XIVth INTERNATIONAL ASTRONAUTICAL CONGRESS

XIVe CONGRÈS INTERNATIONAL D'ASTRONAUTIQUE

Paris, 1963

Editors

EDMOND BRUN and IRWIN HERSEY

1965

GAUTHIER-VILLARS, Paris

PWN—POLISH SCIENTIFIC PUBLISHERS, Warszawa

IV. 12

A CONVERGENCE THEOREM ON THE ITERATIVE SOLUTION OF NON-LINEAR TWO-POINT BOUNDARY-VALUE SYSTEMS

by

R. McGILL and P. KENNETH*

Grumman Aircraft Engng. Corp., Bethpage, Long Island, New York (U.S.A.)

Abstract

The non-linear two-point boundary-value problem occurs quite naturally in studies in flight-mechanics, e.g., in the numerical application of the theory of the calculus of variations and in intercept and rendezvous studies. Computationally, obtaining solutions to this class of problems is fraught with difficulties such that no really satisfactory systematic procedure is currently available. A widely used procedure for approaching non-linear problems in diverse branches of mathematics is the substitution of a sequence of linear problems for the non-linear problem in such a manner that the sequence of solutions to the linear problems approach in a limiting sense the solution of the non-linear problem. By means of an analog to the simple but powerful geometric notion underlying the Newton–Raphson method this procedure is applied to systems of non-linear differential equations with two-point boundary conditions. The theorem proved here yields sufficient conditions for convergence and error bounds (which serve as stopping criteria for automatic machine computation) for the resulting strongly convergent sequence. The proof of the theorem, which shows the convergence to be uniform and rapid (quadratic), is based on the Contraction Mapping Principle, as given for example by LYUSTERNIK, and is in fact an existence and uniqueness proof for the solution of the non-linear system. The motivation for this approach is provided by the fact that the *linear* two-point boundary value problem is considerably more tractable numerically, so that the theorem proved here provides support for what may be a powerful method for solving non-linear boundary value problems by means of high-speed computers. This may be particularly so with regard to systems of higher dimension (of great importance

* The writers wish to express their appreciation to H. J. KELLEY for valuable discussions concerning the nature of the problem and the approach treated here, and to G. PINKHAM for suggesting a simplification of the numerical procedure for solving the linear system. In addition, we are indebted to G. TAYLOR for his able programming of the numerical example. Portions of this work were generated in connection with AFOSR Contract No. AF 49 (638)-1207.

in flight mechanics). A few examples of the application of the method under study here are available for the case of a single non-linear equation (HESTENES, KALABA, *et al.*). However, very little is known concerning the application to systems of higher order. A numerical example of the application of the method to systems of equations is given in this paper.

Introduction

Consider the following non-linear equation:

$$y=f(x)\,.$$

For given y and an approximate solution $x=x_0$, we wish to find x such that this equation is satisfied. Writing

$$F(x)=y-f(x)=0$$

we replace $F(x)$ by

$$F(x_0)+F'(x_0)[x-x_0]\,.$$

Setting this relation to zero we solve the resulting linear equation for x_1, and so forth. In other words we replace the original non-linear problem with a sequence of linear problems. Geometrically we have replaced the curve $f(x)$ with its tangent at $x=x_0$. The following sequence of linear equations (the Newton–Raphson sequence) results:

$$y-f(x_n)-f'(x_n)[x_{n+1}-x_n]=0 \qquad n=0,1,2,\ldots$$

This is, of course, a powerful and commonly used method for extracting roots of non-linear equations, and various sufficient conditions for convergence and error bounds are known. For example, if $f(x)$ is a real valued function defined on the real line and x_0 is an approximate solution to the equation $f(x)=0$ such that: $f'(x_0)\neq 0$, $f''(x)$ exists and is bounded on the interval $\{x_0,\ x_0-2[f(x_0)/f'(x_0)]\}$, i.e., $|f''(x)|\leqslant K$, then the condition $|f(x_0)/f'(x_0)^2|\,K\leqslant\frac{1}{2}$ is sufficient for the convergence of the foregoing sequence $\{x_n\}$ to the unique solution x^* of the equation $f(x)=0$ in the interval $\left[x_0,\ x_0-2\,\dfrac{f(x_0)}{f'(x_0)}\right]$. The error is bounded by

$$|x_{n+1}-x^*|<\frac{K}{|f'(x_n)|-K|x_{n+1}-x_n|}|x_{n+1}-x_n|^2<\frac{2K}{|f'(x_n)|}|x_{n+1}-x_n|^2.$$

This latter inequality serves as a useful stopping criterion for machine computation.

Now consider the following non-linear differential equation with non-homogeneous boundary conditions:

$$F(u'',u',u,x)=0 \qquad \begin{matrix} u(a)=u_a \\ u(b)=u_b \end{matrix}$$

By analogy with the previous case we obtain

$$\begin{aligned} &F(u_n'',u_n',u_n,x)+F_u(u_n'',u_n',u_n,x)[u_{n+1}(x)-u_n(x)] \\ &+F_{u'}(u_n'',u_n',u_n,x)[u_{n+1}'(x)-u_n'(x)] \\ &+F_{u''}(u_n'',u_n',u_n,x)[u_{n+1}''(x)-u_n''(x)]=0\,, \qquad n=0,1,2,\ldots \end{aligned}$$

where $u_0(x)$ is an approximate solution for the non-linear equation. Suppose the original equation may be written as

$$F(u'',u',u,x)=u''-f(u',u,x)=0\,.$$

Then we have $F_u=-f_u$, $F_{u'}=-f_{u'}$ and $F_{u''}=1$, which yields

$$\begin{aligned} u_{n+1}''(x)=f(u_n',u_n,x)&+f_u(u_n',u_n,x)[u_{n+1}(x)-u_n(x)] \\ &+f_{u'}(u_n',u_n,x)[u_{n+1}'(x)-u_n'(x)] \end{aligned}$$

$$\begin{matrix} u_n(a)=u_a \\ u_n(b)=u_b \end{matrix} \qquad n=0,1,2,\ldots$$

As before, we have replaced the non-linear problem with a sequence of linear problems. In what follows we shall generalize this approach to vector valued functions and prove a theorem that yields sufficient conditions for convergence and error bounds. For clarity of presentation, we shall first obtain the result for a single equation of the form $u''(x)=f(u,x)$. After introducing the appropriate generalizations, we prove the theorem for systems. This is followed by a brief discussion of a numerical procedure for obtaining the solutions to the sequence of linear systems. The results of a numerical example, concerning the solution of a problem of rocket flight, are included in the attached Appendix.

Discussion

THEOREM I. Consider the non-linear two-point boundary-value problem,

$$d^2u/dx^2=f(u,x) \qquad u(a)=u_a \qquad u(b)=u_b \tag{1}$$

where $f(u,x)$ has the following properties for $x\varepsilon[a,b]$, $u\varepsilon[u_{ab}-K,\ u_{ab}+K]$; $u_{ab}(x)=(1/b-a)\,[(u_b-u_a)x+bu_a-au_b]$: 1) $f(u,x)$ is continuous; 2) $f_u(u,x)$

$\equiv \partial f/\partial u(u, x)$ exists and is continuous, and 3) $f_u(u, x)$ is M_1 Lipschitzian with respect to u, i.e., $|f_u(u_2, x)-f_u(u_1, x)| \leqslant M_1|u_2-u_1|$.

Define the following sequence of linear differential equations:

$$d^2u_{n+1}/dx^2 = f_u(u_n, x)[u_{n+1}(x)-u_n(x)]+f(u_n, x) \qquad (2)$$
$$u_n(a)=u_a \quad u_n(b)=u_b, \quad n=0, 1, 2, \ldots$$

and $u_0(x)$ an arbitrary continuous function on $[a, b]$ such that $\max_{x\in[a,b]} |u_0(x) - u_{ab}(x)| \leqslant K$.

Then for a sufficiently small interval $[a, b]$ the unique solution to Eq. (1) exists and 1) the sequence $\{u_n(x)\}$ converges uniformly to it; 2) the convergence is at least quadratic; and 3) a bound on the error is given by

$$\max_{x\in[a,b]} |u_{n+1}(x)-u(x)| \leqslant \frac{\alpha}{(1-\alpha)^2(K+1)}\left[\max_{x\in[a,b]} |u_{n+1}(x)-u_n(x)|\right]^2,$$

where α is a positive number given below and $u(x)$ is the solution of Eq. (1).

Proof. It follows from 1, 2, and 3 that there exists M_2, $M_3>0$ such that $|f(u, x)| \leqslant M_2$, $|f_u(u, x)| \leqslant M_3$, and $|f(u_2, x)-f(u_1, x)| \leqslant M_3|u_2-u_1|$. Let $m=\max\{M_1, M_2, M_3\}$ and define the following complete metric space S:

$$S=\{u(x) | u(x) \text{ continuous on } [a, b];\ u(a)=u_a,\ u(b)=u_b;\ \rho(u, u_{ab}) \leqslant K\},$$

where

$$\rho(u_1, u_2) = \max_{x\in[a,b]} |u_1(x)-u_2(x)|.$$

Define the operator A on S:

$$Au(x)=u_{ab}(x)-\int_a^b K(x, s)\{f_u(u, s)[Au(s)-u(s)]+f(u, s)\}\,ds,$$

where $K(x, s)$ is the Green's function

$$K(x, s)=\begin{cases} \dfrac{s-b}{a-b}(x-a) & \text{for} \quad x\leqslant s \\ \dfrac{s-a}{a-b}(x-b) & \text{for} \quad x\geqslant s \end{cases} \Bigg\} \to |K(x, s)| \leqslant \tfrac{1}{4}(b-a).$$

The condition $(m/4)(b-a)^2<1$ is sufficient for the existence of unique Au.[1]

A maps S into S, for arbitrary $u \in S$ implies

$$\rho(Au, u_{ab}) = \max_{x\in[a,b]} |Au(x)-u_{ab}(x)| \leqslant \frac{m}{4}(b-a)^2[\rho(Au, u)+1] \leqslant$$

$$\leqslant \frac{m}{4}(b-a)^2[\rho(Au, u_{ab})+\rho(u, u_{ab})+1]$$

or

$$\rho(Au, u_{ab}) \leqslant \frac{\frac{m}{4}(b-a)^2(K+1)}{1-\frac{m}{4}(b-a)^2} \leqslant K$$

for $(b-a)$ sufficiently small. This implies $Au \in S$.

Furthermore, for arbitrary u_1 u_2, $\in S$ we have

$$Au_1-Au_2 = \int_a^b K(x, s)\{f_u(u_2, s)[Au_2(s)-u_2(s)]-f_u(u_1, s)[Au_1(s)-u_1(s)]$$
$$-[f(u_1, s)-f(u_2, s)]\}\,ds \quad \text{or,}$$
$$Au_1-Au_2 = \int_a^b K(x, s)\{f_u(u_1, s)[Au_2(s)-Au_1(s)]+f_u(u_2, s)[u_1(s)-u_2(s)]$$
$$+[f_u(u_1, s)-f_u(u_2, s)][u_1(s)-Au_2(s)]-[f(u_1, s)-f(u_2, s)]\}\,ds$$

from which,

$$\rho(Au_1, Au_2) \leqslant \frac{m}{4}(b-a)^2[\rho(Au_1, Au_2)+\rho(u_1, u_2)+\rho(u_1, u_2)\rho(u_1, Au_2)$$
$$+\rho(u_1, u_2)] \leqslant \frac{m}{4}(b-a)^2[\rho(Au_1, Au_2)+2(K+1)\rho(u_1, u_2)]$$

or

$$\rho(Au_1, Au_2) \leqslant \frac{\frac{m}{2}(b-a)^2(K+1)}{1-\frac{m}{4}(b-a)^2}\rho(u_1, u_2).$$

For sufficiently small $(b-a)$, $(m/2)(b-a)^2(K+1)/[1-(m/4)(b-a)^2]=\alpha<1$, i.e., A is a contraction mapping.

It follows from the Contraction Mapping Principle[1] that the operator equation $Au=u$ has a unique solution in S, and that the sequence $\{u_n(x)\}$ converges uniformly to it; where $u_{n+1}=Au_n$, $n=0, 1, 2, \ldots$, and $u_0(x)$ is arbitrary in S. Since the sequence $\{u_n(x)\}$ is equivalent to the sequence of solutions to Eqs. (2),[2] and the operator equation $Au=u$ is equivalent to Eq. (1), part 1 of the theorem is proved. Toward quadratic convergence w. f. have for $u(x)=\lim_{n\to\infty} u_n(x)$,

$$|u_{n+1}(x)-u(x)| = \Big|\int_a^b K(x, s\ \{f_u(u_n, s)[u_{n+1}(s)-u_n(s)]$$
$$+[f(u_n, s)-f(u, s)]\}\,ds\Big|$$

or, by the mean value theorem,

$$|u_{n+1}(x)-u(x)|=\Big|\int_a^b K(x,s)\{f_u(u_n,s)[u_{n+1}(s)-u_n(s)]$$
$$+f_u(\bar{u},s)[u_n(s)-u(s)]\}\,ds\Big|$$

where $\bar{u}(s)$ is such that $\rho(\bar{u},u)\leqslant\rho(u_n,u)$. Equivalently,

$$|u_{n+1}(x)-u(x)|=\Big|\int_a^b K(x,s)\{[f_u(u_n,s)-f_u(\bar{u},s)][u_{n+1}(s)-u_n(s)]$$
$$+f_u(\bar{u},s)[u_{n+1}(s)-u(s)]\}\,ds\Big|.$$

But this implies

$$\rho(u_{n+1},u)\leqslant\frac{m}{4}(b-a)^2[\rho(u_n,\bar{u})\rho(u_{n+1},u_n)+\rho(u_{n+1},u)]$$
$$\leqslant\frac{m}{4}(b-a)^2[\rho(u_n,u)\rho(u_{n+1},u_n)+\rho(u_{n+1},u)],$$

and since $\rho(u_{n+1},u_n)\leqslant\rho(u_{n+1},u)+\rho(u_n,u)<2\rho(u_n,u)$, where the latter inequality follows from the contraction property, we have

$$\rho(u_{n+1},u)\leqslant\frac{m}{4}(b-a)^2[2\rho^2(u_n,u)+\rho(u_{n+1},u)]$$

or equivalently,

$$\rho(u_{n+1},u)\leqslant\frac{\frac{m}{2}(b-a)^2}{1-\frac{m}{4}(b-a)^2}\rho^2(u_n,u)=\frac{\alpha}{K+1}\rho^2(u_n,u)$$

which establishes part 2 of the Theorem.

For part 3 we observe that the contraction mapping theorem implies $\rho(u_n,u)\leqslant\frac{1}{1-\alpha}\rho(u_{n+1},u_n)$. This inequality together with the inequality just proved imply

$$\rho(u_{n+1},u)\leqslant\frac{\alpha}{(K+1)(1-\alpha)^2}\rho^2(u_{n+1},u_n)$$

which completes the proof.

For the purpose of generalizing the above result to systems of equations we introduce the following vector notation:

$U''=F(U,x)$; $U(a)=U_a$, $U(b)=U_b$, for the system,

$$\begin{array}{lll} u^{(1)''}=f^{(1)}(u^{(1)},u^{(2)},\ldots,u^{(N)},x) & u^{(1)}(a)=u_a^{(1)}, & u^{(1)}(b)=u_b^{(1)} \\ \vdots & \vdots & \vdots \\ u^{(N)''}=f^{(N)}(u^{(1)},u^{(2)},\ldots,u^{(N)},x) & u^{(N)}(a)=u_a^{(N)}, & u^{(N)}(b)=u_b^{(N)} \end{array}$$

and

$$J(U,x)=\begin{bmatrix} f^{(1)}_{u^{(1)}}(u^{(1)},\ldots,u^{(N)},x),\ldots,f^{(1)}_{u^{(N)}}(u^{(1)},\ldots,u^{(N)},x) \\ \vdots \\ f^{(N)}_{u^{(1)}}(u^{(1)},\ldots,u^{(N)},x),\ldots,f^{(N)}_{u^{(N)}}(u^{(1)},\ldots,u^{(N)},x) \end{bmatrix}.$$

The $f^{(i)}$ are defined on the $N+1$ dimensional closed domain D given by

$$|u^{(i)}-u^{(i)}_{ab}|\leqslant K \quad x\varepsilon[a,b] \quad i=1,\ldots,N$$

and

$$u^{(i)}_{ab}(x)=\frac{1}{b-a}[(u^{(i)}_b-u^{(i)}_a)x+bu^{(i)}_a-au^{(i)}_b].$$

We shall now state and prove the main theorem of this paper.

THEOREM II. Consider the following system of non-linear differential equations and two-point boundary conditions:

$$d^2U/dx^2=F(U,x); \quad U(a)=U_a, \quad U(b)=U_b \tag{3}$$

where the $f^{(i)}(u^{(1)},\ldots,u^{(N)},x)$, $i=1,\ldots,N$, have the following properties on D: 1) the $f^{(i)}(u^{(1)},\ldots,u^{(N)},x)$ are continuous; 2) $f^{(i)}_{u^{(j)}}(u^{(1)},\ldots,u^{(N)},x)$, $j=1,\ldots,N$, exist and are continuous; and 3) the $f^{(i)}_{u^{(j)}}(u^{(1)},\ldots,u^{(N)},x)$ are Lipschitzian with respect to U, i.e., there exists N^2 numbers such that

$$|f^{(i)}_{u^{(j)}}(u_1^{(1)},\ldots,u_1^{(N)},x)-f^{(i)}_{u^{(j)}}(u_2^{(1)},\ldots,u_2^{(N)},x)|\leqslant M_{ij}\sum_{i=1}^{N}|u_1^{(i)}-u_2^{(i)}|$$

for an arbitrary pair of points (U_1,x), (U_2,x) in D. Define the following sequence of systems of linear differential equations:

$$\frac{d^2U_{n+1}}{dx^2}=J(U_n,x)[U_{n+1}(x)-U_n(x)]+F(U_n,x) \tag{4}$$

$$U_n(a)=U_a, \quad U_n(b)=U_b, \quad n=0,1,\ldots$$

and $U_0(x)$ is such that $u_0^{(i)}(x)$, $i=1, \ldots, N$, are continuous on $[a, b]$ and $\max_x |u_0^{(i)}(x)-u_{ab}^{(i)}(x)| \leqslant K$, $i=1, \ldots, N$.

Then for a sufficiently small interval $[a, b]$ the unique solution to system (3) exists, and 1) the sequence $\{U_n(x)\}$ converges strongly to it, 2) the convergence is at least quadratic, and 3) a bound on the error is given by

$$\rho(U_{n+1}, U) \leqslant \frac{\sigma}{(1-\sigma)^2 (NK+1)} \rho^2(U_{n+1}, U_n)$$

where the metric ρ and the number σ are defined below and $U(x)$ is the solution to system (3).

Proof. It follows from 1, 2, and 3 that there exist numbers P_i, R_{ij}, Q_i such that $|f^{(i)}(u^{(i)}, \ldots, u^{(N)}, x)| \leqslant P_i$,

$$\left|f_{u^{(j)}}^{(i)}(u^{(1)}, \ldots, u^{(N)}, x)\right| \leqslant R_{ij}$$

and

$$\left|f^{(i)}(u_1^{(1)}, \ldots, u_1^{(N)}, x) - f^{(i)}(u_2^{(1)}, \ldots, u_2^{(N)}, x)\right| \leqslant Q_i \sum_{i=1}^{N} \left|u_1^{(i)} - u_2^{(i)}\right|.$$

Let $m = \max_{\substack{i=1,\ldots,N \\ j=1,\ldots,N}} \{M_{ij}, R_{ij}, P_i, Q_i\}$ and define the complete metric space S:

$$S = \{U(x) \mid u^{(i)}(x) \text{ continuous on } [a, b];\ u^{(i)}(a) = u_a^{(i)};\ u^{(i)}(b) = u_b^{(i)};$$
$$\max_x |u^{(i)}(x) - u_{ab}^{(i)}(x)| \leqslant K, \quad i=1, \ldots, N\}$$

with the distance function $\rho(U_1, U_2)$ given by

$$\rho(U_1, U_2) = \sum_{i=1}^{N} \max_x \left|u_1^{(i)}(x) - u_2^{(i)}(x)\right|.$$

Define the operator A on S:

$$AU = U_{ab}(x) - \int_a^b K(x, s) \{J(U, s)[AU(s) - U(s)] + F(U, s)\} ds$$

where

$$K(x, s) = \begin{cases} \dfrac{s-b}{a-b}(x-a) & \text{for} \quad x \leqslant s \\ \dfrac{s-a}{a-b}(x-b) & \text{for} \quad x \geqslant s \end{cases} \Bigg\} \to |K(x, s)| \leqslant \tfrac{1}{4}(b-a).$$

The condition $N/(4)\, m(b-a)^2 < 1$ is sufficient for the existence of unique AU. We will now show that A is an into mapping on S. Arbitrary $U \in S$ implies

$$\rho(AU, U_{ab}) = \sum_{i=1}^{N} \max_x \left|Au^{(i)} - u_{ab}^{(i)}\right| \leqslant N \frac{m}{4} (b-a)^2 [\rho(AU, U) + 1]$$

$$\leqslant N \frac{m}{4} (b-a)^2 [\rho(AU, U_{ab}) + \rho(U, U_{ab}) + 1]$$

$$\leqslant N \frac{m}{4} (b-a)^2 [\rho(AU, U_{ab}) + (NK+1)]$$

or

$$\rho(AU, U_{ab}) \leqslant \frac{N \frac{m}{4} (b-a)^2 (NK+1)}{1 - N \frac{m}{4} (b-a)^2} \leqslant K$$

for $(b-a)$ sufficiently small, i.e., $AU \in S$.

Furthermore, for arbitrary U_1 and U_2 in S, we have

$$AU_1 - AU_2 = \int_a^b K(x, s) \{J(U_2, s)[AU_2 - U_2] - J(U_1, s)[AU_1 - U_1]$$
$$+ F(U_2, s) - F(U_1, s)\} ds$$
$$= \int_a^b K(x, s) \{J(U_1, s)[AU_2 - AU_1] + J(U_2, s)[U_1 - U_2]$$
$$+ [J(U_1, s) - J(U_2, s)][U_1 - AU_2]$$
$$+ F(U_2, s) - F(U_1, s)\} ds.$$

This implies

$$\rho(AU_1, AU_2) = \sum_{i=1}^{N} \max_x \left|Au_1^{(i)} - Au_2^{(i)}\right| \leqslant N \frac{m}{4} (b-a)^2 [\rho(Au_2, AU_1)$$
$$+ \rho(U_1, U_2) + \rho(U_1, U_2)\, \rho(U_1, AU_2) + \rho(U_2, U_1)]$$
$$\leqslant N \frac{m}{4} (b-a)^2 [\rho(AU_1, AU_2) + \rho(U_1, U_2) \{\rho(U_1, U_{ab})$$
$$+ \rho(AU_2, U_{ab}) + 2\}]$$
$$\leqslant N \frac{m}{4} (b-a)^2 [\rho(AU_1, AU_2) + 2(NK+1)\, \rho(U_1, U_2)]$$

or

$$\rho(AU_1, AU_2) \leqslant \frac{N \frac{m}{2} (b-a)^2 (NK+1)}{1 - N \frac{m}{4} (b-a)^2} \rho(U_1, U_2).$$

For sufficiently small $(b-a)$, $[N(m/2)\ (b-a)^2(NK+1)]/1-N(m/4)(b-a)^2 = \sigma < 1$, i.e., A is a contraction mapping of S into S.

The contraction Mapping Principle implies that the operator equation $AU=U$ has a unique solution in S, and that the sequence $\{U_n(x)\}$ converges strongly to it; where $U_{n+1}=AU_n$, $n=0, 1, 2, \ldots$, and U_0 is arbitrary in S. Since the sequence $\{U_n(x)\}$ is equivalent to the sequence of solutions to system (4), and the operator equation $AU=U$ is equivalent to system (3), part 1 of the Theorem is proved.

Toward quadratic convergence we have for $U(x)=\lim_{n\to\infty} U_n(x)$,

$$|U_{n+1}(x)-U(x)|=\Big|\int_a^b K(x,s)\,\{J(U_n,s)\,[U_{n+1}(s)-U_n(s)] + [F(U_n,s)-F(U,s)]\}ds\Big|.$$

We replace the last term in the curled bracket with terms of which the following is typical:

$$f^{(J)}(u_n^{(1)},\ldots,u_n^{(N)},s)-f^{(J)}(u^{(1)},\ldots,u^{(N)},s)=f^{(J)}_{u^{(1)}}({}^{(J)}\underline{u}^{(1)},\ldots,{}^{(J)}\underline{u}^{(N)},s)\times$$
$$\times[u_n^{(1)}-u^{(1)}]\ldots+f^{(J)}_{u^{(N)}}({}^{(J)}\underline{u}^{(1)},\ldots,{}^{(J)}\underline{u}^{(N)},s)[u_n^{(N)}-u^{(N)}] \qquad j=1,2,\ldots,N$$

This equality follows from the mean value theorem for functions of several variables, where for each $s\in[a,b]$ ${}^{(J)}\underline{U}$ is a vector on the line segment joining U to U_n. Adding and subtracting appropriate terms, as before, we obtain

$$\rho(U_{n+1},U)\leqslant\frac{m}{4}(b-a)^2[\rho(U_{n+1},U_n)\{\rho(U_n,{}^{(1)}\underline{U})+\ldots+\rho(U_n,{}^{(N)}\underline{U})\} + N\rho(U_{n+1},U)]$$

where $\rho(U_n,{}^{(J)}\underline{U})\leqslant\rho(U_n,U)$, $j=1,2,\ldots,N)$, from which, $\rho(U_{n+1},U)\leqslant N(m/4)(b-a)^2[\rho(U_{n+1},U_n)\ \rho(U_n,U)+\rho(U_{n+1},U)]$. Since $\rho(U_{n+1},U_n)\leqslant\rho(U_{n+1},U)+\rho(U_n,U)<2\rho(U_n,U)$, we have $\rho(U_{n+1},U)\leqslant N(m/4)(b-a)^2\times\times[2\rho^2(U_n,U)+\rho(U_{n+1},U)]$, or $\rho(U_{n+1},U)\leqslant[N(m/2)(b-a)^2/1-N(m/4)\times\times(b-a)^2]\ \rho^2(U_n,U)$, which establishes the quadratic convergence.

Part 3 of the theorem is established by the fact that the contraction mapping theorem implies $\rho(U_n,U)\leqslant[1/(1-\sigma)]\ \rho(U_{n+1},U_n)$. This inequality together with the inequality just proved imply

$$\rho(U_{n+1},U)\leqslant\frac{\sigma}{(1-\sigma)^2(NK+1)}\rho^2(U_{n+1},U_n)$$

which completes the Proof.

Numerical Application

From the point of view of numerical analysis the significance of the foregoing theorem lies in the fact that the non-linear system is relatively intractable whereas the linear system, owing to the principle of superposition, may be reduced to the systematic numerical integration of initial value problems. One approach is as follows:

At the $n+1^{\text{st}}$ stage of the iteration we have the linear system,

$$U''_{n+1}(x)=J(U_n,x)[U_{n+1}(x)-U_n(x)]+F(U_n,x)$$
$$U=(u_1,u_2,\ldots,u_N);\quad U_{n+1}(a)=U_n(a)=U_a;\quad U_{n+1}(b)=U_n(b)=U_b.$$

This is equivalent to the following linear system of order $m=2N$:

$$\dot{X}(t)=A(t)X(t)+B(t)$$
$$X=(x_1,x_2,\ldots,x_m),\quad t\in[a,b]$$
$$x_1(a)=x_{1a}\qquad x_1(b)=x_{1b}$$
$$x_2(a)=x_{2a}\qquad x_2(b)=x_{2b}$$
$$\vdots$$
$$x_{\frac{m}{2}}(a)=x_{\frac{m}{2}a}\qquad x_{\frac{m}{2}}(b)=x_{\frac{m}{2}b}$$

Generate by numerical integration a set $\{X^{(\frac{m}{2}+i)}\}$, $i=1,2,\ldots,m/2$, of solutions of the homogeneous system $\dot{X}(t)=A(t)X(t)$ with initial conditions

$$X^{(m/2+1)}(a)=(0,0,\ldots,0,x_{\frac{m}{2}+1}=1,0,\ldots,0)$$
$$X^{(m/2+2)}(a)=(0,0,\ldots,0,x_{\frac{m}{2}+2}=1,0,\ldots,0)$$
$$\vdots$$
$$X^{(m)}(a)=(0,0,\ldots,0,\ldots,1).$$

Generate a particular solution $X^{(P)}(t)$ of the non-homogeneous system $\dot{X}(t)=A(t)X(t)+B(t)$ with initial conditions,

$$X^{(P)}(a)=(x_{1a},x_{2a},\ldots,x_{\frac{m}{2}a},K_1,K_2,\ldots,K_{\frac{m}{2}})$$

where the K_i, $i=1,2,\ldots,m/2$, are arbitrary.

The solution $X(t)$ of the non-homogeneous system with the prescribed boundary conditions is then given by

$$X(t)=C_{\frac{m}{2}+1}X^{[m/2+1]}(t)+C_{\frac{m}{2}+2}X^{[m/2+2]}(t)+\ldots+C_m X^{(m)}(t)+X^{(P)}(t)$$

where the $m/2$ constants, $C_{\frac{m}{2}+i}$, $i=1, 2, \ldots, m/2$ are determined by the boundary conditions at $t=b$.

Numerical solutions using essentially the method discussed in this paper have been obtained by KALABA[3] and HESTENES[4] for the case of a scalar valued function. KALABA[3] also gives a different proof of convergence for the single non-linear equation case under more restrictive conditions. HESTENES[4] offers the comment that while in general he feels the solution for the case of a single equation may perhaps be obtained more readily by varying arbitrary initial conditions, this may not be true in higher dimensions. Numerical experiments are in process to clarify this issue.

Remarks

The theorems of this paper have been developed for equations of the form $u''(x)=f(u, x)$. However, it should be pointed out that no further essential difficulty is introduced in the handling of equations containing velocity terms, i.e., equations of the form $u''(x)=f(u, u', x)$, except that the dimension of the relevant domain is increased and the corresponding linear sequence now reads

$$u''_{n+1}(x)=f_u(u_n, u'_n, x)\left[u_{n+1}(x)-u_n(x)\right]+f_{u'}(u_n, u'_n, x)\left[u'_{n+1}(x)-u'_n(x)\right]$$
$$+f(u_n, u'_n, x), \quad n=0, 1, 2, \ldots$$

with appropriate generalizations to systems of equations.

We also note that, from the numerical point of view, the approach described herein admits of a choice between numerically integrating a sequence of systems of linear differential equations, or alternatively, obtaining by suitable numerical processes the solution to equivalent sets of integral equations. Both procedures are currently under study.

Conclusions

In this paper we have presented a convergence proof for a proposed method of obtaining numerical solutions to systems of non-linear differential equations with two-point boundary conditions. It is known that the usual

methods of approaching this problem lead to very considerable difficulties in practice. The technique discussed herein is systematic and rapidly convergent, and owing to the quadratic convergence, great care is not required in obtaining starting functions. It is sufficient that the trial functions be not so bad as to cause divergence.

While there exist a few examples of the application of this method to single equations[3, 4] very little is known concerning application to systems of higher dimension. Since the proposed method is a natural analog of the Newton-Raphson scheme, which has proved its efficacy in a plethora of numerical applications, there is reason to believe that the method suggested herein may be a potentially powerful tool in this more abstract setting.

Appendix

What follows is an application of the method discussed in this paper to an intercept problem of rocket flight. The purpose of this example is the demonstration of the computational effectiveness of the method. The sequence of iterates is seen to converge strongly and rapidly to the solution despite the fact that the starting function is of the simplest possible form, namely, the straight line joining the two points in space.

Numerical Example

The problem solved is that of determining the path that a space vehicle must follow in transferring from a specified position 300 miles above the Earth to another specified position 600 miles above the Earth with a fixed transit time. Since the vehicle is assumed to be in coasting flight, the path functions must satisfy the Newtonian differential equations. The perturbing effect of the Moon is included. A schematic diagram, Fig. 1, of the problem is shown below where $X_0(t)=[x_0(t), y_0(t), z_0(t)]$, the starting function, is shown to be the line joining the initial point $X(0)$ to the final point $X(2)$ and $X^*(t)$ is the solution function.

The unit of length is taken to be the radius of the Earth and the principal gravitational constant is normalized to one. This results in a time unit of 805.46 sec.

The system of three second-order non-linear differential equations and two-point boundary conditions that furnish the mathematical description of the problem are

$$\ddot{x}(t) = -K\frac{x}{r^3} + K_M\left(\frac{x_M - x}{\delta^3} - \frac{x_M}{r_M^3}\right)$$

$$\ddot{y}(t) = -K\frac{y}{r^3} + K_M\left(\frac{y_M - y}{\delta^3} - \frac{y_M}{r_M^3}\right) \qquad t \in [0, 2]$$

$$\ddot{z}(t) = -K\frac{z}{r^3} + K_M\left(\frac{z_M - z}{\delta^3} - \frac{z_M}{r_M^3}\right)$$

$$x(0) = 1.076000 \qquad x(2) = 0$$

$$y(0) = 0 \qquad y(2) = 0.576000$$

$$z(0) = 0 \qquad z(2) = 0.997661$$

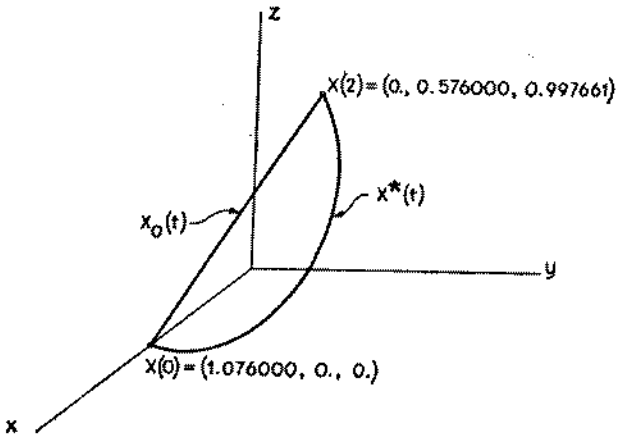

Fig. 1. Schematic diagram

where

$$r = [x^2 + y^2 + z^2]^{\frac{1}{2}}$$

$$r_M = [x_M^2 + y_M^2 + z_M^2]^{\frac{1}{2}}$$

$$\delta = [(x_M - x)^2 + (y_M - y)^2 + z_M - z)^2]^{\frac{1}{2}}.$$

The time interval [0, 2] was subdivided into 100 parts and the necessary numerical integrations carried out by means of a high-speed digital computer (IBM 7094) to an accuracy of seven significant figures. The results are exhibited in the included table where for brevity only six points in time are shown. $X_0(t)$ is the linear starting function; $X_1(t)$ is the first iterate; $X_2(t)$ the second iterate, etc.; and $X^*(t)$ results from the integration of the actual non-linear equations with the initial velocities,

$$\dot{x}(0) = 0.101637$$
$$\dot{y}(0) = 0.472285$$
$$\dot{z}(0) = 0.818022$$

obtained from the final iterate.

The sequence converged, within the accuracy of our computations, in four iterations with

$$\rho(X_1, X_0) = 0.480116$$
$$\rho(X_2, X_1) = 0.133753$$
$$\rho(X_3, X_2) = 0.004375$$
$$\rho(X_4, X_3) = 0.000004$$

where

$$\rho(X_{n+1}, X_n) = \max_t |x_{n+1}(t) - x_n(t)| + \max_t |y_{n+1}(t) - y_n(t)| + \max_t |z_{n+1}(t) - z_n(t)|.$$

As a further check on the over-all accuracy, the perturbing force was set to zero and the final value of the magnitude of the initial velocity was compared with that obtained by the closed form solution for the two-body problem. Within the accuracy of our computations these values were identical.

Table 1

X \ t	0.	0.4	0.8	1.2	1.6	2.0
x_0	1.076000	0.860800	0.645600	0.430400	0.215200	0.
x_1	1.076000	1.015153	0.845061	0.610986	0.323847	0.
x_2	1.076000	1.048799	0.900816	0.657001	0.346085	0.
x_3	1.076000	1.049839	0.902586	0.658550	0.346867	0.
x_4	1.076000	1.049840	0.902587	0.658551	0.346868	0.
x^*	1.076000	1.049840	0.902587	0.658551	0.346868	0.
y_0	0.	0.115200	0.230400	0.345600	0.460800	0.576000
y_1	0.	0.172927	0.324202	0.447591	0.537713	0.576000
y_2	0.	0.184664	0.348339	0.475158	0.553667	0.576000
y_3	0.	0.185100	0.349180	0.476056	0.554172	0.576000
y_4	0.	0.185100	0.349180	0.476057	0.554173	0.576000
y^*	0.	0.185100	0.349180	0.476057	0.554173	0.576000
z_0	0.	0.199532	0.399064	0.598597	0.798129	0.997661
z_1	0.	0.299519	0.561534	0.775250	0.931347	0.997661
z_2	0.	0.319848	0.603341	0.822998	0.958980	0.997661
z_3	0.	0.320602	0.604800	0.824553	0.959854	0.997661
z_4	0.	0.320603	0.604798	0.824555	0.959855	0.997661
z^*	0.	0.320603	0.604798	0.824555	0.959855	0.997661
X \ t	0.	0.4	0.8	1.2	1.6	2.0

References

1 KOLMOGOROV, A. N. and FOMIN, S. V., Elements of the Theory of Functions and Functional Analysis, Vol. 1, pp. 43 ff. (Graylock Press, Rochester, N. Y. 1957).

2 COURANT, R. and HILBERT, D., Methoden der Mathematischen Physik (Julius Springer Verlag, Berlin, 1931) Vol. I, pp. 302 ff.

3 KALABA, R., On non-linear differential equations, the maximum operation, and monotone convergence, *J. Math. and Mech.* 8, 519–574 (1959).

4 HESTENES, M. R., Numerical methods of obtaining solutions of fixed end point problems in the calculus of variations, RM-102, Rand Corp. (August 1949).

IV. 13

MOTION OF AN ARTIFICIAL SATELLITE IN THE EARTH'S GRAVITATIONAL FIELD

by

E. P. AKSENOV

Sternberg Astronomic Institute, Moscow (U.S.S.R.)

Abstract

The question of the analytical theory of the motion of artificial satellites is considered. In developing this theory the generalized problem of two fixed centers is used. When choosing, in a suitable manner, the masses of two fixed centers and the reciprocal distance, the force function of this problem represents the Earth's potential quite accurately, while the solution is reduced to quadratures. Therefore the orbits resulting here are interesting as the possible intermediate orbits when developing the theory of satellite motion. These orbits are subjected to all perturbations known to be principal in the satellites problem and this fact does determine their main advantage relative to the Kepler's orbit.

In the paper these intermediate orbits are considered formerly. There are introduced the generalized Kepler's elements and then the qualitative picture of satellite motion is described. Also the formulae for calculation of spatial positions of the satellite are given. Some particular cases are considered: the motion along the surface of an ellipsoid, circular and elliptical orbits. An example of the real motion of a satellite is considered. Then the problem of calculation of other perturbations is discussed. For this purpose the equations for the osculated elements are derived. The perturbations due to the zonal harmonics of degrees 3, 4, ..., 8, and to one tesseral harmonic of degrees 2 in the series of the terrestial gravitational field, are obtained.

1. Introduction

A considerable number of works on the analytic theory of the movement of artificial satellites have been published lately. But in spite of this there is currently no satisfactory and exact analytic theory of satellite motion. This may be, first of all, because of pure technical difficulties caused by the calculation of a great number of forces affecting the satellite,