3-23-2008

Discrete Variable Methods for Delay-Differential Equations with Threshold-Type Delays

Z. Jackiewicz
Arizona State University

Barbara Zubik-Kowal
Boise State University

This document was originally published by Elsevier in *Journal of Computational and Applied Mathematics.* Copyright restrictions may apply. DOI: 10.1016/j.cam.2008.03.022

Discrete variable methods for delay-differential equations with threshold-type delays

Z. Jackiewicz [a,*], B. Zubik-Kowal [b]

[a] *Department of Mathematics, Arizona State University, Tempe, Arizona 85287*

[b] *Department of Mathematics, Boise State University, 1910 University Drive, Boise, Idaho 83725*

Abstract

We study numerical solution of systems of delay-differential equations in which the delay function, which depends on the unknown solution, is defined implicitly by the threshold condition. We study discrete variable numerical methods for these problems and present error analysis. The global error is composed of the error of solving the differential systems, the error from the threshold conditions and the errors in delay arguments. Our theoretical analysis is confirmed by numerical experiments on threshold problems from the theory of epidemics and from population dynamics.

Key words: Delay-differential systems, threshold conditions, numerical approximations, error analysis
PACS: 65R20, 45L10, 93C22

1 Introduction

In this paper we propose numerical algorithms for the approximate solution of threshold problems in population dynamics and epidemics. In such problems the delay function is not known explicitly and must be determined from appropriate threshold conditions which trigger some events such as, for example, appropriate levels of food supply or accumulated dosage of infection. In the

* The work of this author was partially supported by NSF grant DMS–0509597.
Email addresses: jackiewi@math.la.asu.edu, zubik@math.boisestate.edu (B. Zubik-Kowal).

applications considered in this paper these threshold conditions are defined by integral operators which depend on the history of the solution.

To describe the general setting for such problems denote by y_t the function which depends on past values of y defined by $y_t(s) = y(t+s)$, $-\tau_0 \le s \le 0$, $\tau_0 \ge 0$. Assume that the function

$$f : [0,T] \times \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}^q$$

is continuous and consider the state-dependent delay-differential system

$$\begin{cases} y'(t) = f\Big(t, y(t), y\big(t - \tau(t, y_t)\big)\Big), & t \in [0,T], \\ y(t) = g(t), & t \in [-\tau_0, 0], \end{cases} \tag{1.1}$$

with a given initial fuction g and a threshold-type delay

$$\tau : [0,T] \times C^1([-\tau_0, 0], \mathbb{R}^q) \to \mathbb{R}_+$$

of the form

$$P\Big(t, y_t, \tau(t, y_t)\Big) = m. \tag{1.2}$$

Here, $P : [0,T] \times C([-\tau_0,0], \mathbb{R}^q) \times \mathbb{R}_+ \to \mathbb{R}$ is a given operator and $R_+ = [0, \infty)$. The equation (1.2) is called threshold condition, and m is a given threshold. We are not aware of the existence and uniqueness theorems for general problem (1.1)-(1.2) although some results are known for some of the special cases of this problem which are studied in [5], [6], [7], [8], and [10].

To compute an approximation to the solution y of (1.1) with τ defined by (1.2) we consider numerical scheme of the form

$$y_{n+1} = y_n + h\Phi_h\Big(t_n, y_n, \Psi_h\big(\{y_i\}_{i\le n}\big)\Big), \tag{1.3}$$

for $n = 0, 1, \ldots, N-1$, with the step-size h such that $Nh = T$, and the grid-points $t_n = nh$. The initial values y_n for $n \le 0$ are known from the initial condition imposed on $[-\tau_0, 0]$. Here, Φ_h is an increment function which depends on f and $\Psi_h(\{y_i\}_{i\le n})$ is an approximation to $y(t_n - \tau(t_n, y_{t_n}))$. This approximation is computed in the following way.

(a) We first compute an approximation $c_k(t_n)$ to the value of the delay function $\tau(t_n, y_{t_n})$ given implicitly by (1.2) by some iterative procedure

$$c_{i+1}(t_n) = I(c_i(t_n)), \quad i = 0, 1, \ldots,$$

for example the bisection method, with a given initial value $c_0(t_n) \in [0, \tau_0]$. We stop the iterations when for a given function $R(h)$, such that

$$\lim_{h\to 0} R(h) = 0,$$

the difference between two consecutive iterations is less than or equal to $R(h)$, i.e.,

$$|c_k(t_n) - c_{k-1}(t_n)| \leq R(h)$$

for all $n = 1, 2, \ldots, N$. The approximation to y_{t_n} will be computed by a suitable interpolation formula, compare Section 3.2.

(b) We compute next the approximation $\Psi_h(\{y_i\}_{i\leq n})$ to $y(t_n - c_k(t_n))$. Assume that $\nu \geq 1$ and $\mu \leq n$ are indices such that

$$t_n - c_k(t_n) \in [t_{\mu-\nu}, t_\mu],$$

where, preferably, $t_n - c_k(t_n)$ is located in the middle of this interval. This approximation is then computed as $\eta(t_n - c_k(t_n))$, where η is a polynomial which interpolates to $y_{\mu-i}$ at $t_{\mu-i}$ for $i = 0, 1, \ldots, \nu$. Alternatively, if

$$t_n - c_k(t_n) \in [t_{\mu-1}, t_\mu],$$

$y(t_n - c_k(t_n))$ can be computed by a natural continuous extension already defined on the interval $[t_{\mu-1}, t_\mu]$. Note that, for each $n = 0, 1, \ldots, N-1$, since $\tau(t_n, y_{t_n}) > 0$, we have

$$t_n - \tau(t_n, y_{t_n}) < t_n$$

and, even if $h > \tau(t_n, y_{t_n})$, the value $y(t_n - \tau(t_n, y_{t_n}))$ is approximated by η based only on y_i with $i \leq n$.

If the formula (1.3) is based on an implicit Runge-Kutta method, then the values $y(t_n + c_i h - \tau(t_n + c_i h, y_{t_n + c_i h}))$, with the abscissae $c_i \in [0, 1]$, need to be interpolated. In this case, it may happen that $c_i h > \tau(t_n + c_i h, y_{t_n + c_i h})$. Then, the interpolating polynomial η needs the value y_{n+1}, which makes the method (1.3) implicit. However, the polynomial η is not the only reason of the implicitness of the entire method as the Runge-Kutta method chosen for (1.3) is already implicit. In Section 4, we apply explicit Euler's method to problems from the theory of epidemics and from population dynamics. In this case the method (1.3) is explicit and the value y_{n+1} is not used for the polynomial η.

2 Error estimation

Let $y(t)$ be the exact solution to the problem (1.1)-(1.2) and y_n be approximations determined by steps (a)-(b) described in Section 1. Define the global error at the point t_n by $e_n = y(t_n) - y_n$ and let $\|\cdot\|$ be an arbitrary vector norm in $\mathbb{R}^q$. The following theorem gives an error estimation for $\|e_n\|$.

Theorem 2.1 *Suppose that the method (1.3) satisfies the consistency condi-*

tion

$$\left\| y(t_{n+1}) - y(t_n) - h\Phi_h\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big) \right\| \le ChS(h) \tag{2.1}$$

with a positive constant C *and a positive function* $S(h)$ *such that* $\lim_{h\to 0} S(h)=0$. *Moreover, suppose that*

$$\left\| \Phi_h(t,u_1,v_1) - \Phi_h(t,u_2,v_2) \right\| \le L_1\|u_1-u_2\| + L_2\|v_1-v_2\| \tag{2.2}$$

with constants $L_1, L_2 \ge 0$ *and*

$$\left\| \Psi_h(\{y(t_i)\}_{i\le n}) - \Psi_h(\{y_i\}_{i\le n}) \right\| \le L \max\Big\{ \left\| y(t_i) - y_i \right\| : i \le n \Big\} + s(h) \tag{2.3}$$

with a positive constant L *and a positive function* $s(h)$ *such that* $\lim_{h\to 0} s(h)=0$. *Then*

$$\|e_n\| \le \frac{1}{L_1+LL_2}\Big(e^{(L_1+LL_2)t_n} - 1\Big)\Big(L_2 s(h) + CS(h)\Big), \tag{2.4}$$

$n = 0,1,\ldots,N,$ *as* $h \to 0$.

Proof: Observe that the term $s(h)$ in (2.3) takes into account errors introduced by the approximation y_i to $y(t_i)$. It follows from (2.1) that

$$y(t_{n+1}) = y(t_n) + h\Phi_h\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big) + Q(h), \tag{2.5}$$

with $\|Q(h)\| \le ChS(h)$. Subtracting (1.3) from (2.5) we obtain

$$\begin{aligned} e_{n+1} = e_n + h\bigg(&\Phi_h\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big) - \Phi_h\Big(t_n, y_n, \Psi_h\big(\{y_i\}_{i\le n}\big)\Big)\bigg) \\ &+ Q(h), \end{aligned}$$

which together with (2.2) gives

$$\begin{aligned} \|e_{n+1}\| \le \|e_n\| + hL_1\|e_n\| + hL_2 &\left\| \Psi_h\big(\{y(t_i)\}_{i\le n}\big) - \Psi_h\big(\{y_i\}_{i\le n}\big) \right\| \\ &+ ChS(h). \end{aligned}$$

Application of (2.3) to the third term of the right-hand side of the above inequality results in

$$\begin{aligned} \|e_{n+1}\| &\le (1+hL_1)\|e_n\| + hL_2 L \max_{i\le n}\|e_i\| + hL_2 s(h) + ChS(h) \\ &\le (1 + hL_1 + hL_2 L)\max_{i\le n}\|e_i\| + hL_2 s(h) + ChS(h), \end{aligned}$$

$n = 0, 1, \ldots, N-1$. Introducing the notation

$$\alpha = 1 + hL_1 + hL_2L, \qquad \beta = hL_2 s(h) + ChS(h), \tag{2.6}$$

we obtain

$$\|e_{n+1}\| \le \alpha \max_{i \le n} \|e_i\| + \beta. \tag{2.7}$$

By induction with respect to n, (2.7) implies

$$\|e_n\| \le \beta \frac{1-\alpha^n}{1-\alpha}, \tag{2.8}$$

for all n. increasing Using (2.8) with (2.6) results in

$$\|e_n\| \le \frac{1}{L_1 + L_2L}\left(\left(1 + h(L_1 + L_2L)\right)^n - 1\right)\left(L_2 s(h) + CS(h)\right).$$

Since

$$\left(1 + h(L_1 + L_2L)\right)^n \le e^{nh(L_1+L_2L)},$$

the above inequality leads to (2.4). This completes the proof. □

The error bound (2.4) shows convergence of the method (1.3). The order of the convergence depends on the functions $s(h)$ and $S(h)$, that is, on the choices of Φ_h (integration in time) and Ψ_h (solving the threshold condition (1.2)). For example, for $\Phi_h(t, y, z) = f(t, y, z)$ the numerical scheme (1.3) corresponds to the Euler method composed with the method Ψ_h described by the steps (a)-(b) in Section 1. The next theorem shows that, under certain conditions imposed on f, y and Ψ_h, the consistency condition (2.1) is satisfied for the increment function Φ_h representing Euler's method.

Theorem 2.2 *Suppose that $\Phi_h = f$ satisfies the condition*

$$\|f(t, u, v_1) - f(t, u, v_2)\| \le L_2\|v_1 - v_2\| \tag{2.9}$$

and that the exact solution $y : [-\tau_0, T] \to \mathbb{R}^q$ of (1.1)-(1.2) has its second derivative bounded on $[0, T]$. Let a positive constant C_2 satisfies

$$\max_{t\in[0,T]} \|y''(t)\| \le C_2. \tag{2.10}$$

Moreover, suppose that for each grid-point t_n the method Ψ_h, which performs the steps (a)-(b), satisfies the condition

$$\left\| y\Big(t_n - \tau(t_n, y_{t_n})\Big) - \Psi_h\Big(\{y(t_i)\}_{i\le n}\Big)\right\| \le r(h), \tag{2.11}$$

with a positive function $r(h)$ such that $\lim_{h\to 0} r(h) = 0$. Then

$$\left\| y(t_{n+1}) - y(t_n) - h\Phi_h\Big(t_n, y(t_n), \Psi_h\Big(\{y(t_i)\}_{i\le n}\Big)\Big)\right\| \le \ Ch\Big(h + r(h)\Big), \tag{2.12}$$

with $C = \max\{\frac{1}{2}C_2, L_2\}$.

Proof: For each n the solution y satisfies

$$y(t_{n+1}) = y(t_n) + hf\Big(t_n, y(t_n), y\big(t_n - \tau(t_n, y_{t_n})\big)\Big) + \frac{y''(\xi_n)}{2}h^2,$$

with $\xi_n \in (0, T)$. Here, $y''(\xi_n) = \Big(y_1''(\xi_n), \ldots, y_q''(\xi_n)\Big)^T$. Therefore,

$$\begin{aligned} &y(t_{n+1}) - y(t_n) - h\Phi_h\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big) = \\ &\quad= y(t_{n+1}) - y(t_n) - hf\Big(t_n, y(t_n), y\big(t_n - \tau(t_n, y_{t_n})\big)\Big) + hQ(h) \\ &\quad= \frac{y''(\xi_n)}{2}h^2 + hQ(h) \end{aligned}$$

with

$$Q(h) = f\Big(t_n, y(t_n), y\big(t_n - \tau(t_n, y_{t_n})\big)\Big) - f\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big).$$

Since

$$\Big\|Q(h)\Big\| \le L_2\, r(h)$$

this leads to

$$\Big\|y(t_{n+1}) - y(t_n) - h\Phi_h\Big(t_n, y(t_n), \Psi_h\big(\{y(t_i)\}_{i\le n}\big)\Big)\Big\| \le \frac{1}{2}C_2\, h^2 + hL_2\, r(h)$$

which completes the proof. □

Theorem 2.2 shows that (2.9)-(2.11) imply the consistency condition (2.1) with $S(h) = h + r(h)$. The conditions (2.9)-(2.10) depend on the function f and the solution y. The condition (2.11) and the function $r(h)$ are investigated in the next section.

3 Errors from threshold conditions

3.1 Consistency condition

In this subsection we will show that the property (2.11) is satisfied for the threshold condition (1.2) defined by the integral operator of the form

$$P(t, y_t, \tau) = \int_{-\tau}^{0} p\big(y_t(s)\big)ds, \tag{3.1}$$

where $p : \mathbb{R}^q \to \mathbb{R}$ is a given smooth and positive function and y denotes the solution to (1.1)-(1.2). We assume that the operator Ψ_h is defined by an interpolating polynomial based on the grid-points used for Φ_h. We do not make assumptions about the number of grid-points used for Ψ_h and we denote this number by $l = \nu + 1$, compare (b) in Section 1.

The values of $\tau \in [0, \tau_0]$ are unknown and depend on time t. Let $\tau(t_n, y_{t_n}) > 0$ satisfies (1.2) at $t = t_n$ with the operator P defined by (3.1). To find an approximation to $\tau(t_n, y_{t_n})$ we apply a root-finding numerical method combined with a numerical quadrature for an approximation of the integral in (3.1).

Assume that the chosen quadrature satisfies the relation

$$J\Big(t_n, \{v(t_i)\}_{i \le n}, c\Big) = \int_{-c}^{0} p\Big(v(t_n + s)\Big) ds + R_p(h, t_n, c), \tag{3.2}$$

for any smooth function $v : [-\tau_0, T] \to \mathbb{R}^q$, $n = 1, 2, \ldots, N$, and $c \in [0, \tau_0]$, with the remainder $R_p(h, t_n, c)$ of the quadrature formula which satisfies the conditions

$$\begin{aligned} &\Big|R_p(h, t_n, c)\Big| \le \mathcal{R}_p(h), \quad t_n \in [0, T], \quad c \in [0, \tau_0], \\ &\lim_{h \to 0} \mathcal{R}_p(h) = 0. \end{aligned} \tag{3.3}$$

The properties (3.2) and (3.3) are guaranteed by smoothness of the functions p and v.

Denote by $c_J(t_n)$, $n = 1, 2, \ldots, N$, the roots of the equations

$$J\Big(t_n, \{v(t_i)\}_{i \le n}, c_J(t_n)\Big) = m, \quad n = 1, 2, \ldots, N, \tag{3.4}$$

and by $c_j(t_n)$, $j = 0, 1, 2, \ldots$, the successive approximations to $c_J(t_n)$ determined by the root-finding numerical method applied to (3.4). Let $r_1(h)$ be a chosen positive function such that $\lim_{h \to 0} r_1(h) = 0$ and let k be determined in such a way that

$$|c_k(t_n) - c_J(t_n)| \le r_1(h), \tag{3.5}$$

for $n = 1, 2, \ldots, N$.

Let $r_0(h)$ be a chosen positive function such that $\lim_{h \to 0} r_0(h) = 0$. For $r_0(h)$ and v, the interpolating polynomials u_n, $n = 1, 2, \ldots, N$, are constructed according to the following conditions. Each polynomial u_n is based on l grid points t_i and l values $v(t_i)$ with $i \le n$, i.e., $u_n(t_i) = v(t_i)$ at l grid points t_i chosen arbitrarily within the constraint of the inequality $i \le n$. The number l is chosen in such a way that

$$\max_{t \in V_n} \|u_n(t) - v(t)\| \le r_0(h), \tag{3.6}$$

for all $n = 1, 2, \ldots, N$. Here, V_n is an interval which contains the l grid-points t_i and $t_n - c_k(t_n) \in V_n$.

Having the process of composing the quadature with the root-finding method and with the interpolating polynomials described above, we define the operator Ψ_h by

$$\Psi_h\Big(\{v_i\}_{i \le n}\Big) = u_n\Big(t_n - c_k(t_n)\Big), \tag{3.7}$$

for $n = 1, 2, \ldots, N$, with u_n based on $v_i = v(t_i) \in \mathbb{R}^q$, where the index i is such that $t_i \in V_n$.

For the case of $\Phi_h(t, y, z) = f(t, y, z)$ (Euler's method) choosing $l = 2$ is enough because it results in $V_n = [t_j, t_{j+1}]$, for a certain index j, and since v represents the exact solution y, by (2.10), we have

$$\max_{t \in V_n} \|u_n(t) - v(t)\| = \frac{1}{2} \max_{t \in V_n} \Big|(t - t_j)(t - t_{j+1})\Big| \max_{t \in V_n} \|v''(t)\| \le \frac{C_2}{2} h^2.$$

Therefore, the error of the interpolation is not larger than the error of Euler's method and there is no need to choose $l > 2$. This is confirmed in Section 4, where we apply the method (1.3) with $l = 2$ to problems from population dynamics and theory of epidemics. The order of convergence is presented in Table 1. All the numerical experiments for Table 1 were performed with $l = 2$.

We have the following theorem for the operator Ψ_h needed for the method (1.3).

Theorem 3.1 *Suppose that* y *is a smooth solution to problem (1.1)-(1.2) with the operator* P *defined by (3.1). Suppose that the function* $p : \mathbb{R}^q \to \mathbb{R}_+$, $\mathbb{R}_+ = [0, \infty)$, *is smooth and has a finite number of roots. Suppose that the operator* Ψ_h *is constructed with a quadrature which satisfies (3.2)-(3.3), a root-finding numerical method which satisfies (3.4)-(3.5), and interpolating polynomials which satisfy (3.6). Then there exists a positive function* $r(h)$ *such that* $\lim_{h \to 0} r(h) = 0$ *and the consistency condition (2.11) is satisfied.*

Proof: Let $c_J(t_n)$, $n = 1, 2, \ldots, N$, be the roots of the equations (3.4) based on the values of $y(t_i)$ and let $c_k(t_n)$, $n = 1, 2, \ldots, N$, be their approximations by the root-finding method which satisfies (3.5).

We first estimate the error $|c_k(t_n) - \tau(t_n, y_{t_n})|$. Since

$$\int_{-\tau(t_n, y_{t_n})}^{0} p\Big(y(t_n + s)\Big) ds = m = J\Big(t_n, \{y(t_i)\}, c_J(t_n)\Big),$$

it follows from (3.2)-(3.4) that

$$\left| \int_{-\tau(t_n,y_{t_n})}^{-c_J(t_n)} p\Big(y(t_n+s)\Big)ds \right| = \left| \int_{-\tau(t_n,y_{t_n})}^{0} p\Big(y(t_n+s)\Big)ds - \int_{-c_J(t_n)}^{0} p\Big(y(t_n+s)\Big)ds \right|$$

$$= \left| J\Big(t_n, \{y(t_i)\}, c_J(t_n)\Big) - \int_{-c_J(t_n)}^{0} p\Big(y(t_n+s)\Big)ds \right| \le \mathcal{R}_p(h).$$

This inequality and the relations (3.3) imply that

$$\lim_{h\to 0} \left(\int_{-\tau(t_n,y_{t_n})}^{-c_J(t_n)} p\Big(y(t_n+s)\Big)ds \right) = 0 \tag{3.8}$$

uniformly with respect to $t_n \in [0,T]$. Since the function p is smooth and positive, the relation (3.8) implies the existence of a positive function $r_2(h)$ such that

$$|c_J(t_n) - \tau(t_n, y_{t_n})| \le r_2(h) \quad \text{and} \quad \lim_{h\to 0} r_2(h) = 0. \tag{3.9}$$

From (3.5) and (3.9) we have

$$\begin{aligned} \Big|c_k(t_n) - \tau(t_n, y_{t_n})\Big| &\le \Big|c_k(t_n) - c_J(t_n)\Big| + \Big|c_J(t_n) - \tau(t_n, y_{t_n})\Big| \\ &\le r_1(h) + r_2(h). \end{aligned} \tag{3.10}$$

It follows from (3.10) and (3.6) that

$$\begin{aligned} &\Big\|y\Big(t_n - \tau(t_n, y_{t_n})\Big) - \Psi\Big(\{y(t_i)\}_{i\le n}\Big)\Big\| \\ &\le \Big\|y\Big(t_n - \tau(t_n, y_{t_n})\Big) - y\Big(t_n - c_k(t_n)\Big)\Big\| \\ &+ \Big\|y\Big(t_n - c_k(t_n)\Big) - \Psi_h\Big(\{y(t_i)\}_{i\le n}\Big)\Big\| \\ &\le \max_{s\in[-\tau_0,T]} \|y'(s)\| \Big(r_1(h) + r_2(h)\Big) + r_0(h), \end{aligned}$$

which proves (2.11) with $r(h)$ defined by

$$r(h) = \max_{s\in[-\tau_0,T]} \|y'(s)\| \Big(r_1(h) + r_2(h)\Big) + r_0(h).$$

This completes the proof. □

Theorem 3.1 gives sufficient conditions for (2.11). The function $r(h)$ describes how good the approximations $\Psi_h\Big(\{y_i\}_{i\le n}\Big)$ are for the unknown values $y\Big(t_n - \tau(t_n, y_{t_n})\Big)$ and it depends on three components:

(1) $r_0(h)$ dictated by the interpolation procedure needed for computing the polinomial u_n,
(2) $r_1(h)$ dictated by the iteration process needed for computing the root $c_k(t_n)$,
(3) $r_2(h)$ dictated by the quadrature needed for computing the integral in (3.1).

All of these components can satisfy

$$r_i(h) \le Kh^p, \quad i = 0, 1, 2,$$

with a positive constant K and $p \ge 1$. This can be achieved by choosing an appropriate number l of grid-points for u_n in case $i = 0$, an appropriate number of iterations k in case $i = 1$, and an approriate high order quadrature in case $i = 2$.

3.2 Errors in delay arguments

The next theorem shows that the operator Ψ_h defined by (3.7) satisfies the property (2.3).

Theorem 3.2 *Suppose that the assumptions of Theorem 3.1 are satisfied. Moreover, assume that the composite function $p \circ y$ is bounded from below by a constant $M_0 > 0$, i.e.,*

$$\min_{s \in [-\tau_0, T]} \left| p\Big(y(s)\Big) \right| \ge M_0. \tag{3.11}$$

Let L_i be the weight functions of the Lagrange interpolating polynomials u_n based on the l grid points. Let $\overline{L}$ be a Lipschitz constant for all the weight functions L_i over the interval $[-\tau_0, T]$ and let M be their maximum value, i.e., the Lebesgue constant. Then there exist a positive constant L and a positive function $s(h)$, which satisfy (2.3) and such that $\lim_{h \to 0} s(h) = 0$.

Proof: Let $c_k(t_n)$ be the approximation to the value $\tau(t_n, y_{t_n})$ obtained by the root-finding method described in Section 3 applied to the approximation of the threshold condition (1.2) by the quadrature formula based on the values $y(t_i)$, $i \le n$. Similarly, let $\overline{c}_{\overline{k}}(t_n)$ be obtained by the same process but based on the approximate values y_i, $i \le n$. Let $c_J(t_n)$ and $\overline{c}_J(t_n)$ be the corresponding roots of the equations (3.4) based on $y(t_i)$ and y_i, respectively.

We will first show that there exist a positive constant B and a positive function $b(h)$ such that $\lim_{h\to 0} b(h) = 0$ and

$$\Big|c_k(t_n) - \overline{c}_{\overline{k}}(t_n)\Big| \leq B \max_{i\leq n} \Big\|y(t_i) - y_i\Big\| + b(h), \tag{3.12}$$

for $n = 1, 2, \ldots, N$. Since, for all $n = 1, 2, \ldots, N$, $J(t_n, \{y_i\}_{i\leq n}, \overline{c}_J(t_n))$ and $J(t_n, \{y(t_i)\}_{i\leq n}, \overline{c}_J(t_n))$ are linear combinations of the same number of values from the sets $\{y_i\}_{i\leq n}$ and $\{y(t_i)\}_{i\leq n}$, respectively, with the same constant coefficients, there exists a positive constant $\widetilde{M}$ such that

$$\begin{aligned}&\Big|J\Big(t_n, \{y_i\}_{i\leq n}, \overline{c}_J(t_n)\Big) - J\Big(t_n, \{y(t_i)\}_{i\leq n}, \overline{c}_J(t_n)\Big)\Big| \\ &\qquad \leq \widetilde{M} \max_{i\leq n} \Big\|y_i - y(t_i)\Big\|,\end{aligned} \tag{3.13}$$

for $n = 1, 2, \ldots, N$. Moreover, for (3.4)

$$\begin{aligned}&\Big|J\Big(t_n, \{y_i\}_{i\leq n}, \overline{c}_J(t_n)\Big) - J\Big(t_n, \{y(t_i)\}_{i\leq n}, \overline{c}_J(t_n)\Big)\Big| \\ &\geq -\Big|m - \int_{-c_J(t_n)}^{0} p\Big(y(t_n+s)\Big)ds + \int_{-\overline{c}_J(t_n)}^{0} p\Big(y(t_n+s)\Big)ds \\ &\qquad - J\Big(t_n, \{y(t_i)\}_{i\leq n}, \overline{c}_J(t_n)\Big)\Big| + \Big|\int_{-c_J(t_n)}^{-\overline{c}_J(t_n)} p\Big(y(t_n+s)\Big)ds\Big| \\ &= -\Big|R_p\Big(t_n, h, c_J(t_n)\Big) + R_p\Big(t_n, h, \overline{c}_J(t_n)\Big)\Big| + \Big|\int_{-c_J(t_n)}^{-\overline{c}_J(t_n)} p\Big(y(t_n+s)\Big)ds\Big|.\end{aligned}$$

This together with (3.11) and (3.13) leads to

$$\begin{aligned}M_0\Big|c_J(t_n) - \overline{c}_J(t_n)\Big| &\leq \Big|\int_{-c_J(t_n)}^{-\overline{c}_J(t_n)} p\Big(y(t_n+s)\Big)ds\Big| \\ &\leq \Big|R_p\Big(t_n, h, c_J(t_n)\Big) + R_p\Big(t_n, h, \overline{c}_J(t_n)\Big)\Big| + \widetilde{M} \max_{i\leq n} \Big\|y_i - y(t_i)\Big\|,\end{aligned}$$

which by (3.3) gives

$$\Big|c_J(t_n) - \overline{c}_J(t_n)\Big| \leq \frac{\widetilde{M}}{M_0} \max_{i\leq n} \Big\|y_i - y(t_i)\Big\| + \frac{2}{M_0}\mathcal{R}_p(h).$$

By the inequality (3.5) we obtain

$$\begin{aligned} \Big|c_k(t_n) - \overline{c}_{\overline{k}}(t_n)\Big| &\le \Big|c_k(t_n) - c_J(t_n)\Big| \\ &+ \Big|c_J(t_n) - \overline{c}_J(t_n)\Big| + \Big|\overline{c}_J(t_n) - \overline{c}_{\overline{k}}(t_n)\Big| \\ &\le \frac{\widetilde{M}}{M_0} \max_{i\le n} \Big\|y_i - y(t_i)\Big\| + \frac{2}{M_0}\mathcal{R}_p(h) + 2r_1(h), \end{aligned} \tag{3.14}$$

which shows (3.12). Since Ψ_h is based on interpolation, by (3.14) we have

$$\begin{aligned} &\Big\|\Psi_h\Big(\{y(t_i)\}_{i\le n}\Big) - \Psi_h\Big(\{y_i\}_{i\le n}\Big)\Big\| \\ &= \Big\|\sum_{i\le n} y(t_i)L_i\Big(t_n - c_k(t_n)\Big) - \sum_{i\le n} y_i L_i\Big(t_n - \overline{c}_{\overline{k}}(t_n)\Big)\Big\| \\ &\le \Big\|\sum_{i\le n} y(t_i)L_i\Big(t_n - c_k(t_n)\Big) - \sum_{i\le n} y(t_i) L_i\Big(t_n - \overline{c}_{\overline{k}}(t_n)\Big)\Big\| \\ &+ \Big\|\sum_{i\le n} y(t_i)L_i\Big(t_n - \overline{c}_{\overline{k}}(t_n)\Big) - \sum_{i\le n} y_i L_i\Big(t_n - \overline{c}_{\overline{k}}(t_n)\Big)\Big\| \\ &\le \sum_{i\le n} \overline{L}\|y(t_i)\| |c_k(t_n) - \overline{c}_{\overline{k}}(t_n)| + Ml \max_{i\le n} \|y(t_i) - y_i\| \\ &\le l\overline{L} \max_{s\in[-\tau_0,T]} \Big\|y(s)\Big\| \Bigg(\frac{\widetilde{M}}{M_0} \max_{i\le n} \Big\|y_i - y(t_i)\Big\| + \frac{2}{M_0}\mathcal{R}_p(h) + 2r_1(h)\Bigg) \\ &+ Ml \max_{i\le n} \|y(t_i) - y_i\|. \end{aligned}$$

This shows (2.3) with L and $s(h)$ defined by

$$L = \frac{l\overline{L}\widetilde{M}}{M_0} \max_{s\in[-\tau_0,T]} \Big\|y(s)\Big\| + lM$$

and

$$s(h) = l\overline{L} \max_{s\in[-\tau_0,T]} \Big\|y(s)\Big\| \Big(\frac{2}{M_0}\mathcal{R}_p(h) + 2r_1(h)\Big),$$

which completes the proof. □

Note that, although the assumption (3.11) is significantly used in the proof of Theorem 3.2, it is not used in the proof of Theorem 3.1.

In Section 4, we apply the method (1.3) to problems from population dynamics and theory of epidemics. For these applications the operator Ψ_h is based on the composite trapezoidal rule. Note that, for the integral operator (3.1) computed by the composite trapezoidal rule, the inequality (3.13) is satisfied with a constant $\widetilde{M} = \tau_0$.

By Theorems 2.1, 2.2, 3.1, and 3.2 we have the following corollary.

Corollary 3.3 *Suppose that the function f satisfies the Lipschitz condition*

$$\Big\| f(t, u_1, v_1) - f(t, u_2, v_2) \Big\| \leq L_1 \|u_1 - u_2\| + L_2 \|v_1 - v_2\|.$$

Moreover, suppose that the condition (2.10) and the assumptions of Theorem 3.2 are satisfied. Then the error bound (2.4) holds with $s(h)$ given in Theorem 3.2 and with $S(h) = h + r(h)$, where $r(h)$ is given in Theorem 3.1.

Corollary 3.3 shows that for the choice of $\Phi_h(t, y, z) = f(t, y, z)$ (Euler's method) the composite method (1.3) has order 1 if the iterative scheme and the quadrature applied for the threshold condition (1.2) with (3.1) do not introduce errors larger than of order 1. This is confirmed by Table 1, which is introduced and described in Section 4.

In the next section we apply the numerical technique described in this paper to specific examples from population dynamics and epidemics. To summarize, the numerical algorithm proposed in this paper is based on the following numerical schemes:

(1) A one-step formula (1.3) with increment function Φ_h to advance the step from t_n to t_{n+1}. We will use $\Phi_h(t, y, z) = f(t, y, z)$ which corresponds to the Euler method.
(2) A formula $\Psi_h(\{y_i\}_{i \leq n})$ to compute an approximation to the delayed term $y\Big(t_n - \tau(t_n, y_{t_n})\Big)$ which, in general, requires:
  (a) A suitable interpolation formula based on the given or already computed values y_i, $i \leq n$, to construct a continuous interpolant $u_n(s) \approx y_{t_n}(s)$, $-\tau_0 \leq s \leq 0$. Our algorithm is based on the Lagrange'a interpolation formula.
  (b) An iterative procedure to find an approximation $c_k(t_n)$ to the solution of the threshold condition $P\Big(t_n, u_n, \tau(t_n, u_n)\Big) = m$ or of its suitable approximation $\widetilde{P}\Big(t_n, u_n, \tau(t_n, u_n)\Big) = m$. The algorithm employed in this paper is based on the bisection method.
  (c) A suitable interpolation formula to find an approximation to the solution $y(t_n - c_k(t_n))$. Again this step is based on the Lagrange'a interpolation formula.

4 Applications in population dynamics and theory of epidemics

Example 1. Gourley and Kuang [5] study a new predator-prey model, which is an extension of the model by Aiello and Freedman [1]. The model derived in [5] can be written in the form

$$\begin{cases} x'(t) = \frac{r}{K}\, x(t)\Big(1 - x(t)\Big) - y(t)p\Big(x(t)\Big), \\ y'(t) = be^{-d_j\tau\left(t,x_t\right)}y\Big(t-\tau(t,x_t)\Big)p\bigg(x\Big(t-\tau(t,x_t)\Big)\bigg) - dy(t), \\ x(t) \ = x_0(t), \quad t\in[\alpha,0], \\ y(t) \ = y_0(t), \quad t\in[\alpha,0], \end{cases} \tag{4.1}$$

$t \geq 0$, $\alpha \leq 0$. Here, $x(t)$ is the population of prey at time t and $y(t)$ the population of adult predators. The given initial functions $x_0(t)$ and $y_0(t)$ are nonnegative and continuous on $\alpha \leq t < 0$, and $x(0), y(0) > 0$. The given constant r is the specific growth rate of the prey, K is its caring capacity, and the (given) function $p(x)$ is the adult predators' functional response. The parameters b and d are the adult predators' birth and death rates, respectively. In this model the delay function $\tau(t, x_t)$ which depends on the past history $x(s)$, $s \leq t$, of population of prey is determined from the threshold condition

$$P\Big(t, x_t, \tau(t, x_t)\Big) := \int_{t-\tau(t,x_t)}^{t} p\Big(x(s)\Big)ds = m, \tag{4.2}$$

where $m > 0$ is a given threshold and $p(x)$ is a given differentiable and strictly increasing function.

Numerical approximations to the populations of prey $x(t)$ and predators $y(t)$, and the corresponding unknown delay function $\tau(t, x_t)$ are plotted in Figure 1. The model parameters are $r = 1$, $K = 1$, $b = 10$, $d = 0.5$, $d_j = 1$, $m = 0.2$ and $p(x) = x/(1+0.5x)$. The results were obtained by Euler's method applied with $h = 0.01$ to (4.1) and the composite trapezoidal rule combined with the method of bisection applied to (4.2). For the composite trapezoidal rule we applied 100 grid points. We applied the tolerance of 10^{-3} for the method of bisection. There were no more than 7 iterations per each step. The time of integration was $2.03 \cdot 10^3$ sec.

Example 2. Hoppensteadt and Waltman [7] study a model for the spread of infection. The model is written in the form

$$\begin{cases} S'(t) = -r(t)I(t)S(t),\ t \geq 0, \\ S(0) = S_0, \end{cases} \tag{4.3}$$

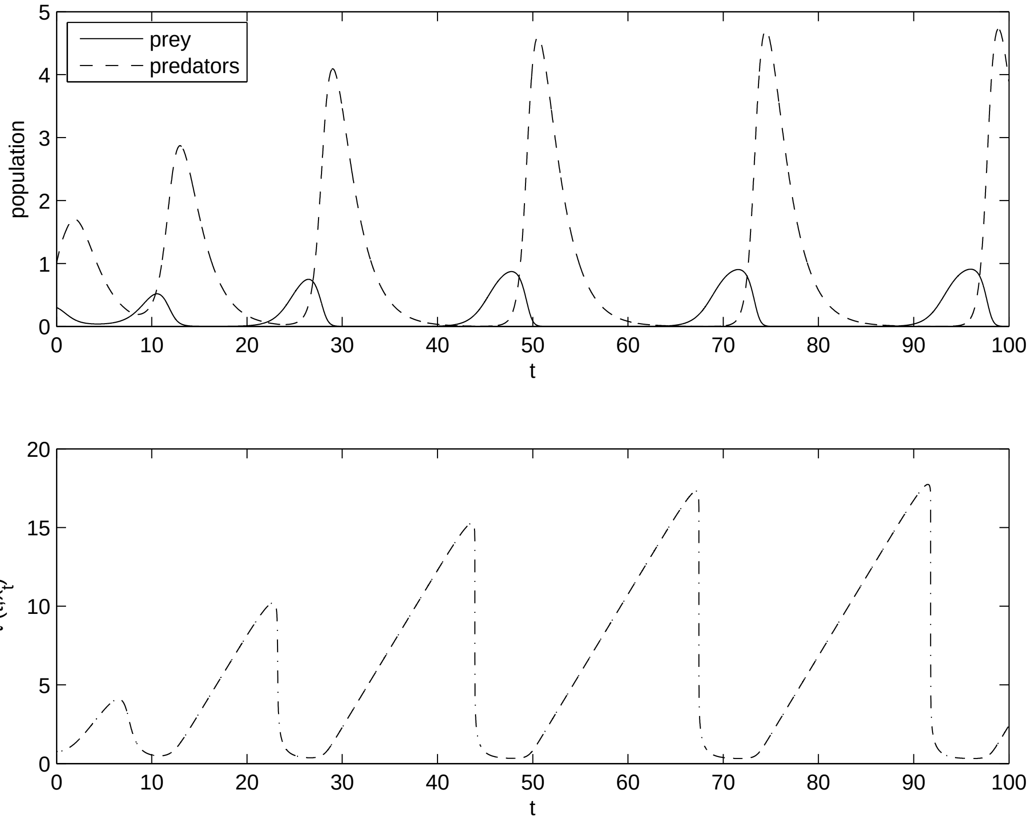

Fig. 1. The populations of prey $x(t)$ and predators $y(t)$, and the corresponding delay function $\tau(t, x_t)$.

with

$$I(t) = \begin{cases} I_0(t), & -\sigma \le t \le t_0, \\ I_0(t) + S_0 - S\Big(\tau(t, I_t)\Big), & t_0 \le t \le t_0 + \sigma, \\ S\Big(\tau(t-\sigma, I_{t-\sigma})\Big) - S\Big(\tau(t, I_t)\Big), & t_0 + \sigma \le t < \infty \end{cases} \tag{4.4}$$

and

$$I_0(t) = \begin{cases} I_0(0) - I_0(t-\sigma), & 0 \le t \le \sigma, \\ 0, & \sigma < t < \infty. \end{cases}$$

Here, $I(t)$ is the number of infectives and $S(t)$ is the number of susceptibles at time $t \ge 0$ in a certain constant population. The number of infectives $I_0(t)$ for $t < 0$ is known and satisfies the conditions $I_0(-\sigma) = 0$ and $I_0(0) = I_0$. Moreover, $r(t) > 0$ and $\rho(t) > 0$ are known proportionality functions and $0 < t_0 < \sigma$ is a unique time which satisfies the threshold condition

$$\int_0^{t_0} \rho(s) I_0(s) ds = m.$$

The function $\tau(t, I_t)$ appearing in (4.4) is unknown and determined from the threshold condition

$$P\Big(t, I_t, \tau(t, I_t)\Big) := \int_{\tau(t,I_t)}^{t} \rho(s)I(s)ds = m. \tag{4.5}$$

This model was numerically solved before by Hoppenstedat and Jackiewicz [6] using the differential form of the threshold condition

$$\tau'(t) = \frac{\rho(t)I(t)}{\rho(\tau(t))I(\tau(t))}, \quad \tau(t_0) = 0, \quad t \geq t_0, \tag{4.6}$$

and by Thompson and Shampine [10] where the threshold time was determined automatically by using an event function of Matlab `dde23` solver, see [9], for delay differential equations. The numerical method presented in this paper deals directly with integral form of the threshold condition (4.5) by fixed point iterations.

The approach proposed in this paper is more general than that considered before in [6] since it does not require that the delay function $\tau(t)$ is differentiable. We can observe that the delay function corresponding to Example 1 has sharp gradients and the corresponding equation (4.6) for $\tau'(t)$ would be stiff.

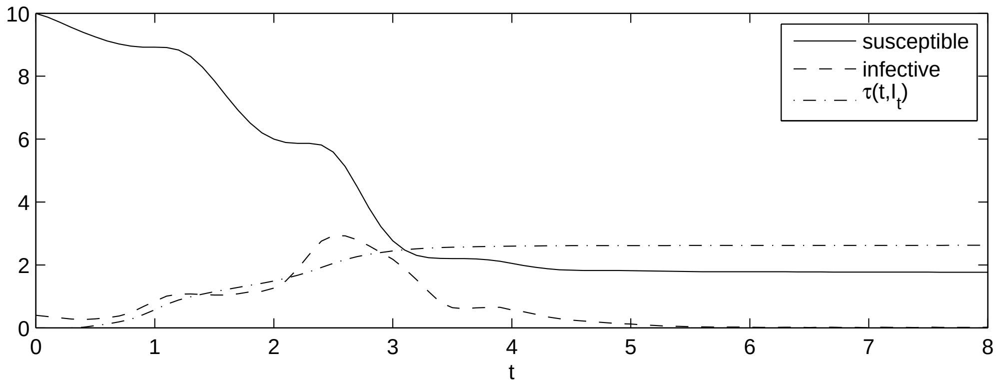

Fig. 2. Infective $I(t)$ and susceptible $S(t)$ population with the corresponding delay function $\tau(t, I_t)$.

Example #	Example 1	Example 2
h	err	err
10^{-1}	$2.84 \cdot 10^{-2}$	$4.02 \cdot 10^{-2}$
10^{-2}	$3.09 \cdot 10^{-3}$	$2.41 \cdot 10^{-3}$
10^{-3}	$3.08 \cdot 10^{-4}$	$2.19 \cdot 10^{-4}$
10^{-4}	$2.81 \cdot 10^{-5}$	$1.97 \cdot 10^{-5}$

Table 1
Errors for Example 1 and 2 at $t = 1$.

Numerical approximations to the infective population $I(t)$, the susceptible population $S(t)$, and the unknown delay function $\tau\Big(t, I_t\Big)$, which solve problem (4.3)-(4.4) with the given function

$$I_0(t) = \begin{cases} 0.4(1+t), \ -1 \le t \le 0, \\ 0.4(1-t), \ 0 \le t \le 1, \\ 0, \qquad\qquad \text{otherwise}, \end{cases}$$

are plotted in Figure 2. The parameter values used for Figure 2 are: $m = 0.1$, $\sigma = 1$, $S_0 = 10$, $t_0 \approx 0.357403$ and the given functions are $\rho(t) = \exp(-t)$ and $r(t) = r_0\Big(1 + \sin(5t)\Big)$ with $r_0 = 0.3$. The results were obtained by Eulers' method applied to (4.3) with the step-size $h = 0.1$, the composite trapezoidal rule applied to the integrals from (4.5), and the method of bisection to find a numerical approximation to the equation (4.5). For the composite trapezoidal rule we applied 20 grid points. We applied the tolerance of 10^{-3} for the method of bisection. There were no more than 13 iterations per each step. The time of integration was 0.76 sec.

The order of the convergence is presented in Table 1 for Example 1 and 2. The errors listed in the table were computed using reference solutions, which were computed with $h = 10^{-5}$, the tolerance 10^{-7} for the method of bisection, and 1000 grid points for the the composite trapezoidal rule. The errors were computed with the corresponding parameters h listed in the first column and with the same tolerance and numbers of grid points as for the reference solutions. Table 1 confirms the order of Euler's method used for the integration in time of both problems.

5 Concluding remarks

We investigated numerical errors of discrete numerical methods for the threshold problems which are applied in the theory of epidemics and population dynamics. Since the problems include systems of delay differential equations and threshold conditions which cannot be solved separately, our error bounds have three components. The first component corresponds with the errors of solving the delay systems, the second component corresponds with the errors from the threshold condition and third component corresponds with the errors in delays. We made numerical experiments for two models: one from the theory of epidemics and other from population dynamics. Our numerical experiments with Euler's method confirm our theoretical estimations.

Future work will address the design of codes for thereshold problems based on adaptations of continuous Runge-Kutta methods [2] and general linear methods in Nordsieck form [3], [4] for ordinary differential equations.

Acknowledgements. The authors wish to express their gratitude to anonymouns referees for their helpful comments.

References

[1] W.G. Aiello and H.I. Freedman, A time-delay model of single species growth with stage structure, Math. Biosci. 101(1990), 139–153.

[2] A. Bellen and M. Zennaro, Numerical Methods for Delay Diffrential Equations, Oxford Science Publications, Clarendon Press, Oxford 2003.

[3] J.C. Butcher, P. Chartier and Z. Jackiewicz, Nordsieck representation of DIMSIMs, Numerical Algorithms 16(1997), 209–230.

[4] J.C. Butcher, P. Chartier and Z. Jackiewicz, Experiments with a variable-order type 1 DIMSIM code, Numerical Algorithms 22(1999), 237–261.

[5] S.A. Gourley and Y. Kuang, A stage structured predator-prey model and its dependence on maturation delay and death rate, J. Math. Biol. 49(2004), 188–200.

[6] F.C. Hoppensteadt and Z. Jackiewicz, Numerical solution of a problem in the theory of epidemics, Appl. Numer. Math. 56(2006), 533–543.

[7] F.C. Hoppensteadt and P. Waltman, A problem in the theory of epidemics, Math. Biosci. 9(1970), 71–91.

[8] H.L. Smith and Y. Kuang, Periodic solutions of differential delay equations with threshold-type delays, Contemp. Math. 129(1992), 153–176.

[9] L.F. Shampine, I. Gladwell and S. Thompson, Solving ODEs with Matlab, Cambridge University Press, Cambridge 2003.

[10] S. Thompson and L.F. Shampine, A note on Zdzislaw's epidemic model, `http://www.radford.edu/~thompson/ffddes/Infection/infection.pdf`