, a process of subdivision may be applied in the hope of identifying smaller rectangles, of the same type, which locate these eigenvalues more closely. It may be noted that, whenever it is known that the eigenvalues of M are real, it is sufficient to

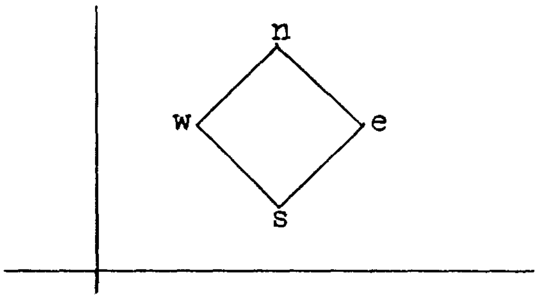

FIG. 1.

calculate $\mathrm{Tr}((M-\zeta I)_\infty)$ at points ζ on the real axis. Then the number of eigenvalues lying in the interval (ζ_1, ζ_2) is

$$S=\frac{\mathrm{Tr}((M-\zeta_1 I)_\infty)-\mathrm{Tr}((M-\zeta_2 I)_\infty)}{2}. \tag{4}$$

II. PROPERTIES OF THE SQUARE ROOT ITERATION

The algorithm (1) for the construction of the sign matrix is suggested by and depends upon the properties of the scalar iteration

$$z_{i+1}=\frac{z_i+z_i^{-1}}{2} \tag{5}$$

for the computation of the square roots of $+1$. It follows from this formula that

$$\frac{z_{i+1}+1}{z_{i+1}-1}=\left(\frac{z_i+1}{z_i-1}\right)^2,$$

so that, if $w_i=(z_i+1)/(z_i-1)$ is the image of z_i under the involutory conformal transformation

$$w=\frac{z+1}{z-1}, \tag{6}$$

then $w_{i+1} = w_i^2$. The square root iteration is thus conformally equivalent to successive squaring. It is known that the transformation (6) maps the imaginary axis $x = 0$ in the z-plane onto the unit circle in the w-plane, and the right (left) half of the z-plane onto the interior (exterior) of the unit circle in the w-plane and vice versa, so that it may be concluded at once that:

(A) If z_i lies in the left (right) half of the z-plane, w_i lies inside (outside) the unit circle in the w-plane. Then w_{i+1} also lies inside (outside) the unit circle, whence z_{i+1} lies in the left (right) half plane.

(B) If z_i lies on the imaginary axis in the z-plane, w_i lies on the unit circle in the w-plane. Then w_{i+1} lies on the unit circle in the w-plane, so z_{i+1} lies on the imaginary axis in the z-plane. A more detailed study of the iteration sequence $\{z_i\}$ in this case shows that it either cycles among the (finitely many) points of a repulsive cycle, or else constitutes an everywhere dense perfect set on the imaginary axis. (See [9, pp. 103–104] for a discussion of the conformally equivalent iteration $w_{i+1} = w_i^2$ in this case.) In practical computation, it is clear that this sequence either cycles indefinitely or else, at some stage, either overflows or underflows the number range of the computer.

(C) If z_0 lies in the left (right) half plane, so that w_0 lies inside (outside) the unit circle, then $\{w_i\}$ is convergent to 0 (∞) and $\{z_i\}$ is correspondingly convergent to -1 ($+1$).

This last observation is the justification for the use of the iteration (1). More specifically, it is easily seen that

$$\frac{z_r + 1}{z_r - 1} = \left(\frac{z_0 + 1}{z_0 - 1}\right)^R = w_0^R,$$

where $R = 2^r$, whence

$$z_r - 1 = \frac{2}{w_0^R - 1}, \qquad z_r + 1 = \frac{2w_0^R}{w_0^R - 1}.$$

If z_0 lies in the left half plane, so that $z_r \to -1$ and $w_0^R \to 0$, then

$$|z_r + 1| = \frac{2|w_0|^R}{|w_0^R - 1|} \leqslant \frac{2|w_0|^R}{1 - |w_0|^R} < \epsilon$$

whenever

$$|w_0|^R < \frac{\epsilon}{\epsilon + 2} < \frac{\epsilon}{2}.$$

Similarly, when z_0 lies in the right half plane, so that $z_r \to +1$ and $w_0^R \to \infty$, then

$$|z_r - 1| = \frac{2}{|w_0^R - 1|} \leqslant \frac{2}{|w_0|^R - 1} < \epsilon$$

whenever

$$|w_0|^{-R} < \frac{\epsilon}{\epsilon + 2} < \frac{\epsilon}{2}.$$

Thus, the rate of convergence of the square root iteration is entirely dependent on the modulus of the initial value w_0. Consideration of the circles of Appolonius (see [1, §3.5, p. 84]) with limit points ± 1—i.e. the images, in the z-plane, of circles $|w| = \rho$—shows that $|w_0|$ is largest when z_0 lies near the imaginary axis. In such cases, convergence may be arbitrarily slow.

It has been suggested [10(b), 8.1] that the iteration (5) may be accelerated by the introduction of real, positive parameters α, β such that (5) becomes

$$z_{i+1} = \frac{\alpha z_i + \beta z_i^{-1}}{\alpha + \beta}, \tag{7}$$

with the possibility that α, β may vary from step to step. To investigate this possibility, it may be noted that (7) implies that

$$\frac{z_{i+1} + 1}{z_{i+1} - 1} = \frac{z_i + 1}{z_i - 1} \times \frac{z_i + \beta/\alpha}{z_i - \beta/\alpha} \tag{8}$$

and that the identity

$$|z + \gamma|^2 - |z - \gamma|^2 = 4\gamma x, \qquad \gamma > 0$$

or

$$\left|\frac{z + \gamma}{z - \gamma}\right|^2 = 1 + \frac{4\gamma x}{|z - \gamma|^2}$$

shows that both factors on the right side of (8) are greater than 1, in modulus, when $x > 0$, and similarly less than 1 when $x < 0$. It may be concluded that the iteration (7) has the properties (A), (B), and (C) detailed above with reference to the iteration (5). By considering the circles of Appollonius referred to above, and noting that they all intersect the unit circle $|z| = 1$ orthogonally, it is seen that

$$\left|\frac{\alpha z/\beta + 1}{\alpha z/\beta - 1}\right| = \left|\frac{z + \beta/\alpha}{z - \beta/\alpha}\right|$$

takes its optimal value—least when $x < 0$, greatest when $x > 0$—when $\alpha|z|/\beta = 1$, or $\beta/\alpha = |z|$. Thus, whenever $z_i \neq 1$, the iteration (7) with $\beta/\alpha = |z_i|$ will be more rapidly convergent than (5).

Finally, it may be noted that alternative iterations to calculate the square root of 1 may be constructed, which have the properties (A), (B), (C) detailed above. For, upon solving the equation

$$\frac{z_{i+1} + 1}{z_{i+1} - 1} = \left(\frac{z_i + 1}{z_i - 1}\right)^r$$

for z_{i+1}, the recursion

$$z_{i+1} = \frac{(z_i + 1)^r + (z_i - 1)^r}{(z_i + 1)^r - (z_i - 1)^r}. \tag{9}$$

is obtained. In case $r = 2$, the iteration (5) is obtained; in case $r = 3$, Bailey's iteration [2]

$$z_{i+1} = \frac{z_i(z_i^3 + 3)}{3z_i^2 + 1}$$

is obtained; in case $r = 4$, the first iterate

$$z_{i+1} = \frac{z_i^4 + 6z_i^2 + 1}{4z_i(z_i^2 + 1)}$$

of (5) is obtained; etc. These functions are probably too elaborate for application to matrix iterations, despite their high orders of convergence.

III. CONSTRUCTION OF THE SIGN MATRIX

The following argument, due, in principle, to Roberts [10(b)], establishes the convergence of the matrix iteration (1) under the most general circumstances, and is reconstructed here for the sake of completeness and for the derivation of several useful corollaries.

DEFINITION. Suppose that $J = V^{-1}MV$ denotes the lower triangular Jordan form of an arbitrary, square, real or complex matrix M, none of whose eigenvalues is zero or pure imaginary. Each column of V is either a right eigenvector of M, or a principal vector of grade greater than one, in the sense of [11, pp. 42–43]. Let J_∞ be the unique diagonal matrix associated with J whose diagonal elements are $+1$ whenever the corresponding diagonal element of J has positive real part, and -1 otherwise. Let the matrix M_∞ be defined by

$$M_\infty = VJ_\infty V^{-1}. \tag{10}$$

THEOREM. *Suppose that matrix M has no zero or purely imaginary eigenvalues. Then the matrix iteration* (1), *begun with $M_0 = M$, is convergent to the matrix M_∞ associated with M by the formula* (10).

Proof. It follows from (1) that, if $J_0 = J$ and $J_{i+1} = (J_i + J_i^{-1})/2$, then $M_j = VJ_jV^{-1}$. Further, all matrices J_j are lower triangular, and their diagonal elements (eigenvalues) are defined, independently, in terms of the eigenvalues of M as starting values, by the scalar iteration (5). Sequences of corresponding eigenvalues of the matrices J_j thus possess the property (C) detailed above.

Since J_∞ may be partitioned, conformably with J, into multiples of the identity matrix, each of appropriate dimension, it follows that M_∞ commutes with M, and with M_j, for each j.

It follows from the definition of M_∞ that the eigenvalues of $M_j + M_\infty$ have nonzero real part for every j; thus the function

$$f_j(M) = (M_j - M_\infty)(M_j + M_\infty)^{-1} \tag{11}$$

is well defined for every j. Moreover, $f_1(M)$ is similar to $(J_1 - J_\infty)(J_1 + J_\infty)^{-1}$, a product of lower triangular matrices, whose diagonal elements, as a consequence of the identity

$$|z+\gamma|^2 - |z-\gamma|^2 = 4\gamma x, \qquad \gamma x > 0,$$

are all less than 1 in absolute value.

The commutativity of M_∞ and M_j leads to the formula

$$f_{j+1}(M) = \left(M_j + M_j^{-1} - 2M_\infty\right)\left(M_j + M_j^{-1} + 2M_\infty\right)^{-1}$$

$$= \left(M_j + M_j^{-1} - 2M_\infty\right)M_j M_j^{-1}\left(M_j + M_j^{-1} + 2M_\infty\right)^{-1}$$

$$= \left(M_j^2 - 2M_\infty M_j + I\right)\left(M_j^2 + 2M_\infty M_j + I\right)^{-1}$$

$$= (M_j - M_\infty)^2 (M_j + M_\infty)^{-2} = [f_j(M)]^2,$$

since, according to the definition (10),

$$M_\infty^2 = I. \tag{12}$$

Since the eigenvalues of $f_1(M)$ are all less than 1 in absolute value, it may be concluded that $f_j(M) \to 0$ as $j \to \infty$.

According to (11)

$$f_j(M) \pm I = (M_j - M_\infty)(M_j + M_\infty)^{-1} \pm (M_j + M_\infty)(M_j + M_\infty)^{-1}$$

$$= (M_j - M_\infty \pm M_j \pm M_\infty)(M_j + M_\infty)^{-1},$$

so that

$$f_j(M) + I = 2M_j(M_j + M_\infty)^{-1},$$

$$f_j(M) - I = -2M_\infty(M_j + M_\infty)^{-1},$$

and

$$2[I - f_j(M)]^{-1} = (M_j + M_\infty)M_\infty^{-1} = (M_j + M_\infty)M_\infty.$$

Therefore

$$[I + f_j(M)][I - f_j(M)]^{-1} = M_j M_\infty^{-1},$$

or

$$M_j = [I + f_j(M)][I - f_j(M)]^{-1} M_\infty,$$

whence $M_j \to M_\infty$ as $j \to \infty$, as anticipated by the choice of notation. ■

COROLLARY 1. *The columns of V, which constitute a basis in n-space and are either eigenvectors of M or principal vectors of M of higher grade, are all eigenvectors of* M_∞.

Thus, for example, the matrix $M_\infty + I$ will have rank equal to the number p of eigenvalues of M, counting multiplicities, with positive real part. The row and column spaces defined by $M_\infty + I$ will be maximal left and right invariant subspaces of M determined by these eigenvalues.

COROLLARY 2. M_∞ *commutes with M, and with each* M_j.

The matrix iteration (1) is designed to construct square roots of the identity, and its convergence may be assessed by comparing M_j^2 with I. Thus, the matrix $E_j = M_j^2 - I$ may be computed for any j, and the Euclidean norm

$$e_j = \|E_j\| = \|M_j^2 - I\|$$

evaluated. It is known, then, that every eigenvalue λ of M_j is such that $|\lambda^2 - 1| < e_j$—i.e. that λ^2 lies in a circle with center $+1$ and radius e_j. From this it may be deduced that every eigenvalue λ of M_j lies in a circle centered at ± 1 with radius $e_j/(1+\sqrt{1-e_j})$. (This is a result of circular arithmetic; see [6, §2, p. 308]). In order, then, that $\mathrm{Tr}(M_j)$ should round correctly to $p - q$, it is sufficient that

$$\frac{ne_j}{1+\sqrt{1-e_j}} < \frac{1}{2}.$$

This inequality may be applied as a criterion to terminate the iteration (1).

The property (B) of the square root iteration shows that convergence of $\{M_j\}$ cannot take place when M has pure imaginary eigenvalues. In this case it is relevant to note that convergence of $\{M_j\}$ implies convergence of $\{\mathrm{Tr}(M_j)\}$ to a real integral limit. A moment's reflection will show that, whether or not $\{M_j\}$ is convergent, $\{\mathrm{Re}[\mathrm{Tr}(M_j)]\}$ will always converge to a real integral limit. When lack of convergence of $\{M_j\}$ suggests the presence of pure imaginary eigenvalues, and j is sufficiently large that $\lim \operatorname{Re}[\operatorname{Tr}(M_j)]$ may be identified and it is practically certain that the remaining eigenvalues are all clustered about ± 1 (and thus removed from the imaginary axis), then the translation which replaces M_j with $\tilde{M}_j = M_j + tI$, where $|t| < 1$ (say $|t| = 0.25$) will give rise to a matrix $\tilde{M}_j$ which can be guaranteed to have no purely imaginary eigenvalues. If $t > 0$, continuation of the iteration from $\tilde{M}_j$ should result in a value for $\lim \operatorname{Re}[\operatorname{Tr}(M_j)] = \lim \operatorname{Tr}(M_j)$ which is greater than the previous limit by the number of pure imaginary eigenvalues of M. If $t < 0$, the same procedure will decrease the previous limit by the same amount.

IV. DETERMINATION OF THE INVARIANT SUBSPACE RELEVANT TO THE EIGENVALUES WITHIN A RECTANGLE

The eigenvalues of $(M - \zeta I)^2$ are $(\lambda - \zeta)^2$, in terms of the eigenvalues λ of M. Consideration of the Jordan form of M shows that the subspace $\mathfrak{M}_\lambda$ spanned by the totality of right eigenvectors and principal vectors of higher grade associated with a particular eigenvalue λ of M is an invariant subspace, of the same type, of $(M - \zeta I)^2$, associated with the eigenvalue $(\lambda - \zeta)^2$. According to Corollary 1 above, this subspace is an invariant subspace, again of the same type, of $\left((M - \zeta I)^2\right)_\infty$. These are the facts which may be applied to determine the number of eigenvalues of M in a rectangle, and the direct sum of the invariant subspaces $\mathfrak{M}_\lambda$ associated with them.

In the simplest case, in which $(\lambda - \zeta)^2$ has nonzero real part for every eigenvalue λ of M and each corner ζ of the rectangle of Figure 1, a sign matrix $\left((M - \zeta I)^2\right)_\infty$ may be associated with each corner. Then, since these matrices are simultaneously diagonalized by the matrix V which transforms M to Jordan form, the corresponding diagonal elements in the canonical forms of $\left((M - \zeta I)^2\right)_\infty$ may be combined by forming

$$S(n, s, e, w) = \tfrac{1}{4}\Big[\left((M - \zeta_w I)^2\right)_\infty + \left((M - \zeta_e I)^2\right)_\infty - \left((M - \zeta_n I)^2\right)_\infty - \left((M - \zeta_s I)^2\right)_\infty\Big]. \tag{13}$$

Examination of the nine cases enumerated in the sketch in Figure 2 shows that the diagonal elements corresponding to eigenvalues of M lying outside the rectangle will cancel out in this sum, while diagonal elements corresponding to eigenvalues lying within the rectangle will total $+1$. Thus, $S(n, s, e, w)$ will be a matrix whose rank equals the number of eigenvalues of M lying within the rectangle, including multiplicities, whose columns span the direct

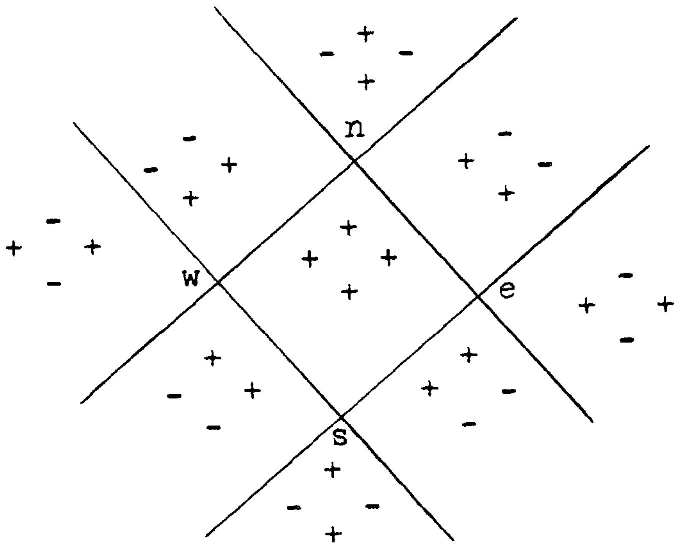

FIG. 2. Signs of the diagonal elements at each corner.

sum of the right invariant subspaces $\mathfrak{M}_\lambda$ of M associated with these eigenvalues, and whose rows span the corresponding left invariant subspace of M. By taking traces of the matrices appearing in (13), the formula (3) is derived.

When, for some point ζ, the matrix $M-\zeta I$ is singular, this point may immediately be identified as an eigenvalue of M. Construction of further rectangles enclosing ζ, and application of the procedures described, will then lead to the invariant subspace associated with the eigenvalue ζ or, if relevant, with a cluster of eigenvalues in a neighborhood of ζ.

According to (2), $(\lambda-\zeta)^2$ has zero real part whenever λ lies on one of the lines which pass through $\zeta=a+ib$ with slopes ± 1. Thus, in the remaining case—in which one or more eigenvalues fall on a side, or a side produced, of the rectangle of Figure 1, but not a corner—two or more of the matrices $(M-\zeta I)^2$ associated with the corners of the rectangle will have one or more pure imaginary eigenvalues different from zero. In such circumstances, the procedures outlined in Section III above may be applied to determine the number of eigenvalues of M lying in the (open) interior of the rectangle. Thus, by making a shift of the side(s) affected, so as to decrease the size of the rectangle, and applying (13) to the sign matrices obtained after the shifts have been made, as described in Section III above, the number of these "interior" eigenvalues may be found, together with the direct sum of the invariant subspaces $\mathfrak{M}_\lambda$ associated with them. In addition, the number of eigenvalues lying on each side of the rectangle may be determined.

This research was supported, in part, by the Natural Sciences and Engineering Research Council of Canada, Grant A 3028.

REFERENCES

1 Lars V. Ahlfors, *Complex Analysis*, McGraw-Hill, New York, 1979.
2 V. A. Bailey, Prodigious calculations, *Austral. J. Appl. Sci.* 3:78–80 (1941).
3 A. N. Beavers, Jr., and E. D. Denman, A computational method for eigenvalues and eigenvectors of a matrix with real eigenvalues, *Numer. Math.* 21:389–396 (1973).
4 Alex. N. Beavers, Jr., and Eugene D. Denman, A new solution method for quadratic matrix equations, *Math. Biosci.* 20:135–143 (1974).
5 Eugene D. Denman and Alex. N. Beavers, Jr., The matrix sign function and computation in systems, *Appl. Math. Comput.* 2:63–94 (1976).
6 I. Gargantini and P. Henrici, Circular arithmetic and the determination of polynomial zeros, *Numer. Math.* 18:305–320 (1972).
7 A. Halbersberg and Y. Bar-ness, Solution of the discrete regulator problem using the sign matrix function, *Electron. Letters* 14:286–287 (1978).
8 C. Hermite, Sur le nombre des racines d'une équation algébrique comprises entre des limites données. Extrait d'une lettre à M. Borchardt, *J. Reine Angew. Math.* 52:39–51 (1856); translation, P. C. Parks, *Internat. J. Control* 26:183–185 (1977).
9 G. Julia, Mémoire sur l'Iteration des Fonctions Rationelles, *J. Math. Pures Appl. Ser* 8 1:47–245 (1918).
10 J. D. Roberts, (a) Linear model reduction and solution of the algebraic riccati equation by use of the sign function, Cambridge Univ. Engineering Report CUED/B-Control/TR 13, 1971; reprinted, *Internat. J. Control* 32:677–687 (1980). (b) The sign function of a matrix applied to linear model reduction and the matrix Lyapunov and Riccati equations, RCS 90, Dept. of Computer Sci., Univ. of Reading, Jan. 1978.
11 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

Received 20 April 1980; revised 26 May 1982