AN APPLICATION OF LINEAR ANALYSIS TO INITIAL VALUE PROBLEMS

by

ALAN GREENWELL LAW

B.A., University of British Columbia, 1958

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

MATHEMATICS

We accept this thesis as conforming to the required standard.

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1961

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of_____

The University of British Columbia,
Vancouver 8, Canada.

Date_____

ABSTRACT

Certain properties of an unknown element u in a Hilbert space are investigated. For u satisfying certain linear constraints, it is shown that approximations to u and error bounds for the approximations may be obtained in terms of functional representers.

The general approximation method is applied to homogeneous systems of ordinary linear differential equations and various formulae are derived. An Alwac III-E digital computer was used to compute optimal approximations and error bounds with the aid of these formulae.

Numerous applications to particular systems are mentioned. On the basis of the numerical results, certain remarks are given as a guide for the numerical application of the method, at least in the framework of ordinary differential equations. From the cases studied it is seen that this can be a practicable method for the numerical solution of differential equations.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to acknowledge the invaluable guidance and assistance extended to me by Dr. C. A. Swanson; without his counsel this thesis would never have materialized.

I should also like to thank Dr. T. E. Hull for his helpful remarks and Dr. C. Clark for his assistance in preparing the final manuscript.

1. INTRODUCTION

The highly diverse field of approximation theory has been considered by such authors as Courant [4], Kantorovich [11], Lanczos [13] and Strutt [18]. Much of the theoretical approach depends on functional analysis. In the last few years, with the advent of high speed computing devices, functional analysis has become of increasing practical importance [1, 3, 6, 7] and, in fact, appears to be developing under the consideration of the capabilities of such devices.

In our research we intended to study methods for solving partial differential equations, with the hope that various classical approaches [11, 13], such as the Rayleigh-Ritz method [4, 5, 11, 18], could be used to advantage with the aid of computers. We also considered more modern methods [2, 5, 7, 10] and it was decided that a method based on that of Golomb and Weinberger [7] deserved further consideration. This method holds much promise since it is pertinent for a variety of approximation problems [7, p. 117] and, also, practical expressions for the maximum error incurred in the approximation may be generated [7, p. 134].

In studying a numerical method for solving differential equations, it is often fruitful to consider the ordinary case before the partial case. Thus, in this thesis, we have restricted our research to a system of ordinary differential equations of the form (4.3). The research required extensive numerical experimentation. This was accomplished with the aid

of an Alwac III-E digital computer at the University of British Columbia Computing Centre; without an electronic computer our numerical procedures could not be utilized.

In order to evaluate the method, we considered it in the framework of simple models (see, e.g., (9.2)). On the basis of the results obtained, some points are clarified so that the numerical success of the process might be increased.

Sections 2 and 3 are devoted to a general development of the theory. It is shown that, for an unknown vector u, which is subject to certain linear constraints, approximations to u and error bounds for the approximations may be obtained in terms of functional representers.

Section 4 deals with an evaluation of the results derived in section 3 for the case of a linear system of differential equations. The unknown vector u is the solution of an initial value problem, where the constraints imposed on u are the initial values. Explicit expressions for the approximation $F_o(\tilde{u})$ and the error estimate E are obtained in terms of initial values and functional representers.

In sections 5 and 6 the pertinent formulae of section 4 are presented in a form which is suitable as a guide for computer programming. Also, a criterion is given for computing polynomials which are orthogonal with respect to the scalar product adopted in this thesis.

In sections 6 and 7 there is some discussion of the numerical results obtained for a single equation and a system of

two equations. These results are considered in determining the details for applying the procedure to the system of section 8.

2. HILBERT SPACE AND BOUNDED LINEAR FUNCTIONALS

Definition 2.1. A (real) Hilbert space $\mathfrak{H}$ is a set of abstract elements u, v, w, ..., called vectors, which satisfy the following conditions:

(i) $\mathfrak{H}$ is a linear vector space over the field of real numbers:

(a) $\mathfrak{H}$ is an Abelian group

(b) $\mathfrak{H}$ admits multiplication by real numbers α, β, ... so that, if $u, v \in \mathfrak{H}$, then

(1) $\alpha u \in \mathfrak{H}$

(2) $(\alpha+\beta)u = \alpha u + \beta u$

(3) $(\alpha\beta)u = \alpha(\beta u)$

(4) $\alpha(u+v) = \alpha u + \alpha v$;

(ii) $\mathfrak{H}$ is a normed space whose norm is derived from a scalar product:

(a) with every pair of elements u, v in $\mathfrak{H}$ there is associated a real number, called the scalar product and denoted by (u, v), in such a way that the following rules are satisfied:

(1) $(\alpha u, v) = \alpha(u, v)$ for every number α

(2) $(u+v, w) = (u, w) + (v, w)$

(3) $(u, v) = (v, u)$

(4) $(u,u) \geq 0$ and $(u,u) = 0$ if and only if $u = 0$;

(b) with every element u in $\mathcal{H}$ there is associated a real number $||u||$, the <u>norm</u> of u, which is defined by

$$||u|| = (u,u)^{1/2} ;$$

(iii) $\mathcal{H}$ <u>is a complete space</u>: any Cauchy sequence of elements in $\mathcal{H}$ converges in the norm to an element of $\mathcal{H}$. Thus, if $\{u_n\}$ is a sequence of elements in $\mathcal{H}$, and $||u_n - u_m|| \longrightarrow 0$ as $n, m \longrightarrow \infty$, then there exists an element u in $\mathcal{H}$ so that $||u - u_n|| \longrightarrow 0$ as $n \longrightarrow \infty$.

It follows from condition (ii) that the norm has the properties of a metric, i.e.,

(a) $||u|| \geq 0$, with equality if and only if $u = 0$ (positive definite property);

(b) $||\alpha u|| = |\alpha| \, ||u||$ (homogeneous property);

(c) $||u+v|| \leq ||u|| + ||v||$ (Minkowski inequality);

for any $u, v \in \mathcal{H}$ and any number α. Properties (a) and (b) are immediate and (c) is a consequence of the well-known [4, p. 49] Schwarz inequality

$$|(u,v)| \leq ||u|| \, ||v||, \quad u,v \in \mathcal{H} .$$

<u>Definition 2.2</u>. A <u>linear functional</u> F is a mapping $u \longrightarrow F(u)$ from $\mathcal{H}$ into the real numbers such that the following conditions are satisfied:

(a) $F(\alpha u+\beta v) = \alpha F(u) + \beta F(v)$, for all $u,v \in \mathcal{H}$ and all numbers α and β

(b) there exists a positive number M such that, for all $u \in \mathcal{H}$,

$$|F(u)| \leq M ||u||.$$

Definition 2.3. The linear functionals $F_1, F_2, \ldots, F_n$, defined on $\mathcal{H}$, are said to be linearly independent if:

$$\sum_{i=1}^{n} \alpha_i F_i(u) = 0 \quad \text{for every} \quad u \in \mathcal{H}$$

implies $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

Remark 2.4. If F is a linear functional on a Hilbert space $\mathcal{H}$, then by the Riesz representation theorem [16, p. 61] there exists an element z in $\mathcal{H}$ such that

$$F(u) = (u,z) \quad \text{for every} \quad u \in \mathcal{H}.$$

Moreover, the representer z is uniquely determined by the functional F.

3. OPTIMAL APPROXIMATION OF LINEAR FUNCTIONALS

Suppose there are given, a priori, $n+1$ linear functionals $F_0, F_1, \ldots, F_n$ which are defined on a known Hilbert space $\mathcal{H}$. We consider an unknown element $u \in \mathcal{H}$ which satisfies

$$F_i(u) = \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{3.1}$$

for fixed known numbers $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ and seek an approximation to the value $F_0(u)$. It shall always be assumed that the

linear functionals $F_o, F_1, \ldots, F_n$ are linearly independent. Under this assumption, it can be shown [20, p. 151] that there exists an element v in $\mathfrak{H}$ such that

$$(3.2) \qquad \begin{cases} F_o(v) = 1 \\ F_i(v) = 0, \quad i = 1, 2, \ldots, n. \end{cases}$$

For any number α, $w = u+\alpha v$ is in $\mathfrak{H}$ and

$$F_i(w) = F_i(u) + \alpha F_i(v) = \varepsilon_i, \quad i = 1, 2, \ldots, n.$$
$$F_o(w) = F_o(u) + \alpha F_o(v) = F_o(u) + \alpha.$$

Thus, $F_o(u)$ can assume arbitrary values for u satisfying (3.1). We wish u to be restricted to some subset $\tilde{\mathfrak{F}}$ of $\mathfrak{H}$ on which the linear functionals $F_o, F_1, \ldots, F_n$ are bounded. Suppose then, that u satisfies (3.1) and

$$(3.3) \qquad ||u|| \leq \lambda$$

where λ is some known number, and let

$$(3.4) \qquad \tilde{\mathfrak{F}} = \{v \in \mathfrak{H} \mid ||v|| \leq \lambda \; ; \; F_i(v) = \varepsilon_i, \quad i = 1,2,\ldots,n\}.$$

Then $F_o, F_1, \ldots, F_n$ are bounded for $v \in \tilde{\mathfrak{F}}$. Let the centre $\tilde{u}$ of the hypercircle $\tilde{\mathfrak{F}}$ be that element which satisfies

$$(3.5) \qquad \begin{cases} ||\tilde{u}|| = \inf\limits_{F_i(v) = \varepsilon_i} ||v|| \\ F_i(\tilde{u}) = \varepsilon_i, \quad i = 1, 2, \ldots, n. \end{cases}$$

The fact that there exists a unique $\tilde{u}$ satisfying conditions (3.5) is well known [16, p. 71]. Since $F_i(u) = \varepsilon_i$, $i = 1,2,\ldots,n$, $||\tilde{u}|| \leq ||u|| \leq \lambda$ and hence $\tilde{u} \in \tilde{\mathfrak{F}}$. By taking the first variation of (v,v) we find that

$$(3.6) \qquad (\tilde{u}, v) = 0$$

for all $v \in \mathcal{S}$, where

$$(3.7) \qquad \mathcal{S} = \{v \in \mathcal{H} \mid F_i(v) = 0, \quad i = 1, 2, \ldots, n\},$$

i.e., $\tilde{u}$ is in the orthogonal complement $\mathcal{S}^{\perp}$ of $\mathcal{S}$ in $\mathcal{H}$. The conditions (3.6) and the second of (3.5) determine $\tilde{u}$ uniquely. Let $\tilde{y}$ be that element of unit norm in $\mathcal{S}$ for which $F_o(v)$ attains its upper bound when v varies under the conditions $||v|| = 1$, $v \in \mathcal{S}$. This element exists and is unique [16, p. 62]. Following Golomb and Weinberger [7, p. 134] we may state

Theorem 1. If u varies over the set $\mathcal{F}$ defined in (3.4), then the range of all possible values of $F_o(u)$ is an interval of length $2F_o(\tilde{y})[\lambda^2 - ||\tilde{u}||^2]^{1/2}$ centered about $F_o(\tilde{u})$.

Thus, the maximum error E incurred in approximating to $F_o(u)$ with $F_o(\tilde{u})$ is given by

$$(3.9) \qquad E = \pm F_o(\tilde{y})[\lambda^2 - ||\tilde{u}||^2]^{1/2} .$$

$F_o(\tilde{u})$ is the optimal approximation to the value $F_o(u)$ in the sense that no smaller interval length can be found.

By (2.4), there exist unique elements $z_o, z_1, \ldots, z_n$ of $\mathcal{H}$ such that

$$(3.10) \qquad F_i(v) = (z_i, v), \quad i = 0, 1, \ldots, n)$$

for every $v \in \mathcal{H}$. Since $F_o, F_1, \ldots, F_n$ are assumed linearly independent, the elements $z_o, z_1, \ldots, z_n$ are linearly indepen-

dent in the sense that no one of them is a linear combination of the others.

The space $\mathcal{S}$, defined in (3.7), can now be characterized as the subspace of all $v \in \mathcal{H}$ which are orthogonal to each of $z_1, z_2, \ldots, z_n$, i.e.,

$$(3.11) \qquad \mathcal{S} = \{v \in \mathcal{H} \mid (v, z_i) = 0, \quad i = 1, 2, \ldots, n\} .$$

Since $\tilde{u} \in \mathcal{S}^{\perp}$, it is a linear combination of $z_1, z_2, \ldots, z_n$. Thus,

$$(3.12) \qquad \tilde{u} = \sum_{j=1}^{n} \alpha_j z_j ,$$

and, using (3.5),

$$(3.13) \qquad \varepsilon_i = F_i(\tilde{u}) = F_i \left(\sum_{j=1}^{n} \alpha_j z_j \right) = \sum_{j=1}^{n} \alpha_j F_i(z_j)$$

$$= \sum_{j=1}^{n} \alpha_j (z_i, z_j).$$

This is a linear algebraic system in $\alpha_1, \alpha_2, \ldots, \alpha_n$. Let $B = (\beta_{ij}) = ((z_i, z_j))$ denote the coefficient matrix of this system. B is non-singular since $z_1, z_2, \ldots, z_n$ are linearly independent [4, p. 62]. If $B^{-1} = (\hat{\beta}_{ij})$ denotes the inverse of B then, from (3.12),

$$\tilde{u} = \sum_{i,j=1}^{n} \hat{\beta}_{ij} \varepsilon_i z_j ,$$

and hence

$$(3.14) \qquad F_o(\tilde{u}) = \sum_{i,j=1}^{n} \varepsilon_i \hat{\beta}_{ij} (z_o, z_j) .$$

Also,

$$(3.15) \qquad (\tilde{u},\tilde{u}) = \sum_{i,j=1}^{n} \hat{\beta}_{ij}\varepsilon_i\varepsilon_j$$

and it can be shown [7, p. 142] that

$$(3.16) \qquad F_o(\tilde{y})^2 = (z_o,z_o) - \sum_{i,j=1}^{n} \hat{\beta}_{ij}(z_o,z_i)(z_o,z_j).$$

4. INITIAL VALUE PROBLEM FOR AN HOMOGENEOUS SYSTEM OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS

In this and all subsequent sections, the Hilbert space $\mathfrak{H}$ will be specialized to the space[1] consisting of N-vectors of the form

$$v = [v^1, v^2, \ldots, v^N],$$

where each component $v^i = v^i(x)$ is an absolutely continuous, single valued function on $0 \leq x \leq 1$, having a Lebesgue square integrable derivative. The inner product in this space is defined by (4.7) (below). We also use $N \times N$ matrices $A = (a_{ij})$

---

[1] To show the completeness of $\mathfrak{H}$ when $N = 1$: every Cauchy sequence $\{F_n\}$ in $\mathfrak{H}$ has the property that the sequence of its derivatives converges in the L^2 norm to a square summable function F [16, p. 59]. Since F is also summable, it is the derivative almost everywhere of an absolutely continuous function G [16, p. 48 and p. 53]. By defining $G(0) = \lim F_n(0)$, G becomes uniquely defined, $G \in \mathfrak{H}$ and $||F_n - G|| \longrightarrow 0$. The proof extends to arbitrary N.

whose elements $a_{ij}(x)$ are piece-wise continuous functions. For any u, v and A we define the following products:

$$(a) \quad u \cdot v = \sum_{i=1}^{N} u^i v^i$$

$$(b) \quad Av = \left[\sum_{i=1}^{N} a_{1i} v^i, \sum_{i=1}^{N} a_{2i} v^i, \ldots, \sum_{i=1}^{N} a_{Ni} v^i \right]$$

$$(c) \quad vA = \left[\sum_{i=1}^{N} a_{i1} v^i, \sum_{i=1}^{N} a_{i2} v^i, \ldots, \sum_{i=1}^{N} a_{iN} v^i \right].$$

In particular, $(v \cdot v)^{1/2}$ is called the norm of the N-vector v and it shall be denoted by $\langle v \rangle$.

Definition 4.1. The derivative of any vector v (or matrix A) is defined to be a vector v' (or matrix A') of the same form whose components (or elements) are the derivatives of the corresponding components (or elements) of the vector v (or matrix A).

We shall assume an operator norm $\langle A \rangle$ such that

$$(4.2) \qquad \langle Av \rangle \leq \langle A \rangle \langle v \rangle, \quad \text{for every } v \in \mathcal{S} .$$

The unknown N-vector u is the solution [12, p. 48] of the problem

$$(4.3) \qquad \begin{cases} u' + A(x)u = 0, & 0 \leq x \leq 1 \\ u(0) = u_o , \end{cases}$$

where u_o is a given vector $[\mu^1, \mu^2, \ldots, \mu^N]$ and the matrix $A = (a_{ij})$ is given. We seek to approximate $u^1(1)$, the first component of the solution vector u evaluated at $x = 1$; i.e., we seek an approximation to $F_o(u)$ where the linear functional

F_o is defined by

$$(4.4) \qquad F_o(v) = v^1(1), \quad v \in \mathfrak{H} .$$

F_o is easily seen to be a (bounded) linear functional; this also follows from (4.12) below.

If there exists a positive integrable function $b(x)$ on $0 \leq x \leq 1$ such that

$$(4.5) \qquad -[A(x)v]\cdot v \leq b(x) \langle v \rangle^2, \quad 0 \leq x \leq 1 \text{ and all } v \in \mathfrak{H} ,$$

then [2, p. 98 and 7, p. 165]

$$(4.6) \qquad \int_o^1 \langle u'(x) \rangle^2 dx \leq \langle u_o \rangle^2 \int_o^1 \langle A(x) \rangle^2 \exp \{2 \int_o^x b(t)dt\} dx .$$

Thus, we introduce the scalar product

$$(4.7) \qquad (v,w) = \int_o^1 v'(x)\cdot w'(x)dx + v(0)\cdot w(0)$$

and (4.6) yields an a priori bound for $||u||$ (see (3.3)). The scalar product (4.7) is chosen for two reasons: first, it induces a positive definite norm $||v||$ and, second, it allows explicit formulae for the representers to be developed (see (4.12)).

Choose $n - N$ linearly independent[2] N-vectors $f_{N+1}, f_{N+2}, \cdots, f_n$ in $\mathfrak{H}$ and define the linear functionals $F_{N+1}, F_{N+2}, \cdots, F_n$ by

---

[2] If $f_{N+1}, f_{N+2}, \cdots, f_n$ are not all linearly independent then the corresponding representers are not all linearly independent and hence the matrix B is singular; i.e., $F_o, F_1, \cdots, F_n$ are linearly dependent.

$$(4.8) \qquad F_i(v) = \int_0^1 f_i(x)\cdot\{v'(x)+A(x)v(x)\}dx, \quad i = N+1,\ldots,n.$$

For any $n \geq N$ and for any choice of $f_i \in \mathfrak{F}$, we have the data

$$(4.9) \qquad F_i(u) = 0, \quad i = N+1, N+2, \ldots, n.$$

In addition, employ the N linear functionals $F_1, F_2, \ldots, F_N$ determined by

$$(4.10) \qquad F_j(v) = v^j(0), \quad j = 1, 2, \ldots, N.$$

From the initial conditions of (4.3) we have

$$(4.11) \qquad F_j(u) = \mu^j, \quad j = 1, 2, \ldots, N.$$

The following explicit formulae for the functional representers $z_0, z_1, \ldots, z_n$ can be readily obtained [7, p. 166]:

$$(4.12) \quad \begin{cases} z_0 = [1+x, 0, 0, \ldots, 0, 0] \\ z_1 = [1, 0, 0, \ldots, 0, 0] \\ z_2 = [0, 1, 0, \ldots, 0, 0] \\ \qquad \cdots \\ z_N = [0, 0, 0, \ldots, 0, 1] \\ z_i = \int_0^x \{f_i(t)+(1+t)f_i(t)A(t)\}dt \\ \qquad\qquad +(1+x)\int_x^1 f_i(t)A(t)dt, \\ \qquad\qquad i = N+1, N+2, \ldots, n. \end{cases}$$

(The equations $F_i(v) = (v, z_i)$ are easily checked using (4.12); since representers are unique, this proves (4.12)).

5. SINGLE EQUATION. COMPUTATIONAL METHODS

In this section we develop explicit expressions for the representers $z_{N+1}, z_{N+2}, \ldots, z_n$ and also for the scalar product (z_i, z_j) of any two representers in the form of power series. These formulae are derived for a single equation but, in view of remarks (5.8), they are also applicable for a system of N equations.

If we assume that the chosen vectors $f_{N+1}, f_{N+2}, \ldots, f_n$ have polynomial components and the coefficient matrix A has polynomial elements, then, because of (4.12), the corresponding representers $z_{N+1}, z_{N+2}, \ldots, z_n$ will have polynomial components; in this case the series (5.1) below will be finite and hence there is no question of convergence. Under this assumption, formulae (5.5) and (5.7) are of finite explicit format and thus are a suitable guide for computer evaluation of (3.14) - (3.16).

Let N = 1 in the previous section. Assume that

$$z_i = \sum_{r=0}^{\infty} \alpha_{ir} x^r, \quad i = 2, 3, \ldots, n. \tag{5.1}$$

Then

$$z_i^{(m)}(0) = m!\alpha_{im}, \quad m = 0, 1, 2, \ldots \tag{5.2}$$

From (4.12)

$$\begin{cases} z_i'(x) = f_i(x) + \int_x^1 f_i(t)A(t)dt \\ z_i''(x) = f_i'(x) - \{f_i(x)A(x)\} \end{cases} \tag{5.3}$$

Differentiating the second member of (5.3) k times and using Leibniz' formula we find that

$$(5.4)\qquad z_i^{(k+2)}(x) = f_i^{(k+1)}(x) - \sum_{j=o}^{\infty} \binom{k}{j} f_i^{(j)}(x) A^{(k-j)}(x),$$

$$k = 0, 1, 2, \ldots .$$

We deduce from (4.12), (5.2), (5.3) and (5.4) that

$$(5.5)\qquad \begin{cases} \alpha_{io} = \int_o^1 f_i(t)A(t)dt, \\ \alpha_{il} = \alpha_{io} + f_i(0), \\ \alpha_{i,k+2} = \frac{1}{(k+2)!}\left\{f_i^{(k+1)}(0) - \sum_{j=0}^{k} \binom{k}{j} f_i^{(j)}(0) A_{(o)}^{(k-j)}\right\}, \end{cases}$$

$$i = 2, 3, \ldots, n; \quad k = 0, 1, 2, \ldots$$

Thus, the coefficients α_{ir} in (5.1) can be computed using (5.5).

Suppose that any two representers z_i, z_j $(i, j = 1, 2, \ldots, n)$ are known in the form (5.1); let

$$(5.6)\qquad z_i = \sum_{r=o}^{\infty} \alpha_{ir} x^r, \quad z_j = \sum_{s=o}^{\infty} \alpha_{js} x^s,$$

where the numbers α_{ir}, α_{js} are known, $r, s = 0, 1, 2, \ldots$. Differentiating (5.6) term by term, substituting in (4.7), then integrating term by term, we find that (4.7) can be expressed in the form

$$(5.7)\qquad (z_i, z_j) = \alpha_{io}\alpha_{jo} + \sum_{r,s=1}^{\infty} \frac{r\alpha_{ir}s\alpha_{js}}{r+s-1}, \quad i,j = 1,2,\ldots,n.$$

Formula (5.5) and (5.7) have been developed for N=1. However, they are applicable to the system (4.3) in view of the following obvious remarks:

$$(5.8)\qquad \begin{cases} \text{(a)} \quad (z_o, z_i) = z_i^1(1); & i = 0, 1, 2, \ldots, n \\ \text{(b)} \quad (z_i, z_j) = z_i^j(0); & i = 0, 1, \ldots, n, \\ & \qquad j = 1, 2, \ldots, N \\ \text{(c)} \quad (z_i, z_j) = \sum_{K=1}^{N} (z_i^K, z_j^K); & i,j = 0, 1, \ldots, n. \end{cases}$$

6. ORTHOGONAL POLYNOMIALS

A set of polynomials $g_1(x)$, $g_2(x)$, ... which are mutually orthogonal with respect to the scalar product (4.7) will be found useful in ensuing sections.

Let $g_1(x) \equiv 1$ and suppose that

$$(6.1)\qquad g_i(x) = \sum_{j=o}^{i-1} \gamma_{ij} x^j, \quad i = 2, 3, \ldots .$$

Impose the conditions

$$(6.2)\qquad \begin{cases} (g_i, g_k) = 0; & k = 1, 2, \ldots, (i-1) \\ (g_i, g_i) \neq 0. \end{cases}$$

When (6.1) is substituted into (6.2) and (5.7) is used (or (4.7)), we obtain a system of (i-1) linear homogeneous equations in $\gamma_{io}, \gamma_{il}, \ldots, \gamma_{i,i-1}$. For each $i = 2, 3, 4, \ldots$, solutions in the form of relatively prime integers are easily found. A set of polynomials (6.1) which satisfy (6.2) is then

$$(6.3)\qquad \begin{cases} g_1 = 1 \\ g_2 = x \\ g_3 = x^2 - x \\ g_4 = 2x^3 - 3x^2 + x \\ g_5 = 5x^4 - 10x^3 + 6x^2 - x \\ \quad \cdot \quad \cdot \quad \cdot \end{cases}$$

7. EXPERIMENTAL RESULTS. SINGLE EQUATION

Suppose the problem is

$$(7.1) \qquad \begin{cases} u' + xu = 0, & 0 \leq x \leq 1 \\ u(0) = 1, \end{cases}$$

and an approximation to $u(1)$ is sought.

The exact solution of (7.1) is $u(x) = \exp(-\frac{1}{2}x^2)$ and $u(1) \approx 0.606531$.

Numerous values for n and choices of functions $f_2, f_3, \ldots, f_n$ were adopted and the resulting approximations $F_o(\tilde{u})$ to $u(1)$ were obtained. The following remarks are based on extensive numerical experimentation.

Remark 1. For any chosen set of functions $f_2, f_3, \ldots, f_{k+1}, \ldots$, let $F_{ok}(\tilde{u})$ denote the approximation $F_o(\tilde{u})$ obtained for the choice of the k functions $f_2, f_3, \ldots, f_{k+1}$. Then, as may be expected for a numerical procedure, the sequence $\{|F_{ok}(\tilde{u})-u(1)|\}$, $k = 1, 2, \ldots$ is not necessarily monotonically convergent.

Remark 2. In some cases the matrix B (see (3.13) et seq.) appeared to be not well-conditioned [8, p. 439, 14, 15, 19]. For example, for $n = 7$ and choosing $f_i = x^{i-2}$, $i = 2,3,\ldots,7$, there was numerical difficulty in inverting B. However, in those cases in which $f_2, f_3, \ldots, f_n$ were all chosen from the set (6.3), the matrix B was well-conditioned.

Remark 3. On the basis of the experimentation, it appeared to

be important that the constant function belong to the set $\{f_2, f_3, \ldots, f_n\}$.

8. EXPERIMENTAL RESULTS. TWO EQUATIONS

Suppose the problem is

$$(8.1) \qquad \begin{cases} u' + Au = 0, & 0 \le x \le 1 \\ u(0) = u_o \end{cases}$$

where $u = [u^1(x), u^2(x)]$, $u_o = [1,0]$ and $A = \begin{bmatrix} 0 & 1 \\ -x & 0 \end{bmatrix}$, and an approximation to $u^1(1)$ is sought. The exact solution [7, p. 166] involves a Bessel function of order 1/3; the value [9] of $u^1(1)$ is 0.83881.

There were numerous experiments with the method for system (8.1). It was found that for the choice

$$(8.2) \qquad f_3 = [g_1, 0],\ f_4 = [0, g_1],\ f_5 = [g_2, 0],\ f_6 = [0, g_2], \ldots$$

the approximations $F_o(\tilde{u})$ to $u^1(1)$ were as follows:

n	$F_o(\tilde{u})$
4	.794006
5	.781422
6	.839162
7	.839143
8	.838793

For fixed n, the approximation obtained for the choice (8.2) was numerically closer to $u^1(1)$ than the one obtained for any other choice for $f_3, f_4, \ldots, f_n$.

9. SYSTEM OF EQUATIONS. EXPERIMENTAL RESULTS

From sections 7 and 8 it is clear that, for satisfactory approximating, the choice of vectors $f_{N+1}, f_{N+2}, \ldots, f_n$ is, by no means, an arbitrary one.

Since $g_i(x)$ is a polynomial of degree (i-1), it can easily be seen that any polynomial has a unique representation as a linear combination of the g_i's. It then follows from the Weierstrass Approximation Theorem [4, p. 65] that the system $\{g_i(x)\}$ of orthogonal polynomials is complete. An immediate generalization is that the system of N-vectors

$$(9.1) \qquad \begin{cases} [g_1,0,\ldots,0],\ [0,g_1,\ldots,0],\ldots,[0,0,\ldots,g_1], \\ [g_2,0,\ldots,0],\ [0,g_2,\ldots,0],\ldots,[0,0,\ldots,g_2], \\ [g_3,0,\ldots,0],\ldots \end{cases}$$

is complete. The members of this system are mutually orthogonal with respect to the scalar product (4.7).

On the basis of the experimental evidence in sections 7 and 8, it could be expected that an appropriate choice of $f_{N+1}, f_{N+2}, \ldots, f_n$ from the set (9.1) could result in (a) a well-conditioned matrix B and (b) a numerically close approximation $F_o(\tilde{u})$ to $F_o(u)$. In view of this, suppose that the problem is

$$(9.2) \qquad \begin{cases} u' + Au = 0, \quad 0 \le x \le 1 \\ u(0) = u_o \end{cases}$$

where $u = [u^1,u^2,u^3]$, $u_o = [1,0,0]$ and $A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ x & 0 & 0 \end{bmatrix}$, and an approximation to $u^1(1)$ is sought.

First we calculate a bound for $||u||$ in order that the error estimate in the approximation may be found.

Let $v = [v^1(x), v^2(x), v^3(x)]$ be any vector in $\mathcal{S}$. Then, for $0 \leq x \leq 1$,

(9.3) $\quad <A(x)v>^2 \leq <v>^2, \quad$ for all $v \in \mathcal{S}$.

By the Schwarz inequality, $|Av \cdot v| \leq <Av><v>$ and, using (9.3), we conclude that

(9.4) $\quad - Av \cdot v \leq <v>^2$.

Thus, we can choose $b(x) \equiv 1$ in (4.5) and, since $<u_o>^2 = 1$ and $<A(x)> \leq 1$ for $0 \leq x \leq 1$, we have from (4.6)

$$\int_o^1 <u'(x)>^2 dx \leq \int_o^1 <A(x)>^2 e^{2x} dx \leq \frac{1}{2} e^2 .$$

Finally, using (4.7), $(u,u) = ||u||^2 \leq \frac{1}{2} e^2 + 1 = 4.694529$. Hence we may take

(9.5) $\quad \lambda^2 = 4.694529$ in (3.9).

Choose $f_4, f_5, f_6, \ldots$ according to

$$\text{(9.6)} \quad \begin{cases} f_4 = [1,0,0], \ f_5 = [0,1,0], \ f_6 = [0,0,1], \\ f_7 = [x,0,0], \ f_8 = [0,x,0], \ f_9 = [0,0,x], \\ f_{10} = [x^2-x,0,0], \ \ldots \end{cases}$$

where the non-zero components are chosen from the set (6.3); thus construct the linear functionals (4.8). Using (5.5) (or (4.12)) the representers $z_4, z_5, \ldots,$ determined by the linear functionals (4.8), may then be found. Thus, in view of (4.12),

we have $z_o, z_1, \ldots, z_n$ as given in Table 1. The symmetric matrix $B = (\beta_{ij}) = ((z_i, z_j))$ can be constructed with the aid of (5.7) and (5.8) (or (4.7)) and is given in Table II. Also, from (5.8)(a),

$$(9.7)\qquad \begin{cases} (z_o, z_o) = 2,\ (z_o, z_1) = 1,\ (z_o, z_2) = 0, \\ (z_o, z_3) = 0,\ (z_o, z_4) = 1,\ (z_o, z_5) = 0, \\ (z_o, z_6) = .833333,\ (z_o, z_7) = .5,\ (z_o, z_8) = 0, \\ (z_o, z_9) = .583333,\ (z_o, z_{10}) = -.166667,\ (z_o, z_{11}) = 0, \\ (z_o, z_{12}) = -.133333, \ \ldots \end{cases}$$

From (9.2):

$$(9.8)\qquad \begin{cases} F_1(u) = 1 \\ F_j(u) = 0, \quad j = 2, 3, 4, \ldots, n. \end{cases}$$

For each $n = 4, 5, 6, \ldots$ and the choice (9.6), we can calculate (3.14) from (9.7), (9.8) and Table II; also (3.15) and (3.16) may be evaluated and, for the bound (9.5), E can be computed according to (3.9). The numerical results obtained are condensed into Table III.

Remark 1. In all the cases considered, the matrix B was well-conditioned.

Remark 2. Because of the accuracy of the calculations, the numbers E of Table III are expected to be gross estimates. For example, for $n = 12$, $F_o(\tilde{y})^2$ in (3.16) is .00000δ, where δ appears to be indeterminate because of the number of significant digits in the computation.

Remark 3. After the results in Table III had been obtained, Dr. Z. A. Melzak pointed out to the author that $u^1(1) \approx .958457$ may be calculated from power series.

To Follow Page 21

TABLE I

Representers Corresponding to the Choice (9.6)

for the Problem (9.2)

$z_0 = [1+x,\ 0,\ 0]$

$z_1 = [1,\ 0,\ 0]$

$z_2 = [0,\ 1,\ 0]$

$z_3 = [0,\ 0,\ 1]$

$z_4 = [x,\ .5x^2-x-1,\ 0]$

$z_5 = [0,\ x,\ .5x^2-x-1]$

$z_6 = [-.166667x^3+.5x+.5,\ 0,\ x]$

$z_7 = [.5x^2,\ .166667x^3-.5x-.5,\ 0]$

$z_8 = [0,\ .5x^2,\ .166667x^3-.5x-.5]$

$z_9 = [-.083333x^4+.333333x+.333333,\ 0,\ .5x^2]$

$z_{10} = [.333333x^3-.5x^2,\ .083333x^4-.166667x^3+.166667x+.166667,\ 0]$

$z_{11} = [0,\ .333333x^3-.5x^2,\ .083333x^4-.166667x^3+.166667x+.166667]$

$z_{12} = [-.05x^5+.083333x^4-.083333x-.083333,\ 0,\ .333333x^3-.5x^2]$

. . .

TABLE II

Matrix B = (z_i, z_j), i, j = 1, 2, ...,

for the Representers of Table I

1										
0	1									
0	0	1								
0	-1	0	2.333333							
0	0	-1	-.5	2.333333						
.5	0	0	.333333	-.5	1.383333					
0	-.5	0	1.208333	-.333333	.125	.716667				
0	0	-.5	-.166667	1.208333	-.333333	-.125	.716667			
.333333	0	0	.25	-.166667	.763888	.1	-.125	.515875		
0	.166667	0	-.391669	.083333	-.058333	-.202779	.024998	-.044447	.071429	
0	0	.166667	.083333	-.391669	.083333	.058333	-.202779	.024998	-.013886	.071429
-.083333	0	0	-.05	.083333	-.229364	-.016667	.058333	-.126288	.008733	-.013886	.043647

To Follow
Page 21

To Follow Page 21

TABLE III

Approximation $F_o(\tilde{u})$ and Error E for (9.2)
With the Choice (9.6)

n	$F_o(\tilde{u})$	E
6	.888889	$\pm$ .528271
7	.885198	$\pm$ .524532
8	.926869	$\pm$ .251036
9	.957037	$\pm$ .067486
10	.957186	$\pm$ .067430
11	.957224	$\pm$ .067353
12	.958447	$\pm$ .002577

10. REFERENCES

[1] Buck, R. C. "Linear spaces and approximation theory", in R. E. Langer, ed., On numerical approximation; proceedings of a symposium conducted by the Mathematics Research Center, United States Army, at the University of Wisconsin, Madison, April 21-23, 1958. Madison: University of Wisconsin Press, 1959, 11-23.

[2] Coddington, E. A. and N. Levinson. Theory of ordinary differential equations. New York: McGraw-Hill Book Co., 1955.

[3] Courant, R. "Remarks about the Rayleigh-Ritz method", in R. E. Langer, ed., Boundary problems in differential equations; proceedings of a symposium conducted by the Mathematics Research Center, United States Army, at the University of Wisconsin, Madison, April 20-22, 1959. Madison: University of Wisconsin Press, 1960, 273-277.

[4] Courant, R. and D. Hilbert. Methods of mathematical physics, vol. 1. New York: Interscience Publishers, 1953.

[5] Forsythe, G. E. and P. C. Rosenbloom. Numerical analysis and partial differential equations. New York: John Wiley and Sons, 1958.

[6] Friedrichs, K. O. Functional analysis and applications. [New York]: New York University, Institute of Mathematical Sciences, 1953.

[7] Golomb, M. and H. F. Weinberger. "Optimal approximation and error bounds", in R. E. Langer, ed., On numerical approximation; proceedings of a symposium conducted by the Mathematics Research Center, United States Army, at the University of Wisconsin, Madison, April 21-23, 1958. Madison: University of Wisconsin Press, 1959, 117-90.

[8] Hildebrand, F. B. Introduction to numerical analysis. New York: McGraw-Hill Book Co., 1956.

[9] Jahnke, E. and F. Emde. Tables of functions with formulae and curves. New York: Dover Publications, S 133.

[10] John, F. Advanced numerical analysis. [New York]: New York University, Institute of Mathematical Sciences, 1956.

[11] Kantorovich, L. V. and V. I. Krylov. Approximation methods of higher analysis. Translated by C. D. Benster. Groningen, The Netherlands: P. Noordhoff, 1958.

[12] Kolmogorov, A. N. and S. V. Fomin. Elements of the theory of functions and functional analysis, vol. 1. Translated by L. F. Boron. Rochester, New York: Graylock Press, 1957.

[13] Lanczos, C. Applied analysis. Englewood Cliffs, New Jersey: Prentice Hall, 1956.

[14] von Neumann, J. and H. H. Goldstine. "Numerical inverting of matrices of higher order", Bulletin of the American Mathematical Society. 53 (November 1947), 1021-99.

[15] von Neumann, J. and H. H. Goldstine. "Numerical inverting of matrices of higher order. II", Proceedings of the American Mathematical Society. 2 (1951), 188-202.

[16] Riesz, F. and B. Sz.-Nagy. Functional analysis. Translated by L. F. Boron. New York: Frederick Ungar, 1955.

[17] Sneddon, I. N. Special functions of mathematical physics and chemistry. New York: Interscience Publishers, 1961.

[18] Strutt, J. W. (Baron Rayleigh). The theory of sound, vol. 1. London: Macmillan, 1894.

[19] Todd, J. "The condition of certain matrices. I", Quarterly Journal of Mechanics and Applied Mathematics. 2 (1949), 469-72.

[20] Zaanen, A. C. Linear analysis. New York: Interscience Publishers, 1953.