ESTIMATES FOR THE ERRORS IN THE RAYLEIGH-RITZ METHOD

FULTON KOEHLER

1. Introduction. Let an eigenvalue problem be given in the form of a homogeneous linear differential equation

$$(1) \qquad Ly = \lambda y,$$

with homogeneous linear boundary conditions, denoted by (C). It is assumed that the parameter λ does not appear in the boundary conditions. The region R of the problem may be of any number of dimensions; the symbol $\int f(P)\,dP$ will mean an integral over R, dP standing for the Euclidean volume element; (f, g) will stand for $\int f(P)g(P)\,dP$; and $\|f\|$, for $(f, f)^{1/2}$. The symbol $\iint F(P, Q)\,dPdQ$ will mean an integral over the Cartesian product of R with itself. We assume that the problem $(1)+(C)$ is positive definite and self-adjoint; that is, $(\phi, L\phi) > 0$ and $(\phi, L\psi) = (L\phi, \psi)$ for any admissible functions ϕ and ψ, an admissible function being a real-valued function, not identically zero, which satisfies the boundary conditions (C) and is continuously differentiable up to derivatives of the order of the operator L. The class of admissible functions will be denoted by $\mathfrak{A}$.

The existence of eigenvalues for the problem $(1)+(C)$ is assumed, and these will be denoted by $\{\lambda_i\}$, $0 < \lambda_1 \le \lambda_2 \le \lambda_3 \cdots$; and the corresponding eigenfunctions, by $\{y_i\}$, chosen so that $(y_i, y_j) = \delta_{ij}$. We assume that a Green's function $G(P, Q)$ exists for the problem $Ly = 0$ with boundary conditions (C), which is symmetric in P and Q and which has the property that, for any continuous function $f(P)$, the function

$$y(P) = \int G(P, Q) f(Q)\,dQ$$

is the unique solution, if it exists, of the equation $Ly = f$ which satisfies (C). We also assume that the integral $\iint G^2(P, Q)\,dPdQ$ is finite and that the integral $\int G^2(P, Q)\,dQ$ is uniformly bounded for all P. Then any admissible function ϕ can be represented by the uniformly convergent series $\sum_{i=1}^{\infty} c_i y_i$, where $c_i = (\phi, y_i)$. Since

Received February 25, 1952.

Pacific J. Math. 3 (1953), 153-164

$$(L\phi, y_i) = (\phi, Ly_i) = \lambda_i c_i,$$

we also have the Parseval equations

$$(\phi, L\phi) = \sum_{i=1}^{\infty} \lambda_i c_i^2$$

and

$$(L\phi, L\phi) = \sum_{i=1}^{\infty} \lambda_i^2 c_i^2.$$

The eigenvalues are characterized by the following minimum property[1]:

$$(2) \qquad \lambda_k = \min \frac{(\phi, L\phi)}{(\phi, \phi)} \qquad [\phi \in \mathfrak{A}, (\phi, y_i) = 0,\ i = 1, 2, \cdots, k-1];$$

and by the following maximum-minimum property[2] in which the functions v_i are any continuous functions:

$$(3) \qquad \lambda_k = \max_{v_i} \min \frac{(\phi, L\phi)}{(\phi, \phi)} \qquad [\phi \in \mathfrak{A}, (\phi, v_i) = 0,\ i = 1, 2, \cdots, k-1].$$

The minimum property (2) forms the basis of the Rayleigh-Ritz method of approximating the eigenvalues and eigenfunctions.

Let $\{\psi_i\}$ be a sequence of independent admissible functions; and let V_n be the class of all functions which are linear combinations of $\psi_1, \psi_2, \cdots, \psi_n$. If we ask for the minimum of $(\phi, L\phi)/(\phi, \phi)$ under the condition that $\phi \in V_n$, we are led to the following nth order equation in λ:

$$(4) \qquad |b_{ij} - \lambda a_{ij}|_1^n = 0,$$

where $a_{ij} = (\psi_i, \psi_j)$ and $b_{ij} = (\psi_i, L\psi_j)$; and, for each root λ of (4), there is a corresponding function $\phi \in V_n$, not identically zero, such that

$$(\phi, L\phi) = \lambda(\phi, \phi).$$

[1] This follows easily from the Parseval equation for $(\phi, L\phi)$. See also [1 vol. I, pp. 345-348] and [8, pp. 10-11].

[2] See [1 vol. I, pp. 351-353] or [8, pp. 12-13].

Let the n roots of (4) be denoted by

$$\mu_i = \mu_i(n) \text{ with } 0 < \mu_1 \le \mu_2 \le \cdots \le \mu_n;$$

and let the corresponding functions be $f_1, f_2, \cdots, f_n$, where

$$f_i = f_i(P) = f_i(P; n),$$

chosen so that $(f_i, f_j) = \delta_{ij}$. We then have, for $k = 1, 2, \cdots, n$:

$$(5) \qquad \mu_k = \min \frac{(\phi, L\phi)}{(\phi, \phi)} \qquad [\phi \in V_n,\ (\phi, f_i) = 0,\ i = 1, 2, \cdots, k-1];$$

and

$$(6) \qquad \mu_k = \max_{v_i} \min \frac{(\phi, L\phi)}{(\phi, \phi)} \qquad [\phi \in V_n,\ (\phi, v_i) = 0,\ i = 1, 2, \cdots, k-1].$$ [3]

2. **Reduction to least square method.** It can easily be seen, by comparing (3) and (6), that $\lambda_k \le \mu_k$; but there is no simple method as yet for estimating the difference $\mu_k - \lambda_k$. We shall derive here an estimate for this difference which, for its application, depends on the solution of another problem in least square approximation.

Let us consider the problem of minimizing the quotient $(L\phi, L\phi)/(\phi, \phi)$ under the condition $\phi \in V_n$. This problem leads to the following equation in λ:

$$(7) \qquad |C_{ij} - \lambda a_{ij}|_1^n = 0,$$

where $C_{ij} = (L\psi_i, L\psi_j)$. Let the roots of (7) be denoted by

$$\nu_k^2 = \nu_k^2(n) \text{ with } 0 < \nu_1 \le \nu_2 \le \cdots \le \nu_n .$$

For each k from 1 to n there is a corresponding function

$$g_k = g_k(P) = g_k(P; n) \in V_n,$$

such that $(g_i, g_j) = \delta_{ij}$ and

[3] Equations (5) and (6) follow from the extremum properties of the eigenvalues and eigenvectors of quadratic forms. See [1, vol. I, pp. 26-27].

$$(8) \qquad (Lg_k, Lg_k) = \nu_k^2 .$$

The numbers ν_k^2 are characterized by the properties:

$$(9) \qquad \nu_k^2 = \min \frac{(L\phi, L\phi)}{(\phi, \phi)} \qquad [\phi \in V_n, (\phi, g_i) = 0, i = 1, 2, \cdots, k-1],$$

and

$$(10) \qquad \nu_k^2 = \max_{v_i} \min \frac{(L\phi, L\phi)}{(\phi, \phi)} \quad [\phi \in V_n, (\phi, v_i) = 0, i = 1, 2, \cdots, k-1].$$

By the Schwarz inequality, we have

$$\frac{(\phi, L\phi)^2}{(\phi, \phi)^2} \leq \frac{(L\phi, L\phi)}{(\phi, \phi)} ;$$

and, therefore, by (6) and (10), we get

$$(11) \qquad \mu_k \leq \nu_k, \; k = 1, 2, \cdots, n.$$

Now let us consider the eigenvalue problem associated with an integral operator over the region R. We assume that the function $K(P, Q)$ is sufficiently regular so as to give rise to a completely continuous operator in the Hilbert space sense, and we write the equation

$$(12) \qquad y(P) = \lambda \int K(P, Q) y(Q) dQ.$$

We also assume that $K(P, Q)$ is symmetric in P and Q and that

$$\iint K(P, Q)\, \phi(P)\, \phi(Q) dP dQ > 0$$

for any continuous function ϕ which is not identically zero. The eigenvalues of (12) are then all positive and will be denoted by $\{l_i\}$, with $0 < l_1 \leq l_2 \leq l_3 \cdots$.

Let $\{w_i\}$ be a complete orthonormal set of continuous functions on R, and let

$$(13) \qquad A_n(P, Q) = \sum_{i,j=1}^{n} \alpha_{ij}\, w_i(P)\, w_j(Q),$$

where

$$\alpha_{ij} = \iint K(P, Q)\, w_i(P)\, w_j(Q)\, dP dQ. \tag{14}$$

Then $A_n(P, Q)$ is the best approximation to $K(P, Q)$ in the L^2 sense over $R \times R$ by a sum of the given form. The integral equation

$$y(P) = \lambda \int A_n(P, Q)\, y(Q)\, dQ \tag{15}$$

will have eigenfunctions of the form $\sum_{i=1}^{n} \beta_i w_i$, and its eigenvalues will be the roots of the equation

$$|\delta_{ij} - \lambda \alpha_{ij}|_1^n = 0, \tag{16}$$

which we shall denote by $u_1 \leq u_2 \leq \cdots \leq u_n$. We can now make an estimate for the differences of corresponding eigenvalues of equations (12) and (15) by using the minimum-maximum principle for the eigenvalues of an integral equation with a symmetric kernel.[4]

Let $z_1, z_2, \cdots, z_n$ be the eigenfunctions of (15), assumed to be orthonormal. Then, letting ϕ be a continuous function subject to the conditions

$$(\phi, \phi) = 1, \ (\phi, z_i) = 0 \qquad (i = 1, 2, \cdots, k-1),$$

we have:

$$\begin{aligned} l_k^{-1} &\leq \max_{\phi} \iint K(P, Q)\, \phi(P)\, \phi(Q)\, dP dQ \\ &\leq \max_{\phi} \iint A_n(P, Q)\, \phi(P)\, \phi(Q)\, dP dQ + \max_{\phi} \iint (K - A_n)\, \phi(P)\, \phi(Q)\, dP dQ. \end{aligned}$$

The first term on the right is u_k^{-1}, and we apply the Schwarz inequality to the second term. Hence

$$l_k^{-1} \leq u_k^{-1} + \epsilon, \tag{17}$$

where

$$\epsilon = \epsilon(n) = \left[\iint (K - A_n)^2\, dP dQ \right]^{1/2}.$$

In order to connect the original differential problem with the least square method for integral operators, we let

[4] This estimate and the method used in its derivation are given by Aronszajn [6] for completely continuous, positive definite operators in Hilbert space.

$$K(P, Q) = \int G(P, X)\, G(X, Q)\, dX, \tag{19}$$

so that $l_k = \lambda_k^2$, and we assume that the functions $\{\psi_i\}$ used in the variational problems are related to the functions $\{w_i\}$ which are used in the least square problem by

$$w_i = L\psi_i \qquad (i = 1, 2, 3, \cdots). \tag{20}$$

Then, since

$$\psi_i(P) = \int G(P, Q)\, w_i(Q)\, dQ,$$

(14) becomes

$$\alpha_{ij} = (\psi_i, \psi_j) = a_{ij},$$

and

$$C_{ij} = (L\psi_i, L\psi_j) = (w_i, w_j) = \delta_{ij}.$$

Hence, equation (16) becomes identical with (7) and $u_k = \nu_k^2$, $k = 1, 2, \cdots, n$. Therefore, from (11) and (17), we obtain

$$\lambda_k^{-2} - \epsilon \le \nu_k^{-2} \le \mu_k^{-2} \le \lambda_k^{-2} \qquad (k = 1, 2, \cdots, n). \tag{21}$$

This inequality shows that, for any fixed k,

$$\lim_{n\to\infty} \nu_k(n) = \lim_{n\to\infty} \mu_k(n) = \lambda_k.$$

The problem of getting an actual estimate on $\mu_k - \lambda_k$ or on $\nu_k - \lambda_k$ is reduced to that of getting an estimate on $\epsilon(n)$. There is probably no general way of treating this problem since the regularity properties of the function $K(P, Q)$ and the possible choices of the sequence $\{w_i\}$ depend on the special nature of a given problem. From the practical point of view, the choice of the sequence $\{w_i\}$ is limited by the fact that the corresponding sequence $\{\psi_i\}$ must lend itself easily to numerical computations. In this paper we shall leave this problem to one side and consider only how estimates can be made in terms of $\epsilon(n)$.

3. **Uniform approximations.** We now take up the problem of uniform approximation of the eigenfunctions.[5] There does not seem to be any simple condition

[5] This problem has been studied in various cases by Courant [2, 3, and 4].

on the sequence $\{\psi_i\}$ which will guarantee uniform convergence of the functions $f_k(P;\,n)$ to corresponding eigenfunctions. However, on the basis of the assumptions already made, we can prove such convergence for the functions $g_k(P;\,n)$. Stated precisely, it will be shown that the difference between g_k and some eigenfunction corresponding to λ_k is, for fixed k, of the order of magnitude $\epsilon^{1/2}$, uniformly over the region R.

The first step in the proof is to establish mean convergence. For this purpose we use the analogue of (2) with $(\phi,\,L\phi)$ replaced by $(L\phi,\,L\phi)$ and λ_k replaced by λ_k^2.[6] Let us assume that

$$\lambda_1 = \lambda_2 = \cdots = \lambda_K < \lambda_{K+1},$$

and consider the function

$$\phi = g_1 - (g_1,\,y_1)y_1 - \cdots - (g_1,\,y_K)y_K.$$

This function satisfies (C) and is orthogonal to $y_1,\,y_2,\,\cdots,\,y_K$; hence

$$(L\phi,\,L\phi) \geq \lambda_{K+1}^2\,(\phi,\,\phi).$$

This gives

$$\nu_1^2 - \lambda_1^2 \sum_{i=1}^{K} (g_1,\,y_i)^2 \geq \lambda_{K+1}^2 \left[1 - \sum_{i=1}^{K} (g_1,\,y_i)^2\right];$$

hence

$$\sum_{i=1}^{K} (g_1,\,y_i)^2 \geq \frac{\lambda_{K+1}^2 - \nu_1^2}{\lambda_{K+1}^2 - \lambda_1^2} = 1 - e_1, \tag{22}$$

where

$$e_1 = e_1(n) = \frac{\nu_1^2 - \lambda_1^2}{\lambda_{K+1}^2 - \lambda_1^2}; \tag{23}$$

and when $\epsilon(n)$ is sufficiently small, e_1 will be less than a fixed constant times $\epsilon(n)$.

[6] This minimum principle can be deduced from the Parsèval equation for $(L\phi,\,L\phi)$.

Because of the multiplicity of the eigenvalue λ_1, the eigenfunctions y_1, $y_2, \cdots, y_\kappa$ are determined only to within an orthogonal transformation. We could equally well take a set $Y_1, Y_2, \cdots, Y_\kappa$ for the first κ eigenfunctions, in which

$$Y_1 = \frac{(g_1, y_1)y_1 + \cdots + (g_1, y_\kappa)y_\kappa}{[(g_1, y_1)^2 + \cdots + (g_1, y_\kappa)^2]^{1/2}} \quad ,$$

so that, from (22),

$$(24) \qquad (g_1, Y_1) \geq (1 - e_1)^{1/2} \geq 1 - e_1 ,$$

where it is assumed that n is so large that $e_1 < 1$.

Let us now assume that it is possible to choose the eigenfunctions $y_1, y_2, \cdots$ in such a way that

$$(25) \quad (g_i, y_i) \geq (1 - e_i)^{1/2} \qquad (i = 1, 2, \cdots, k),$$

where e_i is less than a constant, depending only on i, times $\epsilon(n)$; and that $e_i < 1$ $(i = 1, 2, \cdots, k < n)$. Then

$$(g_i, y_i) \geq 1 - e_i, \ \|g_i - y_i\|^2 \leq 2e_i ,$$

and

$$(26) \ (g_{k+1}, y_i)^2 = (g_{k+1}, y_i - g_i)^2 \leq \|y_i - g_i\|^2 \leq 2e_i \quad (i = 1, 2, \cdots, k).$$

Let

$$\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_{k+M} < \lambda_{k+M+1},$$

and let

$$\phi = g_{k+1} - \sum_{i=1}^{k+M} (g_{k+1}, y_i)y_i .$$

Since ϕ satisfies (C) and is orthogonal to $y_1, y_2, \cdots, y_{k+M}$, we have

$$(L\phi, L\phi) \geq \lambda_{k+M+1}^2 (\phi, \phi);$$

hence

$$\nu_{k+1}^2 - \sum_{i=1}^{k} \lambda_i^2 (g_{k+1}, y_i)^2 - \lambda_{k+1}^2 \sum_{i=k+1}^{k+M} (g_{k+1}, y_i)^2 \geq \lambda_{k+M+1}^2 \left\{ 1 - \sum_{i=1}^{k+M} (g_{k+1}, y_i)^2 \right\}. \tag{27}$$

From (27) we get

$$\sum_{i=k+1}^{k+M} (g_{k+1}, y_i)^2 \geq 1 - \frac{\nu_{k+1}^2 - \lambda_{k+1}^2}{\lambda_{k+M+1}^2 - \lambda_{k+1}^2} - \frac{\sum_{i=1}^{k} (\lambda_{k+M+1}^2 - \lambda_i^2)(g_{k+1}, y_i)^2}{\lambda_{k+M+1}^2 - \lambda_{k+1}^2} = 1 - e_{k+1}, \tag{28}$$

where, by (26),

$$e_{k+1} \leq \frac{\nu_{k+1}^2 - \lambda_{k+1}^2 + 2\sum_{i=1}^{k} (\lambda_{k+M+1}^2 - \lambda_i^2) e_i}{\lambda_{k+M+1}^2 - \lambda_{k+1}^2}. \tag{29}$$

By a suitable orthogonal transformation we can carry the eigenfunctions $y_{k+1}, \cdots, y_{k+M}$ into a new set $Y_{k+1}, \cdots, Y_{k+M}$, where

$$Y_{k+1} = \frac{(g_{k+1}, y_{k+1}) y_{k+1} + \cdots + (g_{k+1}, y_{k+M}) y_{k+M}}{\{(g_{k+1}, y_{k+1})^2 + \cdots + (g_{k+1}, y_{k+M})^2\}^{1/2}}$$

and (28) then becomes

$$(g_{k+1}, Y_{k+1}) \geq (1 - e_{k+1})^{1/2}.$$

We see, therefore, that, for any fixed value of k and for n sufficiently large, the function $g_k(P; n)$ differs in the mean from some eigenfunction corresponding to λ_k by an amount which is less than a constant, depending only on k, times $\epsilon^{1/2}(n)$. In this statement, the phrase "for n sufficiently large" is needed to ensure that $e_i < 1$ and that $\nu_i^2 - \lambda_i^2$ is less than a constant, depending only on i, times $\epsilon(n)$ for $i = 1, 2, \cdots, k$. The latter condition is guaranteed by (21) if $\epsilon(n) < \lambda_i^{-2}$. The actual numerical estimates are obtained from (21), (23), and (29).

Let us now consider the uniform approximation of g_k to y_k under the assumption

$$(g_k, y_k) \geq (1 - e_k)^{1/2} \geq 1 - e_k ,$$

where $e_k = e_k(n) = O(\epsilon(n))$. We have

$$g_k(P) = \int G(P, Q)\, Lg_k(Q)\, dQ,$$

$$y_k(P) = \lambda_k \int G(P, Q)\, y_k(Q)\, dQ.$$

By subtraction and the Schwarz inequality, we get

$$(30) \quad |g_k(P) - y_k(P)| \leq M\|Lg_k - \lambda_k y_k\| = M\{\nu_k^2 - 2\lambda_k^2(g_k, y_k) + \lambda_k^2\}^{1/2}$$

$$\leq M\{\nu_k^2 - 2\lambda_k^2(1 - e_k) + \lambda_k^2\}^{1/2} = M\{\nu_k^2 - \lambda_k^2 + 2\, e_k\, \lambda_k^2\}^{1/2}.$$

where

$$M = \text{l.u.b.}\ \{\int G^2(P, Q)\, dQ\}^{1/2}.$$

Hence, for n sufficiently large, $|g_k(P; n) - y_k(P)|$ is less than a constant, depending only on k, times $\epsilon^{1/2}(n)$.

It is possible to carry through the proof of mean convergence for the functions $f_k(P; n)$ by the same type of argument as is used above for the functions $g_k(P; n)$. The only changes necessary are to replace $(L\phi, L\phi)$ by $(\phi, L\phi)$, λ_i^2 by λ_i, ν_i^2 by μ_i, and g_i by f_i. The argument used for uniform convergence of g_k, however, does not go through for f_k. This is an illustration of a principle which has been discussed by Courant: namely, that in the solution of a differential problem by variational methods, the more weight given to the higher derivatives, the better the results in the way of uniform approximations.

There are, however, some problems in which the functions $f_k(P; n)$ can be shown to give uniform approximations to eigenfunctions with an error that can be estimated in terms of $\epsilon(n)$. We shall consider a class of such problems;

namely, those for which the Green's function $G(P,Q)$ is bounded, say $G(P,Q) \leq B$. This class will include, for example, the usual one-dimensional problems and the two-dimensional problem of the normal modes of vibration of a clamped plate.

We shall use the O notation here since it is obvious how explicit estimates may be obtained from the methods used. The function f_k can be represented by the uniformly convergent series $\sum_{i=1}^{\infty} c_i y_i$, where $c_i = (f_k, y_i)$ and where it is assumed that $c_k = 1 - O(\epsilon)$. The Parseval equation gives

$$\mu_k = (f_k, Lf_k) = \sum_{i=1}^{\infty} c_i^2 \lambda_i .$$

Letting $\sum'$ stand for $\sum_{i=1}^{\infty}$, the term with $i = k$ being omitted, we have

$$\mu_k - \lambda_k = \sum{}' c_i^2 \lambda_i + (c_k^2 - 1)\lambda_k ,$$

from which it follows that

$$\sum{}' c_i^2 \lambda_i = \mu_k - \lambda_k + (1 - c_k^2)\lambda_k = O(\epsilon). \tag{31}$$

We now write

$$f_k - y_k = \sum{}' c_i y_i + (c_k - 1)y_k \tag{32}$$

and estimate the first term on the right side as follows:

$$|\sum{}' c_i y_i|^2 \leq \sum{}' c_i^2 \lambda_i \sum{}' \frac{y_i^2}{\lambda_i} = O(\epsilon),$$

by (31) and from the fact that

$$\sum{}' \frac{y_i^2}{\lambda_i} \leq G(P, P) \leq B.$$

Hence $|f_k - y_k| = O(\epsilon^{1/2})$ uniformly over the region R.

REFERENCES

1. R. Courant and D. Hilbert, *Methoden der mathematischen Physik,* Second Edition, vol. I, 1931; vol. II, 1937.

2. R. Courant, *Über direkte Methoden in der Variationsrechnung und über verwandte Fragen,* Math. Ann. 97 (1927), 711-736.

3. ———, *Über ein convergenzerzeugendes Prinzip der Variationsrechnumg,* Nachr. Akad. Wiss. Göttingen. Math.-Phys. Kl. 1922, 144-150.

4. ———, *Variational methods for the solution of problems of equilibrium and vibrations,* Bull. Amer. Math. Soc. 49 (1943), 1-23.

5. L. Collatz, *Eigenwertprobleme und ihre numerische Behandlung,* Mathematik und ihre Anwendungen in Physik und Technik, Reihe A, Band 19. Akademische Verlagsgesellschaft, Leipzig, 1945.

6. N. Aronszajn, *Studies in Eigenvalue Problems. The Rayleigh-Ritz and Weinstein methods for approximation of eigenvalues. I. Operators in a Hilbert Space.* Dept. of Mathematics, Oklahoma A. & M., Stillwater, Oklahoma.

7. ———, *The Rayleigh-Ritz and A. Weinstein methods for approximation of eigenvalues.* I. *Operators in a Hilbert Space.* II. *Differential Operators.* Proc. Nat. Acad. Sci. U.S.A. 34 (1948), 474-480, 596-601.

8. A. Weinstein, *Étude des spectres des équations aux dérivées partielles de la théorie des plaques élastiques,* Mémorial des Sciences Mathématiques, no. 88, 1937.

UNIVERSITY OF MINNESOTA