2

ORNL-3688
UC-32 – Mathematics and Computers
TID-4500 (32nd ed.)

SOME ALGORITHMS FOR THE CALCULATION OF THE CHARACTERISTIC ROOTS AND VECTORS OF A NORMALIZABLE MATRIX

D. H. Clanton

(Thesis)

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

Printed in USA. Price $1.50. Available from the
Office of Technical Services
U. S. Department of Commerce
Washington 25, D. C.

ORNL-3688

Contract No. W-7405-eng-26

Mathematics Division

SOME ALGORITHMS FOR THE CALCULATION OF THE CHARACTERISTIC ROOTS AND VECTORS OF A NORMALIZABLE MATRIX

D. H. Clanton

Submitted as a thesis to the Graduate Faculty of Auburn University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

OCTOBER 1964

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

ABSTRACT

This thesis presents, for a normalizable matrix A of order n, different algorithms for the development of a sequence of normalized vectors $q^{(\ell)}$ ($\ell = 0, 1, 2, \ldots$) such that the centers of their associated Weinstein disks are the Rayleigh quotients, $q^{(\ell)H} A\, q^{(\ell)}$. If the radii of the Weinstein disks approach zero as a limit, then the Rayleigh quotients associated with the sequence of vectors converge to a characteristic root of A.

In Chapter II, a new vector $q^{(\ell+1)}$ is chosen from a subspace spanned by a set of mutually orthogonal vectors that contains the old vector $q^{(\ell)}$. The radius of the Weinstein disk associated with this vector $q^{(\ell+1)}$ will not be larger than the radius of the one associated with $q^{(\ell)}$ and will, in general, be smaller. The condition that the radii of the Weinstein disks converge to zero is presented.

Algorithms are discussed in Chapter II for the subspace of dimension two. In Chapter IV it is shown that the simplest algorithm from the standpoint of operational count does not always yield vectors whose associated Rayleigh quotients converge to a characteristic root of A.

The algorithm for a normal matrix in Chapter V always yields vectors whose associated Rayleigh quotients converge to a characteristic root of A with one exception. If a Weinstein disk has two or more characteristic roots on its boundary, then the algorithm will not yield a new vector. Any algorithm, however, that belongs to the class of algorithms presented in Chapter II will also have the same exception. In the case of a hermitian matrix, a method is presented that will take care of the exception.

ACKNOWLEDGEMENTS

I wish to express my appreciation to Dr. Alston S. Householder, who suggested and directed this investigation, and to Oak Ridge Institute of Nuclear Studies for a Graduate Fellowship.

TABLE OF CONTENTS

I. INTRODUCTION

An inclusion region for a given matrix A of order n is a region of the complex plane that contains at least one characteristic root of A. Inclusion regions will be considered here only for matrices that are normalizable; that is, they are similar to a normal matrix and hence similar to a diagonal matrix.

Bauer and Householder [1] proved that associated with any nonnull vector q there is a family of inclusion disks which contains a minimum disk. If A is hermitian, the minimum disk was obtained by Weinstein [3]. Henceforth, the minimum disk belonging to a vector q will be called a Weinstein disk. The center of a Weinstein disk is the Rayleigh quotient

$$q^H Aq/q^H q$$

where the superscript "H" signifies the transposed conjugate. If q is a characteristic vector of A, then the Rayleigh quotient is the characteristic root of A that is associated with the vector q.

A class of algorithms for developing a sequence of vectors $q^{(\ell)}$ $(\ell = 0, 1, 2, \ldots)$ such that their associated Weinstein disks are decreasing is presented in Chapter II. Each algorithm is a method of choosing a vector $q^{(\ell + 1)}$ from a subspace of dimension $m(m \leq n)$ which is spanned by a set of mutually orthogonal vectors that contains the vector $q^{(\ell)}$. And the radius of the Weinstein disk associated with the vector $q^{(\ell + 1)}$ will not be larger than the radius of the one associated with $q^{(\ell)}$, and will in general be smaller.

In Chapter III, the subspace from which $q^{(\ell + 1)}$ is chosen is assumed to be a plane. An algorithm is developed in Chapter IV which is most desirable from the standpoint of operational count. In this algorithm, however, the radii of the Weinstein disks may not approach zero in the limit, and this method does not always yield a characteristic root and vector of A.

The algorithm in Chapter V always yields a sequence of Weinstein disks whose radii converge to zero with one exception. If a Weinstein disk contains two or more characteristic roots on its boundary, then the algorithm will not yield a new vector. It is proved, however, that any algorithm which belongs to the class of algorithms in Chapter II will also have this same difficulty regardless of the dimension of the subspace.

The gradient method of Hestenes and Karush [2] for finding the characteristic roots and vectors of a real symmetric matrix is somewhat analogous to the methods presented here. They choose a vector $q^{(\ell + 1)}$ from the plane determined by $q^{(\ell)}$ and $Aq^{(\ell)}$ so that the Rayleigh quotient formed by $q^{(\ell + 1)}$ is always smaller (larger) than the previous ones formed by $q^{(k)}$ $(k = 0, 1, 2, \ldots \ell)$. Thus, they form a monotonic decreasing (increasing) sequence of Rayleigh quotients instead of a monotonic decreasing (increasing) sequence of inclusion disks which the methods presented here use. Their methods, also, give convergence only to an extreme characteristic root of a real symmetric matrix, but the method presented in Chapter V may give convergence to any characteristic root of a normal matrix.

The following definitions and theorems will be used in the remainder of the paper. The Euclidean vector norm,

$$\|v\| = (v^H v)^{\frac{1}{2}} ,$$

and the spectral matrix norm,

$$\|A\| = \max_i (\lambda_i)^{\frac{1}{2}}$$

where $\lambda_i (i = 1, 2, \ldots, n)$ is a characteristic root of the matrix $A^H A$, will be the only norms considered. All vectors v will be column vectors.

Let P be any matrix that diagonalizes A; that is, let

$$A = P \Lambda P^{-1}$$

where

$$\Lambda = \mathrm{diag}\ (\alpha_1, \alpha_2, \ldots, \alpha_n) .$$

Define κ by

$$\kappa = \kappa(A) = g\ell b\ \|P\|\ \|P^{-1}\| \tag{1.1}$$

where P diagnolizes A. Then $\kappa \geq 1$ since

$$1 = \|I\| = \|P\ P^{-1}\| \leq \|P\|\ \|P^{-1}\| = \kappa .$$

For normal matrices $\kappa = 1$ since P can be chosen unitary.

Let v_0 be any nonnull vector. Then the Krylov sequence of vectors is defined by

$$v_i = Av_{i-1} = A^i v_0 .$$

The scalar products

$$\mu_{ij} = v_i^H v_j$$

are called moments. Consequently, $\bar{\mu}_{ij} = \mu_{ji}$.

The following definitions from Bauer and Householder [1] will be used. Two polynomials $\alpha(\lambda)$ and $\beta(\lambda)$ are called mutually orthogonal (with respect to A and v_0) if $\alpha(A)\ v_0$ and $\beta(A)\ v_0$ are orthogonal vectors; that is, if

$$(\beta(A)\ v_0)^H\ (\alpha(A)\ v_0) = 0 \ .$$

The norm of $\alpha(\lambda)$ (with respect to A and v_0) is defined to be $\|\alpha(A)\ v_0\|$.

The sequence of monic polynomials

$$\varphi_0(\lambda) = 1 \ , \quad \varphi_1(\lambda) = \lambda - \frac{\mu_{01}}{\mu_{00}}, \ \dots \ ,$$

$$\varphi_\nu\ (\lambda) = \det \begin{pmatrix} \mu_{00} & \mu_{01} & \cdots & \mu_{0\nu} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{\nu-1,0} & \mu_{\nu-1,1} & \cdots & \mu_{\nu-1,\nu} \\ 1 & \lambda & \cdots & \lambda^\nu \end{pmatrix} \Bigg/ \det M_{\nu-1} \qquad (1.2)$$

where

$$M_{\nu-1} = \begin{pmatrix} \mu_{00} & \mu_{01} & \cdots & \mu_{0,\nu-1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{\nu-1,0} & \mu_{\nu-1,1} & \cdots & \mu_{\nu-1,\nu-1} \end{pmatrix}$$

is a sequence of mutually orthogonal polynomials with respect to A and v_0. For the polynomial $\varphi_\nu(\lambda)$ is orthogonal to λ^σ for all $\sigma < \nu$ since

$$v_0^H A^{\sigma^H} \varphi_\nu(A)\ v_0 = \det\begin{pmatrix} \mu_{00} & \mu_{01} & \cdots & \mu_{0,\nu-1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{\nu-1,0} & \mu_{\nu-1,1} & \cdots & \mu_{\nu-1,\nu-1} \\ \mu_{\sigma 0} & \mu_{\sigma,1} & \cdots & \mu_{\sigma,\nu-1} \end{pmatrix} \Bigg/ \det M_{\nu-1} = 0$$

because, for $\sigma < \nu$, the last row of the above determinant is the same as one of the other rows. Therefore, $\varphi_\nu(\lambda)$ is orthogonal to any polynomial of degree less than ν, and the vectors

$$p_\nu = \varphi_\nu(A)\ p_0$$

where $p_0 = v_0$, are mutually orthogonal.

The following two theorems are proved in [1].

<u>Theorem 1.1</u>. Let the matrix A be normalizable, and let $\alpha(\lambda)$ and $\beta(\lambda)$ be any two polynomials. Then

$$\left\{ \lambda : \left| \frac{\alpha(\lambda)}{\beta(\lambda)} \right| \le \kappa \frac{\|\alpha(A)\ v\|}{\|\beta(A)\ v\|} \right\} \tag{1.3}$$

is an inclusion region, where κ is defined by (1.1) and where v is any nonnull vector.

<u>Theorem 1.2</u>. Among all polynomials of degree ν and with leading coefficient unity, the polynomial $\varphi_\nu\ (\lambda)$ has minimal norm, and hence minimizes the right member of (1.3) for a given $\beta(\lambda)$.

II. ALGORITHMS

Let A be a normalizable matrix and let q be a normalized vector; that is, let

$$q = \frac{\nu}{\|\nu\|}$$

so that $\|q\| = 1$.

In the rest of the paper let $\beta(\lambda) = 1$ and let

$$\alpha(\lambda) = \varphi_1(\lambda) = \lambda - \mu_{01}$$

where $\varphi_1(\lambda)$ is the linear polynomial in the orthogonal set determined by A and q. The inclusion region (1.3) thus becomes

$$\left\{\lambda : |\lambda - \mu_{01}| \le \kappa \|(A - \mu_{01} I) q\| \right\} \tag{2.1}$$

which is an inclusion disk with center μ_{01} and radius

$$\rho = \kappa \|(A - \mu_{01} I) q\|.$$

From Theorem 1.2, it follows that (2.1) is the smallest inclusion region for the vector q and for a linear polynomial, or, in other words, (2.1) is the Weinstein disk belonging to q.

Since Theorem 1.1 states that (2.1) is an inclusion region for any nonnull vector q there may exist a normalized vector $q^{(1)}$ which will yield a radius

$$\gamma = \|(A - \mu_{01} I) q^{(1)}\|$$

such that

$$\gamma < \rho . \tag{2.2}$$

If a $q^{(1)}$ satisfying (2.2) is found, the inclusion disk

$$\left\{\lambda : |\lambda - \mu_{01}| \le \kappa \|(A - \mu_{01} I) q^{(1)}\| \right\} \tag{2.3}$$

will be concentric with the disk (2.1) and will be contained in it.

By Theorem 1.2, the inclusion disk

$$\left\{\lambda : |\lambda - \mu_{01}{}^{(1)}| \le \|(A - \mu_{01}{}^{(1)} I) q^{(1)}\| \right\} \tag{2.4}$$

where

$$\mu_{01}{}^{(1)} = q^{(1)^H} A q^{(1)}$$

will be smaller than the disk (2.3). The disk (2.4) is not in general concentric with the disks (2.1) and (2.3). A later theorem, however, will show that the center of (2.4) will be interior to the disks (2.1) and (2.3).

More generally, let

$$\varphi_1{}^{(\ell)}(\lambda) = \lambda - \mu_{01}{}^{(\ell)} \tag{2.5}$$

be the linear polynomial in the orthogonal set determined by the matrix A and a normalized vector $q^{(\ell)}$. That is $\mu_{01}{}^{(\ell)}$ satisfies the following relation:

$$\mu_{01}{}^{(\ell)} = q^{(\ell)^H} A q^{(\ell)}$$

with $q^{(\ell)^H} q^{(\ell)} = 1$. Henceforth, the superscript "ℓ" on a symbol will mean the symbol is defined in terms of the vector $q^{(\ell)}$.

Obviously, a decreasing sequence of inclusion disks could be attained if a sequence of vectors $q^{(\ell)}$ could be found which yield a decreasing sequence of norms. That is, if there exists a sequence $q^{(\ell)}$ of vectors such that for all ℓ

$$\|(A - \mu_{01}{}^{(\ell+1)} I) q^{(\ell+1)}\| \le \|(A - \mu_{01}{}^{(\ell)} I) q^{(\ell)}\|, \tag{2.6}$$

then the sequence of inclusion disks

$$\left\{\lambda: |\lambda-\mu_{01}{}^{(\ell)}| \leq \kappa \| (A-\mu_{01}{}^{(\ell)}I)q^{(\ell)} \| \right\} \qquad (\ell=0,1,2,\ldots)$$

is decreasing. The following theorem gives a method of determining such a sequence.

Theorem 2.1. Let A be a normalizable matrix of order n, and let q be a normalized vector. If W_0 is an n by m $(n \geq m)$ matrix of orthonormal columns with q as the first column and if

$$q^{(1)} = W_0 w_0 \tag{2.7}$$

where w_0 is the normalized characteristic vector associated with the smallest characteristic root of the positive semidefinite matrix

$$Q_0 = (\varphi_1(A)W_0)^H(\varphi_1(A)W_0), \tag{2.8}$$

then

$$\|\varphi_1{}^{(1)}(A)q^{(1)}\| \leq \|\varphi_1(A)q\| \ .$$

If the smallest characteristic root of Q_0 is a simple root, then equality will hold only if $q^{(1)} = q$.

Proof: Let u be any normalized vector with m elements. Then as u varies the minimum of

$$\|\varphi_1(A)W_0u\| \tag{2.9}$$

cannot be greater than

$$\|\varphi_1(A)q\|$$

since u can be chosen so that

$$q = W_0u. \tag{2.10}$$

Now

$$\|\varphi_1 (A) W_0 u\| = \left[u^H(\varphi_1 (A) W_0)^H (\varphi_1 (A) W_0) u \right]^{\frac{1}{2}}, \qquad (2.11)$$

which is the square root of the Rayleigh quotient formed by the vector u and the hermitian matrix Q_0. Hence, a minimum of (2.11) will be the smallest characteristic root of the matrix Q_0 since the Rayleigh quotient of a hermitian matrix lies on the closed segment between the least and greatest characteristic roots. The minimum will be obtained when

$$u = w_0$$

where w_0 is the normalized, characteristic vector belonging to the smallest characteristic root of Q_0. Consequently,

$$\|\varphi_1 (A) q^{(1)}\| \leq \|\varphi_1 (A) q\| \qquad (2.12)$$

where $q^{(1)}$ is the vector (2.7). But

$$\|\varphi_1^{(1)} (A) q^{(1)}\| \leq \|\varphi_1 (A) q^{(1)}\|$$

by Theorem 1.2. Hence,

$$\|\varphi_1^{(1)} (A) q^{(1)}\| \leq \|\varphi_1 (A) q\| . \qquad (2.13)$$

It follows from (2.10) that equality in (2.12) implies $q^{(1)} = q$ since w_0 is unique if the smallest characteristic root of the matrix Q_0 is a simple root. _A fortiori_, equality in (2.13) implies $q^{(1)} = q$.

Now if

$$q^{(1)} = W_0 w_0 \neq q$$

then, by the use of $q^{(1)}$ in place of q, the method described in the above theorem yields a normalized vector

$$q^{(2)} = W_1 \, w_1$$

with the property

$$\|\varphi_1^{(2)}(A)\, q^{(2)}\| \le \|\varphi_1^{(1)}(A)\, q^{(1)}\| \, .$$

Then $q^{(2)}$ can be used in place of $q^{(1)}$ to obtain a vector $q^{(3)}$. Thus, the method of Theorem 2.1 can be repeated to yield a sequence of vectors whose associated Weinstein disks are monotonic decreasing. If for all ℓ, $q^{(\ell)} \ne q^{(\ell+1)}$, then the radii of the Weinstein disks are strictly monotonic decreasing.

For each choice of the number of columns of W_ℓ and for each method of determining the orthonormal columns of the matrix W_ℓ the above procedure is an algorithm which generates a sequence of normalized vectors

$$q^{(\ell)} = W_{\ell-1} \, w_{\ell-1} \qquad (\ell = 0, 1, 2, \ldots) \tag{2.14}$$

with the initial vector

$$q = q^{(0)} \, .$$

The vectors (2.14) will also satisfy condition (2.6); that is,

$$\|\varphi_1^{(\ell+1)}(A)\, q^{(\ell+1)}\| \le \|\varphi_1^{(\ell)}(A)\, q^{(\ell)}\|$$

for each ℓ. In the remainder of the paper the discussion will be restricted to those algorithms that have the property that the smallest characteristic root of Q_ℓ is a simple root. Equality in (2.6) will thus hold only if $q^{(\ell+1)} = q^{(\ell)}$. A vector $q^{(\ell)}$ which, when used in the algorithm, does not yield a new vector will be called a "stationary vector."

Theorem 2.2. Let A be a normalizable matrix of order n, and let q be a given normalized vector. For the vectors (2.14) and the polynomials (2.5) form the inclusion regions

$$R_\ell = \left\{ \lambda \; : \; |\varphi_1^{(\ell)}(\lambda)| \le \kappa \|\varphi_1^{(\ell)}(A)\, q^{(\ell)}\| \right\} . \tag{2.15}$$

Then (1) as ℓ increases ($\ell = 0, 1, 2, \ldots$), the sequence R_ℓ will decrease as long as

$$q^{(\ell)} \neq q^{(\ell+1)} \; ; \tag{2.16}$$

(2) if

$$\lim_{\ell \to \infty} \|\varphi_1^{(\ell)}(A)\, q^{(\ell)}\| = 0 \; , \tag{2.17}$$

then

$$\lim_{\ell \to \infty} \mu_{01}^{(\ell)} = \alpha_j \tag{2.18}$$

where α_j is a characteristic root of A.

Proof: The proof of (1) follows immediately since the algorithm which gives the vector (2.14) insures that the radius of (2.15) will decrease as long as (2.16) holds.

Since (2.17) implies

$$\lim_{\ell \to \infty} |\varphi_1^{(\ell)}(\lambda)| = \lim_{\ell \to \infty} |\lambda - \mu_{01}^{(\ell)}| = 0$$

it follows that condition (2.18) holds when condition (2.17) is true.

If

$$\|(A - \mu_{01}^{(\ell)} I)\, q^{(\ell)}\| = 0 \,, \tag{2.19}$$

then $q^{(\ell)}$ is a characteristic vector of A since the norm of a vector is zero if and only if the vector is null and since (2.19) implies that

$$(A - \mu_{01}^{(\ell)} I)\, q^{(\ell)} = 0 \,. \tag{2.20}$$

Thus $q^{(\ell)}$ is the characteristic vector of A associated with the characteristic root $\mu_{01}^{(\ell)}$.

Consequently, if the initial vector or a computed vector $q^{(\ell)}$ is a characteristic vector of A, then the algorithm will not yield a new vector since (2.19) would hold. In summary, characteristic vectors of A are stationary vectors.

Theorem 2.3. If $q^{(\ell)}$ is a stationary vector, that is, if

$$q^{(\ell)} = q^{(\ell+1)} = W_{\ell-1}\, w_{\ell-1} \,, \tag{2.21}$$

then the matrix

$$Q^{(\ell)} = \left(\varphi_1^{(\ell)} (A)\, W_{\ell-1}\right)^H \left(\varphi_1^{(\ell)} (A)\, W_{\ell-1}\right) \tag{2.22}$$

is null in the first row and first column, except in the diagonal position. By definition the matrix $W_{\ell-1}$ contains $q^{(\ell)}$ as its first column.

Proof: Since the column vectors of $W_{\ell-1}$ are mutually orthogonal, (2.21) implies that

$$w_{\ell-1} = e_1$$

where e_1 is the first column vector of the identity matrix. Thus e_1 is a characteristic vector of $Q^{(\ell)}$. In other words, the matrix $Q^{(\ell)}$ will be null in the first column except in the diagonal position. The off-diagonal elements of the first row will also be zero since the matrix $Q^{(\ell)}$ is hermitian.

Theorem 2.4. If the algorithm is such that the sequence of matrices W_ℓ ($\ell = 0, 1, 2, \ldots$) has a limit W; that is if

$$\lim_{\ell \to \infty} W_\ell = W , \tag{2.23}$$

then q, the first column of W, is a stationary vector of the algorithm and

$$q = \lim_{\ell \to \infty} q^{(\ell)} . \tag{2.24}$$

Proof: (2.24) is true since each $q^{(\ell)}$ is the first column of W_ℓ. The result in (2.24) implies that the sequence of matrices

$$Q^{(\ell)} = \left(\varphi_1^{(\ell)}(A)\, W_{\ell-1}\right)^H \left(\varphi_1^{(\ell)}(A)\, W_{\ell-1}\right) \qquad (\ell = 1, 2, \ldots)$$

converges to a matrix Q. Thus, there exists a vector w such that

$$\lim_{\ell \to \infty} w_\ell = w \tag{2.25}$$

where w_ℓ is the normalized characteristic vector associated with the smallest characteristic root of $Q^{(\ell)}$. Consequently, by (2.23), (2.24), and (2.25),

$$W\,w = \lim_{\ell \to \infty} W_\ell \lim_{\ell \to \infty} w_\ell = \lim_{\ell \to \infty} (W_\ell\, w_\ell) = \lim_{\ell \to \infty} q^{(\ell)} = q\ .$$

Therefore, q is a stationary vector of the algorithm.

<u>Theorem 2.5</u>. Let A be a normalizable matrix of order n. If the algorithm is such that condition (2.23) holds and all stationary vectors are characteristic vectors of A, then for any initial normalized vector $q^{(0)}$ the algorithm will yield vectors $q^{(\ell)}$ of the form (2.14) whose Rayleigh quotients $\mu_{01}^{(\ell)}$ satisfy

$$\lim_{\ell \to \infty} \mu_{01}^{(\ell)} = \alpha_j \tag{2.26}$$

where α_j is a characteristic root of A.

<u>Proof</u>: Since condition (2.23) holds, from the previous theorem the limiting vector q of the sequence $q^{(\ell)}$ ($\ell = 0, 1, \ldots$) is stationary. By hypothesis q is a characteristic vector of A. Hence, from (2.19) and (2.20) it follows that

$$\lim_{\ell \to \infty} \|\varphi_1^{(\ell)}(A)\, q^{(\ell)}\| = \|\varphi_1(A)\, q\| = 0\ .$$

Then (2.26) is true because of Theorem 2.2.

The importance of Theorem 2.3 is now clear since it gives a criterion for a vector to be stationary.

Examples exist which demonstrate that the Rayleigh quotient $\mu_{01}^{(\ell)}$ does not necessarily have to be in the inclusion region $R_{\ell-2}$. In fact, $\mu_{01}^{(\ell)}$ may converge to a characteristic root that is not in the first inclusion region. However, the following theorem gives a relation between $\mu_{01}^{(\ell)}$ and $R_{\ell-1}$.

Theorem 2.6. The Rayleigh quotient $\mu_{01}^{(\ell)}$ is interior to the disk

$$R_{\ell-1} = \left\{ \lambda : |\lambda - \mu_{01}^{(\ell-1)}| \le \kappa \|(A - \mu_{01}^{(\ell-1)} I)\, q^{(\ell-1)}\| \right\}.$$

Proof: Consider

$$\begin{aligned} |\mu_{01}^{(\ell)} - \mu_{01}^{(\ell-1)}| &= \left[\left(\mu_{01}^{(\ell)} - \mu_{01}^{(\ell-1)}\right)\left(\mu_{10}^{(\ell)} - \mu_{10}^{(\ell-1)}\right)\right]^{\frac{1}{2}} \\ &= \left[\mu_{01}^{(\ell)}\mu_{10}^{(\ell)} - \mu_{01}^{(\ell)}\mu_{10}^{(\ell-1)} - \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell)} + \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell-1)}\right]^{\frac{1}{2}} . \end{aligned} \tag{2.27}$$

Now

$$\|(A - \mu_{01}^{(\ell-1)} I) q^{(\ell)}\| = [\mu_{11}^{(\ell)} - \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell)} - \mu_{10}^{(\ell-1)}\mu_{01}^{(\ell)} + \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell-1)}]^{\frac{1}{2}}. \tag{2.28}$$

From the Cauchy inequality it follows that

$$\mu_{11}^{(\ell)} \ge \mu_{01}^{(\ell)}\, \mu_{10}^{(\ell)} . \tag{2.29}$$

If in (2.27) $\mu_{01}^{(\ell)}\, \mu_{10}^{(\ell)}$ is replaced by $\mu_{11}^{(\ell)}$, then

$$|\mu_{01}^{(\ell)} - \mu_{01}^{(\ell-1)}| \le \left[\mu_{11}^{(\ell)} - \mu_{01}^{(\ell)}\mu_{10}^{(\ell-1)} - \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell)} + \mu_{01}^{(\ell-1)}\mu_{10}^{(\ell-1)}\right]^{\frac{1}{2}}.$$

Thus, by (2.28)

$$|\mu_{01}^{(\ell)} - \mu_{01}^{(\ell-1)}| \le \|(A - \mu_{01}^{(\ell-1)} I)\, q^{(\ell)}\| . \tag{2.30}$$

And by Theorem 2.1,

$$\|(A - \mu_{01}^{(\ell-1)} I)\, q^{(\ell)}\| \le \|(A - \mu_{01}^{(\ell-1)} I)\, q^{(\ell-1)}\| . \tag{2.31}$$

If the right side, therefore, of the inequality (2.30) is replaced by the right side of the inequality (2.31) the result is

$$|\mu_{01}^{(\ell)} - \mu_{01}^{(\ell-1)}| \le \kappa \| (A - \mu_{01}^{(\ell-1)} I) q^{(\ell-1)} \|$$

since $\kappa \ge 1$. Thus, $\mu_{01}^{(\ell)}$ is interior to the disk $R_{\ell-1}$.

III. THE MATRIX W_ℓ WITH TWO COLUMNS

Since the method of finding $R_{\ell+1}$ from $q^{(\ell)} = W_{\ell-1} w_{\ell-1}$ involves finding the smallest characteristic root of

$$Q^{(\ell)} = (\varphi_1^{(\ell)}(A)\ W_\ell)^H\ (\varphi_1^{(\ell)}(A)\ W_\ell) \tag{3.1}$$

from which $w_{\ell-1}$ is determined, the simplest case is that in which $Q^{(\ell)}$ is of order two. Since W_ℓ determines the order of the matrix $Q^{(\ell)}$, let W_ℓ consist of two columns; that is, let

$$W_\ell = \left(q^{(\ell)},\ \ \frac{\pi^{(\ell)}(A)\ q^{(\ell)}}{\rho^{(\ell)}} \right) \tag{3.2}$$

where $\pi^{(\ell)}(\lambda)$ is a polynomial such that

$$q^{(\ell)^H}\ \pi^{(\ell)}(A)\ q^{(\ell)} = 0 \tag{3.3}$$

and where

$$\rho^{(\ell)} = \|\pi^{(\ell)}(A)\ q^{(\ell)}\| \ . \tag{3.4}$$

If the polynomial $\pi^{(\ell)}(\lambda)$ is of degree ν, it can be expressed as a linear combination of the mutually orthogonal (with respect to A and $q^{(\ell)}$) polynomials $\varphi_k^{(\ell)}(\lambda)$ ($k = 1, 2, \ldots, \nu$). Condition (3.3) rules out the appearance of the constant polynomial $\varphi_0^{(\ell)}(\lambda) = 1$ in this expression. Obviously, the most natural choice from a computational standpoint for $\pi^{(\ell)}(\lambda)$ would be the linear polynomial

$$\varphi_1^{(\ell)}(\lambda) = \lambda - \mu_{01}^{(\ell)} \ .$$

Before a discussion of the consequence of this choice, however, theorems that are valid for an arbitray polynomial $\pi^{(\ell)}(\lambda)$ which satisfies condition (3.3) will be presented.

First, it will be desirable to express the elements of the matrix $Q^{(\ell)}$ in terms of the elements of the matrix (3.2) and then to express the elements of the normalized vector w_ℓ in terms of the elements of the matrix $Q^{(\ell)}$ and its smallest characteristic root.

For each ℓ, the set

$$p_\nu^{(\ell)} = \varphi_\nu^{(\ell)}(A)\ q^{(\ell)} \qquad (\nu = 0, 1, 2, \ldots, n - 1)$$

is a set of mutually orthogonal vectors. Hence,

$$\varphi_1^{(\ell)}(A)\ W_\ell = \left(p_1{}^{(\ell)},\ \frac{\pi^{(\ell)}(A)\ p_1{}^{(\ell)}}{\rho^{(\ell)}}\right) \tag{3.5}$$

since matrix polynomials in A are commutative.

If (3.5) is substituted into (3.1), the result is

$$Q^{(\ell)} = \begin{pmatrix} p_1^{(\ell)^H}\ p_1^{(\ell)} & \dfrac{p_1^{(\ell)^H}\ \pi^{(\ell)}(A)\ p_1^{(\ell)}}{\rho^{(\ell)}} \\ \dfrac{p_1^{(\ell)^H}\pi^{(\ell)^H}(A)\ p_1^{(\ell)}}{\rho^{(\ell)}} & \dfrac{p_1^{(\ell)^H}\pi^{(\ell)^H}(A)\ \pi^{(\ell)}(A)\ p_1^{(\ell)}}{\rho^{(\ell)^2}} \end{pmatrix} = \begin{pmatrix} \rho_1^{(\ell)^2} & \sigma_{01}^{(\ell)} \\ \sigma_{10}^{(\ell)} & \sigma_{11}^{(\ell)} \end{pmatrix} \tag{3.6}$$

where

$$\sigma_{ij} = \left(\pi^{(\ell)^i}(A)\ p_1^{(\ell)}\right)^H \left(\pi^{(\ell)^j}(A)\ p_1^{(\ell)}\right) \Big/ \rho^{(\ell)^{i+j}} \tag{3.7}$$

and

$$\rho_1^{(\ell)^2} = p_1^{(\ell)^H} \, p_1^{(\ell)} = \|p_1^{(\ell)}\|^2 \,.$$

Thus $\rho_1^{(\ell)}$ is the radius of the inclusion region R_ℓ, since

$$\rho_1^{(\ell)} = \|p_1{}^{(\ell)}\| = \|\varphi_1^{(\ell)}(A)\; q^{(\ell)}\| \,. \tag{3.8}$$

The smallest characteristic root of $\lambda_1^{(\ell)}$ of $Q^{(\ell)}$ is

$$\lambda_1^{(\ell)} = \frac{\rho_1^{(\ell)^2} + \sigma_{11}^{(\ell)} - \sqrt{(\sigma_{11}^{(\ell)} - \rho_1^{(\ell)^2})^2 + 4\,\sigma_{10}^{(\ell)}\,\sigma_{01}^{(\ell)}}}{2} \tag{3.9}$$

and the normalized characteristic vector w_ℓ of the matrix $Q^{(\ell)}$ corresponding to $\lambda_1^{(\ell)}$ is

$$w_\ell = \begin{pmatrix} \dfrac{\sigma_{01}^{(\ell)}}{\rho_2^{(\ell)}} \\[2ex] - \dfrac{(\rho_1^{(\ell)^2} - \lambda_1^{(\ell)})}{\rho_2{}^{(\ell)}} \end{pmatrix} \tag{3.10}$$

where

$$\rho_2^{(\ell)} = \left[\sigma_{01}^{(\ell)}\;\sigma_{10}^{(\ell)} + \left(\rho_1^{(\ell)^2} - \lambda_1^{(\ell)}\right)^2\right]^{\frac{1}{2}} .$$

Now if the hypothesis in (2) of Theorem 2.2 is valid, then $\lim_{\ell \to \infty} \mu_{01}^{(\ell)} = \alpha_j$ where α_j is a characteristic root of A. In order to determine the characteristic root to which the sequence in (2.15) converges, the behavior of $\mu_{01}^{(\ell)}$ must be investigated. To this end, the remaining theorems in this section are presented.

Let P be any matrix that diagonalizes A; that is, assume that

$$A = P \Lambda P^{-1} \tag{3.11}$$

where

$$\Lambda = \operatorname{diag} (\alpha_1, \alpha_2, \ldots, \alpha_n) .$$

Since P is a nonsingular matrix, there exists a unique vector $v^{(0)}$ such that

$$q^{(0)} = Pv^{(0)} \tag{3.12}$$

where $q^{(0)}$ is the vector (2.14). If A is a normal matrix, then P can be chosen unitary and thus $v^{(0)}$ will be normalized.

The following theorem will be used to express the Rayleigh quotient $\mu_{01}^{(\ell)}$ as a linear combination of the characteristic roots of A and thus to compare the coefficients associated with α_j in $\mu_{01}^{(\ell)}$ with those in $\mu_{01}^{(\ell + 1)}$. Specifically, if the coefficients for a particular α_j increase with ℓ while the coefficients of other α decrease, then $\mu_{01}^{(\ell)}$ converges to that α_j.

<u>Theorem 3.1</u>. Let P be a matrix that satisfies condition (3.11), and let $v^{(0)}$ satisfy condition (3.12) where $q^{(\ell)}$ is a normalized vector. Then the vectors (2.14) can be expressed as

$$q^{(\ell)} = P\, D^{(0)}\, D^{(1)} \ldots D^{(\ell)}\, v^{(0)} \qquad (D^{(0)} = I)$$

where

$$D^{(k)} = (d_1^{(k)}, d_2^{(k)}, \ldots, d_n^{(k)}) \qquad (k = 1, 2, \ldots, \ell)$$

with

$$d_j^{(k)} = \frac{\rho^{(k-1)}\sigma_{01}^{(k-1)} - \left(\rho_1^{2\,(k-1)} - \lambda_1^{(k-1)}\right)\pi^{(k-1)}\left(\alpha_j\right)}{\rho^{(k-1)}\left[\sigma_{01}^{(k-1)}\sigma_{10}^{(k-1)} + \left(\rho_1^{2\,(k-1)} - \lambda_1^{(k-1)}\right)^2\right]^{\frac{1}{2}}} \,. \qquad (3.13)$$

<u>Proof</u>: From (3.2) and (3.12)

$$\begin{aligned} q^{(1)} &= W_0\, w_0 \\ &= \left(q^{(0)}, \frac{\pi^{(0)}(A)\, q^{(0)}}{\rho(0)}\right) w_0 \qquad (3.14) \\ &= \left(Pv^{(0)}, \frac{\pi^{(0)}(A)\, Pv^{(0)}}{\rho(0)}\right) w_0 \,. \end{aligned}$$

Since condition (3.11) implies that

$$\pi^{(\ell)}(A)\, P = P\, \pi^{(\ell)}(\Lambda) \;,$$

it follows that

$$\begin{aligned} q^{(1)} &= \left(Pv^{(0)}, \frac{P\,\pi^{(0)}(\Lambda)\, v^{(0)}}{\rho(0)}\right) w_0 \\ &= P\left(v^{(0)}, \frac{\pi^{(0)}(\Lambda)\, v^{(0)}}{\rho(0)}\right) w_0 \,. \end{aligned} \qquad (3.15)$$

Then by (3.10)

$$q^{(1)} = P\left(\frac{\sigma_{01}^{(0)}\,I}{\rho_2^{(0)}} - \frac{1}{\rho_2^{(0)}}\left(\rho_1^{\overset{2}{(0)}} - \lambda_1^{(0)}\right)\frac{\pi^{(0)}\,(\Lambda)}{\rho^{(0)}}\right)v^{(0)} \tag{3.16}$$

$$= P\;D^{(1)}\;v^{(0)}\;,$$

where

$$D^{(1)} = \mathrm{diag}\;(d_1^{(1)},\; d_2^{(1)},\; \ldots,\; d_n^{(1)}) \tag{3.17}$$

with

$$d_j^{(1)} = \frac{\rho^{(0)}\;\sigma_{01}^{(0)} - \left(\rho_1^{\overset{2}{(0)}} - \lambda_1^{(0)}\right)\pi^{(0)}\left(\alpha_j\right)}{\rho^{(0)}\;\rho_2^{(0)}}$$

$$= \frac{\rho^{(0)}\;\sigma_{01}^{(0)} - \left(\rho_1^{\overset{2}{(0)}} - \lambda_1^{(0)}\right)\pi^{(0)}\left(\alpha_j\right)}{\rho^{(0)}\left[\sigma_{01}^{(0)}\;\sigma_{10}^{(0)} + \left(\rho_1^{\overset{2}{(0)}} - \lambda_1^{(0)}\right)^2\right]^{\frac{1}{2}}}\;.$$

Let

$$v^{(\ell)} = D^{(\ell)}\;v^{(\ell-1)} \qquad (\ell = 1,\, 2,\, \ldots)\;.$$

The results stated in the theorem follow by induction since

$$q^{(\ell)} = P\;D^{(\ell)}\;v^{(\ell-1)}$$

$$= P\;D^{(\ell)}\left[D^{(1)}\;\ldots\;D^{(\ell-1)}\;v^{(0)}\right]$$

$$= P\;D^{(0)}\;\ldots\;D^{(\ell)}\;v^{(0)} \qquad (\ell = 1,\, 2,\, \ldots)$$

because diagonal matrices are commutative and by the hypothesis that

$$q^{(0)} = P\, D^{(0)} .$$

The next theorem is a generalization of the well-known theorem that a Rayleigh quotient found for a normal matrix is a weighted mean of the roots of the matrix.

Theorem 3.2. The Rayleigh quotient $\mu_{01}^{(\ell)}$ formed with a normalizable matrix is a linear combination of the characteristic roots α_j of A, and the coefficients of the roots may be complex. But their sum is one.

Proof: The proof is a direct consequence of the fact that any Rayleigh quotient for a normalizable matrix is

$$v^H\, P\, \Lambda\, P^{-1}\, v/v^H\, v .$$

The sum of coefficients of the α_j's can be obtained by the replacement of Λ by I. This sum is one since

$$v^H\, P\, I\, P^{-1}\, v/v^H\, v = v^H\, v/v^H\, v = 1 .$$

In case A is normal, P can be chosen unitary with

$$v^H\, P = (P^{-1}\, v)^H .$$

For normal matrices the coefficients are thus nonnegative. If A is normal and if

$$v^{(0)} = (\beta_1, \ldots, \beta_n)^T ,$$

then from the previous theorem

$$\mu_{01}^{(\ell)} = \sum_{1}^{n} \alpha_j \; |d_j^{(1)} \, d_j^{(2)} \, \ldots \, d_j^{(\ell)}|^2 \; |\beta_j|^2 \tag{3.18}$$

so that

$$\sum_{1}^{n} |d_j^{(1)} \, d_j^{(2)} \, \ldots \, d_j^{(\ell)}|^2 \; |\beta_j|^2 = 1 \; .$$

As previously indicated, the purpose of the above theorem is to express the coefficients of α_j in terms of $d_j^{(\ell)}$ so that a comparison can be made of the coefficients associated with α_j in $\mu_{01}^{(\ell)}$ and $\mu_{01}^{(\ell+1)}$. Therefore, let

$$P^H \, P = (p_{ij}) \tag{3.19}$$

In consequence,

$$\begin{aligned}\mu_{01}^{(\ell)} &= q^{(\ell)^H} \, A \, q^{(\ell)} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_j \left(d_j^{(1)} \, \ldots \, d_j^{(\ell)} \, \beta_j\right)\overline{\left(d_i^{(1)} \, \ldots \, d_i^{(\ell)} \, \beta_i\right)} p_{ij}\end{aligned}$$

and

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \left(d_j^{(1)} \, d_j^{(2)} \, \ldots \, d_j^{(\ell)} \, \beta_j\right) \overline{\left(d_i^{(1)} \, d_i^{(2)} \, \ldots \, d_i^{(\ell)} \, \beta_i\right)} p_{ij} = 1 \; .$$

If $|d_h^{(k)}| > 1$ and $|d_j^{(k)}| < 1$ $(j = 1, 2, \ldots, h - 1, h + 1, \ldots n$ and $k = 1, 2, \ldots)$, then

$$\lim_{\ell \to \infty} \mu_{01}^{(\ell)} = \alpha_h \; .$$

For a normal matrix A,

$$p_{ij} = \delta_{ij} \ , \quad \text{the Kronecker delta} \ .$$

Obviously, $\mu_{01}^{(\ell)}$ is a weighted mean of the characteristic roots of A.

The next theorem is applicable to any normalizable matrix. Its use, however, is more easily seen if A is a normal matrix since the condition (3.18) implies that if $|d_j^{(\ell)}|^2 > 1$ ($|d_j^{(\ell)}|^2 < 1$) the coefficient of α_j is increased (decreased) with ℓ. In any case, it is advantageous to know when $|d_j^{(\ell)}|^2 \geq 1$ and when $|d_j^{(\ell)}|^2 < 1$.

<u>Theorem 3.3</u>. Let $d_j^{(\ell)}$ ($j = 1, 2, \ldots, n$) be defined as in (3.13), and let

$$f^{(\ell)}(\lambda) = |\pi^{(\ell)}(\lambda)|^2 - \frac{2\rho^{(\ell)}}{{\rho_1^{(\ell)}}^2 - \lambda_1^{(\ell)}} \operatorname{Re}\left[\sigma_{10}^{(\ell)} \pi^{(\ell)}(\lambda)\right] - {\rho^{(\ell)}}^2 \ . \qquad (3.20)$$

Then for each j, one and only one of the following conditions holds:

$$\begin{aligned} &(1) \quad f^{(\ell)}(\alpha_j) > 0 \text{ and } |d_j^{(\ell)}| > 1 \ , \\ &(2) \quad f^{(\ell)}(\alpha_j) < 0 \text{ and } |d_j^{(\ell)}| < 1 \ , \\ &(3) \quad f^{(\ell)}(\alpha_j) = 0 \text{ and } |d_j^{(\ell)}| = 1 \ . \end{aligned} \qquad (3.21)$$

<u>Proof</u>: If (3.13) is substituted into

$$|d_j^{(\ell)}|^2 = 1 \ ,$$

the result is

$$\frac{\left[\rho^{(\ell)}\sigma_{10}^{(\ell)} - \left(\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right)\overline{\pi^{(\ell)}(\alpha_j)}\right]\left[\rho^{(\ell)}\sigma_{01}^{(\ell)} - \left(\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right)\pi^{(\ell)}(\alpha_j)\right]}{\rho^{\overset{2}{(\ell)}}\left[\sigma_{01}^{(\ell)}\,\sigma_{10}^{(\ell)} + \left(\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right)^2\right]} = 1.$$

Hence,

$$\begin{aligned}&\rho^{\overset{2}{(\ell)}}\,\sigma_{10}^{(\ell)}\,\sigma_{01}^{(\ell)} - \rho^{(\ell)}\left[\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right]\left[\sigma_{10}^{(\ell)}\,\pi^{(\ell)}(\alpha_j) + \sigma_{01}^{(\ell)}\,\overline{\pi^{(\ell)}(\alpha_j)}\right] \\ &+ \left[\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right]^2 |\pi^{(\ell)}(\alpha_j)|^2 = \rho^{\overset{2}{(\ell)}}\,\sigma_{01}^{(\ell)}\,\sigma_{10}^{(\ell)} + \rho^{(\ell)}\left(\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}\right)^2.\end{aligned} \tag{3.22}$$

Because the smallest root of a hermitian matrix cannot exceed a diagonal element,

$$\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)} \geq 0\ ; \tag{3.23}$$

and by (3.9), equality holds only if

$$\sigma_{01}^{(\ell)}\,\sigma_{10}^{(\ell)} = 0\ . \tag{3.24}$$

Equality in (3.23) and condition (3.24) imply that $d_j^{(\ell)}$ is not defined; also, equality in (3.23) implies that $f^{(\ell)}(\lambda)$ is not defined. Therefore,

$$\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)} > 0$$

may be assumed.

If (3.22) is simplified, it becomes

$$|\pi^{(\ell)}(\alpha_j)|^2 - \frac{2\,\rho^{(\ell)}\,\mathrm{Re}\left[\sigma_{10}^{(\ell)}\,\pi^{(\ell)}(\alpha_j)\right]}{\rho_1^{\overset{2}{(\ell)}} - \lambda_1^{(\ell)}} - \rho^{\overset{2}{(\ell)}} = 0\ . \tag{3.25}$$

The left side of (3.25) is $f^{(\ell)}(\alpha_j)$. Hence, $|d_j^{(\ell)}|^2 = 1$ implies that $f^{(\ell)}(\lambda) = 0$. Since the steps are reversible, (3) of (3.21) is valid.

The proofs of the other two parts of (3.21) follow immediately since all the above operations either involve multiplying or dividing by a positive number or involve some other operation which does not change the sense of an inequality.

In [1], it is proved that for a normal matrix A and for a nonzero vector v, a polynomial $X(A^H,A)$ which satisfies

$$v^H X(A^H,A)\, v = 0$$

is a separation polynomial. Its locus will separate the complex plane into two inclusion regions for A with the locus common to both.

If A is a normal matrix, the condition (3.3) imposed on the polynomial $\pi^{(\ell)}(\lambda)$ which has been used throughout this discussion implies that $\pi^{(\ell)}(\lambda)$ is a separation polynomial. The next theorem shows that the polynomial (3.20) is a separation polynomial.

<u>Theorem 3.4</u>. If A is a normal matrix, the polynomial (3.20) is a separation polynomial with respect to A and the normalized vector $q^{(\ell)}$.

<u>Proof</u>: The proof consists of showing that

$$q^{(\ell)H} f^{(\ell)}(A) q^{(\ell)} = q^{(\ell)H}\left\{\pi^{(\ell)H}(A)\pi^{(\ell)}(A) - \frac{\rho^{(\ell)2}}{\rho_1^{(\ell)}-\lambda_1^{(\ell)}}\left[\sigma_{10}^{(\ell)}\pi^{(\ell)}(A)+\sigma_{01}^{(\ell)}\pi^{(\ell)}(A)\right] - \rho^{(\ell)2} I\right\} q^{(\ell)} = 0 \,. \tag{3.26}$$

Since condition (3.3) implies that $\pi^{(\ell)}(A)$ is a separation polynomial,

$$q^{(\ell)^H} \pi^{(\ell)^H} (A) \, q^{(\ell)} = \left[q^{(\ell)^H} \pi^{(\ell)} (A) \, q^{(\ell)} \right]^H = 0 \; .$$

Consequently,

$$q^{(\ell)^H} \left[\sigma_{10}{}^{(\ell)} \pi^{(\ell)} (A) + \sigma_{01}{}^{(\ell)} \pi^{(\ell)^H} (A) \right] q^{(\ell)} = 0 \; .$$

The left side of (3.26) thus becomes

$$q^{(\ell)^H} \pi^{(\ell)^H} (A) \, \pi^{(\ell)} (A) \, q^{(\ell)} - \rho^{(\ell)^2} \; .$$

But this expression vanishes since the first term is $\|\pi^{(\ell)}(A) \, q^{(\ell)}\|^2$ and since by definition (3.4)

$$\rho^{(\ell)} = \|\pi^{(\ell)} (A) \, q^{(\ell)}\| \; .$$

IV. AN ALGORITHM WITH A LINEAR POLYNOMIAL

This section contains a discussion of the algorithm with the linear polynomial $\varphi_1^{(\ell)}(\lambda)$ and shows that this algorithm can have stationary vectors that are not characteristic vectors. The Rayleigh quotients $\mu_{01}^{(\ell)}$ may not converge to a characteristic root of the matrix A.

For the matrix

$$W_\ell = \left(q^{(\ell)},\ [\pi^{(\ell)}(A)q^{(\ell)}]/\rho^{(\ell)}\right) \quad ,$$

the linear polynomial

$$\pi^{(\ell)}(\lambda) = \varphi_1^{(\ell)}(\lambda) = \lambda - \mu_{01}^{(\ell)}$$

will give the simplest algorithm from a computational standpoint. Therefore, let

$$W_\ell = \left(q^{(\ell)},\ (A - \mu_{01}^{(\ell)}\ I)q^{(\ell)}/\rho^{(\ell)}\right)$$

where

$$\rho^{(\ell)} = \|(A - \mu_{01}^{(\ell)}\ I)q^{(\ell)}\| = \|p_1^{(\ell)}\| = \rho_1^{(\ell)}$$

by (3.7).

Since $\sigma_{01}^{(\ell)}$ is an off-diagonal element of the second order matrix (3.6), Theorems 2.3, 2.4, and 2.5 show that an important question is whether there are any vectors other than characteristic vectors of A that make $\sigma_{01}^{(\ell)}$ vanish. Since $\sigma_{01}^{(\ell)}$ is the conjugate of the other off-diagonal element, only $\sigma_{01}^{(\ell)}$ needs to be discussed.

Let A be a normal matrix with characteristic roots α_j ($j = 1, 2, 3, \ldots, n$). Now,

$$\sigma_{01}^{(\ell)} = p_1^{(\ell)^H} \left(A - \mu_{01}^{(\ell)} I\right) p_1^{(\ell)}$$

$$= q^{(\ell)^H} \left(A - \mu_{01}^{(\ell)} I\right)^H \left(A - \mu_{01}^{(\ell)} I\right)^2 q^{(\ell)}$$

which is a Rayleigh quotient of the matrix

$$\left(A - \mu_{01}^{(\ell)} I\right)^H \left(A - \mu_{01}^{(\ell)} I\right)^2$$

whose characteristic roots are $\left(\bar{\alpha}_j - \mu_{10}^{(\ell)}\right)\left(\alpha_j - \mu_{01}\right)^2$, $j = 1, 2, \ldots, n$, since A is a normal matrix. Consequently, by Theorem 3.2

$$\sigma_{01}^{(\ell)} = \sum_{j=1}^{n} \sum_{i=1}^{n} |\alpha_j - \mu_{01}^{(\ell)}|^2 \left(\alpha_j - \mu_{01}^{(\ell)}\right) \left(d_j^{(1)} \ldots d_j^{(\ell)} \beta_j\right) \left(\overline{d_i^{(1)} \ldots d_i^{(\ell)} \beta_i}\right) p_{ij}.$$

If the roots of A are real, an answer to the above question is easy to formulate. Let A, therefore, be a hermitian matrix of order n such that

$$A = P \Lambda P^{-1} = P \Lambda P^H$$

where

$$\Lambda = \text{diag } (\alpha_1, \alpha_2, \ldots \alpha_n)$$

with

$$\alpha_1 < \alpha_2 < \ldots < \alpha_n .$$

Since A is hermitian, $\mu_{01}^{(\ell)} = \mu_{10}^{(\ell)}$ and $\sigma_{01}^{(\ell)} = \sigma_{10}^{(\ell)}$. Let

$$\mu_1^{(\ell)} = \mu_{01}^{(\ell)} = \mu_{10}^{(\ell)}$$

and

$$\sigma_1^{(\ell)} = \sigma_{01}^{(\ell)} = \sigma_{10}^{(\ell)}.$$

From (3.18) it follows that if A is hermitian, then

$$\mu_1^{(\ell)} = \sum_1^n \alpha_j \, (d_j^{(1)} \ldots d_j^{(\ell)} \beta_j)^2$$

and

$$\sigma_1^{(\ell)} = \sum_1^n (\alpha_j - \mu_1^{(\ell)})^3 \, (d_j^{(1)} \ldots d_j^{(\ell)} \beta_j)^2$$

where

$$\sum_{j=1}^n (d_j^{(1)} \ldots d_j^{(\ell)} \beta_j)^2 = 1 \, .$$

The previous question of the existence of vectors other than characteristic vectors of A which make $\sigma_1^{(\ell)}$ vanish is equivalent to the question of the existence, for a fixed $\mu_1 \neq \alpha_j$, of vectors

$$\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)^T$$

that satisfy the system of equations

$$\begin{cases} \sum_1^n \gamma_j^2 = 1 \\ \sum_1^n \alpha_j \gamma_j^2 = \mu_1 \\ \sum_1^n (\alpha_j - \mu_1)^3 \gamma_j^2 = 0. \end{cases}$$

For some $\mu_1 \neq \alpha_j$, the answer to the question is in the affirmative, as the following theorem shows.

<u>Theorem 4.1.</u> Let A be a 3 by 3 hermitian matrix with characteristic roots

$$\alpha_1 < \alpha_2 < \alpha_3 \, . \tag{4.1}$$

If $\mu_1 < \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$ and $\mu_1 < \alpha_2$, then there is a real noncharacteristic stationary vector whose Rayleigh quotient is μ_1 for each

$$\frac{\alpha_1 + \alpha_2}{2} \le \mu_1 \le \frac{\alpha_1 + \alpha_3}{2} \tag{4.2}$$

Similarly, if $\mu_1 > \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$ and $\mu_1 > \alpha_2$, then there is a real non-characteristic stationary vector whose Rayleigh quotient is μ_1 for each

$$\frac{\alpha_1 + \alpha_3}{2} \le \mu_1 \le \frac{\alpha_2 + \alpha_3}{2} . \tag{4.3}$$

Proof: Assume that A is translated so that $\mu_1 = 0$. If μ_1 is not a characteristic root of A, then the vector is a noncharacteristic stationary vector if it satisfies the following system of equations:

$$\begin{cases} \sum_1^3 \gamma_i^2 = 1 \\ \sum_1^3 \alpha_i \gamma_i^2 = 0 \\ \sum_1^3 \alpha_i^3 \gamma_i^2 = 0 . \end{cases} \tag{4.4}$$

The determinant of coefficients for this system is

$$\delta = \det \begin{pmatrix} 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1^3 & \alpha_2^3 & \alpha_3^3 \end{pmatrix} = (\alpha_1 + \alpha_2 + \alpha_3)(\alpha_3 - \alpha_2)(\alpha_2 - \alpha_1)(\alpha_3 - \alpha_1),$$

and by Cramer's rule

$$\begin{aligned} &(1) \quad \delta\,\gamma_1^2 = \alpha_2\,\alpha_3\,(\alpha_3^2 - \alpha_2^2) \\ &(2) \quad \delta\,\gamma_2^2 = \alpha_1\,\alpha_3\,(\alpha_1^2 - \alpha_3^2) \\ &(3) \quad \delta\,\gamma_3^2 = \alpha_1\,\alpha_2\,(\alpha_2^2 - \alpha_1^2)\ . \end{aligned} \tag{4.5}$$

Since $\alpha_1 + \alpha_2 + \alpha_3 \neq 0$ and since the characteristic roots are simple, $\delta \neq 0$ and the vector γ exists.

Now assume that $\alpha_1 + \alpha_2 + \alpha_3 > 0$. Then $\delta > 0$.

The assumption that a translation is performed on the matrix A so that $\mu_1 = 0$ means that

$$\alpha_1 < 0 < \alpha_3 \tag{4.6}$$

since Rayleigh quotient of a hermitian matrix lies on the closed segment between the least and greatest characteristic roots. The inequalities (4.6) must be strict if the vector is to be a noncharacteristic vector of A.

If

$$\alpha_2 > 0 \tag{4.7}$$

the equations (4.5) have a real solution if and only if

$$\alpha_1^2 - \alpha_3^2 = (\alpha_1 + \alpha_3)(\alpha_1 - \alpha_3) \leq 0$$

and

$$\alpha_2^2 - \alpha_1^2 = (\alpha_2 - \alpha_1)(\alpha_1 + \alpha_2) \leq 0\ .$$

These two inequalities hold if and only if

$$\alpha_1 + \alpha_3 \geq 0 \tag{4.8}$$

and

$$\alpha_1 + \alpha_2 \leq 0 \tag{4.9}$$

since by hypothesis

$$\alpha_1 - \alpha_3 < 0$$

and

$$\alpha_2 - \alpha_1 > 0 \ .$$

If $\alpha_i - \mu_1$ is substituted for α_i $(i = 1, 2, 3)$ in (4.8) and (4.9) and if the resulting inequalities are solved for μ_1, the results are (4.2).

Now, assume that $\alpha_1 + \alpha_2 + \alpha_3 > 0$; then $\delta > 0$.

If

$$\alpha_2 < 0$$

the equations (4.5) have a real solution if and only if

$$\alpha_3^2 - \alpha_2^2 = (\alpha_3 + \alpha_2)(\alpha_3 - \alpha_2) \geq 0$$

and

$$\alpha_1^2 - \alpha_3^2 = (\alpha_1 + \alpha_3)(\alpha_1 - \alpha_3) \geq 0 \ .$$

These two inequalities hold if and only if

$$\alpha_3 + \alpha_2 \geq 0 \tag{4.10}$$

and

$$\alpha_1 + \alpha_3 \leq 0 \tag{4.11}$$

since by hypothesis

$$\alpha_3 - \alpha_2 > 0$$

and

$$\alpha_1 - \alpha_3 < 0 \ .$$

If $\alpha_i - \mu_1$ is substituted for α_i in (4.10) and (4.11), and if the resulting inequalities are solved for μ_1, the results are (4.3).

If μ_1 satisfies one of the inequalities (4.2) or (4.3), then there exist real vectors which satisfy the system of equations (4.4). This means that if the algorithm yields vectors $q^{(\ell)}$ that converge to one of these vectors, the inclusion regions R_ℓ will converge to a disk with a nonzero radius, and the sequence of Rayleigh quotients $\mu_1^{(\ell)}$ will converge to a number that is not a characteristic root of A. Consequently, this algorithm presents some serious difficulties when it is used to find characteristic roots and vectors.

If A is an nth order hermitian matrix, the system of equations (4.4) will be three in number, but the number of unknowns will be increased to n. There, thus, will be additional noncharacteristic real stationary vectors whose Rayleigh quotients μ_1 satisfy inequalities similar to (4.2) and (4.3).

V. AN ALGORITHM WITH A QUADRATIC POLYNOMIAL

This section presents the properties of the algorithm with the monic quadratic polynomial

$$\pi^{(\ell)} (\bar{\lambda}, \lambda) = (\bar{\lambda} - \mu_{10}^{(\ell)}) (\lambda - \mu_{01}^{(\ell)}) - \rho_1^{(\ell)^2} \tag{5.1}$$

$$= \overline{\varphi_1^{(\ell)} (\lambda)} \; \varphi_1^{(\ell)} (\lambda) - \rho_1^{(\ell)^2}$$

where $\rho_1^{(\ell)}$ is the radius of the inclusion disk

$$|\lambda - \mu_{01}^{(\ell)}| \leq \| (A - \mu_{01}^{(\ell)} I) q^{(\ell)}\| \ . \tag{5.2}$$

Henceforth, the symbol π will denote this algorithm.

If $q^{(\ell)}$ is a stationary vector of the algorithm π then either $q^{(\ell)}$ is a characteristic vector of the given normal matrix A, or the inclusion disk (5.2) has two or more characteristic roots of A on its boundary. A later theorem shows that if $q^{(\ell)}$ is a stationary vector of π, then it is a stationary vector for all algorithms that use the method discussed in Theorem 2.1 to determine $q^{(\ell+1)}$ from $q^{(\ell)}$. That is, from the subspace spanned by a set of mutually orthogonal vectors that contains the old vector $q^{(\ell)}$ a vector $q^{(\ell+1)}$ cannot be chosen that will yield a smaller inclusion disk with $\mu_{01}^{(\ell)}$ as center than the disk (5.2). Thus,

$$\| (A - \mu_{01}^{(\ell)} I) q^{(\ell)}\| \leq \| (A - \mu_{01}^{(\ell)} I) x \| \ / \ \| x \|$$

for all vectors x in the subspace regardless of the dimension of the subspace.

Since

$$q^{(\ell)^H} [\varphi_1^{(\ell)^H} (A) \varphi_1^{(\ell)} (A) - \rho_1^{(\ell)^2} I] q^{(\ell)}$$

$$= q^{(\ell)^H} \varphi_1^{(\ell)^H} (A) \varphi_1^{(\ell)} (A) q^{(\ell)} - \rho_1^{(\ell)^2}$$

$$= \rho_1^{(\ell)^2} - \rho_1^{(\ell)^2}$$

$$= 0 ,$$

the polynomial $\pi^{(\ell)} (\bar{\lambda}, \lambda)$ satisfies condition (3.3).

The next two theorems show that the Rayleigh quotients converge to a characteristic root of A if the radii of the Weinstein disks converge to zero.

Theorem 5.1. Let A be a normal matrix. If the radii $\rho_1^{(\ell)}$ of the inclusion disks (5.2) formed by the algorithm converge to zero, then the sequence of vectors $q^{(\ell)}$ will contain an infinite convergent subsequence that converges to a characteristic vector of A.

Proof: Since by hypothesis

$$\lim_{\ell \to \infty} \rho_1^{(\ell)} = 0$$

the only stationary vector are characteristic vectors. Hence, if the normalized vector $q^{(0)}$ is not a characteristic vector, then the algorithm π will yield a new vector $q^{(1)}$. Then by Theorem 2.1

$$\| \varphi_1^{(1)} (A) q^{(1)} \| < \| \varphi_1^{(0)} (A) q^{(0)} \| .$$

This inequality is strict since $\sigma_1^{(\ell)} \neq 0$ and since condition (3.9) implies that the smallest characteristic root of the matrix (2.8) is a simple root.

For any initial normalized vector $q^{(0)}$, therefore, the algorithm π will yield the vectors of the form (2.14) such that, for all ℓ,

$$\| \varphi_1^{(\ell+1)}(A)\, q^{(\ell+1)} \| < \| \varphi_1^{(\ell)}(A)\, q^{(\ell)} \|$$

as long as $q^{(\ell)}$ is not a characteristic vector of A.

The vectors $q^{(\ell)}$ are in the 2n dimensional compact space. By the Bolzano-Weierstrass theorem, there exists a subsequence, say $[q^{(k)}]$, which will have a limit in the space. There, thus, exists a vector q such that

$$\lim_{k \to \infty} q^{(k)} = q \; . \qquad (5.3)$$

Associated with this sequence of vectors is a sequence of matrices W_k that converges to a matrix W since

$$W_k = (q^{(k)}, \pi^{(k)} (A^H, A)\, q^k)$$

is a continuous function of $q^{(k)}$ because $\pi^{(k)} (A^H, A)$ is a continuous function of $q^{(k)}$. Therefore,

$$\lim_{k \to \infty} W_k = \lim_{k \to \infty} (q^{(k)}, \pi^{(k)} (A^H, A)\, q^{(k)}) = W \; . \qquad (5.4)$$

Since the columns of w_k are orthonormal and since

$$q^{(k+1)} = W_k\, w_k \; ,$$

then

$$w_k = W_k^H\, q^{(k+1)} \; .$$

By (5.3) and (5.4),

$$\lim_{k \to \infty} w_k = \lim_{k \to \infty} W_k^H q^{(k+1)}$$

$$= \lim_{k \to \infty} W_k^H \lim_{k \to \infty} q^{(k+1)}$$

$$= W^H q$$

$$= e_1 .$$

Thus $W e_1 = q$, which implies that q is a stationary vector. Consequently, q is a characteristic vector of A since the hypothesis that the radii $\rho_1^{(\ell)}$ of the inclusion disks converge to zero implies that a stationary vector is a characteristic vector.

The next theorem shows that the Rayleigh quotients formed by the sequence of vectors $q^{(\ell)}$ have a unique limit.

Theorem 5.2. Let A be a normal matrix and let the vectors be formed by the algorithm in the previous theorem. Then the Rayleigh quotients $\mu_{01}^{(\ell)}$ formed by the vectors $q^{(\ell)}$ will converge to a characteristic root α_j of the matrix A. That is,

$$\lim_{\ell \to \infty} \mu_{01}^{(\ell)} = \alpha_j .$$

If α_j is a simple root of A, then

$$\lim_{\ell \to \infty} q^{(\ell)} = q$$

where q is the normalized characteristic vector of A associated with the characteristic root α_j .

Proof: Since every convergent subsequence of the vectors must converge to a characteristic vector of A, the radii $\rho_1^{(\ell)} = \| \varphi_1^{(\ell)}(A)\, q^{(\ell)} \|$ of the Weinstein disks

$$R_\ell = \left\{ \lambda : \; |\lambda - \mu_{01}^{(\ell)}| \leq \rho_1^{(\ell)} \right\} \quad (\ell = 0, 1, \ldots)$$

converge to zero. Hence, every infinite subset of the Rayleigh quotients $\mu_{01}^{(\ell)}$ must converge to a characteristic root of A.

Now, to show that every infinite subset of the Rayleigh quotients converges to the same characteristic root of A, let α_j be within the Weinstein disk R_k and let the diameter of $R_k (2\, \rho_1^{(k)})$ be less than one-half the distance between α_j and the closest characteristic root of A. The diameter of R_k can be arbitrarily small since

$$\lim_{\ell \to \infty} \rho_1^{(\ell)} = 0 \quad .$$

By Theorem 2.6, the Rayleigh quotient $\mu_{01}^{(k+1)}$ must be within R_k. The Weinstein disk R_{k+1} can contain, therefore, only one of the α_j. Similarly, the Weinstein disks R_m $(m = k+2, k+3, \ldots)$ can contain only this same α_j. Consequently, the sequence of Rayleigh quotients $\mu_{01}^{(\ell)}$ $(\ell = 0, 1, 2, \ldots)$ converges to this characteristic root α_j.

If α_j is a simple characteristic root of A, then there is only one normalized characteristic vector associated with α_j. That is,

$$\lim_{\ell \to \infty} q^{(\ell)} = q$$

where q is the normalized characteristic vector associated with α_j.

For simplicity the superscript "ℓ" will be omitted from now on since the remaining theorems do not involve iterations.

Since the matrix $\pi(A^H, A) = \varphi_1^H (A)\, \varphi_1(A) - \rho_1^2 I$ is hermitian, the off-diagonal elements σ_{01} and σ_{10} of the matrix Q(3.1) are equal. Let

$$\sigma_1 = \sigma_{10} = \sigma_{01} .$$

Then by (3.7),

$$\sigma_1 = p_1^H \pi (A^H, A)\, p_1 . \tag{5.5}$$

The next theorem shows that σ_1 is the norm of the vector $\pi(A^H, A)\, q$. If $\sigma_1 = 0$, then by one of the properties of norms

$$\pi(A^H, A)\, q = (\varphi_1^H (A)\, \varphi_1(A) - \rho_1^2 I)\, q = 0 .$$

Thus,

$$\varphi_1^H (A)\, \varphi_1(A)\, q = \rho_1^2\, q . \tag{5.6}$$

Hence, ρ_1^2 is the characteristic root and q is the characteristic vector of $\varphi_1^H (A)\, \varphi_1(A)$.

<u>Theorem 5.3.</u> Let A be a normal matrix. Then

$$\sigma_1 = \| \pi (A^H, A)\, q \| .$$

<u>Proof:</u> From condition (5.5),

$$\sigma_1 = \left\{ p_1^H [(A - \mu_{01} I)^H (A - \mu_{01} I) - \rho_1^2 I]\, p_1 \right\} / \rho$$

$$= [p_1^H (A - \mu_{01} I)^H (A - \mu_{01} I)\, p_1 - \rho_1^4] / \rho \tag{5.7}$$

$$= \left\{ q^H [(A - \mu_{01} I)^H (A - \mu_{01} I)]^2 q - \rho_1^4 \right\} / \rho .$$

By definition (3.4)

$$\rho^2 = \| \pi (A^H, A) q \|^2$$

$$= q^H \pi^H (A^H, A) \pi (A^H, A) q$$

$$= q^H [\pi (A^H, A)]^2 q$$

$$= q^H [(A - \mu_{01} I)^H (A - \mu_{01} I) - \rho_1^2 I]^2 q \tag{5.8}$$

$$= q^H [(A - \mu_{01} I)^H (A - \mu_{01} I)]^2 q - 2\rho_1^2 q^H (A-\mu_{01} I)^H (A-\mu_{01} I) q + \rho_1^4$$

$$= q^H [(A - \mu_{01} I)^H (A - \mu_{01} I)]^2 q - 2\rho_1^4 + \rho_1^4$$

$$= q^H [(A - \mu_{01} I)^H (A - \mu_{01} I)]^2 q - \rho_1^4 .$$

Thus, (5.7) and (5.8) imply that

$$\sigma_1 = \frac{\rho^2}{\rho} = \rho \; \| \pi (A) q \| .$$

<u>Theorem 5.4.</u> Let A be a normal matrix of order n. If q is a normalized stationary vector of the algorithm π, then either q is a characteristic vector of A or the inclusion disk determined by q and A has two or more characteristic roots of A on its boundary.

<u>Proof:</u> Since A is normal, there exists a unitary matrix P such that

$$A = P \Lambda P^H$$

where

$$\Lambda = \mathrm{diag}\, (\alpha_1, \alpha_2, \ldots \alpha_n) .$$

There exists a normalized vector v such that

$$q = Pv \tag{5.9}$$

since q is a normalized vector and since P is a nonsingular matrix.

If (5.9) is substituted into (5.6), then

$$\varphi_1^H (A)\, \varphi_1 (A)\, Pv = \rho_1^2\, Pv. \tag{5.10}$$

If the equation (5.10) is multiplied by P^H, then

$$\varphi_1^H(\Lambda)\, \varphi_1(\Lambda)\, v = \rho_1^2\, v ,$$

since

$$P^H \varphi_1^H (A)\, P\, P^H \varphi_1(A)\, P = \varphi_1^H (\Lambda)\, \varphi_1(\Lambda) .$$

Thus

$$\operatorname{diag} (|\alpha_1 - \mu_{01}|^2, \ldots, |\alpha_n - \mu_{01}|^2)\, v = \rho_1^2 \; v .$$

Consequently, if

$$|\alpha_k - \mu_{01}| = \rho_1$$

and if

$$|\alpha_k - \mu_{01}| \neq |\alpha_i - \mu_{01}| \tag{5.11}$$

for all $k \neq i$, then

$$v = e_k .$$

Hence,

$$\Lambda v = \alpha_k v$$

and

$$APv = \alpha_k P v .$$

Therefore, if condition (5.11) holds, $q = Pv$ is a characteristic vector of A; and if it does not hold, there will be at least two characteristic roots of A on the boundary of the inclusion disk determined by A and q.

If A is a hermitian matrix and if q is a noncharacteristic stationary vector of the algorithm π, the next theorem shows that two characteristic vectors of A are linear combinations of the two orthogonal vectors q and p_1.

Theorem 5.5. Let A be a hermitian matrix, let q be a stationary vector of the algorithm π which is not a characteristic vector of A, and let α_i and α_j be the end points of the inclusion interval determined by A and q. Then α_i and α_j are characteristic roots of A and

$$x_i = \frac{\sqrt{2}}{2} \left(q - \frac{1}{\rho_1} p_1\right) \tag{5.12}$$

and

$$x_j = \frac{\sqrt{2}}{2} \left(q + \frac{1}{\rho_1} p_1\right) \tag{5.13}$$

are characteristic vectors of A corresponding to α_i and α_j.

Proof: By the previous theorem the end points of the inclusion interval determined by A and q are characteristic roots of A since q is a noncharacteristic vector of A and a stationary vector of the algorithm π. Therefore, let α_i and α_j be the left and right end points, respectively. Thus $\alpha_i = \mu_{01} - \rho_1$, and $\alpha_j = \mu_{01} + \rho_1$. Consequently, by (5.12)

$$\begin{aligned} \alpha_i x_i &= (\mu_{01} - \rho_1) \; x_i \\ &= (\mu_{01} - \rho_1) \frac{\sqrt{2}}{2} \left(q - \frac{1}{\rho_1} p_1\right) \qquad (5.14) \\ &= \frac{\sqrt{2}}{2\rho_1} (\mu_{01} - \rho_1)(\rho_1 q - p_1) \;, \end{aligned}$$

and by (5.13)

$$\begin{aligned}\alpha_j x_j &= (\mu_{01} + \rho_1) x_j \\ &= (\mu_{01} + \rho_1) \frac{\sqrt{2}}{2} (q + \frac{1}{\rho_1} p_1) \qquad (5.15) \\ &= \frac{\sqrt{2}}{2\rho_1} (\mu_{01} + \rho_1)(\rho_1 q + p_1) \quad .\end{aligned}$$

Since q is a stationary vector of the algorithm π, then

$$\begin{aligned}(\varphi_1^2 (A) - \rho_1^2 I) q &= \varphi_1 (A) p_1 - \rho_1^2 q \\ &= Ap_1 - \mu_{01} p_1 - \rho_1^2 q \\ &= 0 \ .\end{aligned}$$

Hence,

$$\begin{aligned}Ap_1 &= \mu_{01} p_1 + \rho_1^2 q \qquad (5.16) \\ &= \mu_{01} Aq + (\rho_1^2 - \mu_{01}^2) q \ .\end{aligned}$$

If (5.12) is multiplied by A, then

$$\begin{aligned}Ax_i &= \frac{\sqrt{2}}{2} (Aq - \frac{1}{\rho_1} A p_1) \qquad (5.17) \\ &= \frac{\sqrt{2}}{2\rho_1} (\rho_1 Aq - Ap_1) \quad .\end{aligned}$$

If (5.16) is substituted into (5.17), then

$$\begin{aligned}Ax_i &= \frac{\sqrt{2}}{2\rho_1} [\rho_1 Aq - \mu_{01} Aq - (\rho_1^2 - \mu_{01}^2) q] \qquad (5.18) \\ &= \frac{\sqrt{2}}{2\rho_1} (\mu_{01} - \rho_1) [-Aq + (\rho_1 + \mu_{01}) q]\end{aligned}$$

$$= \frac{\sqrt{2}}{2\rho_1} (\mu_{01} - \rho_1) [-(A - \mu_{01} I) q + \rho_1 q] \qquad (5.18)$$

$$= \frac{\sqrt{2}}{2\rho_1} (\mu_{01} - \rho_1) (\rho_1 q - p_1) \quad .$$

Therefore,

$$Ax_i = \alpha_i x_i$$

since (5.18) and (5.14) are the same. Thus, x_i is the characteristic vector of A corresponding to the characteristic root α_i.

Similarly, x_j is a characteristic vector of A corresponding to the characteristic root α_j.

In the following discussion let C denote the class of algorithms that was discussed in Chapter II.

Theorem 5.6. Let A be a normal matrix, and let q be a stationary vector of the algorithm π. Then q is a stationary vector for every algorithm that belongs to Class C.

Proof: Since q is a stationary vector of the algorithm π,

$$\varphi_1^H (A) \varphi_1(A) q = \rho_1^2 q \quad . \qquad (5.19)$$

For simplicity, let

$$X_i = X_i (A^H, A)$$

where $X_i (A^H, A)$ $(i = 1, 2, \ldots, m-1)$ is a sequence of polynomials that satisfies the following relations:

$$q^H X_i q = 0 \qquad (5.20)$$

$$q^H X_i^H X_j q = 0 \quad . \qquad (5.21)$$

An algorithm that belongs to Class C is determined when a method of determining the columns of the n x m orthonormal matrix W is given. Therefore, let

$$W = (q, \chi_1 q, \chi_2 q, \ldots, \chi_{m-1} q) .$$

Hence,

$$\varphi_1(A)\, W = (\varphi_1(A)\, q,\ \varphi_1(A)\chi_1 q,\ \varphi_1(A)\chi_2 q,\ \ldots,\ \varphi_1(A)\chi_{m-1} q) .$$

Since A is normal, the polynomials $\varphi_1(A)$ and χ_i commute. Thus,

$$\begin{aligned} [\varphi_1(A)\, \chi_i q]^H\, [\varphi_1(A)\, \chi_j q] &= q^H\, \chi_i^H\, \varphi_1^H(A)\, \varphi_1(A)\, \chi_j\, q \\ &= q^H\, \chi_i^H\, \chi_j\, \varphi_1^H(A)\, \varphi_1(A)\, q \\ &= \rho_1^2\, q^H\, \chi_i^H\, \chi_j\, q \\ &= \rho_1^2\, \delta_{ij} \end{aligned} \tag{5.22}$$

by (5.20) and (5.21). Consequently, by (5.22)

$$(\varphi_1(A)\, W)^H\ (\varphi_1(A)\, W) = \rho_1^2\, I .$$

Therefore, q is a stationary vector for every algorithm that belongs to Class C.

The following discussion concerns the application of Theorem 3.3 to the algorithm π.

If $\pi\,(\bar{\lambda},\lambda)$ is substituted for $\pi(\lambda)$ in (3.20), the result is

$$f(\lambda) = |\pi(\bar{\lambda},\ \lambda)|^2 - \frac{2\sigma_1^2}{\rho_1^2 - \lambda_1}\ \pi\,(\bar{\lambda},\ \lambda) - \sigma_1^2$$

$$= \left|\pi(\bar{\lambda}, \lambda) - \frac{\sigma_1^2}{\rho_1^2 - \lambda_1}\right|^2 - \left(\sigma_1^2 + \frac{\sigma_1^4}{(\rho_1^2 - \lambda_1)^2}\right)$$

$$= \left\{\left|\pi(\bar{\lambda}, \lambda) - \frac{\sigma_1^2}{\rho_1^2 - \lambda_1}\right| - \left(\sigma_1^2 + \frac{\sigma_1^4}{(\rho_1^2 - \lambda_1)^2}\right)^{\frac{1}{2}}\right\}\left\{\left|\pi(\bar{\lambda}, \lambda) - \frac{\sigma_1^2}{\rho_1^2 - \lambda_1}\right|\right.$$

$$\left. + \left(\sigma_1^2 + \frac{\sigma_1^4}{(\rho_1^2 - \lambda_1)^2}\right)^{\frac{1}{2}}\right\} .$$

Since the second factor of $f(\lambda)$ is always greater than or equal to zero, the first factor determines the signs of $f(\lambda)$. If the first factor of $f(\lambda)$ is less than zero, then

$$\left|\, |\lambda - \mu_{01}|^2 - \rho_1^2 - \frac{\sigma_1^2}{\rho_1^2 - \lambda_1}\right| < \left(\sigma_1^2 + \frac{\sigma_1^4}{(\rho_1^2 - \lambda_1)^2}\right)^{\frac{1}{2}}$$

which is the annulus

$$\left\{\rho_1^2 + \frac{\sigma_1}{\rho_1^2 - \lambda_1}\left(\sigma_1 - \sqrt{\sigma_1^2 + (\rho_1^2 - \lambda_1)^2}\right)\right\}^{\frac{1}{2}} < |\lambda - \mu_{01}|$$

$$< \left\{\rho_1^2 + \frac{\sigma_1}{\rho_1^2 - \lambda_1}\left(\sigma_1 + \sqrt{\sigma_1^2 + (\rho_1^2 - \lambda_1)^2}\right)\right\}^{\frac{1}{2}}$$

that will be denoted by B.

The circles of the annulus B are concentric with the inclusion region

$$R = \left\{\lambda : |\lambda - \mu_{01}| \leq \kappa \;\|\; \varphi_1(A)\, q \;\| = \kappa\, \rho_1\right\}$$

and the boundary of the inclusion region R is within the annulus B.

Now, $f(\lambda) < 0$ for every λ within the annulus, and $f(\lambda) > 0$ for every λ within the inner disk or outside the outer disk. By Theorem 3.3, $|d_j| > 1$ when the characteristic root α_j is within the inner disk or outside the outer disk, and $|d_j| < 1$ when the characteristic root α_j is in the annulus. Thus, the weight of a characteristic root will be increased if it is within the inner disk or outside the outer disk and the weight decreased if it is in the annulus.

In summary, this study shows different algorithms for the development of a sequence of vectors $q^{(\ell)}$ ($\ell = 0, 1, 2, \ldots$) such that the radii of their associated Weinstein disks are decreasing. If the radii of the Weinstein disks approach zero as a limit, then the Rayleigh quotients associated with the sequence of vectors converge to a characteristic root of the matrix A.

The algorithm with the linear polynomial

$$\varphi_1 (\lambda) = \lambda - \mu_{01}$$

does not always yield vectors whose associated Rayleigh quotients converge to a characteristic root of the matrix A. However, the algorithm π does yield vectors whose associated Rayleigh quotients converge to a characteristic root of a normal matrix A with the exception that is noted in Theorem 5.4. In the case of a hermitian matrix, Theorem 5.5 gives a method to cover the exception.

BIBLIOGRAPHY

1. Bauer, Friedrich L. and A. S. Householder, "Moments and Characteristic Roots," _Numerische Mathematik_, II (1960), 42-53.

2. Hestenes, Magnus R. and William Karush, "A Method of Gradients for the Calculation of the Characteristic Roots and Vectors of a Real Symmetric Matrix," _Journal of Research of the National Bureau of Standards_, XLVII (1951), 45-61.

3. Weinstein, D. H. "Modified Ritz Method," _Proceedings of the National Academy of Science_, XX (1934), 529-532.

ORNL-3688
UC-32 — Mathematics and Computers
TID-4500 (32nd ed.)

INTERNAL DISTRIBUTION

1. Biology Library
2-4. Central Research Library
5. Reactor Division Library
6-7. ORNL — Y-12 Technical Library Document Reference Section
8-27. Laboratory Records Department
28. Laboratory Records, ORNL R.C.
29. G. J. Atta
30. Susie E. Atta
31. R. E. Biggers
32. B. F. Bombay
33. Allan Bostwick (K-25)
34. E. E. Branstetter
35. John A. Carpenter
36. M. H. Davis (K-25)
37. R. E. Funderlic (K-25)
38. D. A. Gardiner
39. A. S. Householder
40. N. C. Hull
41. W. H. Jordan
42. H. W. Joy
43. C. E. Larson
44. M. E. LaVerne
45. F. L. Miller
46. R. V. Miskell
47. C. W. Nestor, Jr.
48. D. C. Ramsey
49. Bert Rust (K-25)
50. M. J. Skinner
51. W. J. Stelzman
52. J. A. Swartout
53. A. M. Weinberg
54. H. D. Wills
55. R. E. Worsham
56. J. H. Zeigler (K-25)
57. Henry Zeldes

EXTERNAL DISTRIBUTION

58-67. Donald H. Clanton, 8 Lisa Drive, Greenville, South Carolina
68. Research and Development Division, AEC, ORO
69-676. Given distribution as shown in TID-4500 (32nd ed.) under Mathematics and Computers category (75 copies — OTS)