AN ITERATIVE METHOD FOR FINDING CHARACTERISTIC VECTORS OF A SYMMETRIC MATRIX

W. KARUSH

1. Introduction. Given a real symmetric linear operator A on a vector space $\mathcal{E}$, we wish to describe a procedure for finding a "minimum" characteristic vector of A, that is, a characteristic vector with least characteristic value, supposing such to exist. The method to be used is, in a general way, the following. Select an initial vector x^0 and a positive integer $s > 1$. Imbed x^0 in an s-dimensional linear subspace $\mathcal{E}^0$ (appropriately selected). Determine the next approximation x^1 as the minimum characteristic vector relative to this subspace (to be defined later). Next, imbed x^1 in an s-dimensional subspace $\mathcal{E}^1$ and determine x^2 as the minimum characteristic vector relative to this subspace. Proceeding in this manner, construct a sequence of subspaces $\mathcal{E}^0, \mathcal{E}^1, \cdots$ of fixed dimension s, with a corresponding sequence of vectors $x^1, x^2, \cdots$. It is to be expected that under appropriate hypotheses the sequence of vectors will converge to a minimum characteristic vector of A.

We shall treat the case when $\mathcal{E}$ is of finite dimension n, and $\mathcal{E}^i$ is chosen as the subspace spanned by the vectors $x^i, Ax^i, A^2x^i, \cdots, A^{s-1}x^i$. We shall establish the desired convergence under these circumstances, the sequence $\{x^i\}$ satisfying at the same time a relation $x^{i+1} = x^i + \eta^i$ with $(x^i, \eta^i) = 0$. The main result is formulated in Theorem 2 of §6. An analogous result holds for a "maximum" characteristic vector.

It is of interest to compare the present iteration method with what might be called Rayleigh-Ritz procedures. In the latter, one fills out the space $\mathcal{E}$ by a judiciously chosen monotone sequence of subspaces

$$\mathcal{E}_1 \subset \mathcal{E}_2 \subset \mathcal{E}_3 \subset \cdots \qquad (\dim \mathcal{E}_i = i)$$

of increasing dimension. One then obtains successive approximations to a minimum vector of A by determining minimum characteristic vectors of the successive

Received May 20, 1950. The preparation of this paper was sponsored (in part) by the Office of Naval Research.

Pacific J. Math. 1 (1951), 233-248.

subspaces. This procedure has the serious computational drawback that to obtain an improved approximation a problem of increased complexity, that is, of higher dimension, must be solved. This restriction is important even in the finite dimensional case where the iteration, in theory, terminates in a finite number of steps. The method of the present paper, however, requires only the solution of a problem of fixed dimension s at each step, the dimension s being chosen from the outset as any desired value. The $\mathcal{E}^i$ form a chain of subspaces in which successive subspaces $\mathcal{E}^i$ and $\mathcal{E}^{i+1}$ overlap in x^{i+1}; in general this chain will be infinite even when $\mathcal{E}$ is finite-dimensional. Thus the method is useful where it is desired to fix beforehand the degree of complexity for all steps; and yet a great many iterations may readily be performed. This is the case with high speed computing machines.

The present procedure may be interpreted as a gradient method; cf [1]. For $s = 2$, in the equation $x^{i+1} = x^i + \eta^i$, η^i is a multiple of the gradient at $x = x^i$ of the function $(x, Ax)/(x, x)$. For $s > 2$, the vector η contains higher order terms. The applicability of the present procedure with $s = 2$ to quadratic functionals in infinite-dimensional spaces has been pointed out to the author by M.R.Hestenes in conversation, and has been outlined by L.V.Kantorovitch [2].

2. Subspaces. Before describing in detail the iteration procedure to be used, and proving its convergence, we find it convenient to formulate some preliminary results. In this section we construct an orthogonal basis for the space spanned by the powers of A operating on a fixed vector x; in the next section we describe the characteristic roots and vectors relative to certain subspaces of this space. We shall encounter polynomials $p_j(\lambda)$ of central importance. In these two sections we shall be treating, essentially, only one level of the iteration. Accordingly, the superscript i denoting the various steps of the iteration will not appear until §4, where we are concerned with the progression from one level to the next.

Let $\mathcal{E}$ denote the n-dimensional space of n-tuples of real numbers; by *vector* we understand always an element of $\mathcal{E}$. We consider a linear operator A on $\mathcal{E}$ which is real and symmetric; that is, one for which Ax is a real vector and

$$(Ax, z) = (x, Az)$$

for arbitrary real vectors x, z. A characteristic number (root, value) of A is a number λ for which there exists a non-null vector y such that

$$Ay = \lambda y.$$

There are n (real) characteristic numbers (counting multiplicities).

With a non-null vector x we associate the number

$$\mu(x) = \frac{(x, Ax)}{(x, x)}$$

and the vector

$$\xi(x) = Ax - \mu(x)x .$$

Let $\lambda_{\min}$ ($\lambda_{\max}$) be the least (greatest) characteristic root of A. It is well known that

$$\text{(1)} \qquad \lambda_{\min} = \min_{x \neq 0} \mu(x) , \qquad \lambda_{\max} = \max_{x \neq 0} \mu(x) , \qquad (x \in \mathcal{E}) .$$

For a non-null vector x we define the subspaces

$$\mathcal{A}_j(x) = (x, Ax, \cdots, A^{j-1}x) \qquad (j = 1, 2, 3, \cdots),$$
$$\mathcal{A}(x) = (x, Ax, A^2x, \cdots) ,$$

where, in each case, the right side of the equation denotes the space spanned by the designated vectors. The space $\mathcal{A}(x)$ is the smallest invariant subspace containing x; denote its dimension by $r = r(x)$. Clearly $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_r = \mathcal{A}$, where "$\subset$" denotes strict inclusion. The space $\mathcal{A}$ contains r independent characteristic vectors of A. We now construct an orthogonal basis for $\mathcal{A}_j$.

LEMMA 1. *Let the vectors ξ_j $(j = 0, 1, \cdots, r)$ be defined by*

$$\text{(2)} \qquad \xi_0 = x , \qquad \xi_1 = A\xi_0 - \mu_0\xi_0 \qquad (\mu_0 = \mu(x)),$$
$$\xi_{j+1} = A\xi_j - \mu_j\xi_j - t_j^2\xi_{j-1} \qquad (\mu_j = \mu(\xi_j)),$$
$$t_j = \frac{|\xi_j|}{|\xi_{j-1}|} \qquad (j = 1, 2, \cdots, r-1) .$$

Then for $j, k = 0, 1, \cdots, r-1$, we have $\xi_j \neq 0$, and

$$\text{(3)} \qquad \mathcal{A}_{j+1}(x) = (\xi_0, \xi_1, \cdots, \xi_j) , \qquad (\xi_j, \xi_k) = 0 ,$$
$$(A\xi_j, \xi_{j+1}) = |\xi_{j+1}|^2 \qquad (j \neq k) .$$

The lemma may be verified directly by induction. We remark that $\xi_r = 0$.

LEMMA 2. *Let the polynomials* $p_j(\lambda)$ $(j = 0, 1, \cdots, r)$ *be defined by*

$$p_0(\lambda) = 1, \qquad p_1(\lambda) = (\lambda - \mu_0), \qquad p_2(\lambda) = (\lambda - \mu_0)(\lambda - \mu_1) - t_1^2,$$

$$p_{j+1}(\lambda) = p_j(\lambda)(\lambda - \mu_j) - t_j^2 p_{j-1}(\lambda) \qquad (j = 1, 2, \cdots, r-1).$$

Suppose $\mathfrak{B}$ *is an invariant subspace containing* x; *write*

$$x = a_1 y_1 + a_2 y_2 + \cdots + a_l y_l \tag{4}$$

in terms of a basis of characteristic vectors of $\mathfrak{B}$. *Then*

$$\xi_j = a_1 p_j(\lambda_1) y_1 + a_2 p_j(\lambda_2) y_2 + \cdots + a_l p_j(\lambda_l) y_l \qquad (j = 0, 1, \cdots, r), \tag{5}$$

where λ_k *is the characteristic number of* y_k.

The lemma follows immediately from the definitions (2).

The polynomials $p_j(\lambda)$ have also been used by C. Lanczos [3].

3. Characteristic values relative to subspaces. Let $\mathfrak{B}$ be an arbitrary (linear) subspace of $\mathfrak{E}$; let π be the operator on $\mathfrak{E}$ which carries any vector into its projection on $\mathfrak{B}$. We define a linear operator $A(\mathfrak{B})$ on $\mathfrak{B}$ to $\mathfrak{B}$ as follows:

$$A(\mathfrak{B})x = \pi(Ax) \qquad (x \in \mathfrak{B}).$$

Then $A(\mathfrak{B})$ is a symmetric operator on $\mathfrak{B}$, since $A(\mathfrak{B}) = \pi A \pi$. *By the characteristic roots and vectors of A relative to the subspace $\mathfrak{B}$, we mean the corresponding quantities of $A(\mathfrak{B})$*. If $\mathfrak{B}$ is invariant, then these quantities are characteristic for A itself. We shall use the following easily verified fact: y is a characteristic vector relative to $\mathfrak{B}$ with characteristic value λ if and only if $y \neq 0$, $y \in \mathfrak{B}$, and $(Ay, z) = \lambda(y, z)$ for $z \in \mathfrak{B}$. By a *minimum characteristic vector* of $\mathfrak{B}$ we shall mean a characteristic vector relative to $\mathfrak{B}$ with least characteristic value. When no confusion can arise we shall omit the qualifying term "relative."

LEMMA 3. *The j characteristic roots relative to the subspace $\mathfrak{A}_j(x)$ are distinct and are given by the solutions of*

$$p_j(\lambda) = 0.$$

Each characteristic vector (relative to $\mathfrak{A}_j$) has a non-null projection on x.

To prove the last statement, suppose that y is a characteristic vector with characteristic value λ. If $(y, x) = 0$, then $(y, Ax) = (Ay, x) = \lambda(y, x) = 0$, and

$(y, A^2x) = (Ay, Ax) = \lambda(y, Ax) = 0, \cdots$, and $(y, A^{j-1}x) = 0$. From the definition of $\mathfrak{A}_j$ it follows that y is orthogonal to this space. But y belongs to this space; hence $y = 0$, a contradiction.

The distinctness of the roots now follows. For if two independent characteristic vectors belong to λ then there is a non-null linear combination orthogonal to x belonging to λ.

To complete the proof we use the basis (3) of $\mathfrak{A}_j$. The matrix representation, call it A_j, of $A(\mathfrak{A}_j)$ relative to this basis has as element in the $(k+1)$st row and $(l+1)$st column:

$$\frac{(A\xi_k, \xi_l)}{|\xi_k|\,|\xi_l|} \qquad (k, l = 0, 1, \cdots, j-1)\,.$$

Using (2) and the second line of (3) we find that

$$A_j = \begin{Vmatrix} \mu_0 & t_1 & 0 & \cdot & \cdot & & \\ t_1 & \mu_1 & t_2 & & & 0 & \\ 0 & t_2 & \mu_2 & \cdot & & & \\ \cdot & & \cdot & \cdot & \cdot & & \\ \cdot & & & \cdot & \cdot & \cdot & \\ & 0 & & & & & \\ & & & & \cdot & \cdot & t_{j-1} \\ & & & & & t_{j-1} & \mu_{j-1} \end{Vmatrix}$$

Thus, the characteristic roots are the roots of the polynomial

$$q_j(\lambda) = |\lambda I_j - A_j|\,,$$

where I_j is the j-rowed square identity matrix. Let $q_0(\lambda) = 1$. Direct calculation shows that $q_1(\lambda) = p_1(\lambda)$, and that the $q_j(\lambda)$ satisfy the same recursion relation as the $p_j(\lambda)$. Hence the two sets of polynomials are identical. This completes the proof.

LEMMA 4. *Let ν_j be the minimum characteristic root relative to $\mathfrak{A}_j$; that is,*

$$\nu_j = \text{min. root of } p_j(\lambda) \qquad (j = 1, 2, \cdots, r)\,.$$

Then

$$\lambda_1 = \nu_r < \nu_{r-1} < \cdots < \nu_1\,, \tag{6}$$

where λ_1 *is the minimum characteristic root of the invariant subspace* $\mathfrak{A}_r$. *Further, each root* σ *of each polynomial* $p_j(\lambda)$ *satisfies*

$$\lambda_{\min} \leq \sigma \leq \lambda_{\max}. \tag{7}$$

The last statement follows at once from Lemma 3 and (1) when we notice that each characteristic root σ is a value of $\mu(z) = (z, Az)/(z, z)$; namely, σ is that value obtained by replacing z by the corresponding characteristic vector.

To prove (6) we apply (1) to the operator $A(\mathfrak{A}_j)$. Using the fact that $(Az, z) = [A(\mathfrak{A}_j)z, z]$ for z in $\mathfrak{A}_j$, we find that

$$\nu_j = \min_{z \neq 0} \mu(z), \qquad (z \text{ in } \mathfrak{A}_j).$$

From $\mathfrak{A}_j \subset \mathfrak{A}_{j+1}$ we infer that the roots are non-increasing. Suppose that $\nu_k = \nu_{k+1}$. Denote the common value by ν. From the recursion formula for the polynomials it follows that

$$p_{k-1}(\nu) = p_{k-2}(\nu) = \cdots = p_0(\nu) = 0,$$

contrary to the definition $p_0(\lambda) \equiv 1$.

LEMMA 5. *The minimum characteristic vector relative to* $\mathfrak{A}_j$ *is given by*

$$z_j = x + \frac{p_1(\nu_j)}{\tau_1^2}\xi_1 + \frac{p_2(\nu_j)}{\tau_2^2}\xi_2 + \cdots + \frac{p_{j-1}(\nu_j)}{\tau_j^2}\xi_{j-1}, \tag{8}$$

where

$$\tau_k = t_1 \cdot t_2 \cdots t_k = \frac{|\xi_k|}{|x|}.$$

More generally, the characteristic vector belonging to an arbitrary root σ is obtained by replacing ν_j by σ on the right in (8). To prove this, let z denote the vector obtained by this substitution. It is sufficient to show that $\eta = Az - \sigma z$ is orthogonal to $\mathfrak{A}_j$; to this end we use the basis in (3). Using the definition of z and the relations (2) and (3), we find that

$$\begin{aligned}(x, \eta) &= (x, Ax) + \frac{p_1(\sigma)}{\tau_1^2}|\xi_1|^2 - \sigma|x|^2 \\ &= [p_1(\sigma) - (\sigma - \mu_0)]|x|^2 = 0,\end{aligned}$$

$$(\xi_l, \eta) = \frac{p_{l-1}(\sigma)}{\tau_{l-1}^2} \cdot |\xi_l|^2 + \frac{p_l(\sigma)}{\tau_l^2} \mu_1 \cdot |\xi_l|^2$$

$$+ \frac{p_{l+1}(\sigma)}{\tau_{l+1}^2} \cdot |\xi_{l+1}|^2 - \nu \frac{p_l(\sigma)}{\tau_l^2} \cdot |\xi_l|^2$$

$$= \frac{|\xi_l|^2}{\tau_l^2} [p_{l+1}(\sigma) - \{p_l(\sigma)(\sigma - \mu_l) - p_{l-1}(\sigma)\, t_l^2\}] = 0$$

for $l = 1, 2, \cdots, j-2$. For $l = j-1$, the term in p_{l+1} does not appear, and we obtain

$$(\xi_{j-1}, \eta) = -\frac{|\xi_{j-1}|^2}{\tau_{j-1}^2} p_j(\sigma) = 0 .$$

This completes the argument.

4. The iteration procedure. We shall henceforth be dealing with a sequence $\{x^i\}$ of vectors; with each vector we associate the quantities described previously for an arbitrary vector x. To indicate dependence upon x^i we shall adjoin the superscript i to the symbols denoting these quantities.

Consider an initial vector $x^0 \neq 0$. By definition r^0 $[= r(x^0)]$ is the dimension of $\mathfrak{A}^0$ $[= \mathfrak{A}(x^0)]$, the smallest invariant subspace containing x^0. Since $\mathfrak{A}^0 = \mathfrak{A}_{r^0}(x^0)$, according to Lemma 3 there are r^0 distinct characteristic roots

$$\lambda_1 < \lambda_2 < \cdots < \lambda_{r^0}$$

relative to $\mathfrak{A}^0$; and the corresponding characteristic vectors can be normalized so that

$$x^0 = y_1 + y_2 + \cdots + y_{r^0} .$$

All vectors considered below will lie in the invariant space $\mathfrak{A}^0$. Henceforth the symbols λ_j and y_j will denote the characteristic quantities of this subspace.

To specify the iteration procedure at hand we require, besides x^0, the selection of a fixed dimension $s > 1$. We remark at this point that the significant case is that for which the dimension of the invariant space $\mathfrak{A}(x^i)$ at every stage exceeds s; that is,

$$(9) \qquad (i = 0, 1, 2, \cdots) .$$

To simplify presentation, unless otherwise stated *it will be assumed that this condition holds*. The trivial case in which (9) fails will be treated at the end of this section.

Consider now the s-dimensional subspace $\mathfrak{A}_s^0 = \mathfrak{A}_s(x^0)$. Relative to this subspace there is, by Lemma 3, a unique minimum characteristic vector $x^0 + \eta^0$ with $(x^0, \eta^0) = 0$; call it x^1. Now form $\mathfrak{A}_s^1 = \mathfrak{A}_s(x^1)$ and select x^2 as the unique minimum characteristic vector relative to this space of the form $x^1 + \eta^1$, $(x^1, \eta^1) = 0$. In general we define x^{i+1} as the minimum characteristic vector $x^i + \eta^i$, $(x^i, \eta^i) = 0$, relative to the subspace $\mathfrak{A}_s^i$. Notice that these subspaces form a chain in which successive subspaces of index i and $i+1$ overlap in x^{i+1}.

LEMMA 6. *The sequence $\{x^i\}$ is given by*

$$(10) \qquad x^{i+1} = x^i + \frac{p_1^i(\nu^i)}{(\tau_1^i)^2}\, \xi_1^i + \cdots + \frac{p_{s-1}^i(\nu^i)}{(\tau_{s-1}^i)^2}\, \xi_{s-1}^i ,$$

where ν^i is the least root of $p_s^i(\lambda)$. Further,

$$(11) \qquad \nu^i = \mu(x^{i+1}) .$$

Also $\{\nu^i\}$ is decreasing; in fact

$$(12) \qquad \lambda_1 < \nu^i = \nu_s^i < \nu_{s-1}^i < \cdots < \nu_1^i = \mu(x^i),$$

where ν_j^i is the minimum zero of $p_j^i(\lambda)$.

By Lemma 3 the minimum characteristic root relative to $\mathfrak{A}_s^i$ is ν_s^i. It follows by the definition of x^{i+1} that the equality (11) holds. The relations (12) follow from Lemma 4, condition (9), and definition. The formula (10) is (8) of Lemma 5 interpreted for $x = x^i$ and $j = s$.

LEMMA 7. *In terms of the characteristic basis of $\mathfrak{A}^0$ we have*

$$(13) \qquad x^i = a_1^i y_1 + a_2^i y_2 + \cdots + a_{r^0}^i y_{r^0} ,$$

$$(14) \qquad \xi_j^i = a_1^i\, p_j^i(\lambda_1) y_1 + a_2^i\, p_j^i(\lambda_2) y_2 + \cdots + a_{r^0}^i\, p_j^i(\lambda_{r^0}) y_{r^0}$$

$$(i = 0, 1, 2, \cdots; \quad j = 0, 1, \cdots, r^i) .$$

where

$$a_k^{i+1} = a_k^i \left\{ 1 + \frac{p_1^i(\nu^i)\, p_1^i(\lambda_k)}{(\tau_1^i)^2} + \cdots + \frac{p_{s-1}^i(\nu^i)\, p_{s-1}^i(\lambda_k)}{(\tau_{s-1}^i)^2} \right\} \tag{15}$$

$$(k = 1, 2, \cdots, r^0)\,.$$

Furthermore, $a_k^0 = 1$ *and*

$$1 = a_1^0 < a_1^1 < a_1^2 < \cdots\,. \tag{16}$$

Formula (14) follows from (13) by Lemma 2; (15) is a consequence of (13), (14) and (10) of Lemma 6. To prove (16) we notice that $p_j^i(\lambda)$ $(j = 1, 2, \cdots, s-1)$ is not zero, and has the same sign, at λ_1 and at ν^i [since by (12) the least root of the polynomial exceeds these values]. Hence each term in braces in (15) is positive; this completes the proof.

We conclude the present section with a consideration of the possible failure of (9). Suppose that for some first value m of i this inequality fails. Then $\mathfrak{A}_s^m$ is an invariant subspace, and the minimum characteristic vector x^{m+1} relative to this subspace is a characteristic vector of A. Thus $\mathfrak{A}_s^{m+1}$ is a one-dimensional invariant subspace containing only multiples of x^{m+1}. It follows that $x^i = x^{m+1}$ for $i \geq m + 1$. But the argument used in establishing (16) shows that $x^i = Ly_1$, $L > 0$, for $i \geq m + 1$. The theorems to be proved in the next two sections now hold trivially. We are thereby justified in the assumption of (9).

5. Convergence in direction. We shall first prove that the sequence $\{x^i\}$ converges in direction; in §6 we shall establish the more troublesome property of convergence in length.

THEOREM 1. *Starting with an initial vector* $x^0 \neq 0$, *and a fixed dimension* $s > 1$, *construct the sequence* $\{x^i\}$ *described above. Then*

$$\lim_{i \to \infty} \frac{x^i}{|x^i|} = \frac{y_1}{|y_1|}\,.$$

Proof. From (12), the sequence $\{\nu^i\}$ is a strictly decreasing sequence bounded from below by λ_1. Hence there is a number $\bar{\nu}$ such that

$$\lim_{i\to\infty} \nu^i = \bar{\nu} \geq \lambda_1 .$$

By (12) the smaller root ν_2^i of the polynomial $p_2^i(\lambda)$ is not less than ν^i. Hence

$$p_2^i(\nu^i) = (\nu^i - \mu_0^i)(\nu^i - \mu_1^i) - (t_1^i)^2 \geq 0,$$

$$(t_1^i)^2 \leq (\nu^{i-1} - \nu^i)(\mu_1^i - \nu^i),$$

since $\mu_0^i = \mu(x^i) = \nu^{i-1}$ [see (2) and (11)]. By (1) there is a constant M, independent of i, such that

$$(t_1^i)^2 \leq M(\nu^{i-1} - \nu^i) . \tag{17}$$

In particular,

$$t^i \longrightarrow 0 \text{ as } i \longrightarrow \infty.$$

Recalling (13), put

$$b_j^i = \frac{a_j^i}{|x^i|} |y_j| .$$

Thus

$$\frac{x^i}{|x^i|} = \sum_{j=1}^{r^0} b_j^i \frac{y_j}{|y_j|}, \quad \sum_{j=1}^{r^0} (b_j^i)^2 = 1 . \tag{18}$$

From (14) and the definition of t_1, we have

$$(t_1^i)^2 = \frac{|\xi_1^i|^2}{|x^i|^2} = (b_1^i)^2[p_1^i(\lambda_1)]^2 + \cdots + (b_{r^0}^i)^2[p_1^i(\lambda_{r^0})]^2 .$$

Since the sum of squares on the right tends to 0, each term must do the same. But $p_1^i(\lambda_j) = (\lambda_j - \mu_0^i) = (\lambda_j - \nu^{i-1}) \longrightarrow (\lambda_j - \bar{\nu})$. From the second equation of (18), it follows that for some index l we have

$$\bar{\nu} = \lambda_l, \quad |b_l^i| \longrightarrow 1, \quad b_j^i \longrightarrow 0 \qquad \text{for } j \neq l .$$

(The last two conditions follow from the distinctness of the λ_j.)

We propose to show that $l = 1$. Suppose $l \neq 1$. Then

$$\frac{|y_l|}{|y_1|} \cdot \frac{|b_1^i|}{|b_l^i|} = \frac{|a_1^i|}{|a_l^i|} \longrightarrow 0 . \tag{19}$$

Using (12), we have

$$\lambda_1 < \lambda_l < \nu^i < \nu_j^i \qquad (j = 1, 2, \cdots, s-1) .$$

It follows that $p_j^i(\lambda)$ has the same sign at $\lambda = \lambda_i, \lambda_1, \nu^i$. Furthermore, since by Lemma 3 this polynomial has only real roots, we have

$$|p_j^i(\lambda_1)| > |p_j^i(\lambda_l)| .$$

Thus in formula (15) each term in braces for the coefficients a_1^i and a_l^i is positive, and each term for a_1^i is not smaller than the corresponding term for a_l^i. Hence, for all i, we have

$$\frac{|a_1^{i+1}|}{|a_l^{i+1}|} \geq \frac{|a_1^i|}{|a_l^i|} \qquad (i = 0, 1, 2, \cdots) .$$

By assumption, $a_k^0 = 1$, $k = 1, 2, \cdots, r^0$. We now have a contradiction to (19). Thus $l = 1$.

Since $a_1^i > 0$ by (16), we have $b_1^i > 0$. Hence

$$b_1^i \longrightarrow 1 , \quad b_j^i \longrightarrow 0 \quad \text{for} \quad j \neq 1 .$$

The theorem now follows from the first equation of (18).

6. The main theorem. Before proving the principal result, Theorem 2, we establish two lemmas.

LEMMA 8. *Let $\mathfrak{B}$ be an invariant subspace with lowest characteristic value λ_1 having multiplicity one. Then for $x \neq 0$ in $\mathfrak{B}$, we have*

$$\mu(x) - \lambda_1 \leq \frac{1}{\lambda_2 - \mu(x)} \cdot \frac{|\xi(x)|^2}{|x|^2} \quad \textit{whenever} \quad \mu(x) < \lambda_2 .$$

Proof. (An alternative proof, applicable to normal matrices, is given by H. Wielandt [4].) Write x in the form (4) where $y_1, y_2, \cdots, y_l$ is a complete set of orthonormal characteristic vectors in $\mathfrak{B}$. We let

$$x^* = x - a_1 y_1 , \quad \mu = \mu(x) , \quad \mu^* = \mu(x^*) ,$$

and

$$\xi = \xi(x) \equiv Ax - \mu x \ , \quad \xi^* = \xi(x^*) = Ax^* - \mu^* x^* \ .$$

From $(x^*, y_1) = 0$, we obtain

$$(\xi^*, y_1) = 0.$$

From this and $(\xi^*, x^*) = 0$, we obtain

$$(\xi^*, x) = (\xi^*, x^* + a_1 y_1) = 0.$$

From the definition of ξ^*, we have

$$\begin{aligned} \xi^* &= Ax - a_1 \lambda_1 y_1 - \mu^* x + a_1 \mu^* y_1 \\ &= \xi - (\mu^* - \mu) x + (\mu^* - \lambda_1) a_1 y_1 \ . \end{aligned}$$

Hence

$$0 = (\xi^*, x) = -(\mu^* - \mu)|x|^2 + (\mu^* - \lambda_1) a_1^2 \ .$$

Also

$$0 \leq (\xi^*, \xi^*) = (\xi^*, \xi) = |\xi|^2 + (\mu^* - \lambda_1)(\lambda_1 - \mu) a_1^2$$

from the definition of ξ. Eliminating a_1^2 from the preceding equation, we obtain

$$(\mu - \lambda_1)(\mu^* - \mu)|x|^2 \leq |\xi|^2 \ .$$

Since $x^* \in \mathfrak{B}$ and x^* is orthogonal to y_1, we have

$$\mu^* \geq \lambda_2 \ .$$

Hence, whenever $\mu < \lambda_2$, the inequality of Lemma 8 follows from the second inequality above.

We shall eventually show that the sequence of lengths $|x^i|$ converges. To do this we shall require a bound on the ratio $|p_j^i(\nu^i)|/\tau_j^i$. This is obtained in the next lemma.

LEMMA 9. *Suppose that for all i we have $s < r^i$. Then there exists a constant K, independent of i and j, such that for i sufficiently large we have*

$$|p_j^i(\nu^i)| \leq K(\tau_j^i)^2 \qquad (j = 1, 2, \cdots, s-1) .$$

Proof. By Theorem 1, we have $\mu(x^i) = \nu^{i-1} \longrightarrow \lambda_1$. Hence we may confine ourselves to i's so large that, say,

$$\nu^{i-1} - \lambda_1 < (1/2)(\lambda_2 - \lambda_1) .$$

Consider first $j = 1$. Apply the inequality of Lemma 8 with $x = x^i$, $\mathfrak{B} = \mathfrak{A}^0$. We find that

$$\mu(x^i) - \lambda_1 \leq \frac{(t_1^i)^2}{\lambda_2 - \mu(x^i)} .$$

By (11), we have

$$|\lambda_1 - \mu(x^i)| \geq |\nu^i - \mu(x^i)| = |p_1^i(\nu^i)|$$

and

$$\frac{1}{\lambda_2 - \mu(x^i)} = \frac{1}{\lambda_2 - \nu^{i-1}} < \frac{2}{\lambda_2 - \lambda_1} .$$

Hence

(20) $$|p_1^i(\nu^i)| \leq K(t_1^i)^2 ,$$

as desired.

Let

$$R_j^i = \frac{|p_j^i(\nu^i)|}{(\tau_j^i)^2} \qquad (j = 1, 2, \cdots, s-1) .$$

The inequality (20) may be written $R_1^i \leq K$. We propose to show that for some constant K_1, independent of i and j, we have

(21) $$R_j^i \leq K_1(R_{j-1}^i)^2 \qquad (j = 2, 3, \cdots, s-1) .$$

This, together with (20), will establish the lemma.

For the remainder of the proof we omit the superscript i. Writing $p_j(\lambda)$ as a product of linear factors, we obtain from (12) and (7) the result that

(22) $$|p_j(\nu)| \leq K_2|\nu - \nu_j| \leq K_2(\nu_j - \lambda_1) .$$

In order to estimate the last difference we make use of the minimum characteristic vector z relative to the subspace $\mathfrak{A}_j = (x^i, Ax^i, \cdots, A^{j-1}x^i)$.

We have

$$\mu(z) = \nu_j \quad .$$

By (12) we may apply the inequality of Lemma 8 with $x = z$ and $\mathfrak{B} = \mathfrak{A}^0$. Thus

$$\nu_j - \lambda_1 \leq \frac{1}{\lambda_2 - \nu_j} \cdot \frac{|\xi(z)|^2}{|z|^2} \tag{23}$$

$$\leq K_3 \frac{|\xi(z)|^2}{|z|^2} \quad ,$$

where

$$\xi(z) = Az - \nu_j z \; .$$

The vector $\xi(z)$ is orthogonal to $\mathfrak{A}_j$ and lies in $\mathfrak{A}_{j+1}$. By (3) the vector is a scalar multiple of ξ_j. To determine the scalar we use (8) and (2). We find that

$$\xi(z) = \frac{p_{j-1}(\nu_j)}{\tau^2_{j-1}} \; \xi_j \; .$$

Since $(\nu^i =)\ \nu < \nu_j < \nu_{j-1}$, the above coefficient of ξ_j does not exceed R_{j-1} $(= R^i_{j-1})$ in absolute value, ν_{j-1} being the least root of the polynomial. Also $|z|^2 \geq |x|^2$, by (8). Thus

$$\frac{|\xi(z)|^2}{|z|^2} \leq R^2_{j-1} \frac{|\xi_j|^2}{|x|^2} = R^2_{j-1} \tau^2_j \; . \tag{24}$$

The combination of (22), (23), and (24) yields the desired inequality (21).

We turn to the main theorem. The theorem has an obvious counterpart for the maximum characteristic vector.

THEOREM 2. *Let A be a real symmetric operator on a real vector space of dimension n. Given an initial vector $x^0 \neq 0$ and a fixed dimension s $(1 < s < n)$, construct a sequence of vectors $\{x^i\}$ as follows: let x^{i+1} be the unique minimum characteristic vector relative to the subspace $\mathfrak{A}_s(x^i)$ of the form $x^i + \eta^i$, with $(x^i, \eta^i) = 0$. Then x^i converges to the minimum characteristic vector in $\mathfrak{A}(x^0)$,* *the smallest invariant subspace containing* x^0. *Further, the vector* x^{i+1} *is given by* (10), *and the least root of* $p_s^i(\lambda)$ *converges to* λ_1, *provided* (9) *holds*. (In the event that condition (9) fails, the sequence $\{x^i\}$ is eventually constant, as remarked in the last paragraph of §4.)

Proof. By Theorem 1, it is sufficient to show that the increasing sequence $|x^i|^2$ converges. It is an easy consequence of (10) that

$$|x^{i+1}|^2 = |x^0|^2 \prod_{k=0}^{i} (1 + c^k) ,$$

where

$$c^k = \left[\frac{p_1^k(\nu^k)}{\tau_1^k}\right]^2 + \cdots + \left[\frac{p_{s-1}^k(\nu^k)}{\tau_{s-1}^k}\right]^2$$

By a well-known theorem on infinite products, to prove the desired convergence it is sufficient to verify that $\Sigma_{k=0}^{\infty} c^k$ converges. By Lemma 9, this requirement is reduced to showing that each of the series $\Sigma_{k=0}^{\infty} (\tau_j^k)^2$ converges. For $j = 1$, this series converges by (17). There is a constant K_1 such that $|Ax| \leq K_1 |x|$. Using this inequality and (2), we obtain

$$|\xi_{j+1}^i| \leq K_2 |\xi_j^i| + (t_j^i)^2 |\xi_{j-1}^i| .$$

Hence we have

$$t_{j+1}^i \leq K_2 + t_j^i .$$

It follows that for all i we have

$$t_j^i \leq K_3 \qquad (j = 2, 3, \cdots, s-1) .$$

The convergence of the remaining series now follows from the convergence for $j = 1$. This completes the proof.

REFERENCES

1. M. R. Hestenes and W. Karush, *A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix*, To appear in J. Research Nat. Bur. Standards.

2. L. V. Kantorovitch, *On an effective method of solving extremal problems for quadratic functionals*, C.R. (Doklady) Acad. Sci. URSS (N.S.) 48 (1945), 455-460.

3. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards **45** (1950), 255-282.

4. H. Wielandt, *Ein Einschliessungssatz für charakteristische Wurzeln normaler Matrizen*, Arch. Math. **1** (1948–1949), 348-352.

UNIVERSITY OF CHICAGO AND
NATIONAL BUREAU OF STANDARDS, LOS ANGELES