FUNCTIONAL ANALYSIS AND GALERKIN'S METHOD

Lamberto Cesari

INTRODUCTION

In the present paper we discuss a process for the existence analysis concerning solutions of linear and nonlinear equations $Kx = y$ in functional spaces (Chapters 1 and 2). The method reduces the problem to the study of a *finite* system of transcendental equations (determining system) in a finite-dimensional Euclidean space; the system in turn is analyzed by considerations based on the topological index of a mapping. The method is connected with Galerkin's method of successive approximations, which is often applied to cases where the existence of an exact solution is not known. In this situation the process may give an answer to two questions: 1. If a certain m^{th} approximation $x^{(m)}$ is known, is it possible to argue whether an exact solution X also exists? 2. If the answer to (1) is in the affirmative, is it possible to give an upper estimate for the difference $X - x^{(m)}$ (error bound)?

In association with these problems we mention Kantorovič's theory [7], [8], [9] for linear equations of the form $Kx = y$, and Keldiš's proof [10] of the convergence of Galerkin's method in linear differential problems with boundary values (for which the exact solution is known to exist). Also, let us mention the work of Schmidt [15] and Bartle [2] on functional equations, and Rothe's analysis [13], [14] of gradient mappings and topological order in Banach's spaces.

In Chapter 1 we discuss a few points of the present approach, remaining, as far as possible, in the frame of a normed linear space (the setting of the first pages of our previous paper [4] was only slightly more general). In Chapter 2, with a view toward applications, we recast the present approach in the frame of a separable, complete, real Hilbert space, and under enough assumptions so as to simplify as much as possible the final statement (Theorem viii of Section 10).

In Chapter 3 we apply the process as discussed in Chapter 2 to a particular problem, namely, the nonlinear ordinary differential equation

$$x'' + x + \alpha x^3 = \beta t \qquad (0 \leq t \leq 1)$$

with homogeneous boundary conditions $x(0) = 0$, $x'(1) + hx(1) = 0$. For $h = 1$, we show that a solution exists for all $|\alpha| \leq 1$, $|\beta| \leq 1$. The process is then applied to the study of the first Galerkin approximation to the solution of the same problem, for particular values of h, α, and β, and numerical error bounds are obtained.

In other papers we shall apply the same process to an existence analysis for nonlinear partial differential equations for smooth or generalized solutions.

In previous papers [4], [5] we have applied the process of the present paper to questions concerning existence, approximation, and error bounds for periodic solutions of periodic (or autonomous) nonlinear ordinary differential equations. The numerical case discussed there provides another exemplification of the process that

---

Received November 26, 1963.

Research partially supported by NSF research grant GP-57 and the Institute of Science and Technology of the University of Michigan, Ann Arbor, Michigan.

is here given in a more general setting. In that example we used the same approach to prove the existence of an exact solution X, first independently from any Galerkin approximation and then in association with the second Galerkin approximation ($m = 2$), to obtain an upper estimate of $x^{(2)} - X$.

H. W. Knobloch [11], [12] applies the approach given in the present paper and in [4], [5] to a qualitative existence analysis of periodic solutions of certain types of second-order (not necessarily linear) ordinary differential equations. The results Knobloch obtains by this approach are similar in form and partially overlapping in extension with others obtained by Barbalat by means of Ważewski's topological method.

C. Borges [3] applies the process of the present paper to obtain existence and error bounds for periodic solutions of nonlinear (periodic, and autonomous) second-order ordinary differential equations in association with the first Galerkin approximation ($m = 1$, method of harmonic balance).

For linear equations derived from the problem of the minimum of integrals of quadratic forms, Galerkin's method is known to coincide with Ritz's method (see [22], for example).

THE m^{th} GALERKIN APPROXIMATION

1. Let S be a separable, real, complete Hilbert space of elements $x, y, \cdots$. Let $x \cdot y$ and $\|x\|$ denote respectively the inner product and norm in S, and let $(\phi_1, \phi_2, \cdots, \phi_n, \cdots)$ be a complete orthonormal system of S. Let $K: S_K \to S$ be a mapping (not necessarily linear) from a subset S_K of S into S. We shall consider an equation of the form

$$Kx = 0. \tag{1}$$

Since K is not necessarily linear, this problem includes equations of the form $Lx - y = 0$.

Often Galerkin's method is applied; that is, approximate solutions to $Kx = 0$ are sought that have the form $x = c_1 \phi_1 + \cdots + c_m \phi_m$ and satisfy the equations

$$K\left(\sum_{j=1}^{m} c_j \phi_j\right) \cdot \phi_k = 0 \qquad (i = 1, \cdots, m). \tag{2}$$

If $c_1^{(m)}, c_2^{(m)}, \cdots, c_m^{(m)}$ is a solution (in case one exists) of this system of m equations in the m unknowns $c_1, \cdots, c_m$, then

$$x^{(m)} = c_1^{(m)} \phi_1 + \cdots + c_m^{(m)} \phi_m$$

is said to be an m^{th} *Galerkin approximate solution* to equation (1). Often the mapping K is replaced by another mapping K_0 (close to K), and the m^{th} Galerkin approximation is obtained by replacing K by K_0 in (2).

CHAPTER 1. NONLINEAR EQUATIONS IN FUNCTIONAL SPACES

2. We shall assume first that S is a normed linear space of elements x and norm $\| \ \|$. Let E, N, $K = E - N$, P, H be mappings, and let S_E, S_N, S_K, S_0, and S_1 be subsets of S satisfying the following hypotheses:

(Ia) $E: S_E \to S$ *is a linear operator (not necessarily bounded),* $N: S_N \to S$ *is an operator (not necessarily linear),* $K: S_K \to S$, *where* $S_K = S_E \cap S_N \neq \emptyset$ *and* $K = E - N$.

(Ib) S *is the (topological) direct sum of the (closed) subspaces* S_0 *and* S_1. *Let* $P: S \to S_0$ *be the (bounded) projection of* S *with null-space* S_1 *and range* S_0.

(Ic) $H: S_1 \to S_1$ *is a linear operator such that*

$$H(I - P)Ex = (I - P)x \quad \textit{for all } x \in S_E, \tag{3}$$

where I *is the identity operator on* S.

In the applications, S may for instance be the class of all continuous functions, or of all functions $x(\alpha)$ $(\alpha \in A)$, L^2-integrable over a domain A, and E may be a differential operator of order M. Then S_E will be a subset of S composed of functions $x(\alpha)$ sufficiently "smooth" so that $Ex(\alpha)$ is defined and is an element of S; that is, $E: S_E \to S$, as assumed in (Ia). The operator N is not necessarily linear and need not be defined on the whole of S. Actually, we shall consider N as operating only on certain subsets of S defined by means of limitations of the form $\|x\| \leq M$, or analogous ones. Below, we shall have to assume that the subsets we need consider are actually contained in S_N. Since $K = E - N$, equation (1) can be written in the form $Ex = Nx$. The assumption (Ic) states essentially that H is a partial left inverse of E (see Remark 1 below).

3. First let us note that $PPx = Px$ for every $x \in S$ and $Px = x$ for every $x \in S_0$. Also, every $x \in S$ admits a unique decomposition $x = x_0 + y$, $x_0 \in S_0$, $y \in S_1$. If I is the identity operator in S, then $I - P: S \to S_1$, and $P(I - P)x = 0$ for each $x \in S$. Also, $H: S_1 \to S_1$, and hence

$$PHx = 0 \quad \text{for every} \quad x \in S_1, \tag{4}$$

$$PH(I - P)x = 0 \quad \text{for every} \quad x \in S. \tag{5}$$

Remark 1. If $H: S \to S$, and H is a left inverse of E that commutes with P, that is, $HEx = x$ for every $x \in S_E$ and $HPx = PHx$ for every $x \in S$, then in S_E

$$H(I - P)E = HE - HPE = HE - PHE = I - P.$$

Besides, by force of the relations

$$HPx = PHx, \qquad H(I - P)x = (I - P)Hx \qquad (x \in S),$$

we deduce that $H: S_0 \to S_0$ and $H: S_1 \to S_1$ in this particular situation. Thus (Ic) holds in this particular situation.

We have already noticed that, by force of (Iabc), equation (1) $Kx = 0$ becomes

$$Ex = Nx. \tag{6}$$

If y is a solution of this equation, then $Ey = Ny$, $y \in S_E \cap S_N$, and, by applying the operator $H(I - P)$, we see, again by force of (Iabc), that

$$H(I - P)Ey = H(I - P)Ny,$$

$$(I - P)y = H(I - P)Ny.$$

If F denotes the operator $F = H(I - P)N$, then $F: S_N \to S_1$, and

$$y = Py + Fy = Py + H(I - P)Ny. \tag{7}$$

If $T = P + F$, then $y = Ty$; that is, y is a fixed point of the transformation $T: S_N \to S$ defined by

$$T = P + H(I - P)N. \tag{8}$$

4. For given constants c, d $(0 < c < d)$, let V be the set

$$V = [x \,|\, x \in S,\ \|Px\| \leq c].$$

For a given element x^* of the set V, let S^* be the set

$$S^* = [x \,|\, x \in S,\ Px = Px^*, \text{ and } \|x\| \leq d]. \tag{9}$$

The set S^* is not empty, since $Px^* \in S^*$. Besides (Iabc), we now need the further assumption that

(Id) *for each element* $x^* \in V$,

$$S^* \subset S_N, \textit{ and } \quad \|Fx\| \leq d - c \textit{ for every } x \in S^*.$$

Let us prove that hypotheses (Iabcd) imply that $T: S^* \to S^*$. Indeed, if $y = Tx$, $x \in S^*$, then

$$y = Tx = Px + Fx = Px + H(I - P)Nx,$$

$$Py = PPx + PH(I - P)Nx = Px = Px^*,$$

$$\|y\| \leq \|Px^*\| + \|Fx\| \leq c + d - c = d.$$

If, in addition, $F|\, S^*$ is a contraction, that is,

$$\|Fx_1 - Fx_2\| \leq k\|x_1 - x_2\| \tag{10}$$

for all x_1, $x_2 \in S^*$ and some constant $k < 1$, then $T: S^* \to S^*$ is also a contraction. Indeed,

$$y_i = Tx_i = Px_i + Fx_i \quad (i = 1, 2,),\ Px_1 = Px_2 = Px^*, \text{ and}$$

$$\|y_1 - y_2\| = \|Fx_1 - Fx_2\| \leq k\|x_1 - x_2\|.$$

By fixed-point theorems of Schauder and Banach, respectively, we obtain the two theorems below.

THEOREM (i). *Under hypotheses* (Iabcd), $T: S^* \to S^*$. *If* S^* *is compact (or* S^* *is complete and* $T(S^*)$ *is compact), then* $T: S^* \to S^*$ *has at least one fixed point* $y = Ty$ *in* S^*.

THEOREM (ii). *Under hypotheses* (Iabcd), *if* $F|\, S^*$ *is a contraction, then* $T: S^* \to S^*$ *is also a contraction. If* S^* *is complete, then* T *has exactly one fixed point* $y = Ty$ *in* S^*.

5. We now need a new assumption, in order to deduce important properties of any possible fixed points of $T: S^* \to S^*$.

(Ie) *For each fixed point* $y = Ty \in S^*$ *of the transformation* T *(if any exist),* $y \in S_E$, $Py \in S_E$, *and the following relations hold:*

$$EPy = PEy, \tag{11}$$

$$EH(I - P)Ny = (I - P)Ny. \tag{12}$$

Assume now that all hypotheses (Iabcde) are satisfied, so that $T: S^* \to S^*$. Let $y = Ty$ be any fixed element (if any exists) of $T|\, S^*$. Then $y = Ty \in S_E \cap S^*$, and

$$y = Ty = Py + Fy = Py + H(I - P)Ny.$$

Since $y \in S_E$ and $Py \in S_E$, we see that $H(I - P)Ny \in S_E$. Applying to both sides the operator E, we successively conclude by (Ie) that

$$Ey = EPy + EH(I - P)Ny,$$

$$Ey = PEy + (I - P)Ny,$$

$$Ey = Ny + P(E - N)y,$$

that is, $Ky = PKy$. If $p = PKy$, then we conclude that any fixed point $y = Ty$ of $T|\, S^*$ satisfies the equation

$$Ky = p, \tag{13}$$

where $p = PKy$ is an element of S_0.

In order to establish the desired connection with Galerkin's method, we shall now restrict the generality of the space S by means of a new assumption:

(If) S *is a separable, real, complete, Hilbert space,* $(\phi_1, \phi_2, \cdots)$ *is a complete orthonormal system in* S, *and* $(\phi_1, \cdots, \phi_m)$ *spans* S_0.

Under the hypotheses (Iabcdef), we now see that

$$p \equiv PKy = (Ky \cdot \phi_1)\phi_1 + \cdots + (Ky \cdot \phi_m)\phi_m.$$

We conclude with the following theorem.

THEOREM (iii). *Under hypotheses* (Iabcdef), *if* $y = Ty$ *is any fixed element of* $T: S^* \to S^*$, *then* y *satisfies the equation* $Kx = 0$ *if and only if the* m *Galerkin-like equations*

$$U_i \equiv Ky \cdot \phi_i = 0 \qquad (i = 1, \cdots, m) \tag{14}$$

hold.

These will be called the *determining equations* of equation (1).

CHAPTER 2. THE EXISTENCE THEOREM

6. In order to prove an existence theorem for solutions of the equation $Kx = 0$, and with a view toward applications, we prefer to recast the previous consideratioı in a different setting. In a way, the new setting is more complete because it takes into consideration an "approximate" solution x_0 and an "approximate" equation $K_0 x = 0$ besides $Kx = 0$. On the other hand, it is from the beginning placed in a real, complete, separable Hilbert space.

Let S be a real, complete, separable Hilbert space, let $(\phi_1, \phi_2, \cdots)$ be a complete orthonormal system in S, let E, N, N_0, K, K_0, P, H be mappings, and S_E, S_N, S_K, S_0, S_1 subsets of S, satisfying the following hypotheses:

(IIa) E: $S_E \to S$ *is a linear operator (not necessarily bounded),* N *and* N_0 *are operators (not necessarily linear) from the same subset* S_N *of* S *into* S, $K = E - N$, $K_0 = E - N_0$, K *and* K_0 *map* S_K *into* S, *and* $S_K = S_E \cap S_N \neq \emptyset$.

(IIb) S *is the (topological) direct sum of the (closed) subspaces* S_0 *and* S_1 *with* S_0 *spanned by* $(\phi_1, \cdots, \phi_m)$ $(1 \leq m < \infty)$ *and* $S_0 \subset S_E$. *Let* P: $S \to S_0$ *be the (bounded) projection of* S *with null-space* S_1 *and range* S_0.

(IIc) This hypothesis is the same as (Ic).

Let x_0 denote any element of $S_E \cap S_N$. Actually, we think of N_0 as an operator "close" to N, and of x_0 as an "approximate" solution of the approximating equation $K_0 x \equiv (E - N_0)x = 0$.

7. We may assume that the Hilbert space S, besides the norm $\| \ \|$, has one or more seminorms $\mu(x)$ (for which the value $+\infty$ is not excluded). We shall assume that N and N_0 are at least defined in a "neighborhood" $\bar{S}_0$ of x_0. Precisely, corresponding to numbers c, d, r, and R_0 with $0 < c < d$ and $0 < r < R_0$, let $\bar{S}_0$ be the set

$$\bar{S}_0 = [x \mid x \in S, \ \|x - x_0\| \leq d, \ \mu(x - x_0) \leq R_0]. \tag{15}$$

We shall assume in (IId) below that $\bar{S}_0 \subset S_N$. Obviously, $\bar{S}_0$ is not empty since it contains x_0.

We could now consider the set

$$[x \mid x \in S, \ \|P(x - x_0)\| \leq c, \ \mu P(x - x_0) \leq r], \tag{16}$$

which is not empty since it contains x_0. Most of the considerations below apply to this situation. Nevertheless, to simplify matters, we shall first assume that x_0 is actually an element of S_0, that is, x_0 is an element of the form

$$x_0 = c_{01}\phi_1 + \cdots c_{0m}\phi_m.$$

Second, we shall restrict the set (16) to elements $x \in S_0$, so that $Px = x$, $Px_0 = x_0$, and $P(x - x_0) = x - x_0$ $(x - x_0 \in S_0)$. In other words, we replace (16) by

(17) $$V = [x \mid x \in S_0,\ \|x - x_0\| \leq c,\ \mu(x - x_0) \leq r].$$

Since $0 < c < d$, $0 < r < R_0$, we see that $V \subset \overline{S}_0$.

Finally, we shall assume, again to simplify matters, that a relation of the form $z \in S_0$, $\|z\| \leq c$, implies $\mu z \leq r$ for some r, and that c and r are so related. This situation actually occurs in the applications we have in view. We could dispense with these three restrictions, but they simplify matters in the last step of the process that we shall describe. Note that with the last convention, V is actually defined by the simple conditions

(18) $$V = [x \mid x \in S_0,\ \|x - x_0\| \leq c],$$

since x, x_0, $x - x_0 \in S_0$, and $\|x - x_0\| \leq c$ imply $\mu(x - x_0) \leq r$.

For each $x^* \in V$ we now consider the set

(19) $$S_0^* = [x \mid x \in S,\ Px = Px^*,\ \|x - x_0\| \leq d,\ \mu(x - x_0) \leq R_0].$$

Note that S_0^* is not empty, since it contains x^*, and also, by comparison with (15), that $S_0^* \subset \overline{S}_0$.

Finally, we shall assume that for each $x^* \in V$ the corresponding set S_0^* is complete. This is a stronger form of an assumption mentioned in Theorem (ii) of Chapter 1.

For the reader's convenience we summarize the assumptions we have made in the present section:

(IId) $x_0 \in S_0$ *is a given element of* S_0, $0 < c < d$, $0 < r < R_0$ *are given numbers so related that* $z \in S_0$, $\|z\| \leq c$ *imply* $\mu(z) \leq r$. *If* V, $\overline{S}_0$ *are the sets defined by* (18), (15), *assume that* $\overline{S}_0 \subset S_N$. *Then* $V \subset \overline{S}_0 \subset S_N$, *and* $S_0^* \subset \overline{S}_0 \subset S_N$ *for every* $x^* \in V$. *Also, assume that for each* $x^* \in V$, *the corresponding set* S_0^* *is complete in the norm* $\| \ \|$.

Note that $x_0 \in S_0 \subset S_E$ by (IIbd), and that $x_0 \in \overline{S}_0 \subset S_N$ by (15) and (IId). Hence, $x_0 \in S_E \cap S_N$, as was assumed at the end of Section 6. Analogously, from the inclusions $V \subset \overline{S}_0 \subset S_E$ and $V \subset S_N$, we deduce that $V \subset S_E \cap S_N$.

8. In analogy with Chapter 1 we denote by F and F_0 the operators

$$F = H(I - P)N, \qquad F_0 = H(I - P)N_0,$$

and by T the operator

$$T = P + F = P + H(I - P)N.$$

Let a, a', b, b' be positive numbers such that

(20) $$\|Fx_0 - F_0x_0\| \leq a, \qquad \mu(Fx_0 - F_0x_0) \leq a', \qquad \|\Delta\| \leq b, \qquad \mu(\Delta) \leq b',$$

where Δ is the element of S defined by

$$x_0 = Px_0 + F_0x_0 + \Delta.$$

Note that, if $\theta = K_0 x_0$ denotes the error with which x_0 satisfies the approximating equation $K_0 x = 0$, then we can successively conclude

$$\theta = K_0 x_0 = Ex_0 - N_0 x_0 \qquad (\theta \in S),$$

$$(I - P)x_0 = H(I - P)Ex_0 = H(I - P)(N_0 x_0 + \theta),$$

$$x_0 = Px_0 + F_0 x_0 + \triangle, \text{ and } \triangle = H(I - P)\theta .$$

Thus, the numbers b, $b' \geq 0$ above are related to the error with which x_0 satisfies the approximating equation $K_0 x = 0$. The closeness of the operators N and N_0 is measured by the numbers a and a'.

We finally assume that the operators N and H satisfy hypotheses of continuity. Precisely, we assume

(IIe) *there exists a constant* $L \geq 0$ *such that, for any pair* x^1 *and* x^2 *in* $\bar{S}_0$,

$$\|Nx^1 - Nx^2\| \leq L\|x^1 - x^2\| . \tag{21}$$

Concerning H we could assume that

$$\|H(I - P)z\| \leq k\|z\|, \quad \mu H(I - P)z \leq k'\|z\| \tag{22}$$

for every $z \in S$. The first inequality is the usual boundedness of H(I - P), and hence of H. The second inequality is actually a property of the seminorm μ, analogous to the one already assumed in (IId). Indeed, if $z \in S$ implies $\mu(s) \leq C\|z\|$, then we could take $k' = Ck$. But all this is far too restrictive. The weaker assumption (IIf) below is sufficient for the proofs of the theorems to follow. Let Z_0 be the set $Z_0 = N(\bar{S}_0) - N(\bar{S}_0)$, in other words, the set of all $z = z^1 - z^2$ with $z^1 = Nx^1$, $z^2 = Nx^2$ $(x^1, x^2 \in \bar{S}_0)$. All we need is the requirement

(IIf) *there exist constants* $k, k' \geq 0$ *such that the inequalities* (22) *hold for all* $z \in Z_0$.

Under hypotheses (IIabcdef),

$$\|Fx^1 - Fx^2\| = \|H(I - P)(Nx^1 - Nx^2)\| \leq kL\|x^1 - x^2\| \tag{23}$$

for all $x^1, x^2 \in \bar{S}_0$; moreover,

$$\|Fx - Fx_0\| = \|H(I - P)(Nx - Nx_0)\| \leq kL\|x - x_0\| \leq kLd, \tag{24}$$

$$\mu(Fx - Fx_0) \leq k'L\|x - x_0\| \leq k'Ld$$

for all $x \in \bar{S}_0$.

9. The first part of assumption (Id) has in the present setting been replaced by the stronger assumption $\bar{S}_0 \subset S_N$ contained in (IId). Now we need an assumption that corresponds to the second part of (Id):

(IIg) *The numbers* a, a', b, b', k, k', c, d, r, *and* R_0 *satisfy the three relations*

$$kL < 1, \quad kLd < d - c - a - b, \quad k'Ld < R_0 - r - a' - b'.$$

We are now in a position to state and prove a theorem that corresponds to Theorem (ii) of Chapter 1:

THEOREM (iv). *Under hypotheses* (IIabcdefg) *and for each* $x^* \in V$, T *maps* S_0^* *into* S_0^* *and is a contraction, and hence it admits one and only one fixed point* $y = Ty \in S_0^*$, *which depends on* x^* *and is therefore a single-valued function* $y = \mathfrak{T}(x^*)$ $(x^* \in V)$; *that is,* $\mathfrak{T}: V \to S$.

First let us prove that T maps S_0^* into S_0^*. For every $x^* \in V$ and every $x \in S_0^*$ we see from (19) that $Px = Px^* = x^* \in S_0$, and, as in Chapter 1, we conclude that

$$y = Tx = Px + H(I - P)x,$$

$$Py = PPx + PH(I - P)x = Px = Px^* = x^*.$$

Also,

$$y = Px + Fx, \qquad x_0 = Px_0 + F_0 x_0 + \triangle,$$

and, by way of comparison,

$$y - x_0 = P(x - x_0) + (Fx - Fx_0) + (Fx_0 - F_0 x_0) - \triangle.$$

Hence, by (17), (24), (20), and (IIg),

$$\begin{aligned} \|y - x_0\| &\leq \|x - x_0\| + \|Fx - Fx_0\| + \|Fx_0 - F_0 x_0\| + \|\triangle\| \\ &\leq c + kLd + a + b \\ &\leq d. \end{aligned}$$

Analogously,

$$\begin{aligned} \mu(y - x_0) &\leq \mu(x - x_0) + \mu(Fx - Fx_0) + \mu(Fx_0 - F_0 x_0) + \mu(\triangle) \\ &\leq r + k'Ld + a' + b' \\ &\leq R_0. \end{aligned}$$

Thus $y = Tx \in S_0^*$ for every $x \in S_0^*$, or $T: S_0^* \to S_0^*$.

Let us prove that $T| S_0^*$ is a contraction. Indeed, for every two elements $x_1, x_2 \in S_0^*$, it follows from (19) and (23) that

$$y_i = Tx_i = Px_i + Fx_i \ (i = 1, 2), \quad Px_1 = Px_2 = Px^* = x^*, \text{ and}$$

$$\|y_1 - y_2\| = \|Fx_1 - Fx_2\| \leq kL\|x_1 - x_2\|,$$

where $kL < 1$. Thus $T| S_0^*$ is a contraction and maps S_0^* into S_0^*, and S_0^* is complete in the norm $\| \ \|$. Thus, the fixed point $y = Tx \in S_0^*$ of $T| S_0^*$ exists and is unique. Theorem (iv) is thereby proved.

THEOREM (v). *Under the hypotheses of Theorem* (iv), $y = \mathfrak{T}(x^*)$ $(x^* \in V)$, *is a continuous function of* x^*.

This proposition is essentially known. Nevertheless, we prove it here for the convenience of the reader. If x^*, x'^* are any two elements of V, then

$$y = \mathfrak{T}(x^*) \in S_0^* \subset \overline{S}_0, \qquad y' = \mathfrak{T}(x'^*) \in S_0'^* \subset \overline{S}_0,$$

$$y = Py + H(I - P)Ny, \qquad y' = Py' + H(I - P)Ny',$$

$$Py = Px^* = x^*, \qquad Py' = Px'^* = x'^*, \qquad y = \mathfrak{T}x^*, \qquad y' = \mathfrak{T}x'^*.$$

Hence,

$$\|y - y'\| \le \|Px^* - Px'^*\| + \|H(I - P)(Ny - Ny')\|,$$

where $y, y' \in \overline{S}_0$. Hence, by force of (23), we successively see that

$$\|y - y'\| \le \|x^* - x'^*\| + kL\|y - y'\|,$$

$$\|\mathfrak{T}x^* - \mathfrak{T}x'^*\| = \|y - y'\| \le (1 - kL)^{-1}\|x^* - x'^*\|.$$

This proves the continuity of $\mathfrak{T}: V \to S$ under the conditions of Theorem (iv).

10. In order to obtain further properties of the mapping $y = \mathfrak{T}(x^*)$, we need two more assumptions, the first of which corresponds to (Ie). We now presuppose (IIabcdefg).

(IIh) *For each* $x^* \in V$, *the fixed point* $y = Ty = \mathfrak{T}(x^*)$ *of* $T|S_0^*$ *is an element of* S_E, *and the relations*

$$EPy = PEy,$$

$$EH(I - P)Ny = (I - P)Ny$$

hold.

(IIi) E *has the property that there exist some elements* $\bar{\phi}_i \in S$ $(i = 1, \cdots, m)$ *such that, for all* $x \in S_E$,

$$Ex \cdot \phi_i = x \cdot \bar{\phi}_i \qquad (i = 1, \cdots, m).$$

Let us note that the requirement $Py \in S_E$ of (Ie) is not explicitly stated in (IIh), since it is a consequence of the stronger requirement $S_0 \subset S_E$ of (IIb). As noted in Section 5, $y \in S_E$, $Py \in S_E$ imply $H(I - P)Ny \in S_E$.

Hypothesis (IIh) implies that $\mathfrak{T}: V \to (S_E \cap S_N)$. Indeed, for each $x^* \in V$, we see that $y = Ty \in S_0^*$, where $S_0^* \subset \overline{S}_0 \subset S_N$ by (IId) and $y \in S_E$ by (IIh). Thus $y \in S_E \cap S_N$, and $\mathfrak{T}: V \to (S_E \cap S_N)$.

If we denote $\bar{\phi}_i$ by $G\phi_i$, then (IIi) becomes

$$Ex \cdot \phi_i = x \cdot G\phi_i \qquad (x \in S_E,\ i = 1, \cdots, m).$$

Thus, G can be extended into a linear operator $G: S_0 \to S$ such that

$$Ex \cdot z = x \cdot Gz \quad (x \in S_E,\ z \in S). \tag{25}$$

In other words, (IIi) can be stated by saying that E admits an "adjoint" operator G satisfying (25) for all $x \in S_E$ and $z \in S_0$.

Under hypothesis (IIi), we shall use the notation

$$e = \left(\sum_1^m \|\bar{\phi}_i\|^2\right)^{1/2} = \left(\sum_1^m \|G\phi_i\|^2\right)^{1/2}.$$

We can now prove the following theorem.

THEOREM (vi). *Under hypotheses* (IIabcdefgh), *for each* $x^* \in V$, *the fixed element* $y = Ty = \mathfrak{T}(x^*)$ *of* $T \mid S_0^*$ *satisfies the relation* $Ky = PKy$, *with*

$$p \equiv PKy = (Ky \cdot \phi_1)\phi_1 + \cdots + (Ky \cdot \phi_m)\phi_m.$$

Hence, $y = \mathfrak{T}(x^*)$ *is a solution of the equation* $Kx = 0$ *if and only if the* m *Galerkin-like equations* $Ky \cdot \phi_i = 0_Z$ $(i = 1, \cdots, m)$ *are satisfied.*

The proof is the same as that in Chapter 1, and we do not repeat it.

As stated in formula (18) of Section 7, V is the set of all finite expressions $x^* = c_1\phi_1 + \cdots + c_m\phi_m$ with $\|x^* - x_0\| \le c$. Let E_m, E'_m be two auxiliary m-dimensional Euclidean spaces whose points will be denoted by $\gamma = (c_1, \cdots, c_m)$ and $u = (u_1, \cdots, u_m)$, respectively. The norms in E_m, E'_m will be denoted as usual by $\|\gamma\|$ and $\|u\|$, respectively. We shall denote by $\phi\colon E_m \to S_0$ the trivial map $x = \phi(\gamma)$ that associates the expression $x = c_1\phi_1 + \cdots + c_m\phi_m$ with the point $\gamma = (c_1, \cdots, c_m)$ of E_m. If Γ denotes the closed ball with center

$$\gamma_0 = (c_{01}, \cdots, c_{0m}) \in E_m$$

and radius c, then $\phi\colon \Gamma \to V$. Note that ϕ is a one-to-one linear map with $\|\phi\gamma\| = \|\gamma\|$, $\|\phi\| = 1$, and $\|\phi^{-1}\| = 1$.

Let $\psi\colon (S_E \cap S_N) \to E'_m$ with $u = \psi(x)$ $(x \in S_E \cap S_N)$ be the map defined by

$$u_i = Kx \cdot \phi_i = K\left(\sum_{s=1}^{\infty} c_s\phi_s\right) \cdot \phi_i \qquad (i = 1, \cdots, m),$$

where

$$x = \sum_{s=1}^{\infty} c_s\phi_s \in S_E \cap S_N \qquad \text{and} \qquad u = (u_1, \cdots, u_m) \in E'_m,$$

so that $u = \phi^{-1}PKx$ $(x \in S_E \cap S_N)$, or $\psi = \phi^{-1}PK$.

THEOREM (vii). *Under hypotheses* (IIabcdefghi) *the mapping*

$$\psi\colon (S_E \cap S_N) \to E'_m$$

is continuous.

Indeed, if $x^1, x^2 \in S_E \cap S_N$, then by Bessel's inequality and the definitions of G and e,

$$\|\psi x^1 - \psi x^2\| = \sum_{i=1}^{m} (|(Kx^1 - Kx^2)\cdot\phi_i|^2)^{1/2}$$

$$= \sum_{i=1}^{m} (|E(x^1 - x^2)\cdot\phi_i - (Nx^1 - Nx^2)\cdot\phi_i|^2)^{1/2}$$

$$= \sum_{i=1}^{m} (|(x^1 - x^2)\cdot G\phi_i - (Nx^1 - Nx^2)\cdot\phi_i|^2)^{1/2}$$

$$\leq (e + L)\|x^1 - x^2\|.$$

This proves the continuity of ψ.

Finally, we shall consider the two maps

$$M = \psi\phi\colon \Gamma \to E_m', \quad \mathfrak{M} = \psi\mathfrak{T}\phi\colon \Gamma \to E_m',$$

where

$$\phi\colon \Gamma \to V \subset (S_E \cap S_N), \qquad \mathfrak{T}\colon V \to (S_E \cap S_N), \qquad \psi\colon (S_E \cap S_N) \to E_m'.$$

Both of the maps $M = \psi\phi$ and $\mathfrak{M} = \psi\mathfrak{T}\phi$ are continuous on Γ.

Each $\gamma \in \Gamma$ determines a unique $\phi(\gamma) = x^* \in V$. Thus for each $\gamma \in \Gamma$, the error function p in Theorem (vi) has the alternate representation

$$p \equiv PKy = \phi\phi^{-1} PK\mathfrak{T}x^* = \phi\psi\mathfrak{T}x^*,$$

where $x^* = \phi(\gamma)$. Hence, $\phi^{-1}p = \psi\mathfrak{T}\phi(\gamma)$. Moreover, $p = 0$ if and only if $\phi^{-1}p = 0$; that is, $p = 0$ if and only if $\psi\mathfrak{T}\phi(\gamma) = 0$.

We shall denote by μ_0 and μ the topological degrees of M and $\mathfrak{M}$ with respect to the origin O of E_m'. If we denote by $\partial\Gamma$ the boundary of Γ, then $\partial\Gamma$ is an (m - 1)-dimensional sphere in E_m, and

$$C_0 = M|\,\partial\Gamma\colon \partial\Gamma \to E_m' \quad \text{and} \quad C = \mathfrak{M}|\,\partial\Gamma\colon \partial\Gamma \to E_m'$$

are singular (m - 1)-cycles whose topological orders with respect to the origin of E_m' are still the numbers $\mu_0 = \mathcal{O}(C_0, O)$, $\mu = \mathcal{O}(C, O)$.

The distance $\|C, C_0\|$ between C and C_0 is the number

$$\|C, C_0\| = \max \|A(\gamma) - A_0(\gamma)\|,$$

where $C\colon u = A(\gamma)$, $C_0\colon u = A_0(\gamma)$ $(\gamma \in \partial\Gamma)$, and the maximum is taken over all $\gamma \in \partial\Gamma$. If

$$\Lambda = \max_{\gamma\in\partial\Gamma} \|\psi\phi(\gamma) - \psi\mathfrak{T}\phi(\gamma)\|,$$

then $\|C, C_0\| \leq \Lambda$.

Let Q denote the minimum distance of the points of the range of C_0 from the origin O of E'_m. It is well known from topology that if

$$\|C, C_0\| < Q,$$

then $\mathscr{O}(C, O) = \mathscr{O}(C_0, O)$. Thus $\Lambda < Q$ implies $\mu = \mu_0$. In this situation, if $\mu_0 \neq 0$, then $\mu \neq 0$ also. Hence, we know from topology that there exists at least one point $\bar{\gamma}$ in the interior of Γ such that $\psi \mathfrak{T} \phi(\bar{\gamma}) = 0$, that is, for which $p = 0$. Therefore $Ky = 0$, and $y = \mathfrak{T}\phi(\bar{\gamma})$ satisfies the original equation $Kx = 0$ exactly. Also, $\mu_0 \neq 0$ implies that there exists a point $\bar{\gamma}_0$ inside Γ such that $\psi\phi(\bar{\gamma}_0) = 0$, that is, which satisfies exactly the equations of the m^{th} Galerkin approximation. We conclude this section with the following theorem.

THEOREM (viii) *(Existence and Approximation Theorem). Under hypotheses* (IIabcdefghi), *if* $\mu_0 \neq 0$ *and*

$$\Lambda < Q, \tag{26}$$

then there exists an exact solution y *of the given equation* $Kx = 0$. *Also,* $y = \mathfrak{T}x^*$ *and* $Py = Px^* = x^*$ *for some* $x^* \in V$. *Moreover,* $y \in S_0^*$, $y \in (S_E \cap S_N)$, *and*

$$\|y - x_0\| \leq d, \qquad \|Py - x_0\| \leq c,$$

$$\mu(y - x_0) \leq R_0, \qquad \mu(Py - x_0) \leq r.$$

Thus the numbers c, R_0 give evaluations of the distance of the exact solution y from the "approximate" solution x_0 of the equation $K_0 x = 0$. For some values x_0 (say $x_0 = 0$), c and R_0 cannot be small, and Theorem (viii) is a mere existence theorem.

11. In applications, verification of the hypotheses (26) that $\Lambda < Q$ may be made along the following lines. We determine first of all convenient upper bounds for the distance $\|C, C_0\|$. To do this, we shall assume that the elements $\phi_1, \phi_2, \cdots$ satisfy the usual relations $E\phi_i + \lambda_i \phi_i = 0$ $(i = 1, 2, \cdots)$ (say, they have been obtained by solving a boundary value problem relative to the linear operator E), and that $|\lambda_i| \to \infty$ as $i \to +\infty$.

Then

$$x_0 = \sum_{i=1}^{m} c_{0i}\phi_i, \qquad x^* = \sum_{i=1}^{m} c_i\phi_i, \qquad y = \sum_{i=1}^{\infty} c_i\phi_i,$$

$$Ex_0 = -\sum_{i=1}^{m} \lambda_i c_{0i}\phi_i, \qquad Ex^* = -\sum_{i=1}^{m} \lambda_i c_i \phi_i.$$

If we put

$$Nx_0 = \sum_{i=1}^{\infty} \gamma_{0i}\phi_i, \qquad Nx^* = \sum_{i=1}^{\infty} \gamma_i^*\phi_i, \qquad Ny = \sum_{i=1}^{\infty} \gamma_i\phi_i,$$

then

$$u_i = Kx^* \cdot \phi_i = (E - N)x^* \cdot \phi_i = Ex^* \cdot \phi_i - Nx^* \cdot \phi_i = x^* \cdot G\phi_i - Nx^* \cdot \phi_i,$$

$$U_i = Ky \cdot \phi_i = (E - N)y \cdot \phi_i = Ey \cdot \phi_i - Ny \cdot \phi_i = y \cdot G\phi_i - Ny \cdot \phi_i,$$

and

$$\begin{aligned} |u_i - U_i| &= |x^* \cdot G\phi_i - y \cdot G\phi_i - Nx^* \cdot \phi_i + Ny \cdot \phi_i| \\ &= |(x^* - y) \cdot G\phi_i - (Nx^* - Ny) \cdot \phi_i| \quad (i = 1, \cdots, m). \end{aligned}$$

Hence,

$$\begin{aligned} \left[\sum_{i=1}^{m} (u_i - U_i)^2\right]^{1/2} &= \left\{\sum_{i=1}^{m} [(x^* - y) \cdot G\phi_i - (Nx^* - Ny) \cdot \phi_i]^2\right\}^{1/2} \\ &\leq \left\{\sum_{i=1}^{m} [(x^* - y) \cdot G\phi_i]^2\right\}^{1/2} + \left\{\sum_{i=1}^{m} [(Nx^* - Ny) \cdot \phi_i]^2\right\}^{1/2} \\ &\leq \|x^* - y\| \left[\sum_{i=1}^{m} \|G\phi_i\|^2\right]^{1/2} + \|Nx^* - Ny\| \\ &\leq e\|x^* - y\| + L\|x^* - y\| \\ &= (e + L)\|y - x^*\|. \end{aligned}$$

In case E is selfadjoint ($E = G$), the first of the two expressions above is zero, and $e + L$ can be simply replaced by L.

Now $y = Py + Fy$, $x_0 = Px_0 + F_0 x_0 + \triangle$, and $Py = x^*$; hence,

$$y - Py = Fy - F_0 x_0 - \triangle,$$

$$\begin{aligned} \|y - x^*\| = \|y - Py\| &\leq \|Fy - Fx_0\| + \|Fx_0 - F_0 x_0\| + \|\triangle\| \\ &\leq kLd + a + b. \end{aligned}$$

Thus

$$\Lambda = \max\left[\sum_{i=1}^{m} (u_i - U_i)^2\right]^{1/2} \leq (e + L)(kLd + a + b).$$

Instead of (26), it is easier to verify that

$$(27) \qquad (e + L)(kLd + a + b) < Q,$$

where Q is the distance between the range of C_0 and the origin in E'_m. If E is selfadjoint, that is, $E = G$, then $e + L$ can be replaced by L in all these formulas.

12. In particular situations both the numbers $\mu_0 = \mathcal{O}(C_0, O)$ and Q can be determined numerically, for a given $c > 0$. Nevertheless, it is convenient to obtain a

lower bound for the number Q appearing in (26) and (27), and together with it a criterion assuring that $\mu_0 = 0$.

Let

$$Kx_0 \cdot \phi \equiv K\left(\sum_{s=1}^{m} c_{0s}\,\phi_0\right)\cdot \phi_i = \eta_i \qquad (i = 1, \cdots, m),$$

the numbers η_i being the errors within which x_0 satisfies the exact equations of the m^{th} Galerkin approximation, and let $\eta = (\eta_1^2 + \cdots + \eta_m^2)^{1/2}$. Now, for $x = x^* \in V$,

$$u_i = Kx^* \cdot \phi_i = K\left(\sum_{s=1}^{m} c_s\,\phi_s\right)\cdot \phi_i \qquad (i = 1, \cdots, m).$$

We assume that these functions of $c_1, \cdots, c_m$ are known and are of class C^2, and that we are able to determine the m^2 numbers

$$m_{ij} = [\partial u_i/\partial c_j]_{c=c_{0i}} \qquad (i, j = 1, \cdots, m).$$

Furthermore, let

$$M = \max\left|\frac{\partial^2 u_i}{\partial c_j \partial c_h}\right|,$$

where the maximum is taken over all $i, j, h = 1, \cdots, m$ and all $\gamma \in \Gamma$, that is, over all $(c_1, \cdots, c_m)$ with

$$\sum_{s=1}^{m} (c_i - c_{0i})^2 \le c^2.$$

If $m = \det[m_{ij}] \neq 0$, then the $(m - 1)$-dimensional sphere

$$\partial\Gamma\colon \sum_{s=1}^{m} (c_i - c_{0i})^2 = c^2,$$

is mapped by

$$\bar{u}_i = \eta_i + \sum_{j=1}^{m} m_{ij}(c_j - c_{0j}) \tag{28}$$

into an $(m - 1)$-dimensional ellipsoid C_∞. To simplify notation, we shall also denote by C_∞ the map (28) restricted to $\partial\Gamma$. The m semiaxes of the ellipsoid C_∞ can be determined by analytic geometry. Indeed, by reversing the linear relations (28) we obtain the formulas

$$c_i - c_{0i} = \sum_{j=1}^{m} \mu_{ij}(\bar{u}_j - \eta_j)$$

and

$$c^2 = \sum_{i=1}^{m} (c_i - c_{0i})^2 = \sum_{j,h=1}^{m} A_{jh}(\bar{u}_j - \eta_j)(\bar{u}_h - \eta_h),$$

where

$$A_{jj} = \sum_{i=1}^{m} \mu_{ij}^2, \qquad A_{jh} = \sum_{i=1}^{m} \mu_{ij}\mu_{ih} \qquad (j \neq h;\ j, h = 1, \cdots, m).$$

If $\sigma_1, \cdots, \sigma_m$ are the (positive) roots of the equation $\det[A_{ih} - \sigma I] = 0$, then the m numbers $\alpha_j = c\sigma_j^{-1/2}$ $(j = 1, \cdots, m)$ are the semiaxes. Let

$$\sigma = \min[\sigma_1^{-1/2}, \cdots, \sigma_m^{-1/2}].$$

By Taylor's formula,

$$u_i = \eta_i + \sum_{j=1}^{m} m_{ij}(c_j - c_{0j}) + \frac{1}{2} \sum_{j,h=1}^{m} d_{ijh}(c_j - c_{0j})(c_h - c_{0h}), \tag{29}$$

where the second derivatives d_{ijh} are computed at some point inside Γ. By comparison with (28), we see that

$$\begin{aligned} |u_i - \bar{u}_i| &= \left| \frac{1}{2} \sum_{j,h=1}^{m} d_{ijh}(c_j - c_{0j})(c_h - c_{0h}) \right| \leq 2^{-1} M \sum_{j,h=1}^{m} |c_j - c_{0j}|\, |c_h - c_{0h}| \\ &= 2^{-1} M \left(\sum_{j=1}^{m} |c_j - c_{0j}| \right)^2 \leq 2^{-1} M \left(\sum_{j=1}^{m} (c_j - c_{0j})^2 \right) \left(\sum_{j=1}^{m} 1^2 \right) \\ &= 2^{-1} Mmc^2 . \end{aligned}$$

Thus

$$\begin{aligned} \left[\sum_{i=1}^{m} (u_i - \eta_i)^2 \right]^{1/2} &\geq \left[\sum_{i=1}^{m} (\bar{u}_i - \eta_i)^2 \right]^{1/2} - \left[\sum_{i=1}^{m} (u_i - \bar{u}_i)^2 \right]^{1/2} \\ &\geq c\sigma - 2^{-1} Mm^{3/2} c^2 , \end{aligned}$$

and, finally,

$$\left(\sum_{i=1}^{m} u_i^2 \right)^{1/2} \geq c\sigma - 2^{-1} Mm^{3/2} c^2 - \eta,$$

where $c\sigma$ is the minimum distance, say Q', of the points of the ellipsoid C_∞ from the origin O of E'_m. If

(30) $$c\sigma - 2^{-1} Mm^{3/2} c^2 - \eta > 0,$$

then $\|C_\infty, C_0\| < Q' = c\sigma$, hence

$$\mu_0 = \mathscr{O}(C_0, O) = \mathscr{O}(C_\infty, O) \neq 0,$$

and it is not necessary to compute μ_0. On the other hand, if (30) is satisfied, then, instead of (27), we may simply verify that

(31) $$(e + L)(kLd + a + b) < c\sigma - 2^{-1} Mm^{3/2} c^2 - \eta .$$

Of course, this type of evaluation may be convenient only for small values of m.

13. For $m = 1$ (the first Galerkin approximation), the reasonings above remain the same. In this case C_0, C, and C_∞ are pairs of real numbers, namely, the values of u, U, and $\bar{u}$ at the points $c_{01} - c$, $c_{01} + c$, and to say $\mathscr{O}(C, O) \neq 0$ simply means that zero separates the pair $u|_{c_1=c_{01}\pm c}$. Also,

$$Q = \min |K(c_1 \phi_1) \cdot \phi_1| \qquad \text{for } c_1 = c_{01} \pm c,$$

and we are required to verify that

$$(e + L)(kLd + a + b) < Q.$$

Concerning (31), note that here there is only one number m_{11}:

$$m_{11} = [\partial u_1/\partial c_1]_{c_1=c_{01}} .$$

Let

$$M = \max |\partial^2 u_1/\partial c_1^2| \qquad \text{for all } c_{01} - c \leq c_1 \leq c_{01} + c .$$

Then

$$Q \geq |m_{11}| c - 2^{-1} Mc^2 - \eta,$$

and we may simply verify that

$$(e + L)(kLd + a + b) < |m_{11}| c - 2^{-1} Mc^2 - \eta .$$

14. A rather typical situation is that of a linear differential operator E of order M, a domain A, and preassigned linear homogeneous boundary conditions (B) on the boundary ∂A of A involving derivatives of order $M_1 < M$. Concerning E we shall assume that the associated linear problem $Ex + \lambda x = 0$ with conditions (B) has a countable system of eigenvalues and eigenfunctions λ_i and ϕ_i, and that $[\phi_1, \phi_2, \cdots]$ is a complete orthonormal system in the Hilbert space $S = L_2(A)$ of all square-integrable functions $x(\alpha)$ $(\alpha \in A)$. By $\cdot$ and $\| \ \|$ we shall denote the inner product and square-norm in $S = L_2(A)$.

Let $\mu(x)$ denote the uniform norm, that is, let

$$\mu(x) = \sup_{\alpha \in A} |x(\alpha)| .$$

Every element $x \in S$ has a Fourier series

$$x(\alpha) \sim c_1 \phi_1 + c_2 \phi_2 + \cdots,$$

and we shall denote by Px the projection operator

$$Px = c_1 \phi_1 + \cdots + c_m \phi_m,$$

for some fixed m.

For S_E we could for instance take the linear set of all functions $x(\alpha)$ $(\alpha \in A)$ such that $x(\alpha)$ is continuous in $A \cup \partial A$, $x(\alpha)$ has all partial derivatives of orders no greater than M_1 and they are continuous in $A \cup \partial A$, $x(\alpha)$ has all partial derivatives of the orders no greater than M and they are continuous in A, $x(\alpha)$ satisfies conditions (B) on ∂A, and Ex is an element of S. We shall assume $\phi_i \in S_E$ $(i = 1, 2, \cdots)$; hence $P: S \to S_E$. Finally, we shall assume that for $x \in S_E$ we can apply the operator E formally to the series

$$x(\alpha) \sim c_1 \phi_1 + c_2 \phi_2 + \cdots,$$

so that

$$Ex \sim - c_1 \lambda_1 \phi_1 - c_2 \lambda_2 \phi_2 - \cdots \qquad (x \in S_E).$$

For each $x \in S$, to $x(\alpha)$ and $(I - P)x(\alpha)$ there correspond the Fourier series

$$x(\alpha) \sim c_1 \phi_1 + c_2 \phi_2 + \cdots,$$

$$(I - P)x(\alpha) \sim c_{m+1} \phi_{m+1} + c_{m+2} \phi_{m+2} + \cdots.$$

We shall let $H: S_1 \to S_1$ be the operator defined by

$$H(I - P)x(\alpha) \sim - c_{m+1} \lambda_{m+1}^{-1} \phi_{m+1} - c_{m+2} \lambda_{m+2}^{-1} \phi_{m+2} - \cdots. \tag{32}$$

Then, for every $x \in S_E$,

$$H(I - P)Ex = (I - P)x, \quad EPx = PEx, \quad EH(I - P)x = (I - P)x.$$

Thus, the parts of hypotheses (IIa), (IIb), (IIc), and also (IIh) concerning E and H hold provided the fixed element $y = Ty$ of each set S_0^* can be proved to belong to S_E. As we noted in Section 5, if this is so, then $y \in S_E$ and $Py \in S_E$ imply

$$H(I - P)Ny \in S_E.$$

Remark. For the problem of periodic solutions of ordinary periodic differential equations (see [4]), say of period $T = 2\pi/\omega$ $(\omega > 0)$, we took in [4] for E the operator $Ex(t) = x''(t)$ $(0 \leq t \leq 2\pi/\omega)$ with eigenvalues $\lambda_1 = 0$, $\lambda_i = (i - 1)^2 \omega^2$ $(i = 2, 3, \cdots)$. Thus, the operator H satisfying (32) is defined in S_1 (not in S) as desired, with $Px = c_1 \phi_1 + \cdots + c_m \phi_m$ and $m \geq 1$.

Remark. The considerations above apply also to generalized solutions, provided the continuity requirements on the derivatives of x are omitted and both the equation and the boundary conditions are understood to be satisfied in a conveniently chosen weak sense. As mentioned in the Introduction, we shall discuss this problem in subsequent papers.

It may occur that H is represented in the form

$$u(\alpha) = Hx = \int_A h(\alpha, \beta)\, x(\beta)\, d\beta, \tag{33}$$

where h is a convenient kernel defined in $A \times A$. Since

$$Px(\alpha) = (x \cdot \phi_1)\phi_1 + \cdots + (x \cdot \phi_m)\phi_m = \int_A x(\beta) \left(\sum_{i=1}^{m} \phi_i(\alpha)\, \phi_i(\beta) \right) d\beta,$$

we see that

$$\begin{aligned} z(\alpha) = H(I - P)x &= \int_A h(\alpha, \beta) \left[x(\beta) - \int_A z(\gamma) \left(\sum_{i=1}^{m} \phi_i(\beta)\, \phi_i(\gamma) \right) d\gamma \right] d\beta \\ &= \int_A k(\alpha, \beta)\, x(\beta)\, d\beta, \end{aligned} \tag{34}$$

where

$$k(\alpha, \beta) = h(\alpha, \beta) - \int_A h(\alpha, \beta) \left(\sum_{i=1}^{m} \phi_i(\beta)\, \phi_i(\gamma) \right) d\gamma .$$

Thus

$$|z(\alpha)| = \left(\int_A k^2(\alpha, \beta)\, d\beta \right)^{1/2} \left(\int_A x^2(\beta)\, d\beta \right)^{1/2} .$$

We may take for k the number

$$k = \left\{ \int\int_{A \times A} k^2(\alpha, \beta)\, d\alpha\, d\beta \right\}^{1/2} ; \tag{35}$$

and, if k is continuous on $A \times A$, we may take for k' the number

$$k' = \operatorname*{Max}_{\alpha \in A} \left\{ \int_A k^2(\alpha, \beta)\, d\beta \right\}^{1/2} , \tag{36}$$

since then

$$\|z\| = \|H(I - P)x\| \leq k\|x\| \qquad \text{and} \qquad \mu(z) = \mu(H(I - P)x) \leq k'\|x\| .$$

Thus k and k' depend on m. It is known that $k \to 0$ as $m \to \infty$, since this simply corresponds to the possibility of approaching the kernel $k(\alpha, \beta)$ in the norm $\| \ \|$. Thus, the main condition $kL < 1$ ensuring that F is a contraction is certainly satisfied for all sufficiently large m. This process for the estimation of k and k' requires only the knowledge of the m eigenfunctions $\phi_1, \cdots, \phi_m$ and of the kernel h.

Assume that the eigenvalues are ordered, $|\lambda_1| \leq |\lambda_2| \leq \cdots$, and that $\lambda_{m+1} \neq 0$ is known. From $E\phi_i + \lambda_i \phi_i = 0$, we deduce $\phi_i + \lambda_i H\phi_i = 0$, that is, $H\phi_i = -\lambda_i^{-1} \phi_i$ $(i \geq m+1)$. Hence, if

$$x(\alpha) \sim c_1 \phi_1 + c_2 \phi_2 + \cdots,$$

then

$$(I - P)x(\alpha) \sim c_{m+1} \phi_{m+1} + \cdots,$$

$$H(I - P)x(\alpha) \sim - c_{m+1} \lambda_{m+1}^{-1} \phi_{m+1} - \cdots,$$

and

$$\|H(I - P)x(\alpha)\| = \left(\sum_{i=m+1}^{\infty} c_i^2 \lambda_i^{-2} \right)^{1/2} \leq |\lambda_{m+1}|^{-1} \left(\sum_{i=m+1}^{\infty} c_i^2 \right)^{1/2} \leq |\lambda_{m+1}|^{-1} \|x\| .$$

Thus we may take for k the value

$$k = |\lambda_{m+1}|^{-1}. \tag{37}$$

If we have lower estimates $\ell_i \leq \lambda_i$ for the eigenvalues λ_i $(i = m+1,\ m+2, \cdots)$ and upper estimates $\mu_i \geq |\phi_i(\alpha)|$ for the corresponding eigenfunctions ϕ_i, then we can obtain another estimate for k'. Indeed,

$$|H(I - P)x(\alpha)| \leq \sum_{i=m+1}^{\infty} |c_i|\ |\lambda_i^{-1}|\ |\phi_i|$$

$$\leq \left(\sum_{i=m+1}^{\infty} \lambda_i^{-2} \phi_i^2 \right)^{1/2} \left(\sum_{i=m+1}^{\infty} c_i^2 \right)^{1/2} \leq \left(\sum_{i=m+1}^{\infty} \ell_i^{-2} \mu_i^2 \right)^{1/2} \|x\| ,$$

and hence we can assume

$$k' = \left(\sum_{i=m+1}^{\infty} \ell_i^{-2} \mu_i^2 \right)^{1/2}, \tag{38}$$

provided this series is convergent. We may take for S_0^* sets of the form

$$S_0^* \equiv [x(\alpha) |\ x(\alpha) \in S,\ \|x(\alpha) - x^*(\alpha)\| \leq d,\ |x(\alpha) - x^*(\alpha)| \leq R_0].$$

We have to prove that each S_0^* is complete in the norm $\|\ \|$. Let $x_n(\alpha)$ be a Cauchy sequence of elements of S_0^*. Then the limit $\bar{x}(\alpha)$ exists in S since $S = L_2(A)$ is complete, and obviously $\|\bar{x}_n(\alpha) - x^*(\alpha)\| \leq d$ implies $\|\bar{x}(\alpha) - x^*(\alpha)\| \leq d$. Convergence in L_2 implies convergence in the mean, and this in turn implies convergence in measure. Also, there exists a sequence $\{n_k\}$ such that $x_{n_k}(\alpha) \to \bar{x}(\alpha)$ as $k \to \infty$, almost everywhere in A. Thus $|x_{n_k}(\alpha) - x^*(\alpha)| \leq R_0$ implies $|\bar{x}(\alpha) - x^*(\alpha)| \leq R_0$ almost everywhere, and we may assume that the relation $|\bar{x}(\alpha) - x^*(\alpha)| \leq R_0$ is satisfied for all $\alpha \in A$.

Once the conditions ensuring that $T| S_0^*$ is a contraction and maps S_0^* into S_0^* are satisfied, $T| S_0^*$ has one and only one fixed element $y = Ty$. Since y satisfies the equation

$$y(\alpha) = Py + H(I - P)Ny = (y \cdot \phi_1)\phi_1 + \cdots + (y \cdot \phi_m)\phi_m + H(I - P)Ny,$$

we see that y is decomposed into the sum of $m + 1$ functions all satisfying the linear boundary conditions (B). Thus y satisfies the same boundary conditions.

CHAPTER 3. A NUMERICAL EXAMPLE

15. We shall study a nonlinear boundary value problem in ordinary differential equations:

$$\text{(39)} \qquad \begin{gathered} x'' + x + \alpha x^3 = \beta t \qquad (0 \leq t \leq 1), \\ x(0) = 0, \quad x'(1) + hx(1) = 0, \end{gathered}$$

where α and β are numerical constants.

We shall prove that for $h = 1$ this problem certainly has a solution for all (α, β) such that $|\alpha| \leq 1$, $|\beta| \leq 1$.

EXISTENCE ANALYSIS

Let $S = L_2(0, 1)$, so that the norm $\| \ \|$ is the square-norm. If $x(t) \in S$ is a bounded function, let us denote by μ the uniform norm

$$\mu(x) = \sup_{x \in [0,1]} |x(t)| .$$

Let $\phi_i(t)$ $(i = 1, 2, \cdots)$ be the eigenfunctions of the familiar boundary value problem

$$x'' + \ell^2 x = 0 \ (0 \leq t \leq 1), \quad x(0) = 0, \quad x'(1) + hx(1) = 0.$$

Then, for $h > 0$,

$$\begin{gathered} \phi_i(t) = \nu_i \sin \ell_i t \qquad (0 \leq t \leq 1), \\ h \tan \ell_i + \ell_i = 0, \quad \nu_i = [2(h^2 + \ell_i^2)(h^2 + \ell_i^2 + h)^{-1}]^{1/2}, \\ (2i - 1)\pi/2 < \ell_i < i\pi \qquad (i = 1, 2, \cdots), \end{gathered}$$

where the numbers ν_i are determined by the condition

$$\int_0^1 \nu_i^2 \sin^2 \ell_i t \, dt = 1 .$$

For $h = 1$,

$$\ell_1 \cong 2.0288, \quad \nu_1 \cong 1.2934, \quad \ell_2 \cong 4.9132, \quad \nu_2 \cong 1.3868,$$
$$1 < \nu_i < 2^{1/2} \quad (i = 1, 2, \cdots), \qquad \nu_i \to 2^{1/2} \text{ as } i \to \infty.$$

It is convenient to assume that E is the linear operator $Ex = x'' + x$, with the homogeneous boundary conditions $x(0) = 0$, $x'(1) + hx(1) = 0$. The normalized eigenfunctions are the functions $\phi_i(t)$ above, and

$$E\phi_i + (\ell_i^2 - 1)\phi_i = 0 \qquad (i = 1, 2, \cdots),$$

that is, the eigenvalues relative to E are the numbers $\lambda_i = \ell_i^2 - 1$ $(i = 1, 2, \cdots)$.

We shall take for S_E the set of all functions $x(t)$ $(0 \leq t \leq 1)$ such that $x(t)$ and $x'(t)$ are absolutely continuous in $[0, 1]$, $x''(t)$ is L^2-integrable in $[0, 1]$, and $x(t)$ satisfies the conditions $x(0) = 0$, $x'(1) + hx(1) = 0$. Thus $S_E \subset S$ and $E: S_E \to S$.

If $x \in S = L_2(0, 1)$ has the Fourier series

(40) $$x(t) \sim \sum c_i \nu_i \sin \ell_i t,$$

where Σ ranges over all $i = 1, 2, \cdots$, then we define H by taking

(41) $$u(t) = Hx = -\sum(\ell_i^2 - 1)^{-1} c_i \nu_i \sin \ell_i t \qquad (0 \leq t \leq 1).$$

For $x \in S$, we note that $\Sigma c_i^2 < +\infty$, and hence, by Schwarz's inequality,

$$\sum |c_i|(\ell_i^2 - 1)^{-1} < +\infty \quad \text{and} \quad \sum |c_i| \ell_i (\ell_i^2 - 1)^{-1} < +\infty.$$

This shows that the series above for $u(t)$ and the series

(42) $$-\sum c_i \ell_i (\ell_i^2 - 1)^{-1} \nu_i \cos \ell_i t,$$

are absolutely and uniformly convergent in $[0, 1]$. Thus $u(t)$ and $u'(t)$ are continuous in $[0, 1]$ and satisfies the boundary conditions. Also, the series for $x - u$, that is, the difference of the two series (40) and (41),

(43) $$\sum c_i \ell_i^2 (\ell_i^2 - 1)^{-1} \nu_i \sin \ell_i t,$$

is L^2-convergent in $[0, 1]$. If s_n and σ_n denote the partial sums of the series (43) and (42), respectively, then $\|s_n - (x - u)\| \to 0$ as $n \to \infty$; hence, by Schwarz's inequality, it is also true that

$$\left| \int_0^t s_n(\alpha)d\alpha - \int_0^t (x(\alpha) - u(\alpha))d\alpha \right| \to 0$$

as $n \to \infty$, uniformly in $0 < t < 1$. This limit relation can be written in the form

$$\left| (\sigma_n(t) - \sigma_n(0)) - \int_0^t (x(\alpha) - u(\alpha))d\alpha \right| \to 0$$

as $n \to \infty$, $0 \leq t \leq 1$; hence

$$u'(t) = u'(0) + \int_0^t (x(\alpha) - u(\alpha))d\alpha \qquad (0 \leq t \leq 1).$$

This implies that $u'(t)$ and $u(t)$ are absolutely continuous in $[0, 1]$ and that $u''(t) = x(t) - u(t)$ almost everywhere. Hence $u''(t) \in S = L_2(0, 1)$. Thus $Hx \in S_E$, and $H: S \to S_E$. Also, since $Eu = u'' + u = x$, we see that $EHx = x$ for every $x \in S$.

16. We define $P: S \to S_E$ by taking $Px = c_1 \phi_1 = (x \cdot \phi_1)\phi_1$ for every $x \in S$. Obviously, $P^2 x = Px$ for every $x \in S$, and

$$H(I - P)Ex = (I - P)x, \qquad EPx = PEx, \qquad EH(I - P)x = (I - P)x$$

for every $x \in S_E$. Note that

$$H(I - P)x = -\sum{}' c_i(\ell_i^2 - 1)^{-1} \nu_i \sin \ell_i t \qquad (x \in S),$$

where Σ' ranges over $i = 2, 3, \cdots$. It follows that

$$\|H(I - P)x\| = \left[\sum{}' c_i^2(\ell_i^2 - 1)^{-2}\right]^{1/2} \leq (\ell_2^2 - 1)^{-1} \left(\sum{}' c_i^2\right)^{1/2} \leq k\|x\| . \tag{44}$$

Since $(\ell_2^2 - 1)^{-1} = 0.043217$, we may take $k = 0.044$. Now

$$\begin{aligned} |H(I - P)x(t)| &\leq \sum{}' |c_i|(\ell_i^2 - 1)^{-1} \nu_i|\sin \ell_i t| \\ &\leq \left(\sum{}'(\ell_i^2 - 1)^{-2} \nu_i^2\right)^{1/2} \left(\sum{}' c_i^2\right)^{1/2} \\ &\leq \left(\sum{}'(\ell_i^2 - 1)^{-2} \nu_i^2\right)^{1/2} \|x\| \\ &\leq k'\|x\| . \end{aligned} \tag{45}$$

Since $(2i - 1)\pi/2 < \ell_i < i\pi$ and $\nu_i^2 < 2$, the numerical factor in (45) is less than

$$\{(\ell_2^2 - 1)^{-2} \nu_2^2 + 2\sum{}'' [((2i - 1)\pi/2)^2 - 1]^{-2}\}^{1/2} ,$$

where Σ'' ranges over $i = 3, 4, \cdots$. Since the quotients

$$[(2i - 1)\pi/2]^2/[((2i - 1)\pi/2)^2 - 1] \qquad (i = 3, 4, \cdots)$$

all lie between 1 and $\xi = (5\pi/2)^2/[(5\pi/2)^2 - 1]$, the same numerical factor in (45) is less than

$$\{(\ell_2^2 - 1)^{-2} \nu_2^2 + 2\xi^2 (2/\pi)^4 (5^{-4} + 7^{-4} + \cdots)\}^{1/2}.$$

This expression is less than 0.06621, and thus we can take $k' = 0.067$ in (45). Thus

$$\|H(I - P)x\| \leq k\|x\|, \qquad \text{and} \qquad \mu H(I - P)x = \max |H(I - P)x| \leq k' \|x\| \tag{46}$$

for all $x \in S$, where $k = 0.044$ and $k' = 0.067$.

17. Let $Nx = -\alpha x + \beta t$ and $K = E - N$; then (39) becomes $Ex = Nx$ or $Kx = 0$. Also, let $F = F_0 = H(I - P)N$; then (39) yields

$$H(I - P)Ex = Fx \quad \text{or} \quad (I - P)x = Fx.$$

Thus

$$x = Px + Fx.$$

If $x_0(t)$ is any approximation of $x(t)$ $(x_0 \in S_E)$, then

$$(47) \qquad Ex_0 = Nx_0 + \theta(t) \quad \text{and} \quad x_0 = Px_0 + Fx_0 + \Delta(t),$$

where $\Delta(t) = H(I - P)\theta(t)$.

For $x = \gamma\phi_1 = \gamma\nu_1 \sin \ell_1 t$, we find that

$$(48) \qquad \begin{aligned} \theta(t;\gamma) &= x'' + x + \alpha x^3 - \beta t \\ &= \gamma(1 - \ell_1^2)\nu_1 \sin \ell_1 t + 4^{-1}\gamma^3 \alpha \nu_1^3 (3 \sin \ell_1 t - \sin 3\ell_1 t) - \beta t, \end{aligned}$$

and we define

$$A(\gamma) \equiv \int_0^1 \theta(t;\gamma)\,\phi_1(t)\,dt, \qquad \Delta(t;\gamma) \equiv H(I - P)\,\theta(t;\gamma).$$

Then

$$(49) \qquad \begin{aligned} A(\gamma) &= (1 - \ell_1^2)\gamma + [(3/4) + (\nu_1^2/32\ell_1)(\sin 4\ell_1 - 2 \sin 2\ell_1)]\nu_1^2 \alpha\gamma^3 \\ &\quad - \nu_1 \ell_1^{-2}(\sin \ell_1 - \ell_1 \cos \ell_1)\beta \\ &\cong -(3.11603)\gamma + (1.36478)\alpha\gamma^3 - (0.56373)\beta. \end{aligned}$$

For $\gamma = \gamma_0 = 0$, that is, for $x = \gamma_0 \phi_1 = x_0(t) \equiv 0$, we see that $\theta(t; 0) = -\beta t$, and

$$Pt = (t \cdot \phi_1)\phi_1 = \nu_1^2 \ell_1^{-2}(\sin \ell_1 - \ell_1 \cos \ell_1) \sin \ell_1 t.$$

Now

$$(50) \qquad \begin{aligned} H \sin \ell_i t &= -(\ell_i^2 - 1)^{-1} \sin \ell_i t \qquad (i = 1, 2, \cdots), \\ H \sin \ell t &= (\ell^2 - 1)^{-1}[-\sin \ell t + (h \sin \ell + \ell \cos \ell)\cdot(h \sin 1 + \cos 1)^{-1} \sin t] \end{aligned}$$

for $\ell \neq 1$ and $\ell \neq \lambda_i$, and $Ht = t - (h+1)(h \sin 1 + \cos 1)^{-1} \sin t$. Finally, we obtain the relation

$$(51) \qquad \begin{aligned} \Delta(t; 0) &\equiv H(I - P)\theta(t; 0) \\ &= -\beta\{t - (h+1)(h \sin 1 + \cos 1)^{-1} \sin t\} \\ &\quad + \nu_1^2 \ell_1^{-2}(\ell_1^2 - 1)^{-1}(\sin \ell_1 - \ell_1 \cos \ell_1) \sin \ell_1 t \\ &\cong -\beta\{t - 1.447415 \sin t + 0.234009 \sin \ell_1 t\}. \end{aligned}$$

We shall take $h = 1$. Numerically, for $t = 0; 0.1; \cdots; 1.0$, respectively, $-\beta^{-1}\Delta(t; 0) = 0, 0.00265; 0.00481; 0.00605; 0.00608; 0.00478; 0.00227; -0.00109; -0.00444; -0.00736; -0.00807$, and $|\beta^{-1}\Delta(t)|_{max} = 0.00816$ at $t = 0.976$. Hence,

$$\|\Delta(t; 0)\| \leq 0.0050|\beta| \quad \text{and} \quad \max|\Delta(t)| \leq 0.0082|\beta| .$$

We shall take $b = 0.0050|\beta|$, $b' = 0.0082|\beta|$. Since $F \equiv F_0$, we take $a = a' = 0$.

18. The requirement $\|x^* - x_0\| \leq c$ with $x^* \in S_0$, $x_0 \equiv 0$ in (17), or $|\gamma| \leq c$, implies $|x^*(t)| = |\gamma\nu_1 \sin \ell_1 t| \leq c\nu_1$. We shall take $r = c\nu_1 = 1.2934c$, so that $\|x^* - x_0\| \leq c$ implies $|x^*(t) - x_0(t)| \leq r$ $(0 \leq t \leq 1)$, and V is then defined by

$$V = [x^* |\ x^* \in S_0,\ \|x^*\| \leq c\ .$$

Since $x_0(t) \equiv 0$, we see that $|x(t)| \leq R_0$ for $|x(t) - x_0(t)| \leq R_0$, and we take $R = R_0$. Now, for any two elements $x_1, x_2 \in S$ with $|x_1| \leq R$ and $|x_2| \leq R$,

$$x_1^3 - x_2^3 = (x_1 + x_1 x_2 + x_2)(x_1 - x_2), \quad \text{and} \quad |x_1^3 - x_2^3| \leq 3R^2|x_1 - x_2| .$$

Thus,

$$\|Nx_1 - Nx_2\| = 3|\alpha|R^2\|x_1 - x_2\| ,$$

and we take $L = 3|\alpha|R^2$.

For each $x^* \in V$ the set S_0^* is defined by

$$S_0^* = [x|\ x \in S,\ \|x\| \leq d,\ |x| \leq R]$$

(recall that $R_0 = R$), and this set is obviously complete in the L_2-norm (see Section 14).

If $y = Ty$ is any fixed element of a set S_0^*, then

$$y = Ty = Py + H(I - P)Ny \quad \text{and} \quad |y(t)| \leq R .$$

Thus y is bounded and measurable, and, by the properties of N and H, y is absolutely continuous together with its first derivative y', and y' is bounded and measurable. By repeating the argument we see that y is of class C^2 (actually of class C^∞). Thus, $Ey = y'' + y$ and y'' are continuous in $[0, 1]$. Finally, y satisfies the boundary conditions $y(0) = 0$, $y'(1) + hy(1) = 1$, since y is the sum of two functions with the same property. Thus $y = Ty \in S_0^*$ is an element of S_E.

19. The relations (IIg) now become

$$0 < c < d, \quad r = 1.2934\,c < R = R_0, \quad c + 0.0050|\beta| \leq [1 - 0.044(3|\alpha|R^2)]d,$$

$$1.2934\,c + 0.0082|\beta| < R - 0.067(3|\alpha|R^2)d, \tag{52}$$

$$L = 3|\alpha|R^2, \quad kL = 0.044(3|\alpha|R^2) < 1 .$$

From the definition (40) of $A(\gamma)$ we see that

$$A'(\gamma) \cong -3.11603 + 4.09434\,\alpha\gamma^2 \quad \text{and} \quad A''(\gamma) \cong 8.18868\,\alpha\gamma .$$

Hence, $A(0) \cong -0.56373\beta$, $A'(0) \cong -3.11603$; and, for $|\gamma| \leq c$, $A''(\gamma) \leq 8.18868\, c\, |\alpha|$. Finally,

$$u_{01} \cong A(-c) = 3.11603\, c - 1.36478\, c^3 |\alpha| - 0.56373\, |\beta| ,$$

$$u_{02} = A(c) \leq -3.11603\, c + 1.36478\, c^3 |\alpha| + 0.56373\, |\beta| .$$

As mentioned above, we shall verify the inequalities $u_{01} > 0 > u_{02}$, and we take $\Omega = \min[|u_{01}|, |u_{02}|]$. Relation (27), or the inequality $L(kLd + 0.0050\,|\beta|) < \Omega$, is now replaced by the inequality

$$\begin{aligned} &3|\alpha|R^2 (0.044\,(3|\alpha|R^2)d + 0.0050\,|\beta|) \\ &\qquad < 3.11603\, c - 0.56373\, |\beta| - 1.36478\, c^3 |\alpha| . \end{aligned} \tag{53}$$

The positive character of the second member then ensures that $u_{01} > 0 > u_{02}$.

For instance, if we take

$$R = 0.5300, \quad |\alpha| \leq 1, \quad |\beta| \leq 1, \quad c = 0.3774, \quad d = 0.4500,$$

then all inequalities in (52) and (53) are certainly satisfied. This proves our contention that the nonlinear boundary value problem (39) has at least one exact solution for $|\alpha| \leq 1$, $|\beta| \leq 1$.

FIRST GALERKIN APPROXIMATION FOR $\alpha = \beta = 1/2$

20. For $\alpha = \beta = 0.5$, we conclude from (49) that

$$A(\gamma) \cong -3.11603\,\gamma + 0.68239\,\gamma^3 - 0.281865,$$

and the Galerkin equation for the first approximation, $A(\gamma) = 0$, yields

$$\gamma = \gamma_1 \cong -0.090610.$$

Since $\gamma_1 \nu_1 \cong -0.11721$, a first Galerkin approximation for problem (39) with $\alpha = \beta = 0.5$ is

$$x(t) = \gamma_1 \nu_1 \sin \ell_1 t = -0.11721 \sin(2.0288\,t) \qquad (0 \leq t \leq 1).$$

This function satisfies equations (47) with errors $\theta(t)$, $\Delta(t)$:

$$Ex_0 = Nx_0 + \theta(t), \qquad x_0 = Px_0 + Fx_0 + \Delta(t),$$

with

$$\Delta(t) = H(I - P)\theta(t).$$

For $x = \gamma\phi_1 = \gamma\nu_1 \sin \ell_1 t$, we see that $Ex - Nx = \theta(t)$, where $\theta(t)$ is given by (48) and

$$\begin{aligned}(I - P)\theta(t) &= \theta(t) - A(\gamma_1)\nu_1 \sin \ell_1 t = \theta(t)\\ &= \gamma_1(1 - \ell_1^2)\nu_1 \sin \ell_1 t + 4^{-1}\gamma_1^3 \alpha \nu_1^3 (3 \sin \ell_1 t - \sin 3\ell_1 t) - \beta t,\end{aligned}$$

since $A(\gamma_1) = 0$.

By using the same formulas (50) above, we obtain the relation

$$\begin{aligned}\Delta(t) = H(I - P)\theta(t) &= \gamma_1 \nu_1 \sin \ell_1 t + 4^{-1}\gamma_1^3 \nu_1^3 \{-3(\ell_1^2 - 1)^{-1} \sin \ell_1 t\\ &+ (9\ell_1^2 - 1)^{-1}[\sin 3\ell_1 t - (h \sin 3\ell_1 + 3\ell_1 \cos 3\ell_1)(h \sin 1 + \cos 1)^{-1} \sin t]\}\\ &- \beta[t - (h + 1)(h \sin 1 + \cos 1)^{-1} \sin t].\end{aligned}$$

For $h = 1$, $\ell_1 \cong 2.0288$, $\nu_1 \cong 1.2934$, $\alpha = 0.5$, $\beta = 0.5$, and $\gamma_1 \cong -0.090619$,

$$\begin{aligned}\Delta(t) \cong &-0.11701337 \sin \ell_1 t - 0.00000558 \sin 3\ell_1 t\\ &- 0.5 t + 0.72343164 \sin t,\end{aligned}$$

and the values of $\Delta(t)$ for $t = 0.0, 0.1, \cdots, 1.0$ are 0, -0.0013, -0.0024, -0.0030, -0.0030, -0.0024, -0.0011, +0.0006, +0.0023, +0.0037, 0.0040, respectively. Also $\max \Delta(t) \leq 0.004094$ at $t \cong 0.0759$. Finally,

$$\|\Delta\| = \left(\int_0^1 \Delta^2(t)\, dt\right)^{1/2} \leq 0.0026, \qquad \max|\Delta(t)| \leq 0.0041,$$

and we may take $b = 0.0026$ and $b' = 0.0041$. Since $F = F_0$, we also know that $a = a' = 0$.

21. The requirement $\|P(x^* - x_0)\| \leq c$, or $|\gamma - \gamma_1| \leq c$, in (17) implies

$$|x^*(t) - x_0(t)| = |(\gamma - \gamma_1)\nu_1 \sin \lambda_1 t| \leq c\nu_1.$$

We shall take $r = c\nu_1 = 1.2934\, c$, so that $\|P(x^* - x_0)\| \leq c$ implies

$$|x^*(t) - x_0(t)| \leq r,$$

and V is then defined by

$$V = [x^* |\ x^* \in P(S),\ \|x^* - x_0\| \leq c].$$

Since, $|x_0(t)| = |\gamma_1 \nu_1 \sin \lambda_1 t| \leq |\gamma_1| \nu_1 \leq 0.11721$, we see that for

$$|x(t) - x_0(t)| \leq R_0,$$

$$|x(t)| \leq |x_0(t)| + R_0 \leq 0.11721 + R_0.$$

We take $R = 0.11721 + R_0$. For numbers c, d $(0 < c < d)$ that are for the moment undetermined, let $\overline{S}_0$ be the set

$$\overline{S}_0 = [x(t) |\ x(t) \in S,\ \|x(t) - x_0(t)\| \leq d,\ |x(t) - x_0(t)| \leq R_0].$$

Then $|x(t)| \leq R$ for every $x(t) \in \overline{S}_0$. Now for any two real numbers x_1, x_2 with $|x_1|$, $|x_2| \leq R$,

$$|x_1^3 - x_2^3| = |(x_1^2 + x_1 x_2 + x_2^2)(x_1 - x_2)| \leq 3R^2|x_1 - x_2|.$$

Hence, for any two elements $x_1, x_2 \in \overline{S}_0$,

$$\|Nx_1 - Nx_2\| \leq 3R^2\|x_1 - x_2\|,$$

and we take $L = 3R^2 = 3(0.11721 + R_0)^2$.

Now, for each $x^* \in V$, the set S_0^* is defined by

$$S_0^* = [x \mid x \in S,\ \|x - x^*\| \leq d,\ |x - x^*| \leq R_0],$$

and each set S_0^* is obviously complete in the L_2-norm (see Section 14).

If $y = Ty$ is any fixed element of a set S_0^*, then

$$y = Ty = Py + H(I - P)Ny \quad \text{and} \quad |y| \leq R.$$

Then y is continuous in $[0, 1]$ together with y', since y is the sum of two functions with the same property. By the same argument, $Ey = y'' + y$ is continuous in $[0, 1]$, and so is $y'' = Ey - y$. Finally, y satisfies the boundary condition $y(0) = 0$, $y'(1) + hy(1) = 0$. Thus $y = Ty$ is an element of S_E.

The relations (IIg) now become

$$c + 0.0026 < (1 - 0.044\,L)d,$$

$$1.2934\,c + 0.0041 \leq R_0 - 0.067\,Ld,$$

$$R = 0.11721 + R_0, \qquad L = 3R^2, \qquad 0.044\,L < 1.$$

From the expression (49) for $A(\gamma)$, we deduce that

$$A'(\gamma) \cong -3.1160 + 2.04717\gamma^2 \quad \text{and} \quad A''(\gamma) \cong 4.09434\gamma.$$

For $\gamma_1 = -0.090610$, we obtain the estimates

$$A'(\gamma_1) \cong -3.0992,$$

$$|A''(\gamma)| \leq 4.09434\,(0.09062 + c) \qquad \text{for } |\gamma - \gamma_1| \leq c.$$

Hence

$$\Omega > 3.0992\,c - 2^{-1}(4.09434)(0.09062 + c)c^2 \geq 3.0992\,c - 2.04717(0.09062 + c)c^2,$$

and relation (27) becomes

$$L(kLd + 0.0026) < \Omega.$$

For $c = 0.00063$, $d = 0.0038$, $R = 0.2$, and $R_0 = 0.08279$, we deduce that

$$c + 0.0026 = 0.0032300 < 0.0037799 \leq (1 - 0.044\,L)d,$$

$$1.2934c + 0.0041 \leq 0.00493 < 0.8276 \leq R_0 - 0.067\,Ld,$$

$$L = 0.12, \quad Lk = (0.12)(0.044) = 0.00528 < 1,$$

$$L(kLd + 0.0026) \leq 0.0003145,$$

$$0.0019520 \leq 3.0992c - 2.04717(0.09062 + c)c^2 < \Omega.$$

Thus an exact solution x(t) of problem (39) *exists,* and

$$\|x(t) + 0.11873 \sin 2.0288t\| \leq d = 0.0038,$$

$$|x(t) + 0.11873 \sin 2.0288t| \leq R_0 \leq 0.083,$$

$$|Px(t) + 0.11873 \sin 2.0288t| \leq c = 0.00063.$$

REFERENCES

1. W. Bader, *Nichtlineare Systeme und ihre mathematische Behandlung,* Nachrichten Techn. Fachber. 21 (1960), 1-11.

2. R. G. Bartle, *Singular points in functional equations,* Trans. Amer. Math. Soc. 75 (1953), 366-384.

3. C. Borges, (a) *Periodic solutions of nonlinear differential equations. Existence and error bounds.* Contributions to Differential Equations, J. Wiley & Sons, Inc., New York, to appear; (b) *On the existence of periodic solutions of nonlinear differential equations near a given first Galerkin approximation.* (to appear).

4. L. Cesari, *Functional Analysis and periodic solutions of nonlinear differential equations,* Contributions to Differential Equations, J. Wiley & Sons, Inc., New York, 1 (1963), 149-167.

5. ———, *Periodic solutions of nonlinear differential systems,* Proceedings Symposium Active Networks and Feedback Systems, Polytechnic Institute Brooklyn, 10 (1960), 545-560.

6. B. G. Galerkin, *Series developments for some cases of equilibrium of plates and beams,* Vestnik Ingen. Teknik 19 (1915), 897-908 (Russian).

7. L. V. Kantorovič, *Functional analysis and applied mathematics,* Uspehi Math. Nauk 3 (1948), no. 6, 89-185, (Russian). English translation: National Bureau of Standards Report no. 1509, 1952, (mimeographed). See particularly pp. 83-92.

8. L. V. Kantorovič and V. I. Krylov, *Approximation methods in higher analysis,* Moscow-Leningrad 1952, (Russian). English translation: P. Noordhoff, Groningen, 1958. See particularly pp. 258-303.

9. L. V. Kantorovič and G. P. Akilov, *Functional analysis and normed spaces,* Gosudarstv. Izdat. Fis.-Mat. Lit., Moscow 1959, (Russian). See particularly pp. 487-540.

10. M. V. Keldyš, *On Galerkin's method of solution of boundary problems,* Izv. Akad. Nauk SSSR Ser. Mat. 6 (1942), 309-330, (Russian).

11. H. W. Knobloch, *Remarks on a paper of L. Cesari on functional analysis and nonlinear differential equations,* Michigan Math. J. 10 (1963), 417-430.

12. ———, *Eine neue Methode zur Approximation periodischer Lösungen nicht-linearer Differentialgleichnungen zweiter Ordnung,* Math. Z. 82 (1963), 177-197.

13. E. H. Rothe, *Critical points and gradient fields of scalars in Hilbert space,* Acta Math. 85 (1951), 73-98.

14. ———, *Gradient mappings,* Bull. Amer. Math. Soc. 59 (1953), 5-19.

15. E. Schmidt, *Über die Auflösung der nichtlinearen Integralgleichung und die Verzweigung ihrer Lösungen,* Math. Ann. 65 (1908), 370-399.

Some further references on Galerkin's method

16. L. Collatz, *Genäherte Berechnung von Eigenwerten,* Z. Angew. Math. Mech. 19 (1939), 224-248, 297-318.

17. ———, *Eigenwertaufgaben mit technischen Anwendungen,* Mat. und ihre Anwend. in Phys. und Tech., Reihe A, Bd. 19, Akad. Verlag, Leipzig, 1949. See particularly pp. 205-244.

18. R. C. DiPrima, *Application of the Galerkin method to a problem in hydrodynamic stability,* Quart. Appl. Math. 13 (1955), 55-62.

19. W. J. Duncan, *Galerkin's method in mechanics and differential equations,* Ministry of Aircraft Prod., London, Aeron. Res. Comm. Rep. and Mem., 1937, no. 1798, 33 pp.

20. ———, *The principles of the Galerkin method,* Ministry of Aircraft Prod., London, Aeron. Res. Comm. Rep. and Mem., 1938, no. 1848, 24 pp.

21. L. S. Leibenson, *Variational methods for the solution of problems of the theory of elasticity,* Moscow-Leningrad, 1943, (Russian).

22. W. E. Milne, *Numerical solution of differential equations,* J. Wiley & Sons, Inc., New York. See particularly pp. 114-118.

23. D. Yu. Panov, *Application of Galerkin's method to some nonlinear problems of the theory of elasticity,* Prikl. Mat. Meh. 3 (1939), 139-142, (Russian).

24. P. F. Papkovitsch, *Zwei Fragen zur Theorie der dünnen elastischen Platten,* Prikl. Mat. Meh. 5 (1941), 359-374.

25. Y. I. Pearlman, *Galerkin's method in calculus of variations and in the theory of elasticity,* Prikl. Mat. Meh. 5 (1941), 345-358 (Russian).

26. A. M. O. Smith, *On the growth of Taylor-Görtler vortices along highly concave walls,* Quart. Appl. Math. 13 (1955), 233-262.

27. B. A. Sokolov, *On the torsion of a shaft of varying width,* Prikl. Mat. Meh. 3 (1937), 153-160, (Russian).

28. L. G. Stepanjantz, *The calculation of laminar boundary layers around bodies of revolution,* Prikl. Mat. Meh. 6 (1942), 317-326, (Russian).

The University of Michigan