

A Method of Gradients for the Calculation of the Characteristic Roots and Vectors of a Real Symmetric Matrix¹

Magnus R. Hestenes and William Karush

Let A be a real symmetric matrix, $\mu(x)$ the Rayleigh quotient formed with a vector x , and $\xi(x)$ the gradient vector of $\mu(x)$. The method of gradients consists in an infinite iteration of the operation $\bar{x} = x - \alpha(x)\xi(x)$. The convergence of the procedure is proved for several choices for $\alpha(x)$, and the rate of convergence is studied extensively for one particular $\alpha(x)$. The directions of the vectors in the sequence are seen to tend to that of the characteristic vector belonging to the lowest characteristic value. The method can be used for a numerical determination of all characteristic vectors and values.

I. Introduction

With a real symmetric matrix

$$A = (a_{ij}), \quad a_{ij} = a_{ji}, \quad i, j = 1, 2, \dots, n,$$

is associated the Rayleigh quotient

$$\mu(x) = \frac{(Ax, x)}{(x, x)}, \quad x \neq 0,$$

whose critical points y are the characteristic vectors of A . The gradient of μ has the direction of²

$$\xi(x) = Ax - \mu x. \quad (1)$$

Suppose now, given a vector x , we wish to modify x to obtain a better approximation \bar{x} to a characteristic vector y_{\min} corresponding to the minimum characteristic root $\lambda_{\min} = \min \mu(x)$. It is natural to form

$$\bar{x} = x - \alpha \xi, \quad \alpha > 0, \quad (2)$$

where α may depend upon x . Similarly, to approximate to a characteristic vector y_{\max} corresponding to $\lambda_{\max} = \max \mu(x)$ we form

$$x = x + \alpha \xi, \quad \alpha > 0.$$

In the present paper we describe several convergent iterative methods based upon this gradient process, and investigate the convergence to the characteristic roots and vectors of A . The results apply to an arbitrary, real symmetric matrix. The methods can be phrased to yield directly y_{\min} or y_{\max} , as one wishes. For convenience we direct our attention mainly to y_{\min} .

In prescribing a gradient method one must specify how the number α is to be chosen at each stage x^i of the iteration. It is shown that the vectors x^i converge to the appropriate characteristic vector if α is

any positive constant (independent of i) less than $2/M$, $M = \lambda_{\max} - \lambda_{\min}$. The bulk of the theory is concerned with this "method of fixed α ." If we impose the stricter requirement $\alpha < 1/M$, we obtain in addition that the gradients ξ^i converge in direction (i. e., the unit vectors $\xi^i/|\xi^i|$ converge) to a characteristic vector; in fact the method will be generalized to yield all of the characteristic vectors of A . As would be expected, the nature of the convergence is essentially geometric.

A second "method of optimum α " is treated in which α , which now depends upon x , is selected in a certain "best" way. In this method the approximations x^i converge to a characteristic vector, but the gradients ξ^i fail to converge in direction.³

The well-known method⁴ of forming powers $A^i x$ can be interpreted as a gradient method in which α is chosen as $-1/\mu$. Here convergence, in general, is to a characteristic vector corresponding to the characteristic root of maximum absolute value. We remark that, commonly, the gradients ξ^i converge in direction to a characteristic vector corresponding to a root of next highest absolute value.

The chief virtue of the gradient methods seems to lie in their simplicity. They are not put forth as rapid procedures for a hand computer, but rather as processes that might be adapted to automatic computing machinery. A survey of methods for calculating characteristic roots and vectors of (more general) matrices may be found in Hammersley.⁵

II. Properties of Symmetric Matrices

In this section we collect for reference some well-known facts on symmetric matrices.

³ An extension of the method of optimum α to more general problems has been outlined by L. V. Kantorovich, On an effective method of solving extremal problems for quadratic functionals, C. R. (Doklady) Acad. Sci., USSR (N. S.), **48**, 455-460 (1945). These results are closely related to some unpublished work of M. R. Hestenes.

⁴ See H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psych. **24**, 417-441 and 498-520 (1933). In this paper, Hotelling treats the symmetric matrix. For the extension to nonsymmetric matrices, see A. C. Aitkin, Studies in practical mathematics II. The evaluation of the latent roots and latent vectors of a matrix, Proc. Roy. Soc. Edinburgh [A] **57**, 269-304 (1937).

⁵ The numerical reduction of non-singular matrix pencils, Phil. Mag. **40**, 783-807 (1949).

¹ The preparation of this paper was sponsored (in part) by the Office of Naval Research.

² For convenience we shall refer to ξ as the gradient.

Consider, for the moment, the space of complex vectors $x = (b_1, b_2, \dots, b_n)$ over the field of complex numbers. A number Λ is called a characteristic root (number, value) of an arbitrary matrix B in case there exists a nonnull vector x such that

$$Bx = \Lambda x.$$

The vector x is called a characteristic vector; we shall say that it belongs to the characteristic root Λ .

For a real symmetric matrix A the characteristic roots are real and the characteristic vectors can be chosen to be real. Accordingly we henceforth limit ourselves to the field of real numbers and to real vectors. We let \mathcal{A} denote the space of real vectors x . We use " (x, y) " for inner product, and " $|x|$ " for length in this space. We note a fundamental property of A :

$$(Ax, y) = (x, Ay) \quad (3)$$

Let the distinct characteristic roots of A be denoted by

$$\Lambda_1 < \Lambda_2 < \dots < \Lambda_\rho, \quad \rho \geq 2,$$

(eliminating the trivial case $\rho = 1$). With each Λ_k is associated the linear subspace \mathcal{M}_k of \mathcal{A} , which is the set of all characteristic vectors belonging to Λ_k , together with the null vector. The dimension of \mathcal{M}_k is the order of Λ_k . The subspaces \mathcal{M}_k are mutually orthogonal and span the space \mathcal{A} .

To each nonnull vector x we attach the number

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

The function μ is homogeneous of degree zero;

$$\mu(bx) = \mu(x), \quad b \neq 0.$$

For a characteristic vector y belonging to a characteristic root Λ ,

$$\mu(y) = \Lambda.$$

We have the following well-known relations between μ and the characteristic roots.

$$\Lambda_1 = \min_x \mu(x), \quad \Lambda_\rho = \max_x \mu(x), \quad x \neq 0$$

or, equivalently,

$$\Lambda_1 = \min_x \mu(x), \quad \Lambda_\rho = \max_x \mu(x), \quad |x| = 1.$$

More generally,

$$\Lambda_k = \min_x \mu(x), \quad x \neq 0 \text{ and orthogonal to } \mathcal{M}_1, \dots, \mathcal{M}_{k-1};$$

$$\Lambda_k = \max_x \mu(x), \quad x \neq 0 \text{ and orthogonal to } \mathcal{M}_{k+1}, \dots, \mathcal{M}_\rho.$$

To study the behavior of subspaces of \mathcal{A} under the matrix A it is convenient to think of A as a

linear operator, without regard for coordinate representation. In general, if B is a symmetric linear operator on a finite dimensional space \mathcal{B} , then the statements of the preceding three paragraphs hold (by "symmetry" is meant the property of (3)). Now suppose that \mathcal{B} is an m -dimensional subspace of \mathcal{A} and let B be the operator A with domain restricted to \mathcal{B} . If \mathcal{B} is invariant under A , i. e., if x belonging to \mathcal{B} implies Ax belongs to \mathcal{B} , then B is a symmetric linear operator on \mathcal{B} . As such, B has m characteristic roots (counting multiplicities), and the roots and vectors of B are roots and vectors of A .

The characteristic roots of the matrix A are the solutions of the polynomial equation

$$|\Lambda I - A| = 0.$$

The multiplicity of a solution Λ_k is precisely the dimension of the subspace \mathcal{M}_k , i. e., the order of Λ_k . We remark that if A is regarded as a linear operator, then this equation, formed with any matrix representation of A , yields the characteristic roots.

In our work we shall be dealing always with a fixed initial vector x^0 . Let x^0 have nonnull projections in $\mathcal{M}_{k_1}, \dots, \mathcal{M}_{k_r}$ and only in these subspaces \mathcal{M}_k . We may represent x^0 uniquely in the form

$$x^0 = a_1^0 y_1 + \dots + a_r^0 y_r \quad (4)$$

$$a_i^0 > 0, \quad |y_j| = 1, \quad y_j \in \mathcal{M}_{k_j}, \quad j = 1, 2, \dots, r \geq 2.$$

(We exclude the trivial case $r = 1$.) Let \mathcal{A}_r be the r dimensional subspace spanned by y_1, \dots, y_r ;

$$\mathcal{A}_r = (y_1, y_2, \dots, y_r).$$

Then \mathcal{A}_r is an invariant subspace under A with orthonormal basis (y_1, y_2, \dots, y_r) . The characteristic roots of A relative to this invariant subspace (i. e. the roots of the linear operator A with domain restricted to \mathcal{A}_r) are

$$\lambda_1 < \lambda_2 < \dots < \lambda_r$$

where

$$\lambda_j = \Lambda_{k_j}, \quad j = 1, \dots, r.$$

In the subspace each root λ_j has order one, and has as corresponding characteristic vector, y_j . In the following pages, after having selected an initial vector x^0 , and thus determined the invariant subspace \mathcal{A}_r , we shall be dealing exclusively with vectors in this subspace.

III. Determination of α

This section is discursive; the theorems of the following sections do not logically depend upon it. A direct calculation shows that the gradient of μ is

$$\frac{2}{(x, x)} [Ax - \mu(x)x],$$

and hence has the direction of ξ , given by eq (1). As remarked, we shall refer to ξ as the gradient. Observe the relations

$$(x, \xi) = 0, \quad (Ax, \xi) = (\xi, \xi). \quad (5)$$

The gradient is the direction in which μ locally increases most rapidly. Thus if we form \bar{x} from x as in eq (2), with $\alpha > 0$ sufficiently small, we should expect $\mu(\bar{x})$ to approximate λ_1 more closely than $\mu(x)$. Beginning now with \bar{x} and, say, $\bar{\alpha}$, we obtain from eq (2) a next approximation $\bar{\bar{x}}$, etc. The question then is to specify α systematically.

By direct computation we find

$$\mu(x) - \mu(\bar{x}) = \frac{\alpha(2 - \alpha s)}{1 + \alpha^2 t^2} t^2 \quad (6)$$

where

$$s = \mu(\xi) - \mu(x), \quad t = \frac{|\xi|}{|x|}.$$

That is,

$$\mu(x) - \mu(\bar{x}) = f(\alpha) t^2, \quad f(\alpha) \equiv f(\alpha, x) = \frac{\alpha(2 - \alpha s)}{1 + \alpha^2 t^2}. \quad (7)$$

A natural requirement on $\alpha(x)$ in order to expect convergence of μ to λ_1 is then

$$f(\alpha(x)) > 0.$$

A possible choice is $\alpha = \text{const.}$ with

$$0 < \alpha < \frac{2}{M}, \quad (8)$$

where

$$M = \Lambda_p - \Lambda_1$$

is the spread of the characteristic roots of A . Since $s \leq M$ it is easily seen that $f(\alpha) > 0$. The method of fixed α , stemming from this observation, is treated in sections 5 and 6 and generalized in sections 10 and 11.

Another possibility for α is that number $\gamma = \gamma(x)$ which maximizes $f(\alpha)$. From $f'(\gamma) = 0$ we obtain

$$t^2 \gamma^2 + s \gamma - 1 = 0, \quad (9)$$

from which

$$\gamma = \frac{2}{s + \sqrt{s^2 + 4t^2}} > 0. \quad (10)$$

Computation shows that $f(\gamma) = \gamma$. Hence

$$\mu(x) - \mu(\bar{x}) = \gamma t^2, \quad \bar{x} = x - \gamma \xi.$$

The method of optimum α , based on this approach, is treated in section 7. We note the following formula:

$$\gamma = \frac{1}{\mu(\xi) - \mu(\bar{x})}. \quad (11)$$

To verify this we write eq (10) as $\gamma(\gamma t^2 + s) = 1$, and note that the quantity in parentheses equals $\mu(\xi) - \mu(x)$.

Remarks on the power method, i. e., forming successive iterates $A^i x$, will be made in section 8. In that section and in the later section 12 modifications of the methods of fixed and optimum α will be suggested.

Before proceeding to any specific gradient method we formulate some general results in the next section.

IV. General Gradient Method

We suppose given a real-valued function γ such that $\gamma(x)$ is defined whenever $x \neq 0$ and x is not a characteristic vector, i. e., whenever $x \neq 0$ and $\xi \neq 0$. We require that

$$f(\gamma(x)) > 0.$$

(See, e. g., the particular function γ determined by eq (10). Beginning with a vector x^0 , expressed in the form (4), we construct the sequence of vectors $\{x^i\}$, $i = 0, 1, 2, \dots$, according to

$$x^{i+1} = x^i - \gamma^i \xi^i, \quad \text{where } \xi^i = \xi(x^i), \quad \gamma^i = \gamma(x^i). \quad (12)$$

This is equivalent to

$$x^{i+1} = (1 + \gamma^i \mu^i) x^i - \gamma^i A x^i, \quad \text{where } \mu^i = \mu(x^i). \quad (13)$$

We must assure ourselves that $\{x^i\}$ is well defined. We note first that if $x^i \neq 0$ and $\xi^i \neq 0$, then x^{i+1} is defined and not null. For, under the hypotheses, x^{i+1} is defined and by eq (12),

$$|x^{i+1}|^2 = |x^i|^2 + \gamma^{i2} |\xi^i|^2.$$

We thereby distinguish two types of sequences. The first, a trivial case, is one such that for some first integer k ,

$$\xi^k = 0.$$

We terminate the induction at $i = k$; the sequence $\{x^i\}$, $i = 0, 1, \dots, k$, is finite with the last vector x^k a characteristic vector. The second type is characterized as a sequence that is not of the first type. In this case $\{x^i\}$ and $\{\xi^i\}$ are well-defined infinite sequences and

$$\xi^i \neq 0, \quad i = 0, 1, 2, \dots$$

To enable us to formulate statements that hold simultaneously for finite and infinite sequences we make the following agreements: (1) if $\{p^i\}$ is an arbitrary finite sequence, then $\lim_{i \rightarrow \infty} p^i$ denotes the last member of the sequence; (2) any statement or formula involving an index is to be limited to meaningful values of the index.

From eq (7),

$$\mu^i - \mu^{i+1} = f(\gamma^i) t^{i2} > 0, \quad t^i = \frac{|\xi^i|}{|x^i|}. \quad (14)$$

Since μ^i is bounded from below, it follows that there is a number ν such that

$$\lim_{i \rightarrow \infty} \mu^i = \nu. \quad (15)$$

The vectors x^i and ξ^i lie within the invariant subspace $\mathcal{A} = (y_1, y_2, \dots, y_r)$; their expansions in terms of the characteristic vectors y_j are readily determined from eq (1) and (12). We find

$$\left. \begin{aligned} x^i &= a_1^i y_1 + a_2^i y_2 + \dots + a_r^i y_r, \\ a_1^i &= \{1 + \gamma^{i-1}(\mu^{i-1} - \lambda_1)\} a_1^{i-1} \\ a_j^i &= \{1 + \gamma^{i-1}(\mu^{i-1} - \lambda_j)\} a_j^{i-1}, \quad j=1, 2, \dots, r, \end{aligned} \right\} \quad (16)$$

and

$$\xi^i = (\lambda_1 - \mu^i) a_1^i y_1 + \dots + (\lambda_r - \mu^i) a_r^i y_r. \quad (17)$$

Thus

$$\left. \begin{aligned} |x^i|^2 &= a_1^{i2} + a_2^{i2} + \dots + a_r^{i2}, \\ |\xi^i|^2 &= (\lambda_1 - \mu^i)^2 a_1^{i2} + \dots + (\lambda_r - \mu^i)^2 a_r^{i2}. \end{aligned} \right\} \quad (18)$$

Theorem 4.1. Suppose γ^i is such that

$$0 < K_1 \leq f(\gamma^i) \text{ and } 0 < \gamma^i,$$

for some constant K_1 . Then

$$\lim_{i \rightarrow \infty} \frac{x^i}{|x^i|} = y_1 \text{ and } \lim_{i \rightarrow \infty} \mu^i = \lambda_1.$$

We limit ourselves to the case when the sequence $\{x^i\}$ is infinite; by appropriate simplification the proof below applies as well to the finite sequence.

From eq (14)

$$t^{i2} \leq \frac{1}{K_1} (\mu^i - \mu^{i+1}). \quad (19)$$

Since $\mu^i \rightarrow \nu$, it follows that

$$t^i \rightarrow 0.$$

Let

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

Then by eq (18)

$$t^{i2} = \frac{|\xi^i|^2}{|x^i|^2} = \sum_{j=1}^r (\lambda_j - \mu^i)^2 b_j^{i2}, \quad \sum_{j=1}^r b_j^{i2} = 1. \quad (20)$$

Consequently

$$\lim_{i \rightarrow \infty} (\lambda_j - \mu^i)^2 b_j^{i2} = 0, \quad j=1, 2, \dots, r.$$

Since $(\lambda_j - \mu^i) \rightarrow (\lambda_j - \nu)$ it follows from the last equation of (20) that for some specific value l of the index j ,

$$\mu^i \rightarrow \lambda_l, \quad b_l^i \rightarrow 1, \quad b_j^i \rightarrow 0 \text{ for } j \neq l.$$

We wish to show that $l=1$. Suppose $l > 1$. Then

$$\frac{b_l^i}{b_l^i} = \frac{|a_l^i|}{|a_l^i|} \rightarrow 0.$$

From the monotonicity of μ^i , we have $\mu^i > \lambda_l > \lambda_1$. From eq (16)

$$\frac{|a_l^{i+1}|}{|a_l^{i+1}|} = \frac{|1 + \gamma^i(\mu^i - \lambda_l)|}{|1 + \gamma^i(\mu^i - \lambda_l)|} \cdot \frac{|a_l^i|}{|a_l^i|} \geq \frac{|a_l^i|}{|a_l^i|}.$$

Hence

$$\frac{|a_l^i|}{|a_l^i|} \geq \frac{|a_l^0|}{|a_l^0|} > 0 \text{ for } i=0, 1, 2, \dots$$

This contradiction shows that $l=1$. We now have

$$\mu^i \rightarrow \lambda_1, \quad \frac{a_1^i}{|x^i|} \rightarrow 1, \quad \frac{a_j^i}{|x^i|} \rightarrow 0, \quad j \neq 1,$$

since

$$a_1^i \geq a_1^0 > 0.$$

The theorem now follows from the first equation of (16).

Theorem 4.2. If, in addition to the hypotheses of theorem 4.1, the sequence γ^i satisfies the condition

$$\gamma^i \leq K_2, \quad K_2 = \text{const.},$$

then the conclusion of that theorem may be strengthened to

$$\lim_{i \rightarrow \infty} x^i = L y_1,$$

where L is a positive constant.

By theorem 4.1 it is sufficient to show that $|x^i| \rightarrow L$. From eq (12),

$$|x^{i+1}|^2 = |x^0|^2 [1 + (\gamma^0 t^0)^2] [1 + (\gamma^1 t^1)^2] \dots [1 + (\gamma^i t^i)^2].$$

It is well known that the product on the right converges if the series $\sum_0^\infty (\gamma^i t^i)^2$ does. From eq (19) the series $\sum t^{i2}$ converges, and hence the first series does, by the boundedness of γ^i . This completes the proof.

V. The Method of Fixed α

In this gradient method we choose $\gamma(x)$ to be a constant α satisfying eq (8); thus,

$$\alpha = \frac{\beta}{M}, \quad 0 < \beta < 2. \quad (21)$$

It follows that

$$f(\alpha) \geq \frac{\alpha(2-\beta)}{1+\alpha^2 t^2}, \quad x \neq 0.$$

From eq (1) we derive

$$|\xi| \leq 2K|x|,$$

where K is the bound of the matrix A . Hence t is bounded from above. Thus $f(\alpha)$ is bounded from below by a positive constant. Theorem 4.2 is now applicable.

Theorem 5.1. Let γ^i be chosen as a constant α satisfying eq (21). Then

$$\lim_{i \rightarrow \infty} x^i = Ly_1, \quad \lim_{i \rightarrow \infty} \mu^i = \lambda_1.$$

For later use we note that

$$\frac{|x^i|}{a_1^i} \rightarrow 1, \quad (22)$$

which was established in the proof of theorem 4.1.

We propose to show now that under a strengthening of condition (21) the gradients ξ^i converge in direction to the second characteristic vector y_2 . The new condition is

$$\alpha = \frac{\beta}{M}, \quad 0 < \beta < 1. \quad (23)$$

Lemma 5.1. Under the condition (23) the sequence $\{x^i\}$ is infinite and

$$a_j^i > 0, \quad \xi^i \neq 0, \quad j = 1, 2, \dots, r; i = 0, 1, 2, \dots$$

From eq (16)

$$a_j^i = \{1 + \alpha(\mu^{i-1} - \lambda_j)\} a_j^{i-1}. \quad (24)$$

By eq (23), the expression in braces is positive. Hence, from $a_j^0 > 0$, we have $a_j^i > 0$. From the second equation (18) and the assumption $r \geq 2$, it follows that $\xi^i \neq 0$. The last inequality assures that $\{x^i\}$ is infinite.

We introduce the following notation

$$\delta_j = 1 - \alpha(\lambda_j - \lambda_1) = 1 - \beta \frac{\lambda_j - \lambda_1}{M}, \quad j = 1, 2, \dots, r. \quad (25)$$

Thus, under the condition (23),

$$0 < \delta_r < \delta_{r-1} < \dots < \delta_2 < \delta_1 = 1.$$

Lemma 5.2 Under the condition (23),

$$\lim_{i \rightarrow \infty} \frac{a_j^{i+1}}{a_j^i} = \delta_j,$$

$$\lim_{i \rightarrow \infty} \frac{a_k^i}{a_j^i} = 0,$$

for $j < k$.

The first equation follows from (24) and the fact that $\mu^i \rightarrow \lambda_1$. To prove the second equation notice that for $j < k$,

$$\frac{a_k^{i+1}}{a_j^{i+1}} \div \frac{a_k^i}{a_j^i} \rightarrow \frac{\delta_k}{\delta_j} \text{ and } 0 < \frac{\delta_k}{\delta_j} < 1.$$

Hence, for any $\epsilon > 0$ there is a constant K and index i_0 such that

$$\frac{a_k^i}{a_j^i} < K \left(\frac{\delta_k}{\delta_j} + \epsilon \right)^i, \quad i \geq i_0.$$

Choose ϵ so that the number in parentheses is less than 1. This completes the proof.

Lemma 5.3 Under the condition (23),

$$\lim_{i \rightarrow \infty} \frac{(\mu^i - \lambda_1) a_1^{i2}}{a_1^{i2}} = (\lambda_2 - \lambda_1), \quad \lim_{i \rightarrow \infty} \frac{|\xi^i|}{a_2^i} = \lambda_2 - \lambda_1.$$

From the orthogonality relation $(x^i, \xi^i) = 0$ and eq (16) and (17) we obtain

$$(\lambda_1 - \mu^i) a_1^{i2} + (\lambda_2 - \mu^i) a_2^{i2} + \dots + (\lambda_r - \mu^i) a_r^{i2} = 0.$$

Divide by a_1^{i2} and take the limit as $i \rightarrow \infty$. The first conclusion of the lemma then follows from lemma 5.2. Using this result we establish the second conclusion in a similar way by means of eq (18).

Theorem 5.2. Let the constant α be chosen to satisfy (23). Then

$$\lim_{i \rightarrow \infty} \frac{\xi^i}{|\xi^i|} = y_2, \quad \lim_{i \rightarrow \infty} \mu(\xi^i) = \lambda_2.$$

Divide both sides of eq (17) by a_2^i . Using lemmas 5.2 and 5.3 we obtain

$$\lim_{i \rightarrow \infty} \frac{\xi^i}{a_2^i} = (\lambda_2 - \lambda_1) y_2.$$

It follows that

$$\frac{\xi^i}{|\xi^i|} = \frac{\xi^i}{a_2^i} \cdot \frac{a_2^i}{|\xi^i|} \rightarrow y_2,$$

as desired. Also,

$$\mu(\xi_1) = \mu \left(\frac{\xi^i}{|\xi^i|} \right) \rightarrow \mu(y_2) = \lambda_2.$$

This completes the proof.

We remark that for the proofs of theorems 5.1 and 5.2 the conditions (21) and (23) could have been relaxed by using $\lambda_r - \lambda_1$ in place of $M = \lambda_r - \lambda_1$.

We conclude this section with a reformulation of theorems 5.1 and 5.2 for convergence to the two highest characteristic vectors y_r, y_{r-1} .

Theorem 5.3. Let the constant α satisfy eq (21). Define the sequence $\{x^i\}$ by the recursion formula

$$x^{i+1} = x^i + \alpha \xi^i.$$

Then

$$\lim_{i \rightarrow \infty} x^i = L y_r, \quad \lim_{i \rightarrow \infty} \mu(x^i) = \lambda_r, \quad L > 0.$$

If, further, α satisfies eq (23,) then

$$\lim_{i \rightarrow \infty} \frac{\xi^i}{|\xi^i|} = -y_{r-1}, \quad \lim_{i \rightarrow \infty} \mu(\xi^i) = \lambda_{r-1}.$$

This result is obtained by replacing A by $B = -A$ in theorems 5.1 and 5.2, and noticing that the characteristic roots become

$$-\lambda_r < -\lambda_{r-1} < \dots < -\lambda_1$$

with corresponding characteristic vectors y_r, y_{r-1}, \dots, y_1 .

VI. Rate of Convergence for Fixed α

In this section we investigate the rate of convergence of the sequences $\{x^i\}$, $\{\xi^i\}$ and related sequences; in the rest of this section we assume eq (23) holds. For convenience we introduce the so-called "ratio" of a sequence of real numbers as a measure of speed of convergence to 0, and develop some elementary properties of sequences with ratios. This notion will also be useful in our later generalization of the method of fixed α .

Let us agree that a sequence $\{b^i\}$ of real numbers will be called positive (negative) if $b^i > 0 (< 0)$ for $i \geq i_0$, i_0 fixed. We shall understand that the sequence is monotonic in case it is monotonic for $i \geq i_0$.

Definition. A sequence $\{b^i\}$ of real numbers will be said to have the ratio κ in case

$$\lim_{i \rightarrow \infty} \frac{b^{i+1}}{b^i} = \kappa \text{ with } \kappa > 0.$$

Necessarily a sequence with a ratio is either positive or negative.

The reason for this definition lies in the next lemma, which is essentially a rephrasing of Lemma 5.2.

Lemma 6.1. The sequences $\{a_j^i\}$, $\{a_k^i/a_j^i\}$ ($j, k = 1, \dots, r$) have ratios $\delta_j, \delta_k/\delta_j$, respectively.

Sequences with ratios resemble geometric progressions. If $\{b^i\}$ has ratio κ , then for an arbitrary $\epsilon > 0$, with $\epsilon < \kappa$, there are numbers $T_1 > 0$, $T_2 > 0$ such that

$$T_1(\kappa - \epsilon)^i \leq |b^i| \leq T_2(\kappa + \epsilon)^i.$$

Thus, if $\kappa < 1$, then b^i tends to zero more rapidly, eventually, than a geometric progression with ratio $\kappa + \epsilon$, and more slowly than the progression with ratio $\kappa - \epsilon$. Accordingly the ratio of a sequence is a measure of the speed of convergence to zero. We shall express our results on rate of convergence in these terms. Notice that when $\kappa < 1$ the sequence $\{b^i\}$ is monotonic; decreasing if the sequence is positive, increasing if negative.

Suppose $\{b^i\}, \{c^i\}$ have ratios κ_1, κ_2 respectively. Then clearly $\{b^i c^i\}$ has ratio $\kappa_1 \kappa_2$ and $\{b^i/c^i\}$ has ratio κ_1/κ_2 . A convergent sequence $\{d^i\}$ with a nonzero limit has ratio 1. Consequently each of the sequences $\{b^i d^i\}, \{b^i/d^i\}$ has ratio κ_1 . The following lemma will be used frequently.

Lemma 6.2. Let $\{b^i\}$ have ratio κ . Suppose $\{d^i\}$ is a sequence such that d^i/b^i has a nonzero limit. Then $\{d^i\}$ has ratio κ .

For, d^i/b^i has ratio 1, and the product $d^i = (d^i/b^i) b^i$ has ratio κ .

If $\kappa \neq 1$, then the sequence of differences $b^i - b^{i+1}$ also has ratio κ . For this we apply the last lemma with $d^i = b^i - b^{i+1}$.

Lemma 6.3. Suppose $\{b^i\}$ has ratio $\delta < 1$. Let

$$B^i = \sum_{j=i}^{\infty} b^j.$$

Then

$$\lim_{i \rightarrow \infty} \frac{B^i}{b^i} = \frac{1}{1 - \delta}.$$

In particular, $\{B^i\}$ has ratio δ .

Clearly we may assume that $\{b^i\}$ is positive. Select an arbitrary ϵ with $0 < \epsilon < \delta$, $\delta + \epsilon < 1$. Then for i sufficiently large, $(\delta - \epsilon)b^i \leq b^{i+1} \leq (\delta + \epsilon)b^i$, $(\delta - \epsilon)^2 b^i \leq b^{i+2} \leq (\delta + \epsilon)^2 b^i$, and in general

$$(\delta - \epsilon)^j b^i \leq b^{i+j} \leq (\delta + \epsilon)^j b^i, \quad j = 0, 1, 2, \dots$$

Hence, summing,

$$\frac{1}{1 - \delta + \epsilon} \leq \frac{B^i}{b^i} \leq \frac{1}{1 - \delta - \epsilon}$$

$$\frac{1}{1 - \delta + \epsilon} \leq \lim_{i \rightarrow \infty} \frac{B^i}{b^i} \leq \overline{\lim}_{i \rightarrow \infty} \frac{B^i}{b^i} \leq \frac{1}{1 - \delta - \epsilon}.$$

Since ϵ is arbitrary, the desired conclusion follows.

Lemma 6.4. Suppose $\{b^i\}$ is a positive sequence with ratio $\delta < 1$. Let

$$Q^i = \prod_{j=i}^{\infty} (1 + b^j).$$

Then

$$\lim_{i \rightarrow \infty} \frac{Q^i - 1}{b^i} = \frac{1}{1 - \delta}.$$

The existence of the infinite product Q^i follows from the fact that the series $\sum_{j=i}^{\infty} b^j$ of positive terms converges. Now

$$\begin{aligned} \prod_{i=1}^{i+l} (1 + b^i) &= 1 + (b^1 + b^{i+1} + \dots + b^{i+l}) \\ &\quad + (b^1 b^{i+1} + \dots) + \dots + (b^1 b^{i+1} \dots b^{i+l}). \end{aligned}$$

When i is sufficiently large for $b^j, j \geq i$, to be positive,

$$\sum_{i=1}^{i+l} b^i \leq \prod_{i=1}^{i+l} (1+b^i) - 1 \leq B^i + (B^i)^2 + \dots$$

Allowing $l \rightarrow \infty$, we obtain

$$B^i \leq Q^i - 1 \leq \frac{B^i}{1-B^i}.$$

Dividing by b^i , letting $i \rightarrow \infty$, and using lemma 6.3 we obtain the desired result.

We are now ready to state the main results of the section.

Theorem 6.1. Let eq (23) hold. Then $\mu^i - \lambda_1$ is a positive sequence with ratio δ_2^2 .

This result is an immediate consequence of lemmas 5.3, 6.1, and 6.2.

From this theorem it follows that $\mu^i - \mu^{i+1}$ is also a sequence with ratio δ_2^2 . Likewise, t^{i^2} is a sequence with the same ratio, as we see from lemma 5.3.

Lemma 6.5. The positive sequence $L - a_1^i$ has ratio δ_2^2 .

The number L is that of theorem 5.1. We have $a_1^i \rightarrow L$. From eq (16),

$$a_1^i = a_1^0 \prod_{k=0}^{i-1} \{1 + \alpha(\mu^k - \lambda_1)\}, \quad L = a_1^0 \prod_{k=0}^{\infty} \{1 + \alpha(\mu^k - \lambda_1)\}.$$

Thus

$$L - a_1^i = a_1^i \left[\prod_{k=i}^{\infty} \{1 + \alpha(\mu^k - \lambda_1)\} - 1 \right].$$

Since $\mu^k - \lambda_1$ has ratio δ_2^2 , the conclusion follows from lemma 6.4.

Theorem 6.2. Let eq (23) hold. Then the sequence $|x^i - Ly_1|$ has ratio δ_2 .

From (16)

$$x - Ly_1 = (a_1^i - L)y_1 + a_2^i y_2 + \dots + a_r^i y_r. \quad (26)$$

Thus

$$|x - Ly_1|^2 = (a_1^i - L)^2 + a_2^{i^2} + \dots + a_r^{i^2}.$$

Hence the left side is a sum of terms having ratios $\delta_2^4, \delta_2^2, \delta_2^2, \dots, \delta_2^2$. Since each of these numbers, other than the second, is strictly smaller than the second, the theorem is established.

Concerning the convergence of $|x^i|$ to L , we remark that $L - |x^i|$ is a sequence of ratio δ_2^2 . The proof is like that of lemma 6.5, with a_1^i replaced by $|x^i|^2$.

Information on the convergence of each component of x^i can be derived from eq (26). We write $x^i = (b_1^i, b_2^i, \dots, b_n^i)$, $y_1 = (c_1, c_2, \dots, c_n)$. Now for fixed j , $b_j^i - Lc_j$ is a linear combination of $a_1^i - L$, a_2^i , \dots , a_r^i , these sequences having respective ratios $\delta_2^2, \delta_2^2, \delta_2^2, \dots, \delta_2^2$. If the j^{th} component of y_2 , the coefficient of a_2^i , is not 0, then $b_j^i - Lc_j$ has ratio δ_2 ; accordingly the difference sequence $b_j^i - b_j^{i+1}$, which is numerically available, has ratio δ_2 . If the j^{th} component of y_2 is 0, then, in general, $b_j^i - Lc_j$ will

have a smaller ratio (assuming the sequence is not identically 0).

For results on the convergence of sequences associated with the second characteristic vector y_2 , we appeal to the forthcoming theorems of section 11. These results may be specialized to the sequence $\{\xi^i\}$ by recognizing that in the notation of that section,

$$x^i = z_1^i, \quad \xi^i = z_2^i, \quad \mu_1^i = \mu(x^i), \quad \mu_2^i = \mu(\xi^i).$$

Also, we agree to interpret δ_{r+1} as 0.

Concerning the rate of convergence of $|\xi^i|$ to 0 we have, from lemma 5.3, that $\{|\xi^i|\}$ is a sequence with ratio δ_2 .

Theorem 6.3. Let eq (23) hold and set $\eta^i = \xi^i / |\xi^i|$.

(1) If $\delta_2^2 > \delta_3$, then $\{\lambda_2 - \mu(\xi^i)\}$ is a positive, monotonic sequence with ratio δ_2^2 , and $\{|\eta^i - y_2|\}$ is a sequence with ratio δ_2 .

(2) If $\delta_2^2 < \delta_3$, then $\{\mu(\xi^i) - \lambda_2\}$ is a positive, monotonic sequence with ratio $(\delta_3/\delta_2)^2$, and $\{|\eta^i - y_2|\}$ is a sequence with ratio δ_3/δ_2 .

(3) If $\delta_2^2 = \delta_3$, then $\mu(\xi^i) - \lambda_2 = O(\delta_2^2)$ and $|\eta^i - y_2| = O(\delta_2)$.

These statements are interpretations of theorems 11.1, 11.2, and their corollaries for $j=2$. From theorem 11.4 we derive the inequality

$$\lambda_2 - \mu(\xi^i) \leq \mu(x^i) - \lambda_1, \quad i=0, 1, 2, \dots,$$

the equality holding just in case $r=2$. From this and the above theorem we obtain the following interesting corollary.

Corollary. If $\delta_2^2 > \delta_3$, then for sufficiently large i ,

$$|\lambda_2 - \mu(\xi^i)| \leq |\lambda_1 - \mu(x^i)|.$$

More generally, when $\lambda_2 - \mu(\xi^i)$ is nonnegative, and i is large, then $\mu(\xi^i)$ is closer to λ_2 than $\mu(x^i)$ is to λ_1 . On the other hand, for case (2) of theorem 6.3, the sequence $\{\mu(\xi^i) - \lambda_2\}$ has ratio $(\delta_3/\delta_2)^2 > \delta_2^2$. But δ_2^2 is the ratio of $\{\mu^i - \lambda_2\}$. Hence in this instance μ^i approaches λ_1 more rapidly than $\mu(\xi^i)$ approaches λ_2 .

Notice that for $r=2$, the inequality $\delta_2^2 > \delta_3$ is automatic. If $r > 2$, then, from theorem 11.3, the inequality $\delta_2^2 > \delta_3$ holds whenever

$$\lambda_3 - \lambda_2 > \lambda_2 - \lambda_1.$$

VII. The Method of Optimum α

In sections 10 and 11 the method of fixed α will be extended to obtain sequences that converge to all of the characteristic vectors y_1, y_2, \dots, y_r . Before proceeding to this generalization we wish to develop first the method of optimum α , so that a comparison of the two methods (and others) may be made at the most advantageous point. The optimum pro-

cedure does not appear to generalize to a larger number of characteristic vectors as simply as the method of fixed α .

According to eq (10) the iteration scheme is now given by eq (12) with

$$\gamma^i = \frac{2}{s^i + \sqrt{s^{i2} + 4t^{i2}}} > 0. \quad (27)$$

The numbers t^i are bounded (see remarks preceding theorem 5.1); so are the numbers s^i . Thus $f(\gamma^i) = \gamma^i$ is bounded from below by a positive constant. Accordingly, theorem 4.1 yields the following lemma.

Lemma 7.1. Let γ^i be given by eq (31). Then the conclusion of theorem 4.1 holds.

In order to apply theorem 4.2 it is necessary to establish the boundedness of the γ^i . This is done in the next lemma.

Lemma 7.2. Let γ^i be given by eq (31). Then

$$\overline{\lim}_{i \rightarrow \infty} \gamma^i \leq \frac{1}{\lambda_2 - \lambda_1}.$$

Since $\mu^i \rightarrow \lambda_1$, it is sufficient, by eq (11), to show that

$$\overline{\lim}_{i \rightarrow \infty} \mu(\xi^i) \geq \lambda_2.$$

Let T denote the inferior limit on the left. Let

$$\eta^i = \frac{\xi^i}{|\xi^i|}.$$

Then $T = \lim_{i \rightarrow \infty} \mu(\eta^i)$. Select a subsequence $\{\eta^k\}$ such that $T = \overline{\lim}_{i \rightarrow \infty} \mu(\eta^k)$. Select a further subsequence $\{\eta^l\}$ that converges to, say, η . From $(\eta^i, x^i) = 0$ and lemma 7.1 we obtain

$$0 = \lim_{i \rightarrow \infty} \left(\eta^i, \frac{x^i}{|x^i|} \right) = (\eta, y_1).$$

Since η belongs to \mathcal{A}_r (space spanned by y_1, \dots, y_r) it follows that $\mu(\eta) \geq \lambda_2$. Hence

$$T = \lim_{i \rightarrow \infty} \mu(\eta^i) = \mu(\eta) \geq \lambda_2,$$

as desired.

Theorem 4.2 now yields the principal result.

Theorem 7.1. Let γ^i be given by eq (27). Then

$$\lim_{i \rightarrow \infty} x^i = L y_1, \quad \lim_{i \rightarrow \infty} \mu^i \rightarrow \lambda_1.$$

That the vectors $\xi^i/|\xi^i|$ do not converge in the method of optimum α is a consequence of the next theorem.

Theorem 7.2. Under the hypotheses of theorem 7.1,

$$(\xi^i, \xi^{i+1}) = 0.$$

Proof. We have,

$$\begin{aligned} \xi^{i+1} &= A x^{i+1} - \mu^{i+1} x^{i+1} \\ &= A x^i - \gamma^i A \xi^i - \mu^{i+1} x^i + \mu^{i+1} \gamma^i \xi^i. \end{aligned}$$

Hence

$$\begin{aligned} (\xi^i, \xi^{i+1}) &= (\xi^i, A x^i) - \gamma^i (A \xi^i, \xi^i) + \mu^{i+1} \gamma^i (\xi^i, \xi^i) \\ &= \{1 - \gamma^i (\mu(\xi^i) - \mu^{i+1})\} (\xi^i, \xi^i) \\ &= 0 \end{aligned}$$

by eq (11).

The last theorem indicates that the coefficients a_i^i of eq (16) converge in an irregular fashion as $i \rightarrow \infty$. Hence a determination of the rate of convergence such as that given in section 5 for the method of fixed α is not to be expected here.

We conclude the section with a modification of theorem 7.1 for convergence to the highest characteristic vector y_r .

Theorem 7.3. Let the sequence $\{x^i\}$ be defined by

$$x^{i+1} = x^i + \gamma^i \xi^i,$$

with

$$\gamma^i = \frac{2}{-s^i + \sqrt{s^{i2} + 4t^{i2}}}.$$

Then

$$\lim_{i \rightarrow \infty} x^i \rightarrow L y_r, \quad L > 0.$$

This result is obtained by applying theorem 7.1 to the matrix $B = -A$.

VIII. Comparison of Methods

Each step in the preceding methods of gradients consists of forming the subspace spanned by x and ξ , i. e., by x and Ax , and choosing in that two-space⁶ a next approximation \bar{x} . In the method of optimum α , the vector \bar{x} is chosen so that $\mu(\bar{x})$ is a minimum in the subspace; in the method of fixed α the particular linear combination $x - \alpha \xi$ is chosen. Superficially, then, it would seem that the first method should give more rapid convergence to λ_1 and y_1 than the second. But the minimizing procedure can only be recognized as an advantage when the vector x^i is common to the two methods. This certainly holds for the initial vector x^0 , but not necessarily beyond this stage; hence the relative merits of the two methods are not evident. It seems reasonable that for low order matrices the procedure of finding an optimum α might be advantageous, while for matrices of higher order the fixed α procedure might be superior. In practice a combination of the two methods would be in order.

The method of fixed α has the advantage of computational simplicity over the alternative method. On the other hand the former requires some advance

⁶ We shall remark later on the possibility of dealing with subspaces of dimension higher than two. Cf. a forthcoming paper by W. Karush, An iterative method for finding characteristic vectors of a symmetric matrix. There the method of optimum α is generalized to subspace of arbitrary (fixed) dimension.

information on the characteristic roots in order to estimate an allowable α ; computation by the method of optimum α demands no such knowledge.

The method of fixed α is a smooth method. The vectors $\xi^i/|\xi^i|$ converge; by a relatively simple extension of the procedure we can obtain convergence to all the characteristic vectors of \mathcal{A} (see section 10). Furthermore the convergence is geometric in nature. The method of optimum α has none of these advantages. The successive gradients are orthogonal, and the coefficients a_2^i, a_3^i, \dots of eq (16) do not tend to 0 smoothly.

Sequences $\{b^i\}$ with ratio $\delta < 1$ provide examples of "linear" convergence. If the ratio $b^{i+1}/(b^i)^2$ has a finite limit, then the sequence converges in a "quadratic" fashion. In the method of fixed α we have convergence of a linear type. It is possible to procure quadratic convergence by modifying the method, but the price paid would be the solution of a system of linear equations at each step of the iteration.

A combination of the two gradient methods can be used to advantage when the characteristic roots λ_1, λ_2 are close and relatively isolated from the other roots. Suppose we begin with the fixed α iteration and continue up to a certain stage. The last estimate x and its gradient ξ may not be good individual approximations to y_1 and y_2 but the pair may provide an excellent approximation to the plane of y_1 and y_2 . One application of the optimum α procedure at this stage should then yield a good approximation \bar{x} to y_1 . The iteration may be started again with \bar{x} . This technique has worked well in several numerical examples.

Since $t = |\xi|/|x| \rightarrow 0$, it is clear that fewer and fewer significant figures will be retained in ξ as the iteration proceeds in a numerical calculation. Nevertheless, ξ and $\mu(\xi)$ supply useful approximations to y_2 and λ_2 . For example, after y_1 and λ_1 have been obtained, the approximation ξ can be used as the initial vector x^0 in the calculation of y_2 and λ_2 by a new use of the method of gradients.

The above remarks apply equally well to the highest vector y_r . One of the advantages of the gradient methods is that they can be applied directly to either end of the scale of characteristic roots.

A single step (2) of the gradient procedure carries one from

$$x = a_1 y_1 + a_2 y_2 + \dots + a_r y_r$$

to

$$\bar{x} = \{1 + \alpha(\mu - \lambda_1)\} a_1 y_1 + \{1 + \alpha(\mu - \lambda_2)\} a_2 y_2 + \dots$$

Thus if any characteristic number λ_j is known, then orthogonality of \bar{x} to y_j can be achieved by choosing⁷

$$\alpha = \frac{1}{\lambda_j - \mu}.$$

⁷ The same result can be obtained by taking $\bar{x} = \mathcal{A}x - \lambda_j x$; the advantage of the above procedure is that it fits into the iteration scheme (12).

This observation may be used to maintain (approximate) orthogonality which may be theoretically assured but may be gradually lost in computation due to round-off errors. For example, in beginning with ξ as an initial vector for obtaining y_2 (see second paragraph above), one must maintain orthogonality to y_1 during the iteration process. To do this, the procedure (12) may be interspersed with steps in which $\gamma^i = (\lambda_1 - \mu^i)^{-1}$. Similarly, one may maintain orthogonality to any number of characteristic vectors, or induce orthogonality to such vectors, provided only that one knows the corresponding characteristic roots (relatively accurately). The desired ends are achieved within the framework of the scheme (12) by particular choices of γ^i . Notice that if $r-1$ of the r characteristic roots λ_j are known then the characteristic vector belonging to the remaining root can be obtained in $r-1$ steps.

It was remarked after theorem 6.1 that $\Delta\mu^i = \mu^{i+1} - \mu^i$ and t^i are sequences with ratio δ_2^2 . Hence observation of the ratios $\Delta\mu^{i+1}/\Delta\mu^i$ or t^{i+1}/t^i can lead to an estimate of δ_2^2 . Since $\delta_2 = 1 - \alpha(\lambda_2 - \lambda_1)$, and α is known, another estimate of λ_2 may be made, alternative to the estimate $\mu(\xi)$. Furthermore, if a sequence c^i has a limit C , and the difference $c^i - C$ is a sequence with a ratio $\delta < 1$, then a transformation for speeding the convergence of c^i is

$$\begin{aligned} \bar{c}^i &= c^i + \frac{\Delta c^i}{1 - \delta^i}, \quad \left(\delta^i = \frac{\Delta c^{i+1}}{\Delta c^i} \right) \\ &= \frac{c^{i+1} - \delta^i c^i}{1 - \delta^i} \\ &\approx \frac{c^{i+1} - \delta c^i}{1 - \delta}. \end{aligned}$$

With an estimate of δ_2 , this formula may be used to improve μ^i and the components of x^i (see remarks following theorem 6.2).

A certain degree of flexibility may be introduced in the method of fixed α by replacing the constant α by a variable α^i ,

$$\alpha^i = \frac{\beta^i}{M}, \quad (28)$$

which satisfies

$$0 < \beta_1 \leq \beta^i \leq \beta_2 < 2, \quad (\beta_1, \beta_2 \text{ constants}). \quad (29)$$

The conclusion of theorem 5.1 remains valid, since this choice of α^i does not violate the assumptions of theorem 4.2.

The convergence of the gradients as in theorem 5.2 may also be justified provided one imposes the additional restrictions

$$\beta_2 < 1 \text{ and } \frac{\beta_2}{\beta_1} < \frac{\lambda_3 - \lambda_1}{\lambda_2 - \lambda_1}. \quad (30)$$

Under these conditions statements on rate of con-

vergence, appropriately formulated, can be established.

For comparison purposes we state here, without proof, some properties of the "power" method. The iteration formula is

$$x^{i+1} = Ax^i.$$

Suppose now that there is one characteristic root of maximum absolute value; denote it by λ_k . Then

$$\lim_{i \rightarrow \infty} \frac{x^{2i}}{|x^{2i}|} \rightarrow y_k, \quad \lim_{i \rightarrow \infty} \mu(x^i) \rightarrow \lambda_k.$$

(In the first limit the sequence with odd indices tends to $-y_k$ if $\lambda_k < 0$, and to y_k if $\lambda_k > 0$.) It was pointed out in the introduction that the power method is a gradient method with

$$\gamma^i = -\frac{1}{\mu^i}.$$

In this form the iteration formula is

$$z^{i+1} = \frac{1}{\mu^i} Ax^i, \quad \mu^i = \mu(z^i),$$

$$\text{i. e., } z^{i+1} = (\mu^0 \mu^1 \dots \mu^i)^{-1} x^{i+1},$$

and the convergence result is

$$\lim_{i \rightarrow \infty} z^i = K y_k, \quad K = \text{const.}$$

(provided μ^i is never 0). Thus μ^i , as well as $|x^i|$, may be used as a normalization factor.

The power method leads directly to an estimate of λ_k , and hence to an upper bound $2|\lambda_k|$ of M . Thus a rough application of this method may be used as a preliminary step in the method of fixed α in order to determine an allowable value of the constant α .

Suppose now that λ_l is a characteristic root of next highest absolute value. Concerning rate of convergence we remark that $\mu^i - \lambda_k$ is a sequence of ratio $(\lambda_l/\lambda_k)^2$. If $-\lambda_l$ is not a characteristic root, then $\xi^{2i} = \xi(x^{2i})$ tends in direction to

$$\frac{(\lambda_l - \lambda_k)}{|\lambda_l - \lambda_k|} y_l;$$

the sequence with odd indices tends to the negative of this vector if $\lambda_l < 0$ otherwise to the same vector.

We conclude this section with a remark on the method of relaxations.⁸ Any method which begins with a vector x and applies a sequence of transformations to it so that the resulting sequence of vectors x^i converges to a characteristic vector, must induce the gradients $\xi(x^i)$ to tend to the null vector. Thus any artful modification of a vector x to a new vector \bar{x} which brings the gradient closer to the null vector

would be a plausible procedure in a numerical calculation. The full skill and intuition of the computer may come into play in varying the vector at any stage to produce a better approximation. One systematic procedure is to modify a single component of x in such a way as to make a single component of the gradient vanish. Other devices may be used; the best trick at any point depends upon the information available at that stage. This is the flexible approach of the method of relaxations. Clearly it is not easily adaptable to automatic computation.

IX. Invariant Subspace \mathcal{A}_r

Before turning to the generalization of the method of fixed α we shall develop here some properties of the invariant subspace \mathcal{A}_r . We will encounter certain polynomials $p_j(\lambda)$ which have been considered by Lanczos,⁹ and some of the results of this section will overlap his work.

In place of the notation x^0 we shall use z_1 to denote a given vector with expansion

$$z_1 = a_1 y_1 + a_2 y_2 + \dots + a_r y_r \quad (31)$$

as in eq (4). The space $\mathcal{A}_r = (y_1 \dots y_r)$ may be characterized as in the following lemma.

Lemma 9.1. The vectors $z_1, Az_1, \dots, A^{r-1}z_1$ span the space \mathcal{A}_r .

Let \mathcal{B} be the space spanned by the vectors $A^k z_1$, $k=0, 1, 2, \dots$. Then this space is the smallest invariant subspace which contains the vector z_1 . Since z_1 is in \mathcal{A}_r , and \mathcal{A}_r is invariant, it follows that the relation $\mathcal{B} \subset \mathcal{A}_r$ holds.

Suppose that \mathcal{B} were a proper subset of the invariant space \mathcal{A}_r . Then \mathcal{B} would be spanned by a proper subset of the characteristic vectors y_1, y_2, \dots, y_r of \mathcal{A}_r . Consequently the vector z_1 , being in \mathcal{B} , would be orthogonal to at least one of these characteristic vectors, say y_k . Thus, from eq (31), $a_k = (z_1, y_k) = 0$. This contradicts the assumption $a_k > 0$ (see eq 4).

We now have $\mathcal{B} = \mathcal{A}_r$. Consequently, $\dim \mathcal{B} = \dim \mathcal{A}_r = r$. The lemma now follows from the fact that $\dim \mathcal{B}$ is the least integer h such that the vectors $z_1, Az_1, \dots, A^{h-1}z_1$ span \mathcal{B} .

We now define the vectors z_j recursively as follows. Let

$$\left. \begin{aligned} t_j &= \frac{|z_j|}{|z_{j-1}|} \\ \mu_j &= \mu(z_j). \end{aligned} \right\} \quad (32)$$

Then

$$\left. \begin{aligned} z_2 &= Az_1 - \mu_1 z_1 \\ z_3 &= Az_2 - \mu_2 z_2 - t_2^2 z_1 \\ &\dots \dots \dots \\ z_{j+1} &= Az_j - \mu_j z_j - t_j^2 z_{j-1}, \quad j=2, 3, \dots, r. \end{aligned} \right\} \quad (33)$$

⁸ Cooper, J. L. B., The Solution of Natural Frequency Equations by Relaxation Methods, Qtrly. Appl. Math., vol. 6, p. 179 (1948).

⁹ "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators," J. Research NBS 45, 255 (1950) RP 2133

The vectors $z_1, z_2, \dots, z_r, z_{r+1}$ are well-defined and

$$z_j \neq 0, \quad j=1, 2, \dots, r.$$

For suppose, z_1, \dots, z_k are defined and none is null, $k < r$. Then by eq (33) z_{k+1} is defined and is a linear combination of $A^k z_1, A^{k-1} z_1, \dots, A z_1, z_1$ with the coefficient of $A^k z_1$ equal to one. Further, if $k < r-1$, then $z_{k+1} \neq 0$ by lemma 9.1. This establishes the assertion.

By induction it is easy to prove that

$$\begin{aligned} (z_j, z_j) &= (z_j, A z_{j-1}), & j=2, \dots, r+1, \\ (z_j, z_k) &= 0, & j \neq k, \quad j, k=1, 2, \dots, r+1. \end{aligned} \quad (34)$$

Lemma 9.2. Let \mathcal{A}_j be the space spanned by $z_1, A z_1, \dots, A^{j-1} z_1$, i. e.,

$$\mathcal{A}_j = (z_1, A z_1, \dots, A^{j-1} z_1), \quad j=1, \dots, r.$$

Then (z_1, \dots, z_j) is an orthogonal basis for \mathcal{A}_j . Also,

$$z_{r+1} = 0.$$

By lemma (9.1), \mathcal{A}_j has dimension j . From the above, the nonnull vectors (z_1, \dots, z_j) are orthogonal, and lie in \mathcal{A}_j ; hence they span \mathcal{A}_j .

The final conclusion of the lemma follows from the fact that z_{r+1} is orthogonal to z_1, \dots, z_r and hence to \mathcal{A}_r . But from the invariance of \mathcal{A}_r , z_{r+1} belongs to \mathcal{A}_r . Thus $z_{r+1} = 0$.

Given an arbitrary vector z_1 the integer r of eq (31) can be determined from the above lemma; r is that integer such that z_{r+1} is the first z_j that vanishes. Notice that except for $j=r$, \mathcal{A}_j is not the space spanned by y_1, \dots, y_j .

We consider now the expansions of the z_j in terms of the y_j . From eq (31) and (33),

$$z_2 = (\lambda_1 - \mu_1) a_1 y_1 + (\lambda_2 - \mu_1) a_2 y_2 + \dots + (\lambda_r - \mu_1) a_r y_r.$$

Thus, if we put $p_1(\lambda) = \lambda - \mu_1$ we have

$$z_2 = p_1(\lambda_1) a_1 y_1 + p_1(\lambda_2) a_2 y_2 + \dots + p_1(\lambda_r) a_r y_r.$$

Again from eq (33),

$$\begin{aligned} z_3 &= \{p_1(\lambda_1)(\lambda_1 - \mu_2) - t_2^2\} a_1 y_1 + \dots \\ &+ \{p_1(\lambda_r)(\lambda_r - \mu_2) - t_2^2\} a_r y_r. \end{aligned}$$

Hence, if we put $p_2(\lambda) = p_1(\lambda)(\lambda - \mu_2) - t_2^2$ we have

$$z_3 = p_2(\lambda_1) a_1 y_1 + p_2(\lambda_2) a_2 y_2 + \dots + p_2(\lambda_r) a_r y_r.$$

In general, if we define the polynomials $p_j(\lambda)$ by

$$\begin{aligned} p_0(\lambda) &= 1, \quad p_1(\lambda) = \lambda - \mu_1, \quad p_j(\lambda) = p_{j-1}(\lambda)(\lambda - \mu_j) \\ &- t_j^2 p_{j-2}(\lambda), \quad j=2, \dots, r \end{aligned} \quad (35)$$

we have the expansions

$$\begin{aligned} z_{j+1} &= p_j(\lambda_1) a_1 y_1 + p_j(\lambda_2) a_2 y_2 + \dots + p_j(\lambda_r) a_r y_r, \\ j &= 0, 1, \dots, r. \end{aligned} \quad (36)$$

From eq (36) with $j=r+1$, and lemma 9.2 we see that the roots of

$$p_r(\lambda) = 0$$

are the characteristic numbers $\lambda_1, \lambda_2, \dots, \lambda_r$. This is a special case of the forthcoming theorem 9.2.

In lemma 9.2 we defined the subspaces \mathcal{A}_j of \mathcal{A}_r . In order to develop some properties of these subspaces we state without proof several well known properties of symmetric linear operators.

Let \mathcal{B} denote a linear space of real vectors. Let B denote a symmetric linear operator on \mathcal{B} . Consider an l -dimensional subspace \mathcal{B}' of \mathcal{B} ; let π be the projection operator into \mathcal{B}' . The operator πB with domain restricted to \mathcal{B}' is a symmetric operator on \mathcal{B}' ; by the characteristic roots and vectors of B relative to \mathcal{B}' we mean the corresponding quantities of this operator. An alternative characterization is given by the property that v is a characteristic vector (relative to \mathcal{B}') belonging to the characteristic root κ if and only if v belongs to \mathcal{B}' , $v \neq 0$, and

$$(x, Bv) = \kappa(x, v), \quad x \text{ in } \mathcal{B}'.$$

Also, the smallest and greatest roots relative to \mathcal{B}' are respectively the minimum and maximum of $\mu(x)$ on \mathcal{B}' . If \mathcal{B}' is an invariant subspace of \mathcal{B} then $\pi B = B$ on \mathcal{B}' and the characteristic roots and vectors relative to \mathcal{B}' are such in the whole space \mathcal{B} ; this is the case with \mathcal{A}_r as a subspace of \mathcal{A} .

If u_1, u_2, \dots, u_l is an orthonormal basis for \mathcal{B}' then the operator πB relative to this basis has the $l \times l$ matrix representation

$$B' = (A u_j, u_k), \quad j, k=1, 2, \dots, l.$$

The characteristic roots relative to \mathcal{B}' are the roots of

$$|\lambda I - B'| = 0.$$

We quote a well-known result. (The first conclusion of the theorem below holds without the restriction that the κ_j be distinct.)

Theorem 9.1. Let B be a symmetric linear operator on an m -dimensional real linear space \mathcal{B} ; let \mathcal{B}' be an l -dimensional subspace of \mathcal{B} . Suppose that B has distinct characteristic values $\kappa_1 < \kappa_2 < \dots < \kappa_m$ with corresponding characteristic vectors v_1, v_2, \dots, v_m . Let the characteristic roots and vectors of B relative to \mathcal{B}' be respectively $\kappa'_1 \leq \kappa'_2 \leq \dots \leq \kappa'_l$ and v'_1, v'_2, \dots, v'_l . Then

$$\kappa_1 \leq \kappa'_1, \quad \kappa_2 \leq \kappa'_2, \quad \dots, \quad \kappa_l \leq \kappa'_l.$$

The equality holds for all indices if and only if

$$v_1=v'_1, \quad v_2=v'_2, \quad \dots, \quad v_l=v'_l;$$

in this case \mathcal{B}' is an invariant subspace.¹⁰

We return to the subspaces \mathcal{A}_j . The characteristic roots and vectors of A relative to \mathcal{A}_j are those of the symmetric operator $B_j=\pi_j A$ on \mathcal{A}_j , where π_j is the projection on \mathcal{A}_j . Since z_1, \dots, z_j is an orthogonal basis of \mathcal{A}_j the operator B_j has the matrix representation

$$\left(\frac{(Az_l, z_m)}{|z_l||z_m|} \right), \quad l, m=1, 2, \dots, j.$$

We calculate this matrix.

Lemma 9.3. For a given j ,

$$\begin{aligned} (Az_j, z_k) &= |z_j|^2 && \text{when } k=j-1 \\ &= \mu_j |z_j|^2 && \text{when } k=j \\ &= |z_{j+1}|^2 && \text{when } k=j+1 \\ &= 0 && \text{otherwise.} \end{aligned}$$

We remark that for $j=1$, the first equation is to be omitted. The first and third equations follow from the first equation of (34). The second follows from the definition of μ_j . The last is a consequence of the orthogonality relations of (34) and the fact that Az_j is a linear combination of z_{j+1} , z_j and z_{j-1} by (33).

It follows from this lemma that B_j has the matrix representation

$$B_j = \begin{bmatrix} \mu_1 & t_2 & 0 & & & \\ t_2 & \mu_2 & t_3 & & & \\ 0 & t_3 & \mu_3 & & & \\ & & & \ddots & & \\ & & & & \ddots & 0 \\ 0 & & & & & t_j \\ & & & & 0 & t_j & \mu_j \end{bmatrix}$$

Lemma 9.4. Let $q_0(\lambda)=1$ and $q_j(\lambda)$ be the characteristic polynomial of B_j ,

$$q_j(\lambda) = |\lambda I - B_j|, \quad j=1, 2, \dots, r.$$

Then

$$p_j(\lambda) = q_j(\lambda), \quad j=0, 1, \dots, r.$$

Using the matrix representation above it is easy to see by direct calculation that

$$q_1(\lambda) = \lambda - \mu_1, \quad q_j(\lambda) = (\lambda - \mu_j) q_{j-1}(\lambda) - t_j^2 q_{j-2}(\lambda).$$

The lemma now follows from the fact that these recursion relations are identical with eq (35).

Theorem 9.2. The roots $\nu_1 \leq \nu_2 \leq \dots \leq \nu_j$ of

$$p_j(\lambda) = 0, \quad j=1, 2, \dots, r,$$

are the characteristic roots of A relative to the subspace \mathcal{A}_j . Furthermore,

$$\lambda_1 \leq \nu_1, \quad \lambda_2 \leq \nu_2, \quad \dots, \quad \lambda_j \leq \nu_j,$$

the equality holding for all indices if and only if $j=r$.

This theorem follows from theorem 9.1 with $\mathcal{B}=\mathcal{A}$, $B=A$. The last statement follows from the fact that \mathcal{A}_j , $j < r$, is not invariant.

X. Extension of Method of Fixed α

We return now to the sequence $\{x^i\}$ of eq (12) with $\gamma^i = \alpha = \text{const}$. As in theorem 5.2 we impose the condition (23)¹¹ on α . We rename x^i as z_i^i and apply the results of the previous section to each z_i^i . The expansion (31) becomes (16),

$$z_1^i = a_1^i y_1 + a_2^i y_2 + \dots + a_r^i y_r.$$

By lemma 5.1 each a_i^i is positive; accordingly, the space \mathcal{A}_i associated with z_1^i is independent of i and coincides with our original space \mathcal{A} . We recognize z_2^i as our original ξ^i .

The polynomials of eq (35) now depend upon the index i ; we denote them by $p_i^i(\lambda)$. For each i they are defined in terms of

$$\mu_i^i = \mu(z_i^i), \quad j=1, 2, \dots, r,$$

$$t_i^i = \frac{|z_i^i|}{|z_{i-1}^i|}, \quad j=2, \dots, r.$$

It is convenient to introduce

$$\bar{p}_0(\lambda) = 1,$$

$$\bar{p}_j(\lambda) = (\lambda - \lambda_j)(\lambda - \lambda_{j-1}) \dots (\lambda - \lambda_1),$$

$$j=1, 2, \dots, r.$$

The vectors $(z_1^i, z_2^i, \dots, z_r^i)$ comprise a sequence of orthogonal bases for the space \mathcal{A} . We show that, when normalized, this sequence converges to the fixed basis y_1, \dots, y_r .

Theorem 10.1. Let the constant α satisfy (23) and let the initial vector $z_1^0 = x^0$ be given by (4). Determine the infinite sequences $\{z_1^i\}, \{z_2^i\}, \dots, \{z_r^i\}$ by (33) and

$$z_1^{i+1} = z_1^i - \alpha z_2^i.$$

¹⁰ This result follows from the minimax principle for characteristic values. Cf. Courant and Hilbert "Methoden der Mathematischen Physik," I, 2d ed, Berlin (1931), pp. 27-29.

¹¹ In this section and the next we assume without further comment that this condition holds.

$$\beta_1 + \beta_2 + \dots + \beta_k = -1$$

$$\bar{p}_1(\lambda_1)\beta_1 + \bar{p}_1(\lambda_2)\beta_2 + \dots + \bar{p}_1(\lambda_k)\beta_k = -\bar{p}_1(\lambda_{k+1})$$

$$\dots$$

$$\bar{p}_{k-1}(\lambda_1)\beta_1 + \bar{p}_{k-1}(\lambda_2)\beta_2 + \dots + \bar{p}_{k-1}(\lambda_k)\beta_k = -\bar{p}_{k-1}(\lambda_{k+1}).$$

In the second equation, $\bar{p}_1(\lambda)$ is a polynomial of degree 1; by adding λ_1 times the first equation to the second we may reduce the second to

$$\lambda_1\beta_1 + \lambda_2\beta_2 + \dots + \lambda_k\beta_k = -\lambda_{k+1}.$$

In the third equation, $\bar{p}_3(\lambda)$ is a second-degree polynomial; a linear combination of the preceding two rows will reduce this equation to

$$\lambda_1^2\beta_1 + \lambda_2^2\beta_2 + \dots + \lambda_k^2\beta_k = -\lambda_{k+1}^2.$$

Continuing in this way, we arrive at the equivalent system,

$$\beta_1 + \beta_2 + \dots + \beta_k = -1$$

$$\lambda_1\beta_1 + \lambda_2\beta_2 + \dots + \lambda_k\beta_k = -\lambda_{k+1}$$

$$\dots$$

$$\lambda_1^{k-1}\beta_1 + \lambda_2^{k-1}\beta_2 + \dots + \lambda_k^{k-1}\beta_k = -\lambda_{k+1}^{k-1}.$$

The solutions β_j are immediately expressible in terms of the function

$$V_k(b_1, b_2, \dots, b_k) \equiv \prod_{\substack{l < m \\ l, m = 1, 2, \dots, k}} (b_m - b_l).$$

From the identity

$$V_k(b_1, b_2, \dots, b_k) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ b_1 & b_2 & \dots & b_k \\ \dots & \dots & \dots & \dots \\ b_1^{k-1} & b_2^{k-1} & \dots & b_k^{k-1} \end{vmatrix}$$

it follows that

$$\beta_1 = (-1)^k \frac{V_k(\lambda_2, \dots, \lambda_k, \lambda_{k+1})}{V_k(\lambda_1, \lambda_2, \dots, \lambda_k)}$$

$$\beta_2 = (-1)^{k-1} \frac{V_k(\lambda_1, \lambda_3, \dots, \lambda_k, \lambda_{k+1})}{V_k(\lambda_1, \lambda_2, \dots, \lambda_k)}$$

$$\dots$$

$$\beta_k = -\frac{V_k(\lambda_1, \dots, \lambda_{k-1}, \lambda_{k+1})}{V_k(\lambda_1, \lambda_2, \dots, \lambda_k)}.$$

The L_j may now be computed from eq (40).

XI. Rate of Convergence

We shall establish some results on rate of convergence for the sequences of the previous section. The terminology is that of section 6. Recall the definition of $\delta_j, j=1, 2, \dots, r$, in eq (25). It is convenient to set

$$\delta_{r+1} = 0.$$

We remark that for the lengths $|z_j^i|$ we have at once, from the corollary to theorem 10.1 and lemma 6.2, that $\{|z_j^i|\}$ is a sequence with ratio δ_j .

Theorem 11.1. For a fixed integer $j, j=2, 3, \dots, r$, consider the sequence $\{\mu(z_j^i)\}$. (1) If $\delta_j^2 > \delta_{j-1}\delta_{j+1}$, then $\{\lambda_j - \mu(z_j^i)\}$ is a positive, monotonic sequence with ratio $(\delta_j/\delta_{j-1})^2$. (2) If $\delta_j^2 < \delta_{j-1}\delta_{j+1}$, then $\{\mu(z_j^i) - \lambda_j\}$ is a positive, monotonic sequence with ratio $(\delta_{j+1}/\delta_j)^2$.

We first assume $j < r$. Then from eq (36)

$$\begin{aligned} \mu(z_j^i) - \lambda_j &= \frac{(z_j^i, Az_j^i) - \lambda_j |z_j^i|^2}{|z_j^i|^2} \\ &= \frac{1}{|z_j^i|^2} \{ (\lambda_1 - \lambda_j) [p_{j-1}^i(\lambda_1) a_1^i]^2 + \dots \\ &\quad + (\lambda_{j-1} - \lambda_j) [p_{j-1}^i(\lambda_{j-1}) a_{j-1}^i]^2 \\ &\quad + (\lambda_{j+1} - \lambda_j) [p_{j-1}^i(\lambda_{j+1}) a_{j+1}^i]^2 + \dots \} \\ &= \frac{a_j^{i^2}}{|z_j^i|^2} = \left\{ b_1^i \frac{a_j^{i^2}}{a_1^{i^2}} + \dots + b_{j-1}^i \frac{a_j^{i^2}}{a_{j-1}^{i^2}} \right. \\ &\quad \left. + b_{j+1}^i \frac{a_j^{i^2}}{a_{j+1}^{i^2}} + \dots \right\} \end{aligned}$$

where, by the corollary to theorem 10.1, the b_k^i have finite limits.

Now suppose $\delta_j^2 > \delta_{j-1}\delta_{j+1}$. We write

$$\begin{aligned} (\mu(z_j^i) - \lambda_j) \frac{a_{j-1}^{i^2}}{a_j^{i^2}} &= \frac{a_j^{i^2}}{|z_j^i|^2} \left\{ b_1^i \frac{a_{j-1}^{i^2}}{a_1^{i^2}} + \dots + b_{j-2}^i \frac{a_{j-1}^{i^2}}{a_{j-2}^{i^2}} \right. \\ &\quad \left. + b_{j-1}^i + b_{j+1}^i \left(\frac{a_{j-1}^i a_{j+1}^i}{a_j^{i^2}} \right)^2 + \dots \right\}. \end{aligned} \quad (41)$$

The factors of b_k^i (except for b_{j-1}^i) may be recognized as sequences of ratio less than 1, by lemma 6.1; hence these sequences tend to 0. Hence, by the above-mentioned corollary, the left side of eq (41) has the limit

$$-(\lambda_j - \lambda_{j-1}) \frac{\bar{p}_{j-2}^2(\lambda_j)}{\bar{p}_{j-2}^2(\lambda_{j-1})} < 0.$$

Using lemma 6.2 we deduce the first conclusion of the theorem.

Suppose next that $\delta_j^2 < \delta_{j-1}\delta_{j+1}$. Then we write

$$(\mu(z_j^i) - \lambda_j) \frac{a_j^{i^2}}{a_{j+1}^{i^2}} = \frac{a_j^{i^2}}{|z_j^i|^2} \left\{ b_1^i \left(\frac{a_j^{i^2}}{a_1^i a_{j+1}^i} \right)^2 + \dots \right. \\ \left. + b_{j-1}^i \left(\frac{a_j^{i^2}}{a_{j-1}^i a_{j+1}^i} \right)^2 + b_{j+1}^i + b_{j+2}^i \frac{a_{j+2}^{i^2}}{a_{j+1}^{i^2}} + \dots \right\}.$$

All terms tend to 0 except b_{j+1}^i . Hence the left side has the limit

$$(\lambda_{j+1} - \lambda_j) \frac{\bar{p}_{j-1}^2(\lambda_{j+1})}{\bar{p}_{j-1}^2(\lambda_j)} > 0.$$

From this we deduce the second conclusion of the theorem.

There remains the case $j=r$. In this instance only case (1) of the theorem is possible. The proof is identical with that given above for that case, except that no terms b_{j+1}^i and higher appear.

Theorem 11.1 leaves open the possibility of the equality $\delta_j^2 = \delta_{j-1} \delta_{j+1}$. The likelihood of this condition holding in a numerical instance is very slight, but the problem is subtler than the instances of inequality and thereby has theoretical interest. We establish the following corollary.

Corollary. Suppose $\delta_j^2 = \delta_{j-1} \delta_{j+1}$. Then

$$\mu(z_j^i) - \lambda_j = O(\delta_j^2 / \delta_{j-1}^2).$$

To prove the corollary we require the following lemma.

Lemma 11.1. If $\delta_j^2 = \delta_{j-1} \delta_{j+1}$ then the sequence

$$\left\{ \frac{a_{j-1}^i a_{j+1}^i}{a_j^{i^2}} \right\}$$

has a finite positive limit.

We remark that although the sequence in question has ratio 1 it does not follow from this that it has a limit.

By eq (16)

$$\frac{a_{j-1}^i a_{j+1}^i}{a_j^{i^2}} = \frac{a_{j-1}^0 a_{j+1}^0}{a_j^{0^2}} \cdot \prod_{k=0}^{i-1} \frac{\{1 - \alpha(\lambda_{j-1} - \mu^k)\} \{1 - \alpha(\lambda_{j+1} - \mu^k)\}}{\{1 - \alpha(\lambda_j - \mu^k)\}^2} \quad (42)$$

where $\mu^k = \mu(z_j^k)$. The general term of the product may be written

$$\frac{[\delta_{j-1} + \alpha(\mu^k - \lambda_1)] [\delta_{j+1} + \alpha(\mu^k - \lambda_1)]}{[\delta_j + \alpha(\mu^k - \lambda_1)]^2}.$$

Letting

$$\beta^k = \mu^k - \lambda_1, \quad K_1 = \frac{\alpha}{\delta_{j-1}}, \quad K_2 = \frac{\alpha}{\delta_{j+1}}, \quad K_3 = \frac{\alpha}{\delta_j}$$

the general term becomes

$$\frac{(1 + K_1 \beta^k)(1 + K_2 \beta^k)}{(1 + K_3 \beta^k)^2} = 1 + b^k$$

where

$$b^k = \beta^k(K_1 + K_2 - 2K_3) + \text{higher terms in } \beta^k.$$

The coefficient of β^k is positive. For,

$$K_1 + K_2 - 2K_3 = \frac{\alpha}{\delta_j} \left(\frac{\delta_j}{\delta_{j-1}} + \frac{\delta_j}{\delta_{j+1}} - 2 \right) \\ = \frac{\alpha}{\delta_j} \left(\rho + \frac{1}{\rho} - 2 \right) \quad \left(\rho = \frac{\delta_j}{\delta_{j-1}} = \frac{\delta_{j+1}}{\delta_j} < 1 \right) \\ = \frac{\alpha \rho}{\delta_j} (\rho - 1)^2 > 0.$$

It follows that for sufficiently large k , b^k is positive. Furthermore, by theorem 6.1 and lemma 6.2, b^k is a sequence with ratio δ_j^2 . Hence $\sum_k b^k$ converges;

thus the product $\prod(1 + b^k)$, and hence eq (42), converges to a positive limit as $i \rightarrow \infty$.

We turn to the proof of the corollary. As in the proof of the theorem we form eq (41). We see that all terms tend to 0 except the terms in b_{j-1}^i and b_{j+1}^i . The first of these has a finite limit, and, by lemma 11.1, the second does, too. This completes the proof.

Notice that for the corollary we cannot deduce that $\mu(z_j^i) - \lambda_j$ is a sequence with a ratio; this is because we do not know that the right side of eq (41) has a nonzero limit. In fact, it seems likely that this limit is zero.

Theorem 11.2. Let j be a fixed integer, $j=2, 3, \dots, r$. Set $u_j^i = \frac{z_j^i}{|z_j^i|}$.

(1) If $\delta_j^2 > \delta_{j-1} \delta_{j+1}$, then $|u_j^i - y_j|$ has ratio δ_j / δ_{j-1} .

(2) If $\delta_j^2 < \delta_{j-1} \delta_{j+1}$, then $|u_j^i - y_j|$ has ratio δ_{j+1} / δ_j .

From the corollary to theorem 10.1 we see that

$$\frac{p_{j-1}^i(\lambda_j) a_j^i}{|z_j^i|} - 1$$

tends to 0. We wish an estimate of the rate of convergence.

By eq (36),

$$|z_j^i|^2 / [p_{j-1}^i(\lambda_j) a_j^i]^2 - 1 = 1 / [p_{j-1}^i(\lambda_j)]^2 \{ [p_{j-1}^i(\lambda_1) a_1^i / a_j^i]^2 + \dots + [p_{j-1}^i(\lambda_{j-1}) a_{j-1}^i / a_j^i]^2 + [p_{j-1}^i(\lambda_{j+1}) a_{j+1}^i / a_j^i]^2 + \dots \}.$$

Denote by c^{i2} the first term on the left. Under case (1) of the theorem we multiply both sides by

$$d_j^i = (a_{j-1}^i / a_j^i)^2.$$

As in the proof of theorem 11.1 we find that $(c^{i^2}-1)d_j^i$ has a nonzero limit. Noting that $1-(1/c^i) = (c^{i^2}-1)/c^i(1+c^i)$ and using $c^i \rightarrow 1$, we find that

$$\left(1 - \frac{1}{c^i}\right) d_j^i$$

has a nonzero limit. (43)

On the other hand, under case (2) we find, by a similar argument that

$$\left(1 - \frac{1}{c^i}\right) d_{j+1}^i \quad (44)$$

has a nonzero limit.

We have

$$\begin{aligned} |w_j^i - y_j|^2 = & \frac{1}{|z_j^i|^2} \left\{ [p_{j-1}^i(\lambda_1) a_1^i]^2 + \dots \right. \\ & + [p_{j-1}^i(\lambda_{j-1}) a_{j-1}^i]^2 + \left(\frac{1}{c^i} - 1 \right)^2 \\ & \left. + [p_{j-1}^i(\lambda_{j+1}) a_{j+1}^i]^2 + \dots \right\}. \end{aligned}$$

Consider condition (1) of the theorem. Multiply both sides of the above equation by d_j^i . From eq (43) and the corollary to theorem 10.1 we obtain that

$$|w_j^i - y_j|^2 d_j^i \quad (45)$$

has a nonzero limit.

This proves the first part of the theorem. Consider next condition (2). This time we multiply both sides of the above equation by d_{j+1}^i . Using eq (44) we find

$$|w_j^i - y_j| d_{j+1}^i$$

has a nonzero limit. This completes the proof of the theorem.

Corollary. Suppose $\delta_j^2 = \delta_{j-1} \delta_{j+1}$. Then

$$|w_j^i - y_j| = O(\delta_j / \delta_{j-1}).$$

Using lemma 11.1 we show as in the proof of the above theorem that the left side of eq (43) has a finite limit. Proceeding further as in the theorem we show that the left side of eq (45) also has a finite limit. This completes the argument.

The following result is of interest in connection with the preceding theorems.

Theorem 11.3. Suppose $r \geq 3$. Let j be an integer with $2 \leq j \leq r-1$. If

$$\lambda_{j+1} - \lambda_j > \lambda_j - \lambda_{j-1},$$

then

$$\delta_j^2 > \delta_{j-1} \delta_{j+1}.$$

Let $q_j = (\lambda_j - \lambda_1)/M$. From the definition of the δ_j ,

$$\begin{aligned} \delta_j^2 - \delta_{j-1} \delta_{j+1} &= (1 - \beta q_j)^2 - (1 - \beta q_{j-1})(1 - \beta q_{j+1}) \\ &= \beta^2 [q_j^2 - q_{j-1} q_{j+1}] + \beta D, \end{aligned}$$

where

$$\begin{aligned} D &= (q_{j+1} - q_j) - (q_j - q_{j-1}) \\ &= \frac{1}{M} [(\lambda_{j+1} - \lambda_j) - (\lambda_j - \lambda_{j-1})]. \end{aligned}$$

We find that

$$\begin{aligned} \delta_j^2 - \delta_{j-1} \delta_{j+1} &= \beta^2 [(q_j - q_{j-1})^2 - q_{j-1} D] + \beta D \\ &= \beta^2 (q_j - q_{j-1})^2 + (\beta - \beta^2 q_{j-1}) D. \end{aligned}$$

Since $0 < \beta < 1$, and $0 \leq q_{j-1} < 1$, it follows that the coefficient of D is positive. Hence the left side is positive if D is. This concludes the proof.

Notice that if $j=r$, the conclusion of the theorem holds, for, $\delta_{r+1}=0$. Also, if $r=2$, then $j=2$ is the only meaningful value of j and the conclusion likewise holds.

Theorem 9.2 enables us to prove an interesting result for the errors $\lambda_j - \mu_j^i$.

Theorem 11.4. Let

$$e_j^i = \lambda_j - \mu_j^i, \quad j=1, 2, \dots, r; \quad i=0, 1, 2, \dots$$

Then for each i ,

$$e_1^i + e_2^i + \dots + e_r^i < 0, \quad j=1, 2, \dots, r-1,$$

$$e_1^i + e_2^i + \dots + e_r^i = 0.$$

Consider a fixed j , $j=1, 2, \dots, r$, and a fixed i . As in theorem 9.2 let v_1, \dots, v_j denote the roots of $p_j^i(\lambda)=0$. From the definition of the j th degree polynomial $p_j^i(\lambda)$ we can verify directly that the coefficient of λ^{j-1} is $-(\mu_1^i + \dots + \mu_j^i)$. Hence

$$\sum_{k=1}^j v_k = \sum_{k=1}^j \mu_k^i.$$

From theorem 9.2,

$$\sum_{k=1}^j e_k^i = \sum_k \lambda_k - \sum_k \mu_k^i = \sum_k \lambda_k - \sum_k v_k = \sum_k (\lambda_k - v_k) \leq 0,$$

the equality holding just in case $j=r$.

XII. Concluding Remarks

Associated with a vector z_1 is the chain of subspaces

$$\mathcal{A}_1 \subset \mathcal{A}_2 \subset \dots \subset \mathcal{A}_r$$

of section 9, together with the corresponding polynomials $p_j(\lambda)$, whose zeros give the characteristic roots of A relative to \mathcal{A}_j . Let $v_{1,j}$ denote a minimum solution of $p_j(\lambda)=0$. It is a consequence of theorem 9.1 that

$$\nu_{1,1} \geq \nu_{1,2} \geq \dots \geq \nu_{1,r} = \lambda_1.$$

Similarly, if $\nu_{2,j}$ denotes the next smallest solution, then

$$\nu_{2,2} \geq \nu_{2,3} \geq \dots \geq \nu_{2,r} = \lambda_2.$$

Analogous results hold for the higher characteristic roots. Thus, the characteristic roots of A can be found by successively finding the zeros of $p_2(\lambda)$, $p_3(\lambda)$, etc.¹² The procedure may be followed, for example, after a good approximation z_1^i has been obtained by, say, one of the gradient schemes.

The subspaces and their polynomials may be used in other ways in conjunction with the method of gradients. It was pointed out earlier that in the gradient methods we pass in each step from z_1 to a next approximation \bar{z}_1 which lies in the subspace \mathcal{A}_2 associated with z_1 . One might consider choosing \bar{z}_1 in \mathcal{A}_3 or even a higher-dimensional subspace. The extra labor involved might be justified, or even essential, in an ill-conditioned problem in which several roots are clustered about λ_1 (or λ_r).

To illustrate, suppose that $\lambda_1, \lambda_2, \lambda_3$ are close together and reasonably isolated from the other roots. After a certain number of iterations it would be expected that the subspace \mathcal{A}_3 spanned by z_1, z_2, z_3 would contain only the characteristic vectors y_1, y_2, y_3 to a good approximation. At this stage, although the z 's might not be good individual approximations to the y 's, the characteristic vectors u_1, u_2, u_3 relative to \mathcal{A}_3 would provide such approximations. Further accuracy for y_1 would then be obtained by continuing the iteration process with u_1 .

We proceed to derive the formulas for the characteristic roots $\nu_1 \leq \nu_2 \leq \nu_3$ and characteristic vectors u_1, u_2, u_3 relative to \mathcal{A}_3 . The roots ν_k are the solutions of

$$p_3(\lambda) = (\lambda - \mu_1)(\lambda - \mu_2)(\lambda - \mu_3) - t_2^2(\lambda - \mu_3) - t_3^2(\lambda - \mu_1) = 0.$$

This cubic is most easily solved by introducing the new variable ρ and the constants μ, σ_k as follows:

$$\bar{\mu} = -(\mu_1 + \mu_2 + \mu_3), \quad \sigma_k = \mu_k - \bar{\mu}, \quad k = 1, 2, 3,$$

$$\lambda = \bar{\mu} + \rho.$$

The equation to be solved is now

$$\rho^3 + b\rho + c = 0,$$

where

$$b = \sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3 - t_2^2 - t_3^2, \quad c = t_2^2\sigma_3 + t_3^2\sigma_1 - \sigma_1\sigma_2\sigma_3.$$

The solutions may now be determined rapidly by, say, Newton's method. To find u_1 , the characteristic vector relative to \mathcal{A}_3 belonging to ν_1 , we write

$$u_1 = z_1 + \alpha z_2 + \beta z_3$$

and attempt to find α and β . To do this observe first that since u_k is a characteristic vector relative to \mathcal{A}_3 belonging to ν_k ,

$$(u, Au_k - \nu_k u_k) = 0, \quad u \text{ in } \mathcal{A}_3, \quad k = 1, 2, 3. \quad (46)$$

Using eq (46) with $k = 1$, and $u = z_1, u = z_3$ in turn we find, by eq (34),

$$0 = (\mu_1 - \nu_1)|z_1|^2 + \alpha|z_2|^2$$

$$0 = \alpha|z_3|^2 + \beta u_3|z_3|^2 - \beta \nu_1|z_3|^2.$$

Thus

$$\alpha = \frac{\nu_1 - \mu_1}{t_2^2}$$

$$\beta = \frac{\alpha}{\nu_1 - \mu_3},$$

so that

$$u_1 = z_1 + \frac{\nu_1 - \mu_1}{t_2^2} \left(z_2 + \frac{1}{\nu_1 - \mu_3} z_3 \right).$$

The formulas for u_2 and u_3 are obtained by replacing ν_1 by ν_2 and ν_3 , respectively.

Suppose the lower characteristic roots $\lambda_1, \lambda_2, \dots$ and the higher characteristic roots $\lambda_r, \lambda_{r-1}, \dots$ have been accurately calculated. The intermediate roots and vectors can then be calculated by the gradient procedure (12) as follows. We apply the procedure to an initial vector x^0 , inducing (if necessary) and maintaining orthogonality to the vectors y_1, y_2, \dots and y_r, y_{r-1}, \dots by more or less frequent selection of γ^i as $1/(\lambda_j - \mu^i)$, with λ_j ranging over the known roots (see section 8). Use of the subspaces $\mathcal{A}_2, \mathcal{A}_3, \dots$ may come into play as described above. Notice that as more roots are found the constant α in the method of fixed α may be chosen larger.

We remark that independent characteristic vectors belonging to a multiple characteristic root can be determined by varying the initial vector x^0 .

We conclude by pointing out that the results of theorem 10.1 remain valid if the constant α satisfying eq (23) is replaced by the variable value of eq (28) where, in addition to eq (29) and (30), the condition

$$\frac{\beta_2}{\beta_1} < \min_{j=3,4,\dots,r} \frac{\lambda_j - \lambda_1}{\lambda_{j-1} - \lambda_1}$$

is imposed. The purpose of this restriction is to guarantee that $a_k^i/a_j^i \rightarrow 0$ for $j < k$. The results on rate of convergence in section 11 may be modified to fit the new conditions.

LOS ANGELES, May 4, 1950.

¹² This is one of the methods developed by Lanczos.