Vernon L. Bakke; Zdzisław Jackiewicz

The numerical solution of boundary-value problems for differential equations with state dependent deviating arguments

Aplikace matematiky, Vol. 34 (1989), No. 1, 1–17

Persistent URL: http://dml.cz/dmlcz/104330

THE NUMERICAL SOLUTION OF BOUNDARY-VALUE PROBLEMS FOR DIFFERENTIAL EQUATIONS WITH STATE DEPENDENT DEVIATING ARGUMENTS

V. L. Bakke, Z. Jackiewicz

(Received January 20, 1987)

Summary. A numerical method for the solution of a second order boundary value problem for differential equation with state dependent deviating argument is studied. Second-order convergence is established and a theorem about the asymptotic expansion of global discretization error is given. This theorem makes it possible to improve the accuracy of the numerical solution by using Richardson extrapolation which results in a convergent method of order three. This is in contrast to boundary value problems for ordinary differential equations where the use of Richardson extrapolation results in a method of order four.

Keywords: Boundary value problem, deviating argument, Richardson extrapolation, convergence of higher order.

AMS Classification: 34K10.

1. INTRODUCTION

In this paper we consider the second order boundary-value problem with state-dependent deviating argument

$$(1)\qquad x''(t) = f(t, x(t), x(\tau(t, x(t)))) \,, \quad t \in [a, b] \,,$$

$$x(t) = \begin{cases} \varphi(t) \,, & t \leqq a \\ \psi(t) \,, & t \geqq b \end{cases} .$$

Here, $f\colon [a, b] \times \mathbb{R}^2 \to \mathbb{R}$, $\tau\colon [a, b] \times \mathbb{R} \to \mathbb{R}$ are continuous and φ and ψ are given initial functions. We do not require that τ is a delay; it can as well be of advanced type. Additional assumptions relative to problem (1) will be given in the next section. Such equations have applications in control theory (further details and references may be found in [7]).

The existence and uniqueness of solutions to (1) was investigated by Grimm and

The work of the second author was partially supported by the Arkansas Science and Technology Authority under grant ASTA 86-B-0072.

Schmitt in [5] and [6]. Similar problems were also treated by Chocholaty and Slahor [3].

It is the purpose of this paper to present a simple difference method for the numerical solution of (1). It will be assumed throughout that a unique solution x to (1) exists. Let $h > 0$ be a given step size and define the grid $t_j = a + jh$, $j = 0, 1, \dots, n + 1$, $(n + 1) h = b - a$. We also define $t_{-1} = a - h$ and $t_{n+2} = b + h$. The numerical method is obtained by approximating the second derivative by the difference operator of the second order, and approximating the solution at non-grid points by piecewise cubic interpolation. Denoting the approximate solution by x_h the resulting method is given by

$$\text{(2)} \qquad x_h(t_{i-1}) - 2x_h(t_i) + x_h(t_{i+1}) = h^2 f(t_i, x_h(t_i), x_h(\tau(t_i, x_h(t_i)))) ,$$

$$x_h(t_i + rh) = R_i(r) := \sum_{j=0}^{3} P_j(r)\, x_h(t_{i-1+j}) ,$$

$i = 1, 2, \dots, n$, $r \in (0, 1]$, where

$$P_0(r) = -\tfrac{1}{6}(r^3 - 3r^2 + 2r) , \quad P_1(r) = \tfrac{1}{2}(r^3 - 2r^2 - r + 2) ,$$

$$P_2(r) -\tfrac{1}{2}(r^3 - r^2 - 2r) , \qquad P_3(r) = \tfrac{1}{6}(r^3 - r) ,$$

are Lagrange fundamental polynomials of cubic interpolation. It is assumed that $x_h(t) = \varphi(t)$ for $t \leqq a$ and $x_h(t) = \psi(t)$ for $t \geqq b$.

In section 2 we show that method (2) is well defined and that the solution to (2) can be obtained by the method of successive approximations. In section 3 we prove a convergence theorem and give a bound on the global error containing local discretization and local interpolation error. In section 4 we prove the existence of one term in the asymptotic expansion of the global discretization error. This fact can be used to improve the accuracy of numerical solution by Richardson extrapolation. Finally, in section 5, the results of this paper are illustrated by some numerical examples.

The numerical solution of problem (1) with τ independent of x was investigated before by de Nevers and Schmitt [4] using the shooting method and by Chocholaty and Slahor [3] using an iterative technique.

2. EXISTENCE AND UNIQUENESS

It is assumed throughout the paper that:

H 1: There exists a constant B such that $|f(t, x, y)| \leqq B$ for $t \in [a, b]$, $x, y \in \mathbb{R}$.

H 2: The function f is Lipschitz-continuous with respect to the second and third argument with constants L_1 and L_2, respectively.

H 3: The function τ is Lipschitz-continuous with respect to the second argument with constant P.

H 4: The functions φ and ψ are Lipschitz-continuous with constants L_φ and L_ψ, respectively.

We can require H 1 without loss of generality since we assumed the existence of a unique solution to problem (1).

It will be convenient for our purposes to use the following convention. With any vector

$$Y_h = [Y_h^1, Y_h^2, \dots, Y_h^n]^{\mathrm{T}}$$

we will associate the function y_h such that $y_h(t) = \varphi(t)$, $t \leqq a$, $y_h(t) = \psi(t)$, $t \geqq b$, and for $t \in [t_j, t_{j+1}]$, $j = 0, 1, \dots, n$, y_h is a cubic polynomial interpolating to y_h^i at t_i, $i = j - 1, j, j + 1, j + 2$. Similarly, if such a function y_h is given we define

$$Y_h := [y_h(t_1), y_h(t_2), \dots, y_h(t_n)]^{\mathrm{T}}\,.$$

In the sequel we denote by lower case letters with subscript h the functions an by uppercase letters with subscript h the vectors related by the described convention. We also define

$$F(Y_h) := \begin{bmatrix} h^2 f(t_1, y_h(t_1), y_h(\tau(t_1, y_h(t_1)))) - \varphi(a) \\ h^2 f(t_2, y_h(t_2), y_h(\tau(t_2, y_h(t_2)))) \\ \vdots \\ h^2 f(t_n, y_h(t_n), y_h(\tau(t_n, y_h(t_n)))) - \psi(b) \end{bmatrix}$$

Where Y_h and y_h are related by the convention described above. Now the method (2) can be written in vector form

$$A_n X_n = F(X_h) \tag{3}$$

where A_n is the $n \times n$ tridiagonal matrix given by

$$A_n = \begin{bmatrix} -2 & 1 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & \dots & 0 & 0 \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ 0 & & 0 & 0 \dots & 1 & -2 \end{bmatrix}$$

For any $V = [v_1, v_2, \dots, v_n]^{\mathrm{T}}$ put $\|V\|_\infty = \max\{|v_i|: 1 \leqq i \leqq n\}$, and for $A \in \mathbb{R}^{n^2}$ denote by $\|A\|_\infty$ the corresponding matrix norm. We will need the following two lemmas.

Lemma 1. $\|A_n^{-1}\|_\infty \leqq \dfrac{(n + 1)^2}{8}\,.$

Proof. This follows easily from the explicit representation of the inverse matrix A_n^{-1}:

$$(A_n^{-1})_{i,j} = r_{i,j} = \begin{cases} \dfrac{i(n + 1 - j)}{n + 1}\,, & i \leqq j \\ r_{i,j} & j < i \end{cases} \tag{4}$$

(see [2]). □

Lemma 2. *Assume* H 1 *and* H 4. *Let X_h be given and define Y_h as the solution of*

$$A_n Y_h = F(X_h)\,. \tag{5}$$

Then

$$|y_h(t) - y_h(s)| \leqq \begin{cases} Q|t - s|\,, & \text{if } t, s \text{ are grid points from } [a, b] \\ \frac{10}{3} Q|t - s|\,, & \text{otherwise,} \end{cases} \tag{6}$$

where

$$Q = \max\left\{L_\varphi, L_\psi, B(b - a) + \frac{|\varphi(a) - \psi(b)|}{b - a}\right\}.$$

Proof. Observe first that in view of H 4 the Lemma is true for $t, s \leqq a$ and $t, s \geqq b$. Denote the ith row of A_n^{-1} by $(A_n^{-1})_i$. Then

$$|y_h(t_{i+1}) - y_h(t_i)| = |((A_n^{-1})_{i+1} - (A_n^{-1})_i\, F(X_h)|$$

for $i = 1, 2, \ldots, n - 1$. From (4) we have

$$(A_n^{-1})_{i+1} - (A_n^{-1})_i = \frac{1}{n + 1}(-1, -2, \ldots, -i, n - i, n - i - 1, \ldots, 1).$$

Hence,

$$|y_h(t_{i+1}) - y_h(t_i)| = \frac{1}{n + 1}\left|\sum_{j=1}^{i} h^2(-j) f(t_j, x_h(t_j), x_h(\tau(t_j, x_h(t_j)))) + \right.$$

$$\left. + \sum_{j=i+1}^{n} h^2(n + 1 - j) f(t_j, x_h(t_j), x_h(\tau(t_j, x_h(t_j)))) + \varphi(a) - \psi(b)\right| \leqq$$

$$\leqq \frac{h^2 B}{n + 1}\left[\sum_{j=1}^{i} j + \sum_{j=1}^{n-i}\right] + \frac{|\varphi(a) - \psi(b)|}{n + 1} =$$

$$= \frac{h^2 B}{2(n + 1)}(i^2 + i + (n - i)(n + 1 - i)) + \frac{h}{b - a}|\varphi(a) - \psi(b)| \leqq$$

$$\leqq h\left[\frac{hB}{2(n + 1)}(n + 1)\, n + \frac{|\varphi(a) - \psi(b)|}{b - a}\right] \leqq$$

$$\leqq h\left[\frac{B(b - a)}{2} + \frac{|\varphi(a) - \psi(b)|}{b - a}\right] \leqq hQ\,,$$

where we have used $nh = n(b - a)/(n + 1) < b - a$. The first equation in the system (5) is given by

$$-2y_h(t_1) + y_h(t_2) = h^2 f(t_1, x_h(t_1), x_h(\tau(t_1, x_h(t_1)))) - y_h(t_0)\,,$$

and it follows that

$$|y_h(t_1) - y_h(t_0)| \leqq |y_h(t_2) - y_h(t_1)| + h^2 B \leqq$$

$$\leqq h\left[\frac{B(b - a)}{2} + \frac{|\varphi(a) - \psi(b)|}{b - a} + hB\right] \leqq hQ\,.$$

By a similar argument we may also show that

$$|y_h(t_n) - y_h(t_{n+1})| \leqq hQ .$$

Assume now that $t, s \in [t_i, t_{i+1}]$, $i = 0, 1, \ldots, n$, and at least one point is not a grid point. We have $t = t_i + rh$, $s = t_i + \bar{r}h$, and

$$x_h(t_i + rh) - x(t_i + \bar{r}h) = R_i(r) - R_i(\bar{r}) = (r - \bar{r})\, R_i'(\xi)$$

for some ξ between t and s. It follows after straightforward calculations that

$$x_h(t_i + rh) - x(t_i + \bar{r}h) = (r - \bar{r})\left[\left(\tfrac{1}{2}\xi^2 + \xi - \tfrac{1}{6}\right)\left(x_h(t_{i+2}) - x_h(t_{i+1})\right) + \right.$$
$$\left. + \left(-\xi^2 - \xi + \tfrac{5}{6}\right)\left(x_h(t_{i+1}) - x_h(t_i)\right) + \left(\tfrac{1}{2}\xi^2 + \tfrac{1}{3}\right)\left(x_h(t_i) - x_h(t_{i-1})\right)\right] .$$

Consequently, since

$$\max\left\{\left|\tfrac{1}{2}\xi^2 + \xi - \tfrac{1}{6}\right| + \left|-\xi^2 - \xi + \tfrac{5}{6}\right| + \left|\tfrac{1}{2}\xi^2 + \tfrac{1}{3}\right| : \xi \in [0, 11]\right\} = \tfrac{10}{3}$$

we obtain

$$|x(t_i + rh) - x(t_i + \bar{r}h)| \leqq \tfrac{10}{3}|r - \bar{r}|\, Qh = \tfrac{10}{3}Q|t - s| ,$$

which proves the Lemma for $t, s \in [t_i, t_{i+1}]$. The repeated use of the triangle inequality proves the Lemma for any $t, s \in [a, b]$. □

Now we show that (3) has a unique solution and that this solution can be computed by the method of successive approximations. Denote by D the constant

$$D = \sup\left\{\sum_{j=0}^{3} |P_j(r)| : r \in [0, 1]\right\} .$$

and consider the condition

$$\text{H 5:} \qquad \frac{(b - a)^2}{8}\left[L_1 + L_2\left(\tfrac{10}{3}\, QP + D\right)\right] < 1 .$$

This condition is similar to one of the conditions given in [6]. We have the following theorem.

Theorem 1. *Assume that the conditions* H 1 – H 5 *hold. Then the system* (3) *has a unique solution* X_h. *Moreover, this solution can be computed by the method of successive approximations defined by*

$$A_n X_n^{k+1} = F(X_h^k) ,$$

$k = 0, 1, \ldots,$ *with arbitrary starting vector* X_h^0.

Proof. It follows from Lemma 2 that each x_h^k corresponding to X_h^k satisfies a Lipschitz condition of the form (6). Observe that for $k \geqq 1$,

$$\|x_h^{k+1} - X_h^k\|_\infty \leqq \|A_n^{-1}\|_\infty\, \|F(X_h^k) - F(X_h^{k-1})\|_\infty ,$$

and

$$\|F(X_h^k) - F(X_h^{k-1})\|_\infty \leqq \max_{1 \leqq j \leqq n} h^2(L_1|x_h^k(t_j) - x_h^{k-1}(t_j)| + L_2\alpha_j),$$

where

$$\alpha_j = |x_h^k(\tau(t_j, x_h^k(t_j))) - x_h^{k-1}(\tau(t_j, x_h^{k-1}(t_j)))|.$$

Using Lemma 2 and the fact that

$$x_h^m(t_i + rh) = \sum_{j=0}^{3} P_j(r)\, x_h^m(t_{i-1+j}),$$

$m = k - 1, k$, we obtain

$$\begin{aligned}\alpha_j \leqq{}& |x_h^k(\tau(t_j, x_h^k(t_j))) - x_h(\tau(t_j, x_h^{k-1}(t_j)))| + \\ &+ |x_h^k(\tau(t_j, x_h^{k-1}(t_j))) - x_h^{k-1}(\tau(t_j, x_h^{k-1}(t_j)))| \leqq \\ &\leqq \tfrac{10}{3}Q|\tau(t_j, x_h^k(t_j)) - \tau(t_j, x_h^{k-1}(t_j))| + \|x_h^k - x_h^{k-1}\|_\infty \leqq \\ &\leqq (\tfrac{10}{3}QP + D)\, \|X_h^k - X_h^{k-1}\|_\infty.\end{aligned}$$

Thus, it follows that

$$\|F(X_h^k) - F(X_h^{k-1})\|_\infty \leqq h^2(L_1 + L_2(\tfrac{10}{3}QP + D))\, \|X_h^k - X_h^{k-1}\|_\infty,$$

and using Lemma 1, we obtain the inequality

$$\|X_h^{k+1} - X_h^k\|_\infty \leqq \frac{(b-a)^2}{8}\, (L_1 + L_2(\tfrac{10}{3}QP + D))\, \|X_h^k - X_h^{k-1}\|_\infty.$$

As a consequence of H 5 the sequence $\{X_h^k\}_{k=0}^\infty$ converges. It is also clear that the limit X_h of this sequence is a solution of (3), and that x_h corresponding to X_h satisfies a Lipschitz condition of the form (6) with the same constant Q.

To prove uniqueness, suppose there is another solution Y_h of (3). Then

$$\|F(X_h) - F(Y_h)\|_\infty \leqq h^2 \max_{1 \leqq j \leqq n} (L_1|x_h(t_j) - y_h(t_j)| + L_2\beta_j),$$

where

$$\begin{aligned}\beta_j ={}& |x_h(\tau(t_j, x_h(t_j))) - y_h(\tau(t_j, y_h(t_j)))| + \\ &+ |x_h(\tau(t_j, y_h(t_j))) - y_h(\tau(t_j, y_h(t_j)))|,\end{aligned}$$

and if $\tau(t_j, y_h(t_j)) \notin [a, b]$, then

$$\beta_j \leqq \tfrac{10}{3}QP|x_h(t_j) - y_h(t_j)|.$$

Otherwise, $\tau(t_j, y_h(t_j)) \in (t_v, t_{v+1}]$, for some $v = 0, 1, \ldots, n$. Putting $r = (\tau(t_j, y_h(t_j)) - t_v)/h$ and using the cubic interpolation formula we obtain

$$\beta_j \leqq \tfrac{10}{3}QP|x_h(t_j) - y_h(t_j)| + \sum_{\mu=0}^{3} |P_\mu(r)|\, |x_h(t_{v-1+\mu}) - y_h(t_{v-1+\mu})|.$$

In either case, we have

$$\beta_j \leqq (\tfrac{10}{3}QP + D)\, \|X_h - Y_h\|_\infty,$$

$j = 1, 2, \ldots, n$, and it follows that

$$\|X_h - Y_h\|_\infty \leqq \|A_n^{-1}\|_\infty h^2(L_1 + L_2(\tfrac{10}{3}QP + D)) \|X_h - Y_h\|_\infty \leqq$$
$$\leqq \frac{(b-a)^2}{8} (L_1 + L_2(\tfrac{10}{3}QP + D)) \|X_h - Y_h\|_\infty .$$

In view of H 5, the conclusion follows. □

3. CONVERGENCE ANALYSIS

In this section we show that the method (3) is convergent, and that the convergence is of second order. A bound on the global error is also derived.

Define the local discretization error at the point t_i by the relation

$$h^2 \eta(t_i) = x(t_{i-1}) - 2x(t_i) + x(t_{i+1}) - h^2 f(t_i, x(t_i), x(\tau(t_i, x(t_i)))) ,$$

$i = 1, 2, \ldots, n$, where x is the solution of (1). It is easy to check that $\eta(t_i) = O(h^2)$ as $h \to 0$ uniformly in t_i. Let us denote by $\xi(t)$ the error of the piecewise cubic interpolation for $x(t)$, i.e.,

$$\xi(t_i + rh) = x(t_i + rh) - \sum_{j=0}^{3} P_j(r)\, x(t_{i-1+j}) ,$$

$i = 0, 1, \ldots, n$, $r \in (0, 1]$. Define

$$e(t; h) = x(t) - x_h(t) ,$$
$$E_h = [e(t_1; h), e(t_2; h), \ldots, e(t_n; h)]^{\mathrm{T}} ,$$
$$\theta_h = [\eta(t_1), \eta(t_2), \ldots, \eta(t_n)]^{\mathrm{T}}$$

and

$$\Omega_h = \max \{|\xi(t)| : t \in [a, b]\} .$$

We have the following convergence result.

Theorem 2. *Assume that the conditions* H 1 – H 5 *are satisfied. Then*

$$\|e(t; h)\|_\infty \leqq WD\|\theta_h\|_\infty + (WDL_2 + 1)\, \Omega_h ,$$

where

$$W = (b-a)^2/(8 - (b-a)^2 (L_1 + L_2(\tfrac{10}{3}QP + D))) .$$

In particular, the method (3) *is convergent and the order of convergence is two.*

Proof. Let $X = [x(t_1), x(t_2), \ldots, x(t_n)]^{\mathrm{T}}$. Since

$$E_h = A_n^{-1}(F(X) - F(X_h) + h^2\theta_h) ,$$

it follows that

$$\|E_h\|_\infty \leqq \|A_n^{-1}\|_\infty \|F(X) - F(X_h)\|_\infty + h^2\|A_n^{-1}\|_\infty \|\theta_h\|_\infty .$$

We have

$$\|F(X) - F(X_h)\|_\infty \leqq h^2 \max_{0 \leqq j \leqq n} (L_1|x(t_j) - x_h(t_j)| + L_2\delta_j),$$

Where

$$\begin{aligned}\delta_j &= |x(\tau(t_j, x(t_j))) - x_h(\tau(t_j, x_h(t_j)))| \leqq \\ &\leqq |x(\tau(t_j, x(t_j))) - x_h(\tau(t_j, x(t_j)))| + \\ &+ |x_h(\tau(t_j, x(t_j))) - x_h(\tau(t_j, x_h(t_j)))|.\end{aligned}$$

As in the proof of Theorem 1 if

$$\tau(t_j, x(t_j)) \notin [a, b], \quad \text{then} \quad \delta_j \leqq \tfrac{10}{3}QP|x(t_j) - x_h(t_j)|.$$

Otherwise, $\tau(t_j, x(t_j)) \in (t_v, t_{v+1}]$, for some $v = 0, 1, \ldots, n$. Putting $r = (\tau(t_j, x(t_j)) - t_v)/h$, using the cubic interpolation formula, and adding and subtracting the term $\sum_{\mu=0}^{3} P_\mu(r)\, x(t_{v-1+\mu})$ we obtain

$$\begin{aligned}\delta_j &\leqq |x(\tau(t_j, x(t_j))) - \sum_{\mu=0}^{3} P_\mu(r)\, x(t_{v-1+\mu})| + \\ &+ |\sum_{\mu=0}^{3} P_\mu(r)\, x(t_{v-1+\mu}) - \sum_{\mu=0}^{3} P_\mu(t)\, x_h(t_{v-1+\mu})| + \\ &+ \tfrac{10}{3}QP|x(t_j) - x_h(t_j)| \leqq \Omega_h + (\tfrac{10}{3}QP + D)\, \|E_h\|_\infty.\end{aligned}$$

Thus,

$$\|E_h\|_\infty \leqq \frac{(b-a)^2}{8}(L_1 + L_2(\tfrac{10}{3}QP + D))\, \|E_h\|_\infty + \frac{(b-a)^2}{8}(\|\theta_h\|_\infty + L_2\Omega_h),$$

and in view of the condition H 5 we obtain

$$\|E_h\|_\infty \leqq W(\|\theta_h\|_\infty + L_2\Omega_h).$$

To get a bound on $e(t; h)$ observe that

$$\begin{aligned}|e(t_i + rh; h)| &\leqq |x(t_i + rh) - \sum_{j=0}^{3} P_j(r)\, x(t_{i-1+j})| + \\ &+ |\sum_{j=0}^{3} P_j(r)\, x(t_{i-1+j}) - \sum_{j=0}^{3} P_j(r)\, x_h(t_{i-1+j})| \leqq D\|E_h\|_\infty + \Omega_h,\end{aligned}$$

and the theorem follows. □

Remark. Observe that in the error estimate given in Theorem 2 $\|\theta_h\|_\infty = O(h^2)$ and $\Omega_h = O(h^4)$ as $h \to 0$. Therefore, we could use piecewise linear interpolation to approximate the solution between the grid points and still maintain the second-order convergence. The corresponding method with piecewisa linear interpolation is convergent under the weaker condition that

$$\frac{(b-a)^2}{8}(L_1 + L_2(QP + 1)) < 1$$

and the error estimate is

$$\|e(t; h)\|_\infty \leqq \tilde{W}\|\theta_h\|_\infty + (\tilde{W}L_2 + 1)\,\Omega_h\,,$$

where Ω_h is the error of linear interpolation and

$$\tilde{W} = (b - a)^2/(8 - (b - a)^2\,(L_1 + L_2(QP + 1)))$$

(compare [1]). The advantage of using piecewise cubic interpolation instead of piecewise linear lies in the fact that the resulting method possesses one term in the asymptotic expansion of the global discretization error which will allow us tu upgrade the accuracy of the numerical solution by Richardson extrapolation. This point is discussed in the next section.

4. ASYMPTOTIC BEHAVIOR OF THE GLOBAL DISCRETIZATION ERROR

In this section we prove the existence of one term in the asymptotic expansion of the global discretization error. We have the following.

Theorem 3. *Assume that* $f \in C^2$, $\tau \in C^2$, *and that* H 1 – H 5 *hold. Then*

$$x(t) - x_h(t) = h^2\,e(t) + \mathrm{O}(h^3)\,, \quad h \to 0\,,$$

where the function e is the solution of the boundary-value problem

$$\begin{aligned} e''(t) = {} & \frac{\partial f}{\partial x}(t, x(t), x(\tau(t, x(t)))\,e(t) + \\ & + \frac{\partial f}{\partial y}(t, x(t), x(\tau(t, x(t))))) \left[\frac{\partial t}{\partial x}(t, x(t))\,x'(\tau(t, x(t)))\,e(t) + \right. \\ & \left. + \, e(\tau(t, x(t)))\right] + \tfrac{1}{12}\,x^{(4)}(t)\,, \\ & e(t) = 0\,, \quad t \leqq a\,, \\ & e(t) = 0\,, \quad t \geqq b\,. \end{aligned} \tag{7}$$

Here, $\partial f/\partial x$ *and* $\partial f/\partial y$ *stand for the derivative of* f *with respect to the second and third argument respectively, and* $\partial\tau/\partial x$ *stands for the derivative of* τ *with respect to the second argument.*

Proof. It is easy to check that the local discretization error η of the method (3) has the form

$$\eta(t_i) = \frac{h^2}{12}\,x^{(4)}(t_i) + \mathrm{O}(h^4)\,, \quad h \to 0\,. \tag{8}$$

For any $t \in [a, b]$ define

$$e_h(t) = \frac{x(t) - x_h(t)}{h^2}\,.$$

Subtracting (3) from

$$x(t_{i-1}) - 2x(t_i) + x(t_{i+1}) = h^2 f(t_i, x(t_i), x(\tau(t_i, x(t_i)))) + h^2\,\eta(t_i)\,,$$

and using (8) we obtain

$$e_h(t_{i-1}) - 2e_h(t_i) + e_h(t_{i+1}) = [f(t_i, x(t_i), x(\tau(t_i, x(t_i))))$$

$$- f(t_i, x_h(t_i), x_h(\tau(t_i, x_h(t_i))))] + \frac{h^2}{12}\,x^{(4)}(t_i) + O(h^4)\,,$$

$i = 1, 2, \ldots, n$. To estimate the expression in brackets observe that

$$x_h(t_i) = x(t_i) - h^2\,e_h(t_i)\,,$$

$$\tau(t_i, x_h(t_i)) = \tau(t_i, x(t_i) - h^2\,e_h(t_i)) = \tau(t_i, x(t_i)) - h^2\,e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i)) + O(h^4),$$

$$x_h(\tau(t_i, x_h(t_i))) = x(\tau(t_i, x_h(t_i))) - h^2\,e_h(\tau(t_i, x_h(t_i))) =$$

$$= x(\tau(t_i, x(t_i)) - h^2\,e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i)) + O(h^4)) -$$

$$- h^2\,e_h(\tau(t_i, x(t_i)) - h^2\,e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i)) + O(h^4)) =$$

$$= x(\tau(t_i, x(t_i))) - h^2\,e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i))\,x'(\tau(t_i, x(t_i))) - h^2[e_h(\tau(t_i, x(t_i))) -$$

$$- h^2\,e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i))\,e_h'(\tau(t_i, x(t_i)))] + O(h^4) =$$

$$= x(\tau(t_i, x(t_i))) - h^2[e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i))\,x'(\tau(t_i, x(t_i))) + e_h(\tau(t_i, x(t_i)))] + O(h^3)$$

where the last equality follows from the fact that

$$e_h'(t) = O\left(\frac{1}{h}\right), \quad h \to 0\,.$$

Consequently,

$$f(t_i, x(t_i), x(\tau(t_i, x(t_i)))) - f(t_i, x_h(t_i), x_h(\tau(t_i, x_h(t_i)))) =$$

$$= h^2\{e_h(t_i)\frac{\partial f}{\partial x}(t_i, x(t_i), x(\tau(t_i, x(t_i)))) + [e_h(t_i)\frac{\partial\tau}{\partial x}(t_i, x(t_i))\,x'(\tau(t_i, x(t_i))) +$$

$$+ e_h(\tau(t_i, x(t_i)))]\frac{\partial f}{\partial y}(t_i, x(t_i), x(\tau(t_i, x(t_i))))\} + O(h^3)\,,$$

and

$$(9) \qquad e_h(t_{i-1}) - 2e_h(t_i) + e_h(t_{i+1}) = h^2 e_h(t_i) \frac{\partial f}{\partial x}(t_i, x(t_i), x(\tau(t_i, x(t_i)))) +$$

$$+ \left[e_h(t_i) \frac{\partial \tau}{\partial x}(t_i, x(t_i))\, x'(\tau(t_i, x(t_i))) + e_h(\tau(t_i, x(t_i))) \right] \times$$

$$\times \frac{\partial f}{\partial y}(t_i, x(t_i), x(\tau(t_i, x(t_i)))) + \tfrac{1}{12} x^{(4)}(t_i)\} + \mathrm{O}(h^3)\,.$$

We also have

$$(10) \qquad e_h(t_i + rh) = \sum_{j=0}^{3} P_j(r)\, e_h(t_{i-1+j}) + \mathrm{O}(h^2)$$

and we can regard (9) and (10) as a perturbed version of the method (3) applied to the boundary-value problem (7) with perturbation of order $\mathrm{O}(h)$ in (9) and of order $\mathrm{O}(h^2)$ in (10). The nonperturbed methods reads

$$\tilde{e}_h(t_{i-1}) - 2\,\tilde{e}(t_i) + \tilde{e}_h(t_{i+1}) = h^2 \left\{ \tilde{e}_h(t_i) \frac{\partial f}{\partial x}(t, x(i), x(\tau(t, x(t)))) + \right.$$

$$+ \left[\tilde{e}_h(t_i) \frac{\partial \tau}{\partial x}(t_i, x(t_i))\, x'(\tau(t_i, x(t_i))) + \right.$$

$$\left. \left. + \tilde{e}_h(\tau(t_i, x(t_i))) \right] \frac{\partial f}{\partial y}(t_i, x(t_i), x(\tau(t_i, x(t_i)))) + \tfrac{1}{12} x^{(4)}(t_i) \right\}$$

$$\tilde{e}_h(t_i + rh) = \sum_{j=0}^{3} P_j(r)\, \tilde{e}_h(t_{i-1+j})\,,$$

and it follows from the convergence theorem that

$$e(t) - \tilde{e}_h(t) = \mathrm{O}(h^2)$$

as $h \to 0$ uniformly in t. Using similar arguments as in the proof of Theorem 2 we also have

$$\tilde{e}_h(t) - e_h(t) = \mathrm{O}(h)\,.$$

Consequently,

$$e(t) - e_h(t) = \mathrm{O}(h)\,,$$

or

$$x(t) - x_h(t) = h^2 e(t) + \mathrm{O}(h^3)\,,$$

$h \to 0$, which is our claim. □

Theorem 3 provides a theoretical basis for the use of Richardson extrapolation to improve the accuracy of the numerical solution. Using standard arguments it follows that

$$\tilde{x}_h(t) := \tfrac{1}{3}(4\, x_h(t) - x_{2h}(t))$$

is an approximation of order three to the solution x of (1). This is in contrast to the numerical solution of two-point boundary value problems for differential equations with deviating argument which does not depend on the state

$$\begin{aligned} (11) \qquad x''(t) &= f(t, x(\tau(t)))\,, \quad t \in [a\,, b] \\ x(t) &= \varphi(t)\,, \quad t \leqq a \\ x(t) &= \psi(t)\,, \quad t \geqq b\,, \end{aligned}$$

where the similar procedure leads to the method of order four. This is a consequence of the fact that in the case of (11) the asymptotic expression of the global discretization error reads

$$x(t) - x_h(t) = h^2\, e(t) + O(h^4)\,,$$

where e is the solution of the boundary-value problem

$$e''(t) = \frac{\partial f}{\partial x}(t, x(t), x(\tau(t)))\, e(t) + \frac{\partial f}{\partial y}(t, x(t), x(\tau(t)))\, e(\tau(t)) + \tfrac{1}{12}x^{(4)}(t)\,,$$

$$e(t) = 0\,, \quad t \leqq a$$

$$e(t) = 0\,, \quad t \geqq b\,.$$

The proof of this fact is similar to the proof of Theorem 3, compare also the corresponding result for two-point boundary-value problems for ordinary differential equations in [8]. Thus the dependence of τ in (1) on the solution x is responsible for the loss of one order of accuracy.

5. NUMERICAL EXAMPLES

The results derived above are illustrated by the following examples.

Example 1 ([3, 4])

$$\begin{aligned} x''(t) &= -(1/16)\sin(x(t)) - (t+1)\,x(t-1) + t\,, && 0 \leqq t \leqq 2\,, \\ x(t) &= t - 1/2\,, && t \leqq 0\,, \\ x(t) &= -1/2 && t \geqq 2\,, \end{aligned}$$

Example 2 ([6]).

$$\begin{aligned} x''(t) &= x(t) - \lambda x(M\sin(x(t)))\,, && 0 \leqq t \leqq T \\ x(t) &= 0\,, && t \leqq 0 \\ x(t) &= 1\,, && t \geqq T \end{aligned}$$

This example is solved for $\lambda = \frac{1}{2}$, $M = \frac{1}{2}$ and $T = 2$.

Example 3.

$$\begin{aligned} x''(t) &= -x(t)\,x^2(t^2) - \sin(t)\cos^2(t^2)\,, && 0 \leqq t \leqq \pi/2 \\ x(t) &= \quad 0 && t \leqq 0 \\ x(t) &= \sin(t) && t \geqq \pi/2 \end{aligned}$$

The theoretical solution is $x(t) = \sin(t)$.

Example 4.

$$\begin{aligned} x''(t) &= x(\ln(t)) + x(t) - t\,, && 1 \leqq t \leqq 2 \\ x(t) &= \exp(t)\,, && t \leqq 1 \\ x(t) &= \exp(t)\,, && t \geqq 2 \end{aligned}$$

The theoretical solution is $x(t) = \exp(t)$.

Example 5.

$$\begin{aligned} x''(t) &= -2\sqrt{(x(2t))}\,, && 0 \leqq t \leqq 1/2 \\ x(t) &= \quad 1\,, && t \leqq 0 \\ x(t) &= \quad \cos^2(t)\,, && t \geqq 1/2 \end{aligned}$$

The theoretical solution is $x(t) = \cos^2(t)$

Example 6.

$$\begin{aligned} x''(t) &= -2\,x(t) - x(t)\ln(x(2t))\,, && 0 \leqq t \leqq 1 \\ x(t) &= \quad 1\,, && t \leqq 0 \\ x(t) &= \exp(-t^2)\,, && t \geqq 1 \end{aligned}$$

The theoretical solution is $x(t) = \exp(-t^2)$.

The system (3) was solved by the method of successive approximations described above, with x_h^0 chosen to be the straight line joining $(a, \varphi(a))$ and $(b, \psi(b))$ over the interval $[a, b]$, and with $x_h^0(t) = \varphi(t)$, $t \leqq a$, $x_h^0(t) = \psi(t)$, $t \geqq b$. The iterations were terminated after the norm of the difference between two successive approximations was less than h^2.

The results are displayed in tables 1–6 below. In all cases $h = (b - a)/n$, where $2^N = n$. The values of x_h are those calculated by the algorithm using cubic interpolation, and y_h is used to denote the values obtained by Richardson extrapolation.

Example 1 was solved be DeNevers and Schmitt [4] by the shooting technique and by Chocholaty and Slahor [3] using an iterative method. The accuracy of our results displayed in Table 1 compares favorably with that of DeNevers and Schmitt for the same stepsize h. Both Tables 1 and 2 show $E_h(T_i) := |x_h(T_i) - X_{2h}(T_i)|$ and $R_h(T_i) := |y_h(T_i) - y_{2h}(T_i)|$, $i = 1, 2, 3$, where $T_i = a + i(b - a)/4$. Since the solutions are known for examples 3~6, the actual errors, $e_h(T_i) := |x(T_i) - x_h(T_i)|$, and $r_h(T_i) := |x(T_i) - y_h(T_i)|$ are displayed.

All computations were performed in double precision on the Amdahl 370/V6-II Computer at the University of Arkansas.

References

[1] *V. L. Bakke, Z. Jackiewicz:* A note on the numerical computation of solutions to second order boundary value problems with state dependent deviating arguments. University of Arkansas Numerical Analysis Technical Report 65110-1, June, 1985.

[2] *B. Chartres, R. Stepleman:* Convergence of difference methods for initial and boundary value problems with discontinuous data. Math. Comp., v. 25, 1971, pp. 724—732.

[3] *P. Chocholaty, L. Slahor:* A numerical method to boundary value problems for second order delay-differential equations. Numer. Math., v. 33, 1979, pp. 69—75.

[4] *K. De Nevers, K. Schmitt:* An application of the shooting method to boundary value problems for second order delay equations. J. Math. Anal. Appl., v. 36, 1971, pp. 588—597.

[5] *L. J. Grimm, K. Schmitt:* Boundary value problems for delay-differential equations. Bull. Amer. Math. Soc., v. 74, 1968, pp. 997—1000.

[6] *L. J. Grimm, K. Schmitt:* Boundary value problems for differential equations with deviating arguments. Aequationes Math., v. 4, 1970, p. 176—190.

[7] *G. A. Kamenskii, S. B. Norkin, L. E. El'sgol'ts:* Some directions of investigation on the theory of differential equations with deviating arguments, (Russian). Trudy Sem. Tear. Diff. Urav. Otklon. Arg., v. 6, pp. 3—36.

[8] *H. B. Keller:* Numerical methods for two-point boundary-value problems, Blaisdel Publishing Company, Waltham 1968.

Table 1. Example 1.

N	$E_h(T_1)$	$R_h(T_1)$	$E_h(T_2)$	$R_h(T_2)$	$E_h(T_3)$	$R_h(T_3)$
2	·559E-1	·146E-2	·120E-0	·298E-2	·779E-1	·261E-2
3	·151E-1	·323E-3	·323E-1	·646E-3	·215E-1	·933E-3
4	·353E-2	·385E-3	·758E-2	·772E-3	·466E-2	·100E-2
5	·117E-2	·181E-3	·247E-2	·363E-4	·192E-2	·479E-4
6	·279E-3	·385E-5	·591E-3	·777E-5	·443E-3	·101E-4
7	·668E-4	·777E-6	·142E-3	·561E-5	·103E-3	·203E-5
8	·161E-4	·155E-6	·343E-4	·310E-6	·243E-4	·403E-6
9	·392E-5	·310E-7	·835E-5	·610E-7	·577E-5	·800E-7
10	·956E-6	·290E-7	·204E-5	·582E-7	·138E-5	·751E-7
11	·217E-6		·467E-6		·289E-6	

Table 2. Example 2.

N	$E_h(T_1)$	$R_h(T_1)$	$E_h(T_2)$	$R_h(T_2)$	$E_h(T_3)$	$R_h(T_3)$
2	·348E-1	·443E-2	·513E-1	·752E-2	·382E-1	·674E-2
3	·538E-2	·383E-2	·719E-2	·527E-2	·449E-2	·343E-2
4	·152E-2	·114E-2	·216E-2	·155E-2	·145E-2	·101E-2
5	·476E-3	·354E-3	·625E-3	·476E-3	·393E-3	·308E-3
6	·146E-3	·106E-3	·201E-3	·144E-3	·132E-3	·930E-4
7	·432E-4	·271E-4	·577E-4	·366E-4	·366E-4	·236E-4
8	·953E-5	·144E-5	·129E-4	·195E-5	·855E-5	·126E-5
9	·130E-5	·915E-6	·178E-5	·124E-5	·119E-5	·798E-6
10	·361E-6	·279E-6	·481E-6	·376E-6	·301E-6	·243E-6
11	·118E-6		·161E-6		·107E-6	

Table 3. Example 3.

N	$e_h(T_1)$	$r_h(T_1)$	$e_h(T_2)$	$r_h(T_2)$	$e_h(T_3)$	$r_h(T_3)$
2	·485E-2	·712E-2	·971E-2	·135E-1	·980E-2	·119E-1
3	·649E-2	·423E-3	·125E-1	·885E-3	·114E-1	·136E-2
4	·130E-2	·310E-3	·248E-2	·583E-3	·183E-2	·432E-3
5	·934E-4	·176E-4	·184E-3	·339E-4	·141E-3	·368E-4
6	·365E-4	·226E-5	·714E-4	·428E-5	·628E-4	·374E-5
7	·108E-4	·647E-6	·210E-4	·122E-5	·184E-4	·107E-5
8	·319E-5	·178E-6	·618E-5	·336E-6	·543E-5	·288E-6
9	·931E-6	·551E-7	·180E-5	·105E-6	·157E-5	·924E-7
10	·274E-6	·153E-7	·528E-6	·289E-7	·462E-6	·253E-7
11	·800E-7	·439E-8	·154E-6	·832E-8	·435E-6	·727E-8
12	·529E-8		·447E-7		·391E-7	

Table 4. Example 4.

N	$e_h(T_1)$	$r_h(T_1)$	$e_h(T_2)$	$r_h(T_2)$	$e_h(T_3)$	$r_h(T_3)$
2	·135E-2	·430E-3	·196E-2	·609E-3	·170E-2	·430E-3
3	·138E-4	·212E-3	·337E-4	·298E-3	·102E-3	·211E-3
4	·161E-3	·441E-4	·232E-3	·624E-4	·184E-3	·441E-4
5	·734E-4	·207E-4	·104E-3	·293E-4	·790E-4	·207E-4
6	·281E-5	·210E-5	·421E-5	·296E-5	·422E-5	·210E-5
7	·227E-5	·453E-6	·327E-5	·640E-6	·263E-5	·453E-6
8	·908E-6	·212E-6	·130E-5	·300E-6	·996E-6	·212E-6
9	·679E-7	·458E-7	·997E-7	·649E-7	·899E-7	·458E-7
10	·174E-7	·215E-7	·237E-7	·304E-7	·119E-7	·215E-7
11	·117E-7	·218E-8	·169E-7	·308E-8	·131E-7	·218E-8
12	·456E-8		·273E-8		·491E-8	

Table 5. Example 5.

N	$e_h(T_1)$	$r_h(T_1)$	$e_h(T_2)$	$r_h(T_2)$	$e_h(T_3)$	$r_h(T_3)$
2	·217E-3	·731E-5	·279E-3	·581E-5	·199E-3	·295E-5
3	·488E-4	·694E-5	·655E-4	·571E-5	·475E-4	·286E-5
4	·699E-5	·220E-5	·121E-4	·179E-5	·975E-5	·893E-6
5	·340E-5	·696E-7	·436E-5	·567E-7	·311E-5	·284E-7
6	·797E-6	·691E-7	·104E-5	·567E-7	·755E-6	·283E-7
7	·147E-6	·221E-7	·219E-6	·181E-7	·168E-6	·907E-8
8	·534E-7	·688E-9	·684E-7	·565E-9	·487E-7	·282E-9
9	·128E-7	·688E-9	·167E-7	·565E-9	·120E-7	·282E-9
10	·269E-8	·220E-9	·375E-8	·181E-9	·277E-9	·940E-10
11	·839E-9	·685E-11	·107E-8	·563E-11	·762E-9	·281E-11
12	·205E-9		·264E-9		·188E-9	

Table 6. Example 6.

N	$e_h(T_1)$	$r_h(T_1)$	$e_h(T_2)$	$r_h(T_2)$	$e_h(T_3)$	$r_h(T_3)$
2	·593E-3	·230E-2	·134E-2	·227E-2	·151E-2	·112E-1
3	·187E-3	·306E-3	·203E-2	·286E-3	·121E-2	·139E-3
4	·238E-3	·455E-4	·294E-3	·444E-4	·199E-3	·220E-4
5	·256E-4	·699E-5	·405E-4	·691E-5	·332E-4	·343E-5
6	·115E-5	·862E-5	·488E-5	·855E-5	·572E-5	·425E-5
7	·675E-5	·108E-5	·763E-5	·107E-5	·462E-5	·534E-6
8	·875E-6	·168E-6	·110E-5	·167E-6	·753E-6	·832E-7
9	·942E-7	·263E-7	·150E-6	·261E-7	·126E-6	·130E-7
10	·339E-8	·326E-7	·179E-7	·323E-7	·218E-7	·161E-7
11	·253E-7	·409E-8	·287E-7	·406E-8	·175E-7	·202E-8
12	·325E-8		·133E-8		·286E-8	

Souhrn

NUMERICKÉ ŘEŠENÍ OKRAJOVÝCH ÚLOH PRO DIFERENCIÁLNÍ ROVNICE SE STAVOVĚ ZÁVISLÝMI ODKLONĚNÝMI ARGUMENTY

V. L. Bakke, Z. Jackiewicz

V článku se studuje numerická metoda řešení okrajové úlohy pro diferenciální rovnici 2. řádu se stavově závislým odkloněným argumentem. Je dokázána konvergence 2. řádu a podána věta o asymptotickém rozvoji globální diskretizační chyby. Tato věta umožňuje zlepšit přesnost numerického řešení použitím Richardsonovy extrapolace, která vede ke konvergenční metodě 3. řádu. Situace se liší od okrajových problémů pro obyčejné diferenciální rovnice, kde užití Richardsonovy extrapolace vede k metodě 4. řádu.

Резюме

ЧИСЛЕННОЕ РЕШЕНИЕ КРАЕВЫХ ЗАДАЧ ДЛЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ЗАВИСЯЩИМИ ОТ СОСТОЯНИЯ ОТКЛОНЯЮЩИМИСЯ АРГУМЕНТАМИ

V. L. Bakke, Z. Jackiewicz

В статье изучается численный метод решения краевой задачи для дифференциального уравнения второго порядка с зависящим от состояния отклоняющимся аргументом.

Доказаны сходимость второго порядка и теорема об асимптотическом разложении глобальной ошибки дискретизации. Эта теорема позволяет повысить точность численного решения при помощи экстраполяции Ричардсона, ведущей к сходящемуся методу третьего порядка. Ситуация отличается от краевых задач для обыкновенных дифференциальных уфавнений, где использование экстраполяции Ричардсона приводит к методу 4-го порядка.

Authors' address: Prof. *V. L. Bakke*, Prof. *Z. Jackiewicz*, Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR 72701.