JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 27, 417–433 (1969)

# On Shooting Methods for Boundary Value Problems

M. R. OSBORNE

*Computer Centre, Australian National University*

*Submitted by Richard Bellman*

## 1. INTRODUCTION

Shooting methods, in which the numerical solution of a boundary value problem is found by integrating an appropriate initial value problem, have been the subject of a number of recent papers (for example, Roberts and Shipman [1]-[3]), and a book (Keller [4]). The attraction of these methods lies in the availability on most computers of reasonably adequate subroutines for the numerical solution of initial value problems [5], and the lack of an alternative systematic approach to the problem.

Let the system of $n$ differential equations be

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) \tag{1.1}$$

subject to the boundary conditions

$$B_1\mathbf{x}(a) + B_2\mathbf{x}(b) = \mathbf{c}, \tag{1.2}$$

where $B_1$ and $B_2$ are $n \times n$ matrices. Provided $\mathbf{f}$ satisfies the usual conditions, the problem expressed by Eqs. (1.1) and (1.2) has a solution provided $| b - a |$ is small enough and $B_1 + B_2$ is nonsingular, and the limiting form of this solution as $b \to a$ is just the solution to the initial value problem with $\mathbf{x}(a) = (B_1 + B_2)^{-1} \mathbf{c}$. In this case we say the conditions (1.2) are compatible. This is the case that is considered in detail in [4], and it appears to be the only one in which the existence of a solution can be guaranteed except for certain special equations of higher order. In other cases it is necessary to assume that the solution to the system (1.1), (1.2) is in some sense well determined, and this assumption is basic to our approach.

The principal difficulties in the use of shooting methods are caused

(i) by instability of the initial value problem for the system of differential equations (and this is not to be confused with instability of the numerical method), and

(ii) by the requirement for good starting values in the most common methods used for the iterative solution of nonlinear problems.

It is hoped that this paper makes contributions to both these problems.

The plan of the paper is as follows. In the next section a formal solution is given in the linear case for the problem of finding a vector of initial values such that the solution to the initial value problem for Eq. (1.1) satisfies also the boundary conditions (1.2). In Section 3 the problem of instability is considered for linear systems, and it is shown that the so called multiple shooting methods can be used for solving problems which are well determined in a certain sense and for which conventional shooting methods are unsuitable. This result complements the treatment given in [4], and an example given by Holt [6] is solved by this method. In Section 4 the reduction to an initial value problem is carried out for a nonlinear system, and it is shown that the problem reduces to that of solving a system of nonlinear equations. In practice these equations are usually solved by Newton's method, and this is described in Section 5. Here the correction at each stage is found by integrating a system of linear equations so that the considerations of Section 3 apply. The main problem is to achieve convergence from starting values which do not require too much care in their selection, and the suggested approach is illustrated by several numerical examples considered difficult by other authors.

All computations reported were carried out using single precision arithmetic on an IBM system 360/50 computer. The programs were written in FORTRAN, and the RKGS subroutine (see [5]) was used to integrate the differential equations. This subroutine uses an absolute accuracy criterion for adjusting steplength, and this was set to $10^{-4}$ for all calculations.

## 2. Reduction to an Initial Value Problem in the Linear Case

Consider the linear system of differential equations

$$\frac{d\mathbf{x}}{dt} + A(t)\,\mathbf{x} = \mathbf{r}(t). \tag{2.1}$$

The fundamental matrix of this system satisfies

$$\frac{dX}{dt} + A(t)\,X = 0 \tag{2.2}$$

subject to the initial condition

$$X(a) = E, \tag{2.3}$$

where the square matrix $E$ is assumed to have its full rank. Let $\mathbf{v}(t)$ be any solution to Eq. (2.1), then the general solution can be given in the form (particular integral plus complementary function)

$$\mathbf{x}(t) = \mathbf{v}(t) + X(t)\,\mathbf{d}, \tag{2.4}$$

where $\mathbf{d}$ is an arbitrary vector. If Eq. (1.2) is to hold, then $\mathbf{d}$ must satisfy

$$\{B_1 E + B_2 X(b)\}\,\mathbf{d} = \mathbf{c} - B_1 \mathbf{v}(a) - B_2 \mathbf{v}(b). \tag{2.5}$$

NOTE. If $E$ is chosen to be the unit matrix then $\mathbf{d}$ is just the correction that must be added to $\mathbf{v}(a)$ to give the correct initial value of $\mathbf{x}(t)$. In this case we speak of a *shooting method.*

An alternative reduction can be based on the use of the adjoint equation (Goodman and Lance [7])

$$\frac{dY}{dt} - YA(t) = 0. \tag{2.6}$$

If Eq. (2.1) is premultiplied by $Y$, and Eq. (2.6) postmultiplied by $\mathbf{x}$, then we obtain by adding the two equations

$$\frac{d}{dt}(Y\mathbf{x}) = Y\mathbf{r}(t) \tag{2.7}$$

whence

$$Y(b)\,\mathbf{x}(b) - Y(a)\,\mathbf{x}(a) = \int_a^b Y\mathbf{r}(t)\,dt; \tag{2.8}$$

setting

$$Y(b) = B_2 \tag{2.9}$$

we have

$$B_2\mathbf{x}(b) = Y(a)\,\mathbf{x}(a) + \int_a^b Y\mathbf{r}(t)\,dt$$

whence

$$\{B_1 + Y(a)\}\,\mathbf{x}(a) = \mathbf{c} - \int_a^b Y\mathbf{r}(t)\,dt. \tag{2.10}$$

REMARK. That the two methods are closely equivalent is readily seen by considering the relation

$$\frac{d}{dt}(YX) = 0. \tag{2.11}$$

However there can be substantial differences in the work required to extract numerical values.

To see this note that by the rank condition on $E$ no column can vanish so the determination of $X$ requires $n$ forward integrations. However, if any row of $B_2$ is identically zero then so is the corresponding row of $Y$, and the corresponding integration need not be carried out. Of course, the condition

$$Y(a) = B_1 \tag{2.12}$$

could have been used in integrating Eq. (2.6), in which case an equation for $\mathbf{x}(b)$ results. The appropriate choice is that for which the work required to integrate equation (2.6) is least.

The methods outlined above extend readily to the multipoint boundary condition

$$\sum_{i=1}^{p} B_i \mathbf{x}(t_i) = \mathbf{c}. \tag{2.13}$$

If the representation (2.4) is used for the solution then

$$\left\{\sum_{i=1}^{p} B_i X(t_i)\right\} \mathbf{d} = \mathbf{c} - \sum_{i=1}^{p} B_i \mathbf{v}(t_i). \tag{2.14}$$

The adjoint equation can also be used. Let $Y_i(t)$ satisfy equation (2.6) and the initial condition

$$Y_i(t_i) = B_i \tag{2.15}$$

then

$$B_i \mathbf{x}(t_i) = Y_i(a)\,\mathbf{x}(a) + \int_a^{t_i} Y_i \mathbf{r}(t)\,dt, \tag{2.16}$$

giving

$$\left\{\sum_{i=1}^{p} Y_i(a)\right\} \mathbf{x}(a) = \mathbf{c} - \sum_{i=1}^{p} \int_a^{t_i} Y_i \mathbf{r}(t)\,dt. \tag{2.17}$$

Yet another class of problem is the two-point boundary value problem supplemented by the interface conditions

$$\mathbf{x}^-(t_i) = \mathbf{x}^+(t_i) + \boldsymbol{\xi}_i, \qquad i = 1, 2,\ldots q, \tag{2.18}$$

where $a < t_1 < t_2 \cdots < t_q < b$. Let $\mathbf{v}_i(t_{i-1}) = 0$ and $X_i(t_{i-1}) = I$, then in the interval $t_{i-1} \leqslant t \leqslant t_i$ we can write (we define $a = t_0$, $b = t_{q+1}$)

$$\mathbf{x}(t) = \mathbf{v}_i(t) + X_i(t)\,\mathbf{d}_i\,. \tag{2.19}$$

Substituting this expression into the conditions (1.2) and (2.18) leads to the system of equations for the $\mathbf{d}_i$, $i = 1, 2, \ldots q+1$

$$\begin{bmatrix} B_1 & & & B_2 X_{q+1}(b) \\ X_1(t_1) - I & & & \\ & X_2(t_2) - I & & \\ \cdots & \cdots & \cdots & \cdots \\ & & & X_q(t_q) - I \end{bmatrix} \begin{bmatrix} \mathbf{d}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{d}_{q+1} \end{bmatrix} = \begin{bmatrix} \mathbf{c} - B_2 \mathbf{v}_{q+1}(b) \\ \boldsymbol{\xi}_1 - \mathbf{v}_1(t_1) \\ \vdots \\ \boldsymbol{\xi}_q - \mathbf{v}_q(t_q) \end{bmatrix}. \tag{2.20}$$

NOTE. (i) Although Eq. (2.20) can be recast using the adjoint differential equation, this appears to be an example where this is of little use.

(ii) When $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \cdots = \boldsymbol{\xi}_q = 0$, then this problem reduces to the two-point boundary value problem already considered. In this case Eq. (2.20) provides a statement of the problem equivalent to that of Eq. (2.5), and we refer to this as a *multiple shooting formulation.*

(iii) With the specialization of $\mathbf{v}_i(t_{i-1})$ and $X_i(t_{i-1})$ given above, we have $\mathbf{d}_i = \mathbf{x}(t_{i-1})$ so that Eq. (2.19) can be written

$$\mathbf{x}(t_i) = \mathbf{v}_i(t_i) + X_i(t_i)\,\mathbf{x}(t_{i-1}) \tag{2.21}$$

(assuming $\boldsymbol{\xi}_i = 0$, $i = 1, 2, \ldots, q$). This equation can be interpreted as an accurate finite difference formulation of Eq. (2.1) (see also [4], example (3.3.6)). This observation would appear to provide an excellent starting point for a comparison of finite difference and shooting methods.

## 3. THE STABILITY PROBLEM

The principal difficulty in computing numerical solutions to initial value problems is instability of the differential equation. We say a solution $\mathbf{x}(t)$ of the differential equation is unstable $(\epsilon, K)$ if there exists another solution $\mathbf{v}(t)$ such that

$$\text{(i)} \qquad \|\mathbf{x}(a) - \mathbf{v}(a)\| \leqslant \epsilon \tag{3.1}$$

and

$$\text{(ii)} \qquad \|\mathbf{x}(t^*) - \mathbf{v}(t^*)\| \geqslant K \tag{3.2}$$

for at least one $t^*$, $a < t^* \leqslant b$. (We assume the use of the maximum norm and the corresponding subordinate matrix norm unless otherwise stated.) In practical applications $\epsilon$ would be of the order of the rounding error, say $10^{-7}(\max(1, \|\mathbf{x}(a)\|))$ on an IBM system 360, while $K$ would be of the order of $\max_{a \leqslant t \leqslant b} \|\mathbf{x}(t)\|$. In this case the smallest value of $t^*$ such that the inequality (3.2) is true would be an upper bound to the interval on which any information could be obtained about $\mathbf{x}(t)$ by forward integration using a stable numerical procedure.

EXAMPLE (3.1). Consider the system (with $k > 0$)

$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ k^4 & 0 & 0 & 0 \end{bmatrix} \mathbf{x} \tag{3.1}$$

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{3.2}$$

The fundamental matrix is

$$X(t) = .5 \begin{bmatrix} \alpha + \gamma & (\beta + \delta)/k & (\alpha - \gamma)\,k & (\beta - \delta)/k^3 \\ k(\beta - \delta) & \alpha + \gamma & (\beta + \delta)/k & (\alpha - \gamma)\,k^2 \\ k^2(\alpha - \gamma) & k(\beta - \delta) & \alpha + \gamma & (\beta + \delta)/k \\ k^3(\beta + \delta) & k^2(\alpha - \gamma) & k(\beta - \delta) & \alpha + \gamma \end{bmatrix}, \tag{3.3}$$

where $\alpha = \cosh(kt)$, $\beta = \sinh(kt)$, $\gamma = \cos(kt)$, and $\delta = \sin(kt)$, and it may be noted that

$$e^{-kt}X(t) \to .25 \begin{bmatrix} 1 \\ k \\ k^2 \\ k^3 \end{bmatrix} [1\; k^{-1} k^{-2} k^{-3}] \tag{3.4}$$

as $kt \to \infty$.

To illustrate the perturbation of an exact integration scheme by round off error the results of computing $\mathbf{x}(t)$ numerically by means of the scheme

$$\mathbf{x}(a + p\Delta t) = X(\Delta t)\,\mathbf{x}(a + (p - 1)\,\Delta t) \tag{3.5}$$

for initial vectors chosen so that the first component of $\mathbf{x}(t)$ should be, respectively $e^{kt}$, $e^{-kt}$, $\cos(kt)$, and $\sin(kt)$ are given in Table 3.1. The integrations were carried out with $\Delta t = .1$ for $k = .2$ and 2, and the solutions are tabulated at $p = 0, 10, 20, \ldots, 100$. It will be seen that the calculation of the increasing solutions is perfectly satisfactory, but that the other calculations can lead to erroneous results.

TABLE 3.1

RESULTS FOR EXAMPLE 3.1

| $t$ | (I) | (II) | (III) | (IV) |
|---|---|---|---|---|
| | | $k = .2$ | | |
| 0 | .1000 E1 | .1000 E1 | .1000 E1 | 0.0 |
| 1 | .1221 E1 | .8187 E0 | .9801 E0 | .1987 E0 |
| 2 | .1492 E1 | .6703 E0 | .9211 E0 | .3894 E0 |
| 3 | .1822 E1 | .5488 E0 | .8253 E0 | .5646 E0 |
| 4 | .2225 E1 | .4493 E0 | .6967 E0 | .7173 E0 |
| 5 | .2718 E1 | .3679 E0 | .5403 E0 | .8415 E0 |
| 6 | .3320 E1 | .3012 E0 | .3624 E0 | .9320 E0 |
| 7 | .4055 E1 | .2466 E0 | .1700 E0 | .9854 E0 |
| 8 | .4953 E1 | .2019 E0 | −.2918 E−1 | .9995 E0 |
| 9 | .6049 E1 | .1653 E0 | −.2272 E0 | .9738 E0 |
| 10 | .7388 E1 | .1354 E0 | −.4161 E0 | .9093 E0 |
| | | $k = 2$ | | |
| 0 | .1000 E1 | .1000 E1 | .100 E1 | 0.0 |
| 1 | .7389 E1 | .1353 E0 | −.4161 E0 | .9093 E0 |
| 2 | .5460 E2 | .1834 E−1 | −.6536 E0 | −.7568 E0 |
| 3 | .4034 E3 | .2655 E−2 | .9604 E0 | −.2795 E0 |
| 4 | .2981 E4 | .1647 E−2 | −.1434 E0 | .9884 E0 |
| 5 | .2202 E5 | .9722 E−2 | −.8236 E0 | −.5509 E0 |
| 6 | .1627 E6 | .7151 E−1 | .9581 E0 | −.5876 E0 |
| 7 | .1202 E7 | .5284 E0 | .9810 E0 | .6136 E0 |
| 8 | .8885 E7 | .3904 E1 | .5281 E1 | −.3073 E1 |
| 9 | .6565 E8 | .2885 E2 | .4676 E2 | −.2133 E2 |
| 10 | .4851 E9 | .2131 E3 | .3410 E3 | −.1511 E3 |

As each of the columns of $X(t)$ correspond to increasing solutions of Eq. (3.1), the above remarks indicate that for large $kt$ (say of order 15 or more on an IBM system 360) the computed approximation (say $\bar{X}(t)$) will agree with $X(t)$ given by Eq. (3.4) to high relative accuracy. By Eq. (2.5), the equations determining the initial vector have matrix $B_1 + B_2\bar{X}(b)$, and this not only has large elements, but it is nearly singular to working accuracy.

The numerical difficulties described above are extreme, and unfortunately the situations in which they occur are not uncommon. The problem here is that certain solutions to the differential equation are increasing so rapidly in magnitude that they cause all information concerning the other solutions to be lost in calculations carried out using finite precision. This is a somewhat different aspect of the problem of instability to that illustrated in table 3.1. Of course information can be lost by solutions decreasing in magnitude relative to others, and this would be more difficult to determine. However the relation

$$\det (X(t)) = \det (X(a)) \exp \left( \int_a^t \operatorname{trace} (A)\, dt \right) \tag{3.6}$$

ensures that rapid decrease must to some extent be balanced by rapid increase, and provides a justification for using the size of $\| X(t) \|$ to indicate instability problems.

When a stable numerical integration scheme is used, then the error can, in general, be bounded by an expression of the form

$$E(t) \leqslant K_1(\Delta t)^s \, e^{K_2 |t-a|}, \tag{3.7}$$

where $K_1$, $K_2$, $s$ are positive constants (for linear multistep methods see for example Henrici [8]). The stability of the method is indicated by the exponential term which determines the error growth depending on $t$ but not on $\Delta t$ so that it is independent of the number of steps required to reach $t$. Equation (3.7) can be interpreted in two ways: (i) that arbitrarily high accuracy can be obtained by working with small enough $\Delta t$ (requiring arithmetic facilities of arbitrarily high precision), and (ii) that reasonable accuracy can be obtained for moderate values of $\Delta t$ provided $| t - a |$ is small enough. It is stressed that Eq. (3.7) is valid for differential equations unstable in the sense used here.

Although $\| X(b) \|$ is a function of the problem, the magnitude of $\| X_i(t_i) \|$ in the interface formulation is also a function of the number and disposition of the interface points which can clearly be chosen to restrict its growth (as $\| X_i(t_{i-1}) \| = 1$). From Eq. (3.7), a suitable distribution of the interface points will ensure that the elements of $X_i$ can be computed to desired accuracy. Thus, given $\delta > 0$ and $\gamma > 1$, there is a multiple shooting formulation of the boundary value problem such that

$$\text{(i)} \quad \| \bar{M} \| \leqslant \gamma \qquad \text{and} \qquad \text{(ii)} \quad \| M - \bar{M} \| < \delta,$$

where $M$ is the matrix of the equations determining the vector of initial values (Eq. (2.20)), and $\bar{M}$ is the computed approximation to it.

Remark. In general it is not difficult to arrange for conditions (i) and (ii) to be satisfied. Most general purpose subroutines attempt to satisfy (ii), and this only makes sense if some condition similar to (i) is imposed. In the subroutines in [5] this must be done through the use of a special user provided output routine.

If now the solution to this multiple shooting problem is well determined, and if a stable method is used to solve the set of linear equations (2.20) for the vector of initial values, then it can be expected that this computation will give results which are a good approximation of the true solution.

EXAMPLE (3.2). In Holt [6] the differential equation

$$\frac{d^2y}{dt^2} - (1 + t^2)y = 0$$

$$y(0) = 1, \qquad y(L) = 0$$

is considered. According to Holt the solution cannot be obtained for $L > 3.5$ by conventional shooting methods. In our treatment the differential equation was turned into a first-order system, and a multiple shooting formulation was used with 50 equispaced interface points in $0 \leqslant t \leqslant 10.2$ ($\Delta t = .2$). In Table 3.2 values of $x_1(t)$, $x_2(t)$, and $\| X(t + \Delta t) \|$ are tabulated for $t = 0(1)10$ and $X(t) = I$. The solution obtained is in excellent agreement with that given by Holt.

TABLE 3.2

RESULTS FOR EXAMPLE 3.2

| $t$ | $x_1$ | $x_2$ | $\| X \|$ |
|---|---|---|---|
| 0 | .1000 E1 | −.1128 E1 | 1.22 |
| 1 | .2593 E0 | −.4250 E0 | 1.50 |
| 2 | .3455 E−1 | −.8355 E−1 | 2.24 |
| 3 | .1987 E−2 | −.6565 E−2 | 3.50 |
| 4 | .4590 E−4 | −.1945 E−3 | 5.38 |
| 5 | .4188 E−6 | −.2138 E−5 | 8.03 |
| 6 | .1409 E−8 | −.8685 E−8 | 11.62 |
| 7 | .1821 E−11 | −.1300 E−10 | 16.42 |
| 8 | .8825 E−15 | −.7169 E−14 | 22.75 |
| 9 | .1597 E−18 | −.1455 E−17 | 31.05 |
| 10 | .1058 E−22 | −.1106 E−21 | 41.86 |

The method used for solving the linear equations is of particular importance. Here it would seem to be attractive to partition the matrix of Eq. (2.20) and express each of the $\mathbf{d}_i$ in terms of $\mathbf{d}_1$. However this procedure is just the unstable recurrence relation illustrated in example (3.1), and we have already argued that this should be avoided. Note that if this elimination could be carried out exactly Eq. (2.5) results. Thus ordinary shooting methods can be thought of as resulting from the application of a potentially unstable elimination process to Eqs. (2.20). Both complete and partial pivoting strategies have been used successfully to solve the system of linear equations. It is particularly important that partial pivoting is successful as it is not difficult to organize this method to economise on storage. This is an important consideration in using multiple shooting techniques even for quite small values of $n$.

The most usual application of multiple shooting methods has been to particular problems for second-order equations where to achieve stable computations the equation is integrated from each boundary point and matched at a suitable interior point (for example Fox [9]). Morrison, Riley and Zancanaro [10] suggest multiple shooting methods as an elaboration of this device, and in both [4] and [10] the device of reversing the direction of computation is suggested. While it is not difficult to include this here, it has not been used in example (3.2) which is typical of the type of problem to which this device applies; and it is suggested that the interface points be chosen (i) to limit the size of $\| X_i \|$, and (ii) to provide a convenient tabulation of the solution. This second consideration again reflects the close connection between multiple shooting methods and finite difference methods.

## 4. Reduction to an Initial Value Problem, the Nonlinear Case

As the ordinary shooting method is a special case of multiple shooting methods only the latter are considered here.

Let the solution to equation (1.1) satisfying the initial condition

$$\mathbf{x}(t_{i-1}) = \mathbf{d}_i \tag{4.1}$$

be written $\mathbf{x}(t, t_{i-1}, \mathbf{d}_i)$ or, more briefly, $\mathbf{x}_i(t)$, then the condition that Eqs. (1.2) and (2.18) be satisfied gives

$$B_1\mathbf{d}_1 + B_2\mathbf{x}_{q+1}(b)\,\mathbf{c} \tag{4.2}$$

$$\mathbf{x}_i(t_i) = \mathbf{d}_{i+1} + \xi_i\,, \qquad i = 1, 2, \ldots, q \tag{4.3}$$

a set of nonlinear equations for $d_1, d_2, \ldots d_{q+1}$.

In this case instability of the differential equation is a factor in determining the accuracy to which the $\mathbf{x}_i(t_i)$ can be computed. If we assume that $\mathbf{f}$ is twice continuously differentiable as a function of its arguments, and if we set

$$\gamma_i = \max_{t_{i-1} \leqslant t \leqslant t_i} \| Q \|,$$

where

$$Q_{ij} = \frac{\partial fi}{\partial x_j}(t, \mathbf{x}),$$

then, by using the inequality

$$\frac{d}{dt}\|\mathbf{x}-\mathbf{y}\| \leqslant \left\|\frac{d}{dt}(\mathbf{x}-\mathbf{y})\right\| \leqslant \left\|\overline{\frac{df}{dx}}\right\| \|\mathbf{x}-\mathbf{y}\|, \tag{4.4}$$

where the bar indicates that an appropriate mean value is taken and $\mathbf{y}$ satisfies Eq. (1.1), it is easy to show that if

$$\|\mathbf{x}(t_{i-1}) - \mathbf{y}(t_{i-1})\| \leqslant \epsilon \tag{4.5}$$

then

$$\|\mathbf{x}(t_i) - \mathbf{y}(t_i)\| \leqslant \epsilon e^{\gamma_i(t_i - t_{i-1})} + K\epsilon^2(t_i - t_{i-1}), \tag{4.6}$$

where $K$ is a constant. Thus, as in the previous section, a suitable choice of the interface points will ensure that accurate approximations to Eqs. (4.2) and (4.3) can be computed.

## 5. Solution of the Nonlinear Equations

The most frequently used technique for solving the nonlinear Eqs. (4.2) and (4.3) is Newton's method (or a variant in which derivatives are estimated by finite differences). Let the current approximation to the solution of Eqs. (4.2) and (4.3) be denoted by $\mathbf{d}_i^{(j)}$, $i = 1, 2,..., q + 1$, then the corresponding corrections $\mathbf{e}_i^{(j)}$ given by Newton's method satisfy the linear equations

$$B_1\mathbf{e}_1^{(j)} + B_2 J_{q+1}^{(j)}\mathbf{e}_{q+1}^{(j)} = \mathbf{c} - B_1\mathbf{d}_1^{(j)} - B_2\mathbf{x}_{q+1}^{(j)}(b) \tag{5.1}$$

and

$$J_i^{(j)}\mathbf{e}_i^{(j)} - \mathbf{e}_{i+1}^{(j)} = \boldsymbol{\xi}_i + \mathbf{d}_{i+1}^{(j)} - \mathbf{x}_i^{(j)}(t_i), \qquad i = 1, 2,..., q, \tag{5.2}$$

where the matrix $J_i^{(j)}$ has elements

$$\frac{\partial(\mathbf{x}_i^{(j)})_p}{\partial(\mathbf{d}_i)_s}(t_i).$$

To compute $J_i^{(j)}$ note that it is equal to $X(t_i)$ where $X(t)$ satisfies the differential equation

$$\frac{dX}{dt} = Q(t)\,X \tag{5.3}$$

and the initial condition

$$X(t_{i-1}) = 1, \tag{5.4}$$

where $Q(t)$ is the matrix with components $(\partial f_p/\partial x_s)(t, \mathbf{x}_i^{(j)}(t))$. It will be noted that the $J_i^{(j)}$ are computed by solving systems of linear differential equations, and that they play a role formally identical to that of the $X_i$ in the linear case. In particular we must here choose the interface points to ensure that the $J_i^{(j)}$ are reasonably bounded in norm. By comparing the inequality (4.4) with the result of taking norms in Eq. (5.3), it will be seen that a choice of interface points that ensures that the $\| J_i^{(j)} \|$ are reasonably bounded can be expected to be suitable for the calculation of Eqs. (4.2) and (4.3).

The convergence of Newton's method is guaranteed only for a sufficiently good first approximation. However a useful weakening of this often stringent requirement can frequently be obtained by the following device [11]. Let $R(\mathbf{d})$ be defined by

$$R(\mathbf{d}) = \| \mathbf{c} - B_1\mathbf{d}_1 - B_2\mathbf{x}_{q+1}(b) \|_E^2 + \sum_{i=1}^{q} \| \boldsymbol{\xi}_i + \mathbf{d}_{i+1} - \mathbf{x}_i(t_i) \|_E^2 , \tag{5.5}$$

where the norm is the euclidean vector norm. Then $R$ has a minimum of zero when $\mathbf{d}$ satisfies Eqs. (4.2) and (4.3), and the problem of solving these equations is equivalent to finding the minimum of $R$. The result that justifies introducing $R$ is that, provided the matrix of Eqs. (5.1) and (5.2) is nonsingular and $\mathbf{e}$ does not vanish identically then for some $w > 0$

$$R(\mathbf{d}^{(j)} + w\mathbf{e}^{(j)}) < R(\mathbf{d}^{(j)}). \tag{5.6}$$

This result is usually expressed by saying that the Newton correction $\mathbf{e}$ is downhill for minimizing $R$, and in this case we take

$$\mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} + w\mathbf{e}^{(j)} \tag{5.7}$$

as the approximation to the solution for the next step of the iteration.

Note that ultimately we expect to be able to use the full Newton correction so that 1 is a sensible initial estimate for $w$. If this value is compatible with the inequality (5.6), it is accepted; otherwise $w$ must be reduced and bisection provides a particularly simple strategy.

EXAMPLE 5.1. In [6] the following boundary value problem is given as a difficult example. It is also used by Roberts and Shipman [2]. The problem is to solve the system of differential equations

$$\begin{aligned}
\frac{dx_1}{dt} &= x_2 \\
\frac{dx_2}{dt} &= x_3 \\
\frac{dx_3}{dt} &= -\,1.55x_1x_3 + .1x_2^2 + 1 - x_4^2 + .2x_2 \\
\frac{dx_4}{dt} &= x_5 \\
\frac{dx_5}{dt} &= -\,1.55x_1x_5 + 1.1x_2x_4 + .2x_4 - .2
\end{aligned}$$

subject to the boundary conditions

$$x_1(0) = x_2(0) = x_4(0) = 0,$$
$$x_2(b) = 0, \qquad x_4(b) = 1.$$

This example corresponds to a boundary layer problem, and it is desirable for $b$ to be large compared with the boundary layer thickness. Roberts and Shipman [2], not without difficulty, computed the solution for $b = 11.2$.

In the calculations reported here a multiple shooting approach was used with nine equispaced interface points in $0 \leqslant t \leqslant 10$. The results are reported in table 5.1 for three different choices of $\mathbf{d}_i^{(0)}$, $i = 1, 2, \ldots, 10$.

(i) $\mathbf{d}_i^{(0)}$ estimated roughly from the curves gives in [6],

(ii) $\mathbf{d}_i^{(0)} = 0$, and

(iii) $\mathbf{d}_i^{(0)} = (-1., 0, 0, 1., 0)^T$.

The results give the iteration number and the value of $R$ and $w$ for each set of $\mathbf{d}^{(0)}$.

TABLE 5.1

RESULTS FOR EXAMPLE 5.1

| I.C. | (i) | | (ii) | | (iii) | |
|---|---|---|---|---|---|---|
| ICOUNT | W | R | W | R | W | R |
| 0 | | .1162 E2 | | .1382 E2 | | .2000 E1 |
| 1 | 1. | .6133 E1 | .78 E–2 | .1316 E2 | .5 | .7537 E0 |
| 2 | 1. | .2019 E–1 | .78 E–2 | .1309 E2 | .5 | .2695 E0 |
| 3 | 1. | .8138 E–5 | .625 E–1 | .1299 E2 | 1. | .1237 E–1 |
| 4 | 1. | .3235 E–9 | .625 E–1 | .9287 E1 | 1. | .7987 E–2 |
| 5 | .5 | .1144 E–9 | .5 | .8636 E1 | 1. | .3527 E–5 |
| 6 | | | .5 | .7271 E1 | 1. | .6900 E–9 |
| 7 | | | .5 | .4711 E1 | 1. | .1417 E–9 |
| 8 | | | .125 | .4468 E1 | | |
| 9 | | | .125 | .4426 E1 | | |
| 10 | | | .156 E–1 | .4352 E1 | | |

From Table 5.1 it will be seen that no convergence was obtained for the second set of initial conditions (the crudest guess), while excellent convergence was obtained for the others. The convergence for the third set is of particular interest because the values given to each $\mathbf{d}_i^{(0)}$ can be interpreted as an estimate of the free stream values. This information is generally available in problems of this type. It is clear that, by comparison with the difficulies reported in [2], the use of the multiple shooting method has simplified the choice of the initial approximation to **d** even when, as in this case, comparatively few interface points are used.

EXAMPLE 5.2. This problem was suggested to the author by Dr G. N. Lance, and derives from a similarity solution to the flow between two infinite rotating discs [12]. The system of differential equations is

$$\frac{dx_1}{dt} = -2x_2$$

$$\frac{dx_2}{dt} = x_3$$

$$\frac{dx_3}{dt} = x_1x_3 + x_2^2 - x_4^2 + k$$

$$\frac{dx_4}{dt} = x_5$$

$$\frac{dx_5}{dt} = 2x_2x_4 + x_1x_5$$

and the boundary conditions are

$$x_1(0) = x_2(0) = 0, \qquad x_4(0) = 1,$$

$$x_1(b) = x_2(b) = 0, \qquad x_4(b) = s.$$

This problem corresponds to a generalized eigenvalue problem [4] as $k$ has to be determined so that all six boundary conditions are satisfied. It can be treated in the same way as the previous example by replacing $k$ by a new dependent variable $x_6$ and adding the extra equation $(dx_6/dt) = 0$. The constant $s$ appearing in the boundary conditions depends on the ratio of the speeds of rotation of the two discs, while $b$, the distance between the discs, corresponds to the square root of the Reynolds number for the problem (the scaling of the problem is described in [12]).

Calculations were carried out for three of the cases considered in [12] ($s = .5, 0, -.3$). Again a multiple shooting method was employed using nine equispaced interface points in $0 < t < b$, and in this case the spacing of these points determined the Reynolds number for the problem. Initially $\mathbf{d} = 0$ was taken as a first approximation, and $R$ and $w$ are tabulated in Table 5.2 for Reynolds numbers of 81 and 324. It is clear from these results that the calculations increase in difficulty as $b$ is increased, and the modified Newton method here shows its value.

TABLE 5.2

RESULTS FOR EXAMPLE 5.2

| b | 9 | | | | | | 18 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $s$ | .5 | | 0 | | −.3 | | .5 | | 0 | | −.3 | |
| 1 | W | R | W | R | W | R | W | R | W | R | W | R |
| 0 | | .1250E1 | | .1000E1 | | .1090E1 | | .1450E1 | | .1000E1 | | .1090E1 |
| 1 | .5 | .6123E0 | .5 | .4084E0 | .5 | .3950E0 | .25 | .9229E0 | .25 | .6799E0 | .25 | .7056E0 |
| 2 | 1. | .2270E0 | 1. | .4209E–1 | 1. | .4916E–1 | .0625 | .8636E0 | .5 | .3371E0 | .5 | .3389E0 |
| 3 | 1. | .2863E–2 | 1. | .7991E–3 | 1. | .3946E–2 | .0625 | .8162E0 | 1. | .4135E–1 | 1. | .4174E–1 |
| 4 | 1. | .2336E–6 | 1. | .7431E–4 | 1. | .4336E–3 | .125 | .7665E0 | 1. | .4669E–2 | 1. | .8051E–4 |
| 5 | 1. | .8985E–12 | 1. | .4238E–9 | 1. | .1903E–5 | .25 | .7315E0 | 1. | .9992E–5 | .625E–1 | .7620E–4 |
| 6 | 1. | .4654E–12 | 1. | .1773E–11 | 1. | .1441E–10 | .5 | .4144E0 | 1. | .8429E–5 | .5 | .4884E–4 |
| 7 | | | | | 1. | .3535E–12 | 1. | .8156E–2 | 1. | .2979E–9 | 1. | .1393E–4 |
| 8 | | | | | | | 1. | .2258E–4 | 1. | .3032E–12 | .25 | .1211E–4 |
| 9 | | | | | | | 1. | .3056E–10 | | | .3125E–1 | .1178E–4 |
| 10 | | | | | | | 1. | .1775E–11 | | | .2441E–3 | .1177E–4 |

An interesting feature of these calculations is that the results for $s = 0$, $b = 18$ do not correspond to those given in [12]. The results in [12] were duplicated by recomputing the solution for $b = 9$, and then solving problems for $b = 9.9, 11.7, 14.4$, and 18, at each stage using the solution to the previous problem as an initial estimate for $\mathbf{d}$. This device was also used to compute the solution for $s = -.3$, $b = 18$ which did not give satisfactory convergence in Table 5.2, and to follow the new solution found for $s = 0$, $b = 18$ to smaller values of $b$. This second set of calculations became very difficult for $b < 15$ and the solution was here changing very rapidly. This new solution appears to correspond to the solution found by Rogers and Lance [13] for the flow due to a single rotating disc.

The technique whereby the solution to the problem of interest is found by solving a sequence of problems depending on a parameter, using the solution for one value of the parameter as an initial estimate for the solution at the next, has occurred quite frequently in the literature. In this context it has been called the method of continuation by Roberts and Shipman [2], [3] who apply it in solving the problem in example 5.1, and it has been used in precisely similar fashion to solve a closely related problem by Rogers and Lance [13]. The method of continuation is always available when the boundary conditions are compatible. This is not the case in any of the examples considered here, so that difficulties could possibly occur for values of $b$ which are too small.

## 6. In Conclusion

For linear boundary value problems it is suggested that the interface points be chosen so that the computed approximation $\overline{M}$ to the matrix $M$ of the multiple shooting formulation satisfies

$$\text{(i)}\quad \|\overline{M}\| \leqslant \gamma \qquad \text{and} \qquad \text{(ii)}\quad \|M - \overline{M}\| \leqslant \delta,$$

where $\delta$ is an appropriate absolute accuracy parameter and $\gamma > 1$. In the examples considered a partial pivoting strategy has proved satisfactory in solving the resulting set of linear equations, and this is significant because this technique can be programmed to economize on storage.

Similar considerations have proved satisfactory in determining the interface points in the nonlinear case. Here they are chosen to ensure that fundamental solution matrices of the variational equation (5.3) are suitably bounded in norm. The integration of this equation is a necessary step in the application of Newton's method.

In applying Newton's method it is often necessary to choose the initial approximation carefully to ensure convergence, and example 5.1 is one in which multiple shooting has proved less exacting than ordinary shooting in this respect. A modification to Newton's method which improves its convergence behaviour has been exemplified, and a method of continuation indicated. It can be seen that the combination of a suitably chosen multiple shooting technique with the modified Newton's method and a method of continuation provides a powerful problem solving tool.

## References


1. S. M. Roberts and J. S. Shipman. The Kantarovich theorem and two-point boundary value problems. *IBM J. Res. Develop.* **10** (1966), 402–406.
2. S. M. Roberts and J. S. Shipman. Continuation in shooting methods for two-point boundary value problems. *J. Math. Anal. Appl.* **18** (1967), 45–58.
3. S. M. Roberts and J. S. Shipman. Justification for the continuation method in two-point boundary value problems. *J. Math. Anal. Appl.* **21** (1968), 23–30.
4. H. B. Keller. "Numerical Methods for Two-Point Boundary Value Problems." Blaisdell, 1968.
5. System/360 Scientific Subroutine Package Programmers Manual.
6. J. F. Holt. Numerical solution of nonlinear two-point boundary problems by finite difference methods. *Comm. ACM* **7** (1964), 366–373.
7. T. R. Goodman and G. N. Lance. The numerical solution of two-point boundary value problems. *MTAC* **10** (1956), 82–86.
8. P. Henrici. "Error Propogation for Difference Methods." Wiley, New York, 1963.
9. L. Fox. "The Numerical Solution of Two-Point Boundary Problems in Ordinary Differential Equations." O.U.P., 1957.
10. D. D. Morrison, J. D. Riley, and J. F. Zancanaro. Multiple shooting methods for two-point boundary problems. *Comm. ACM* **5** (1962), 613–614.
11. J. Kowalik and M. R. Osborne. "Methods for Unconstrained Optimisation Problems." Elsevier Press, New York, 1968.
12. G. N. Lance and M. H. Rogers. The symmetric flow of a viscous fluid between two infinite rotating discs. *Proc. Roy. Soc.* **266** (1962), 109–121.
13. M. H. Rogers and G. N. Lance. The rotationally symmetric flow of a viscous fluid in the presence of an infinite rotating disc. *J. Fluid Mech.* **7** (1960), 617–631.