CONVERGENCE OF A METHOD OF SOLVING LINEAR PROBLEMS[1]

W. KARUSH

1. Introduction. We are concerned with the solution of two problems associated with a linear operator A. First, the characteristic value problem

$$(1) \qquad\qquad Ay = \lambda y$$

for the determination of the characteristic values λ and the characteristic vectors y; second, the linear equation problem

$$(2) \qquad\qquad (A - \lambda I)x = b, \qquad\qquad b \neq 0,$$

for the determination of x, given the number λ and the vector b (I is the identity operator). Lanczos [3][2] has described an interesting iterative method for the solution of these problems which appears to be effective for numerical calculation. It is our purpose to consider the convergence and rate of convergence of the method, in the Hilbert space sense, for a bounded self-adjoint operator.

The procedure for obtaining the solution may be described as follows. Let $b \neq 0$ be a given initial vector, arbitrary for problem (1), equal to the right side of (2) for problem (2). Let

$$(3) \qquad\qquad \mathcal{K}_i = (b, Ab, \cdots, A^{i-1}b),$$

i.e., the linear subspace spanned by the indicated vectors. Let $\mathcal{K}$ be the invariant subspace which is the closure of the linear subspace spanned by all non-negative powers $A^i b$; symbolically

$$(4) \qquad\qquad \mathcal{K} = (b, Ab, \cdots, A^i b, \cdots).$$

Let π_i be the projection operator onto $\mathcal{K}_i$. Then to solve (1) and (2) we replace the operator A by the operator $\pi_i A$ on $\mathcal{K}_i$, solve the corresponding finite-dimensional problem, and allow i to approach ∞. That is, (1) and (2) are approximated respectively by

$$(5) \qquad\qquad \pi_i A y = \lambda y \qquad\qquad \text{on } \mathcal{K}_i$$

and

$$(6) \qquad\qquad (\pi_i A - \lambda I)x = b \qquad\qquad \text{on } \mathcal{K}_i.$$

[2] Numbers in brackets refer to the list of references at the end of the paper.

We shall establish the convergence of the solutions of (5) to a solution of (1) for isolated characteristic values; the convergence to the characteristic value is monotonic. This result applies only to solutions of (1) lying in $\mathfrak{K}$, of course. To prove the convergence of the solutions of (6) to the solution of (2) we make the assumption that A is completely continuous. Under this latter hypothesis we are also able to show that the rate of convergence to the solution for each of the problems (1) and (2) is faster than any geometric progression.

One of the advantages of solving (1) and (2) by means of (5) and (6) is the simple algebraic form that the successive approximations take. For full details, as well as numerical examples, see the paper of Lanczos where self-adjointness is not necessarily assumed. Our description of the method in the preceding paragraph is somewhat different from Lanczos' formulation. The equivalence of the two procedures is shown in the final section of this paper.

Some additional remarks are in order. First, the present method should be contrasted with a procedure studied by the author in an earlier paper [2] in which only the characteristic value problem (1) was considered. In that paper a sequence of subspaces of *fixed finite dimension* was employed, each subspace depending upon the previous one, instead of the sequence of subspaces $\mathfrak{K}_i$ of increasing dimension described in (3). Further, the method of [2] is not immediately applicable to the linear equation problem (2). The second remark is that the present paper is related to the work of Rellich [4] in that we deal with successive approximations $\pi_i A$ to A. However it appears that the direct use here of his results would lead to no essential simplification. Finally, attention should be drawn to a paper of R. C. T. Smith [5] concerned with the calculation of characteristic values of infinite matrices. The matrix is approximated by its finite segments, that is, (1) is approximated by equation (5) with $\mathfrak{K}_i = (e_1, e_2, \cdots, e_i)$, e_j being the sequence with 1 in the jth place and zeros elsewhere. Explicit error estimates for the characteristic values are given.

2. **Assumptions.** In this section we shall specify the notation and assumptions which are to remain in force in the remainder of the paper.

We are given a bounded self-adjoint linear operator A on a (not necessarily separable) real Hilbert space $\mathfrak{R}$ of infinite dimension. With a given vector b we form the subspaces (3) and (4). Since our method of solution never leads us out of the invariant closed (separable) space $\mathfrak{K}$, we henceforth understand by A *an operator on* $\mathfrak{K}$

(i.e., from $\mathfrak{IC}$ to $\mathfrak{IC}$), unless otherwise stated. Thus we obtain only the characteristic solutions of (1) which belong to $\mathfrak{IC}$; further solutions would be obtained by varying b. The hypotheses we shall make on A will be hypotheses holding for A as on $\mathfrak{IC}$; this is only a matter of convenience since it would be a simple matter to impose the conditions on A as on $\mathfrak{R}$ and add a mild condition on b to guarantee the satisfaction of the hypotheses of the theorems.

We understand always by $\pi_i A$ an operator on the subspace $\mathfrak{IC}_i$. As such $\pi_i A$ is self-adjoint; for, with x and y in $\mathfrak{IC}_i$, $(\pi_i A x, y) = (A x, y) = (x, A y) = (x, \pi_i A y)$.

We shall make use of the Rayleigh quotient, which is defined by

$$(7) \qquad\qquad \mu(x) = \frac{(x, A x)}{(x, x)}, \qquad\qquad x \neq 0.$$

For x in $\mathfrak{IC}_i$ we have $(x, \pi_i A x)/(x, x) = (x, A x)/(x, x)$, so that the Rayleigh quotient for the finite-dimensional operator $\pi_i A$ has the form (7) independent of i, provided x is restricted to $\mathfrak{IC}_i$. As is well known the characteristic vectors of (5) are the minimax points for the function μ on $\mathfrak{IC}_i$ [1, pp. 27–29]. Certain other standard properties of μ will be taken as known.

Finally we assume as a matter of convenience that

$$\text{dimension } (\mathfrak{IC}_i) = i.$$

That is, we assume that $\mathfrak{IC}$ is not finite-dimensional. If the contrary held, then, for some k, $\mathfrak{IC} = \mathfrak{IC}_k$ and the method would terminate with an exact solution at the kth step.

3. **The characteristic value problem.** It is easy to show that the vector b has a non-null projection on every characteristic manifold of A, and that every such manifold has dimension one, i.e., every characteristic number is simple [2]. A similar statement holds for the operator $\pi_i A$, whose i characteristic values we denote by

$$(8) \qquad\qquad \lambda_{1i} > \lambda_{2i} > \cdots > \lambda_{ii},$$

and whose corresponding characteristic vectors we denote by

$$y_{1i}, y_{2i}, \cdots, y_{ii}$$

and which are taken to satisfy the conditions

$$(9) \qquad\qquad |y_{ji}| = 1, \quad (b, y_{ji}) > 0, \qquad j = 1, 2, \cdots, i.$$

For the next theorem we shall require the following assumption on the (real) spectrum S of A.

(10) S is the union of disjoint sets S_1, S_2 where S_1 consists of a

finite number of isolated characteristic values

$$\lambda_1 > \lambda_2 > \cdots > \lambda_m, \qquad\qquad m \geqq 1,$$

and S_2 is a closed set lying below λ_m.

This partitioning may be possible in many ways. The characteristic vectors of A corresponding to the values in S_1 we denote by

$$y_1, y_2, \cdots, y_m$$

where

(11) $$|y_j| = 1, \quad (b, y_j) > 0, \qquad j = 1, 2, \cdots, m.$$

THEOREM 1. *Let the spectrum of A satisfy condition* (10). *For each j, $j=1, 2, \cdots, m$, determine the infinite sequence $\{y_{ji}\}$ as above. Then $\{\lambda_{ji}\}$ is a monotonic increasing sequence, and*

$$\lim_{i \to \infty} y_{ji} = y_j, \qquad \lim_{i \to \infty} \lambda_{ji} = \lambda_j, \qquad j = 1, 2, \cdots, m.$$

An entirely analogous result holds if the finite set S_1 lies below, instead of above, the remainder S_2 of the spectrum.

To make the proof we suppose first that the conclusion has been established for $j=1, 2, \cdots, k-1$, $2 \leqq k \leqq m$; we show its validity for $j=k$. Observe that $\lambda_{ki} = \mu(y_{ki})$ is a nondecreasing sequence, as follows from the minimax principle just referred to for characteristic values in finite-dimensional spaces. Further, the sequence is bounded by λ_1, since

$$\lambda_1 = \max \mu(x), \qquad\qquad x \neq 0, \ x \in \mathcal{K}.$$

Let $\bar{\lambda}$ be the limit of the sequence; we wish to show that $\bar{\lambda} = \lambda_k$.

Since $y_k \in \mathcal{K}$, it follows from the definition of $\mathcal{K}$ that there is a sequence of vectors x_i, $x_i \in \mathcal{K}_i$, such that $x_i \to y_k$ as $i \to \infty$. We may write

$$x_i = \sum_{j=1}^{k-1} \beta_{ji} y_{ji} + r_i, \qquad r_i \in \mathcal{K}_i, \ (r_i, y_{ji}) = 0.$$

Take the inner product of each side of this equation with y_{ki} and allow i to tend to ∞. Then using the assumed limit relations and the orthogonality between characteristic vectors we find that $\beta_{ji} \to 0$. Thus r_i tends to y_k, since x_i does. Hence $\mu(r_i) \to \mu(y_k) = \lambda_k$. But $\mu(r_i) \leqq \mu(y_{ki})$ since y_{ki} maximizes $\mu(x)$ for $x \in \mathcal{K}_i$ and x orthogonal to y_{ji}, $j=1, 2, \cdots, k-1$. Thus

$$\lambda_k \leqq \bar{\lambda}.$$

To prove the reverse inequality we write

$$(12) \qquad y_{ki} = \sum_{j=1}^{k} \alpha_{ji} y_j + z_i, \qquad z_i \in \mathfrak{IC}, \ (z_i, y_j) = 0.$$

From (9) and (11)

$$(13) \qquad 1 = \sum_{j=1}^{k} \alpha_{ji}^2 + |z_i|^2.$$

We also note the relation

$$(14) \qquad \lambda_k - \mu(z_i) \geqq \gamma > 0,$$

which follows from the fact that the least upper bound of the complement in $\mathfrak{S}$ of the finite set $(\lambda_1, \lambda_2, \cdots, \lambda_k)$ is the maximum of $\mu(x)$ for $x \in \mathfrak{IC}$, $(x, y_j) = 0$, $j = 1, 2, \cdots, k$. We take the inner product of each side of (12) with (y_{ni}), $n = 1, 2, \cdots, k-1$, and find that $\alpha_{ni} \to 0$ as $i \to \infty$. From (12) and the definition (7) of $\mu(x)$ we find that

$$\mu(y_{ki}) = \sum_{j=1}^{k} \alpha_{ji}^2 \mu(y_j) + \mu(z_i)|z_i|^2,$$

that is,

$$(15) \qquad \lambda_{ki} = \sum_{j=1}^{k} \alpha_{ji}^2 \lambda_j + \mu(z_i)|z_i|^2.$$

Substituting for $|z_i|^2$ from (13) we obtain

$$(16) \qquad \begin{aligned} \lambda_{ki} &= \sum_{j=1}^{k} \alpha_{ji}^2 (\lambda_j - \mu(z_i)) + \mu(z_i) \\ &\leqq \mu(z_i) < \lambda_k. \end{aligned}$$

Thus

$$\bar{\lambda} \leqq \lambda_k,$$

which completes the proof of $\bar{\lambda} = \lambda_k$.

To show the convergence of the vectors write (16) in the form

$$\alpha_{ki}^2 = 1 + \frac{1}{\lambda_k - \mu(z_i)} \left((\lambda_{ki} - \lambda_k) - \sum_{j=1}^{k-1} \alpha_{ji}^2 (\lambda_j - \mu(z_i)) \right).$$

Taking the limit and using (14) we obtain $\alpha_{ki}^2 \to 1$. From (13), $z_i \to 0$. The sequence α_{ki} cannot have the limit point -1. If it did, a subsequence (denoted by the same subscript i) would converge to -1. The corresponding subsequence y_{ki} by (12) would converge to $-y_k$,

hence $(b, y_{ki}) \to -(b, y_k) < 0$ by (11), contrary to (9). Thus the original sequence α_{ki} tends to 1 and, by (12), $y_{ki} \to y_k$.

The proof will be complete if we justify the conclusion of the theorem for $j=1$. This justification is a duplication of the above proof with obvious simplifications. We omit the details.

4. **Rate of convergence.** In order to prove the convergence theorem for the solution of problem (2) we shall require some information on the rate of the convergence established in Theorem 1. This information is of independent interest as well. We now suppose that A is completely continuous, although the results of this section could be formulated somewhat more generally. We shall show that the convergence to the solution of problem (1) is geometric with arbitrarily small ratio.

With A completely continuous the spectrum of A (omitting 0) consists of two sequences of characteristics values

$$\lambda_1 > \lambda_2 > \cdots , \qquad\qquad \lambda_j > 0,$$
$$\lambda_1' < \lambda_2' < \cdots , \qquad\qquad \lambda_j' < 0,$$

not both of which are finite or vacuous. Either infinite sequence has the limit 0. In the sequel we shall speak as if both sequences are infinite. If one is finite or vacuous, then certain obvious notational changes should be made. Every characteristic value is simple and we denote corresponding characteristic vectors by

$$y_1, y_2, \cdots ,$$
$$y_1', y_2', \cdots ,$$

normalized in length and direction by (11).

For problem (5) it will be convenient to denote the characteristic values (8) arranged in reverse order by

$$\lambda_{1i}' < \lambda_{2i}' < \cdots < \lambda_{ii}',$$

with a similar notation for characteristic vectors. By Theorem 1 and the remark following it we have, for every positive integer j,

$$y_{ji} \to y_j, \; \lambda_{ji} \to \lambda_j, \; y_{ji}' \to y_j', \; \lambda_{ji}' \to \lambda_j'.$$

THEOREM 2. *Suppose that A is completely continuous. Let δ be an arbitrary number with $0 < \delta < 1$. Then for each characteristic number λ_k (λ_k') there is a constant K (K') independent of i such that*

$$\left| \lambda_k - \lambda_{ki} \right| \leqq K\delta^i \qquad (\left| \lambda_k' - \lambda_{ki}' \right| < K'\delta^i).$$

We give the proof for a positive value λ_k; for a negative value we need only notice that the proof is entirely analogous or alternatively we may transform to the operator $-A$.

We choose positive integers m, m' so that

$$(17) \qquad \max\left(\frac{\lambda_{m+1}}{\lambda_k}, \ \frac{|\lambda'_{m'+1}|}{\lambda_k}\right) < \delta.$$

We write

$$(18) \qquad b = \sum_{j=1}^{m} \alpha_j y_j + \sum_{j=1}^{m'} \alpha'_j y'_j + z$$

with z orthogonal to the subspace

$$\mathcal{L} = (y_1, \cdots, y_m, y'_1, \cdots, y'_m).$$

By (11) we have

$$(19) \qquad \alpha_j > 0, \qquad \alpha'_j > 0.$$

Also

$$(20) \qquad b = \sum_{j=1}^{k-1} \alpha_{ji} y_{ji} + z_i, \qquad z_i \in \mathcal{K}_i, \ (z_i, y_{ji}) = 0.$$

Let $q(\lambda)$ be an arbitrary polynomial of degree $i-1$. From (18)

$$(21) \qquad x_i \equiv q(A)b = \sum_{j=1}^{m} \alpha_j q(\lambda_j) y_j + \sum_{j=1}^{m'} \alpha'_j q(\lambda'_j) y'_j + q(A)z, \ x_i \in \mathcal{K}_i.$$

Now $q(A)b = q(\pi_i A)b$. Hence, from (20), $x_i = \sum_{j=1}^{k-1} \alpha_{ji} q(\pi_i A) y_{ji} + q(\pi_i A) z_i$, so that

$$(22) \qquad x_i = \sum_{j=1}^{k-1} \alpha_{ji} q(\lambda_{ji}) y_{ji} + \bar{z}_i, \qquad \bar{z}_i \in \mathcal{K}_i, \ (\bar{z}_i, y_{ji}) = 0.$$

We choose

$$q(\lambda) = \lambda^s \prod_{j=1, j \neq k}^{m} (\lambda - \lambda_j) \prod_{j=1}^{m'} (\lambda - \lambda'_j) \prod_{j=1}^{k-1} (\lambda - \lambda_{ji}),$$

$$s = i - m - m' - k + 1.$$

For i sufficiently large q is well-defined and $q(\lambda_k) \neq 0$. By (22), we have $x_i = \bar{z}_i$. Hence

$$(23) \qquad \mu(x_i) \leqq \lambda_{ki},$$

since $\lambda_{ki} = \max \mu(x)$ for $x \in \mathcal{3C}_i$, $x \neq 0$ orthogonal to y_{ji}, $j = 1, 2, \cdots$, $k-1$. By (21)

$$x_i = \alpha_k q(\lambda_k) y_k + r_i, \qquad r_i = q(A)z.$$

Thus

$$\mu(x_i) = \frac{\alpha_k^2 q^2(\lambda_k)\lambda_k + \mu(r_i)\,|\,r_i\,|^2}{\alpha_k^2 q^2(\lambda_k) + |\,r_i\,|^2},$$

(24) $$\lambda_k - \mu(x_i) = \frac{(\lambda_k - \mu(r_i))\,|\,r_i\,|^2}{\alpha_k^2 q^2(\lambda_k) + |\,r_i\,|^2} \leqq K_1 \frac{|\,r_i\,|^2}{q^2(\lambda_k)}.$$

To estimate the last ratio we use the fact that z is orthogonal to $\mathcal{L}$. From the spectral resolution of A,

$$r_i = \int_S q(\lambda)dE(\lambda)z = \int_{\lambda'_{m'+1}}^{\lambda_{m+1}} q(\lambda)dE(\lambda)z.$$

Hence

$$|\,r_i\,|^2 = \int_{\lambda'_{m'+1}}^{\lambda_{m+1}} q^2(\lambda)d\,|\,E(\lambda)z\,|^2 \leqq q^2(\bar{\lambda})\,|\,z\,|^2, \qquad \lambda'_{m'+1} \leqq \bar{\lambda} \leqq \lambda_{m+1}.$$

Utilizing the definition of $q(\lambda)$ and (17) we obtain from the preceding inequality, for i sufficiently large,

$$\frac{|\,r_i\,|^2}{q^2(\lambda_k)} \leqq |\,z\,|^2 \frac{q^2(\bar{\lambda})}{q^2(\lambda_k)} \leqq K_2\left(\frac{\bar{\lambda}^2}{\lambda_k^2}\right)^i \leqq K_3 \delta^i.$$

Combining this with (24) yields

$$\lambda_k - \mu(x_i) \leqq K\delta^i.$$

Finally, from (23), $\lambda_k - \lambda_{ki} \leqq \lambda_k - \mu(x_i)$. The conclusion of the theorem now follows from the fact that λ_{ki} approaches λ_k monotonically from below.

For the convergence of the characteristic vectors a similar estimate holds. We state the result without proof, for we shall not require it in the sequel. The proof is made by re-doing the proof of Theorem 1 with the added result in Theorem 2.

THEOREM 3. *Under the hypotheses of Theorem 2 there is for each characteristic vector* y_k (y'_k) *a constant* K (K') *independent of* i *such that*

$$|\,y_k - y_{ki}\,| \leqq K\delta^i \qquad (\,|\,y'_k - y'_{ki}\,| \leqq K'\delta^i).$$

5. The linear equation problem. We turn now to the solution of problem (2), where the given number λ is assumed to be in the resolvent set of A and the given vector b is, of course, to be used in the construction of the spaces (3) and (4). Then a unique solution

$$(25) \qquad x_0 = (A - \lambda I)^{-1} b$$

in $\mathfrak{IC}$ exists. As an aside we remark that if A is regarded as an operator on the original space $\mathfrak{R}$, then the assumption that λ is in the resolvent set of A as on $\mathfrak{R}$ implies that λ is in the resolvent set of A as on $\mathfrak{IC}$. But the reverse implication does not hold; for example, b may be orthogonal to a particular characteristic manifold of A as on $\mathfrak{R}$. Thus, imposing our condition on A as on $\mathfrak{IC}$ represents a weaker assumption.

As in the preceding section we suppose that A is completely continuous, for we shall require the results of Theorem 2. Since λ is in the resolvent set of A, we have

$$\lambda \neq 0$$

and by Theorem 1 it follows that for i sufficiently large the number λ is not characteristic for $\pi_i A$. Thus equation (5) has a unique solution x_i, i.e.,

$$(26) \qquad \pi_i(A - \lambda I) x_i = b, \qquad\qquad x_i \in \mathfrak{IC}_i.$$

We shall establish the convergence of x_i to the solution (25) of equation (2), and estimate the rate of convergence.

Let

$$\phi_i(\lambda) = \lambda^i + \gamma_1 \lambda^{i-1} + \cdots + \gamma_i = \prod_{j=1}^{i} (\lambda - \lambda_{ji})$$

be the characteristic polynomial of $\pi_i A$ (for simplicity we do not indicate the dependence of the coefficients γ on the index i). We have $0 = \phi_i(\pi_i A)b = \pi_i A^i b + \gamma_1 A^{i-1} b + \cdots + \gamma_i b = \phi_i(A)b - (A^i b - \pi_i A^i b)$. Thus $\phi_i(A)b = A^i b - \pi_i A^i b$, so that if we let

$$(27) \qquad\qquad \xi_i = \phi_i(A)b,$$

then $\xi_i \in \mathfrak{IC}_{i+1}$ and $\pi_i \xi_i = 0$. From (27)

$$(28) \qquad A^i b = \xi_i - \gamma_1 A^{i-1} b - \cdots - \gamma_i b.$$

If we represent the solution x_i in the form

$$(29) \qquad x_i = \beta_0 b + \beta_1 A b + \cdots + \beta_{i-1} A^{i-1} b,$$

then Lanczos [3] has given the solution

$$\beta_{i-1} = -\frac{1}{\phi_i(\lambda)}, \qquad \beta_{i-2} = -\frac{\lambda + \gamma_1}{\phi_i(\lambda)}, \cdots,$$

$$(30) \qquad \beta_1 = -\frac{\lambda^{i-2} + \gamma_1 \lambda^{i-3} + \cdots + \gamma_{i-2}}{\phi_i(\lambda)},$$

$$\beta_0 = -\frac{\lambda^{i-1} + \gamma_1 \lambda^{i-2} + \cdots + \gamma_{i-1}}{\phi_i(\lambda)}.$$

To verify this, we substitute (29) into $b - (A - \lambda I)x_i$, using (30) for the values of β and replacing $A^i b$ by the expression on the right of (28). The result is

$$b - (A - \lambda I)x_i = \xi_i/\phi_i(\lambda).$$

Since $\pi_i \xi_i = 0$, this equation tells us that x_i is the solution of (26); further

$$(31) \qquad r_i \equiv b - (A - \lambda I)x_i = \frac{\phi_i(A)b}{\phi_i(\lambda)}.$$

We now state the convergence theorem.

THEOREM 4. *Suppose that A is completely continuous. Let $\lambda \neq 0$ be noncharacteristic for A, and let x_0 be the unique solution in $\mathcal{IC}$ of equation (2). Let δ be an arbitrary number with $0 < \delta < 1$. Then for sufficiently large i equation (6) has a unique solution x_i in $\mathcal{IC}_i$, and there exists a constant K independent of i such that*

$$(32) \qquad \left| x_i - x_0 \right| \leqq K\delta^i.$$

In particular, $\lim_{i \to \infty} x_i = x_0$.

We have already remarked that there exists a unique x_i for sufficiently large i. To establish (32) we first choose $\alpha > 0$ so that

$$(33) \qquad \alpha < \left| \lambda \right| \quad \text{and} \quad \frac{2\alpha}{\left| \lambda \right| - \alpha} < \delta.$$

Using the notation of the preceding section choose the indices m and m' so that λ_{m+1} and $\lambda'_{m'+1}$ lie in the closed interval $[-\alpha, \alpha]$. We express b in the form (18) and consider the remainder r_i of (31). We have

$$(34) \quad r_i = \frac{\phi_i(A)b}{\phi_i(\lambda)} = \sum_{j=1}^{m} \frac{\phi_i(\lambda_j)}{\phi_i(\lambda)} \alpha_j y_j + \sum_{j=1}^{m'} \frac{\phi_i(\lambda'_j)}{\phi_i(\lambda)} \alpha'_j y'_j + \frac{\phi_i(A)z}{\phi_i(\lambda)}.$$

Let

$$\sigma = 2^{-1} \min \left| \lambda - \nu \right| > 0, \qquad M = 2 \max \left| \nu \right| > 0 \quad \text{for } \nu \in \mathsf{S}.$$

Choose δ_1, $0 < \delta_1 < 1$, so that

$$\delta_1 M / \sigma < \delta.$$

We now consider the coefficient $\phi_i(\lambda_j)/\phi_i(\lambda)$ in (34). Apply Theorem 2 to λ_j for the value δ_1 to obtain an appropriate constant K_1. Then for sufficiently large i we obtain, by writing ϕ_i in factored form,

$$\left| \frac{\phi_i(\lambda_j)}{\phi_i(\lambda)} \right| \leqq \left| \lambda_j - \lambda_{ji} \right| \frac{M^{i-1}}{\sigma^i} \leqq \frac{K_1}{M} \left(\frac{\delta_1 M}{\sigma} \right)^i \leqq K \delta^i.$$

Similarly, we obtain the same inequality for λ_j'.

Finally we treat the last term on the right of (34). We have

$$\left| \frac{\phi_i(A)z}{\phi_i(\lambda)} \right| = \left| \frac{1}{\phi_i(\lambda)} \int_{-\alpha}^{\alpha} \phi_i(\nu) dE(\nu) z \right| \leqq \left| \frac{\phi_i(\bar{\lambda})}{\phi_i(\lambda)} \right| \cdot \left| z \right|, \quad \bar{\lambda} \text{ in } [-\alpha, \alpha].$$

For i sufficiently large all roots λ_{ki}, $k = m+1$, $m+2$, $\cdots$, and λ_{ki}', $k = m'+1$, $m'+2$, $\cdots$, lie in $[-\alpha, \alpha]$. Hence, writing $\phi_i(\lambda)$ in factored form, we find

$$\left| \frac{\phi_i(\bar{\lambda})}{\phi_i(\lambda)} \right| \leqq \left(\frac{M}{\sigma} \right)^{m+m'} \left(\frac{2\alpha}{\left| \lambda \right| - \alpha} \right)^{i-(m+m')} \leqq K \delta^i$$

by (33). From the preceding inequalities and equation (34),

$$\left| r_i \right| \leqq K \delta^i.$$

From (31) $x_i = B(b - r_i) = x_0 - B r_i$ where $B = (A - \lambda I)^{-1}$ is a bounded operator. Thus

$$\left| x_i - x_0 \right| \leqq \left| B r_i \right| \leqq K \delta^i$$

for an appropriate constant K. This completes the proof.

6. **Equivalence with procedure of Lanczos.** In order to show that our method is identical with that of Lanczos we shall reproduce some of his formulas here (see [3, §§VII and X]). His notation has been modified slightly to conform with ours; in particular, his $\mu = 1/\lambda$ becomes our λ. Beginning with the initial vector $b = b_0$, constants α_j, β_j, polynomials $p_j(\lambda)$, and vectors b_j are defined recursively by

$$(35) \quad \begin{aligned} &\alpha_j = \mu(b_j), \qquad \beta_j = \left| b_{j+1} \right|^2 / \left| b_j \right|^2, \\ &p_0(\lambda) \equiv 1, \; p_1(\lambda) = \lambda - \alpha_0, \; p_{j+1}(\lambda) = (\lambda - \alpha_j) p_j(\lambda) - \beta_{j-1} p_{j-1}(\lambda), \\ &b_j = p_j(A) b_0, \qquad b_j \in \mathcal{K}_{j+1}. \end{aligned}$$

These quantities are defined in the order b_j, α_j, β_{j-1}, $p_{j+1}(\lambda)$, b_{j+1}. Let us show first that the polynomial $p_i(\lambda)$ is the characteristic polynomial of the operator $\pi_i A$ on $\mathfrak{IC}_i$. Lanczos shows (pp. 266–267) that b_i is that vector of the form $A^i b - z$, $z \in \mathfrak{IC}_i$, which has minimum length. Clearly this vector must be given by

$$b_i = A^i b_0 - \pi_i A^i b_0.$$

Now let $\phi_i(\lambda)$ be the characteristic function of $\pi_i A$. Then $0 = \phi_i(\pi_i A) b_0 = -A^i b_0 + \pi_i A^i b_0 + \phi_i(A) b_0 = -b_i + \phi_i(A) b_0$. Hence, by the last equation (35), $\phi_i(A) b_0 = p_i(A) b_0$, so that $\phi_i(\lambda) \equiv p_i(\lambda)$, since $\mathfrak{IC}_i$ has dimension i. Thus Lanczos' approximation to the characteristic numbers of A, namely the roots of $p_i(\lambda)$, is identical with our approximation.

His approximation to the jth characteristic vector (see his equation (107)) is

$$(36) \qquad y = \sum_{k=0}^{i-1} \frac{p_k(\lambda_j)}{|b_k|^2} b_k,$$

where λ_j is the jth root of $p_i(\lambda) = 0$. We need to verify that this vector is the jth characteristic vector of $\pi_i A$, i.e., that

$$(37) \qquad \pi_i A y = \lambda_j y_j.$$

To this end we rewrite the recursion formula for the polynomials $p(\lambda)$ in the form

$$\lambda p_k(\lambda) = p_{k+1}(\lambda) + \alpha_k p_k(\lambda) + \beta_{k-1} p_{k-1}(\lambda), \qquad p_{-1}(\lambda) \equiv 0.$$

Replacing λ by A and operating on b_0 leads to

$$
\begin{aligned}
& A b_0 = b_1 + \alpha_0 b_0, \\
(38) \qquad & A b_k = b_{k+1} + \alpha_k b_k + \beta_{k-1} b_{k-1}, \qquad k = 1, 2, \cdots, i - 2, \\
& \pi_i A b_{i-1} = \alpha_{i-1} b_{i-1} + \beta_{i-2} b_{i-2}.
\end{aligned}
$$

Substituting the value of y in (36) into the left side of (37) and reducing the resulting expression to a linear combination of the b_k's by means of (38) leads to the right side of (37). The required verification is thereby accomplished.

Finally, we need to show that his approximation to the solution of equation (2) is the solution of equation (6). With appropriate notational changes his approximation is (see his equation (102))

$$(39) \qquad x = -\frac{1}{p_i(\lambda)} \sum_{k=0}^{i-1} p_{i-k-1}(\lambda) b_k$$

where

$$\bar{p}_0(\lambda) \equiv 1, \qquad \bar{p}_1(\lambda) = \lambda - \alpha_{i-1},$$

$$\bar{p}_k(\lambda) = (\lambda - \alpha_{i-k})\bar{p}_{k-1}(\lambda) - \beta_{i-k}\bar{p}_{k-2}(\lambda).$$

Again, direct substitution of the value of x in (39) into the left side of (6) and use of (38) and the recursion relation for the polynomials $\bar{p}(\lambda)$ leads to the right side of (6), establishing the desired result.

References

1. R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, vol.1, 2d ed., Berlin, 1931.

2. W. Karush, *Determination of the extreme values of the spectrum of a bounded self-adjoint operator*, Proceedings of the American Mathematical Society vol. 2 (1951) pp. 980–989.

3. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research NBS vol. 45 (1950) pp. 255–282.

4. F. Rellich, *Störungstheorie der Spektralzerlegung*. II, Math. Ann. vol. 113 (1936) pp. 677–685.

5. R. C. T. Smith, *The approximate solution of equations in infinitely many unknowns*, Quart. J. Math. vol. 18 (1947) pp. 25–52.

NATIONAL BUREAU OF STANDARDS, LOS ANGELES AND
    UNIVERSITY OF CHICAGO