Adv Comput Math (2022) 48:76
https://doi.org/10.1007/s10444-022-09992-w

A new framework for polynomial approximation to differential equations

Luigi Brugnano[1] · **Gianluca Frasca-Caccia**[2] · **Felice Iavernaro**[3] · **Vincenzo Vespri**[1]

Received: 15 May 2022 / Accepted: 18 October 2022

Abstract

In this paper, we discuss a framework for the polynomial approximation to the solution of initial value problems for differential equations. The framework is based on an expansion of the vector field along an orthonormal basis, and relies on perturbation results for the considered problem. Initially devised for the approximation of ordinary differential equations, it is here further extended and, moreover, generalized to cope with constant delay differential equations. Relevant classes of Runge-Kutta methods can be derived within this framework.

Keywords Ordinary differential equations · Delay differential equations · Orthogonal polynomials · Local Fourier expansion · Polynomial approximations · Runge-Kutta methods

Mathematics Subject Classification (2010) 65L05 · 65L03 · 65L06 · 65P10

Communicated by: Martin Stynes

✉ Felice Iavernaro
felice.iavernaro@uniba.it

Luigi Brugnano
luigi.brugnano@unifi.it

Gianluca Frasca-Caccia
gfrascacaccia@unisa.it

Vincenzo Vespri
vincenzo.vespri@unifi.it

[1] Università di Firenze, Florence, Italy

[2] Università di Salerno, Salerno, Italy

[3] Università di Bari, Bari, Italy

Published online: 14 November 2022

1 Introduction

In this paper, we shall deal with the definition of a framework to discuss polynomial approximations to the solution of initial value problems for ordinary differential equations (ODEs),

$$\dot{y}(t) = f(t, y(t)), \qquad t \in [t_0, T], \qquad y(t_0) = y_0 \in \mathbb{R}^m, \tag{1}$$

and delay differential equations (DDEs) in the form,

$$\begin{aligned} \dot{y}(t) &= f(t, y(t), y(t-\tau)), \qquad t \in [t_0, T], \qquad y(t_0) = y_0, \\ y(t) &= \phi(t), \qquad\qquad\qquad\quad t \in [t_0 - \tau, t_0), \end{aligned} \tag{2}$$

where $\tau > 0$ is a constant delay and, usually, $y_0 = \phi(t_0)$. In the sequel, we shall always assume that f and ϕ are suitably regular in their respective arguments. As is well known, the two problems are related in many ways but, at the same time, have quite different features, which reflect on their numerical solution. We refer, e.g., to the comprehensive monograph [27], concerning (1), and [5] (see also [20]) for (2).

In more detail, in this paper we shall fully develop a novel framework for deriving numerical methods for solving (1), which is then extended to cope with (2).

The framework we are interested in relies on a local expansion of the vector field in (1) along an orthonormal basis. Such basis will be, in the present case, the Legendre polynomial basis $\{P_j\}_{j\geq 0}$:

$$P_j \in \Pi_j, \qquad \int_0^1 P_i(x) P_j(x) \mathrm{d}x = \delta_{ij}, \qquad i, j = 0, 1, \ldots, \tag{3}$$

where, as is usual, Π_j is the vector space of polynomials of degree j, and δ_{ij} is the Kronecker symbol. The idea is actually not new: early use of this approach are, for example, Hulme [29, 30], Bottasso [7], and Betsch and Steinmann [6]; it is also at the basis of the energy-conserving class of Runge-Kutta methods named HBVMs [12] (see also the monograph [9] and the review paper [10]).

The approach that we shall pursue has been initially devised in [14], where the target was problem (1), and its potentialities have been disclosed by using HBVMs as spectral methods in time for efficiently solving highly oscillatory problems [19] and, subsequently, Hamiltonian PDEs [11]. A corresponding error analysis is given in [3]. Moreover, this allows deriving a formulation of HBVMs as continuous-stage Runge-Kutta methods [1, 2].

Starting from this background, in this work we carry out a complete perturbation analysis of problems (1) and (2), and set up a unique and comprehensive framework to deal with the numerical solution of both problems by exploiting the same discretization procedure. In more detail, the truncated Fourier series may be interpreted as a projection of the differential problem onto a finite dimensional vector space, leading to a new, numerically easy-to-handle, differential problem. This latter may be regarded as a perturbation of the original one, so that the perturbation analysis turns out to be crucial to understand how the solutions of the two problems are related. At the best of our knowledge, the perturbation results for problem (2) are new, and provide a powerful general tool of analysis. That the same framework may cover problems of different nature constitute, in our opinion, a specific advancement in this

field, and reveals its potentialities to deal with other classes of problems (which will be the subject of future investigations).

With this premise, the paper is organized as follows: Section 2 concerns the result pertaining to the ODE case; Section 3 contains the corresponding results for the DDE case; Section 4 contains some numerical tests for the DDE case, involving methods which are new in this setting; at last, some concluding remarks and possible developments are reported in Section 5.

2 The ODE case

Without loss of generality, we shall consider problem (1) in the simpler form:

$$\dot{y}(t) = f(y(t)), \qquad t \in [t_0, T], \qquad y(t_0) = y_0 \in \mathbb{R}^m. \tag{4}$$

Having fixed the mesh

$$t_n = t_0 + nh, \qquad n = 0, \ldots, N, \qquad h = \frac{T - t_0}{N}, \tag{5}$$

we formally set, for $n = 1, \ldots, N$:

$$\hat{\sigma}_n(ch) := y(t_{n-1} + ch) \equiv \hat{\sigma}(t_{n-1} + ch), \qquad c \in [0, 1], \tag{6}$$

the restriction of the solution of problem (4) to the time interval $[t_{n-1}, t_n]$ (the function $\hat{\sigma}(t) \equiv y(t)$ is introduced for notational purposes). Consequently, $\hat{\sigma}_n$ satisfies the differential equation

$$\dot{\hat{\sigma}}_n(ch) = \sum_{j \geq 0} P_j(c)\gamma_j(\hat{\sigma}_n), \qquad c \in [0, 1], \qquad \hat{\sigma}_n(0) = y(t_{n-1}), \tag{7}$$

so that,

$$\hat{\sigma}_n(ch) = y(t_{n-1}) + h \sum_{j \geq 0} \int_0^c P_j(x)\mathrm{d}x\, \gamma_j(\hat{\sigma}_n), \qquad c \in [0, 1], \tag{8}$$

and, by virtue of (3),

$$y(t_n) = y(t_{n-1}) + h\gamma_0(\hat{\sigma}_n) \equiv \hat{\sigma}(t_n), \tag{9}$$

where, in general, for any suitably regular function $z : [0, h] \to \mathbb{R}^m$,

$$\gamma_j(z) := \int_0^1 P_j(\zeta) f(z(\zeta h))\mathrm{d}\zeta. \tag{10}$$

We now look for a piecewise polynomial approximation $\sigma(t)$, to the solution of (4), such that, setting for $n = 1, \ldots, N$,

$$\sigma_n(ch) \equiv \sigma(t_{n-1} + ch), \qquad c \in [0, 1], \tag{11}$$

its restriction to the time interval $[t_{n-1}, t_n]$, $\sigma_n \in \Pi_s$ and satisfies the differential equation:

$$\dot{\sigma}_n(ch) = \sum_{j=0}^{s-1} P_j(c)\gamma_j(\sigma_n), \qquad c \in [0,1], \qquad \sigma_n(0) = y_{n-1}, \tag{12}$$

obtained by truncating the infinite series in (7) to a finite sum, with

$$y_n := \sigma_n(h) \equiv \sigma(t_n). \tag{13}$$

Consequently, σ_n can be formally written as:

$$\sigma_n(ch) = y_{n-1} + h\sum_{j=0}^{s-1} \int_0^c P_j(x)\mathrm{d}x\, \gamma_j(\sigma_n), \qquad c \in [0,1], \tag{14}$$

and (compare with (9)),

$$y_n = y_{n-1} + h\gamma_0(\sigma_n). \tag{15}$$

2.1 Preliminary results

We here provide a few preliminary results, which will be needed to derive the main ones in the following subsections. Some of them are taken from [14] but we also report them here, for sake of completeness.

Theorem 1 *Let $G : [0, h] \to V$, with V a vector space, admit a Taylor expansion at 0. Then, for all $j \geq 0$:*

$$\int_0^1 P_j(\zeta)G(\zeta h)\mathrm{d}\zeta = O(h^j).$$

Proof By virtue of (3), one has:

$$\begin{aligned}\int_0^1 P_j(\zeta)G(\zeta h)\mathrm{d}\zeta &= \int_0^1 P_j(\zeta)\sum_{k\geq 0}\frac{G^{(k)}(0)}{k!}(\zeta h)^k\mathrm{d}\zeta\\ &= \sum_{k\geq 0}\frac{G^{(k)}(0)}{k!}h^k\int_0^1 P_j(\zeta)\zeta^k\mathrm{d}\zeta\\ &= \sum_{k\geq j}\frac{G^{(k)}(0)}{k!}h^k\int_0^1 P_j(\zeta)\zeta^k\mathrm{d}\zeta = O(h^j).\end{aligned}$$

□

As a straightforward consequence, setting $G(\zeta h) := f(z(\zeta h))$, the following result is proved.

Corollary 1 *With reference to (10), one has: $\gamma_j(z) = O(h^j)$.*

Let us denote by

$$y(t) \equiv y(t, \xi, \eta) \tag{16}$$

the solution of the problem (compare with (4)):

$$\dot{y}(t) = f(y(t)), \qquad t \in [\xi, T], \qquad y(\xi) = \eta \in \mathbb{R}^m. \tag{17}$$

Hereafter, for sake of brevity, we may use either one of the two notations in (16), depending on the needs. The following theorem contains standard perturbation results w.r.t. all the arguments (see, e.g., [27, Section I.14]).

Theorem 2 *With reference to the solution* (16) *of problem* (17)*, one has:*

$$a)\ \frac{\partial}{\partial t} y(t) = f(y(t)), \qquad b)\ \frac{\partial}{\partial \eta} y(t) = \Phi(t, \xi), \qquad c)\ \frac{\partial}{\partial \xi} y(t) = -\Phi(t, \xi) f(\eta),$$

where $\Phi(t, \xi)$ *is the solution of the variational problem*

$$\dot{\Phi}(t, \xi) = F(y(t))\Phi(t, \xi), \qquad t \in [\xi, T], \qquad \Phi(\xi, \xi) = I \in \mathbb{R}^{m \times m},$$

having set

$$F(y) = \frac{\partial}{\partial y} f(y). \tag{18}$$

From this theorem, the following result readily follows where, hereafter, $|\cdot|$ will denote any convenient vector norm.

Corollary 2 *With reference to* (17)*, and assuming that* $\xi \in [t_{n-1}, t_n]$*, one has:*

$$y(t, \xi, \eta + \delta\eta) = y(t, \xi, \eta) + \Phi(t, \xi)\delta\eta + (t - \xi)O(|\delta\eta|^2), \qquad t \in [t_{n-1}, t_n].$$

2.2 Main results (ODE case)

With reference to (5)–(15), we are now in the position of stating the results concerning the approximation error at the grid points,

$$y(t_n) - y_n \equiv \hat{\sigma}_n(h) - \sigma_n(h), \qquad n = 1, \ldots, N, \tag{19}$$

and, more in general, on each subinterval $[t_{n-1}, t_n]$:

$$\delta\sigma_n(ch) := \hat{\sigma}_n(ch) - \sigma_n(ch) \equiv \hat{\sigma}(t_{n-1} + ch) - \sigma(t_{n-1} + ch), \qquad c \in [0, 1]. \tag{20}$$

For the first step of the approximation procedure, the following theorem holds true, the proof being similar to that of [14, Theorem 1].

Theorem 3 *With reference to* (19) *and* (20)*, one has:*

$$y(t_1) - y_1 = O(h^{2s+1}), \qquad \|\delta\sigma_1\| := \max_{c \in [0,1]} |\hat{\sigma}_1(ch) - \sigma_1(ch)| = O(h^{s+1}).$$

Proof By virtue of Corollary 1 and Theorem 2, one has:

$$\begin{aligned}
\hat{\sigma}_1(ch) - \sigma_1(ch) &= y(t_0 + ch, t_0, y_0) - y(t_0 + ch, t_0 + ch, \sigma_1(ch)) \\
&= y(t_0 + ch, t_0, \sigma_1(0)) - y(t_0 + ch, t_0 + ch, \sigma_1(ch)) \\
&= \int_{ch}^{0} \frac{\mathrm{d}}{\mathrm{d}t} y(t_0 + ch, t_0 + t, \sigma_1(t)) \mathrm{d}t \\
&= \int_{ch}^{0} \left[\frac{\partial}{\partial \xi} y(t_0 + ch, \xi, \sigma_1(t)) \Big|_{\xi = t_0 + t} + \frac{\partial}{\partial \eta} y(t_0 + ch, t_0 + t, \eta) \Big|_{\eta = \sigma_1(t)} \dot{\sigma}_1(t) \right] \mathrm{d}t \\
&= \int_{0}^{ch} \Phi(t_0 + ch, t_0 + t) \left[f(\sigma_1(t)) - \dot{\sigma}_1(t) \right] \mathrm{d}t \\
&= h \int_{0}^{c} \Phi(t_0 + ch, t_0 + \zeta h) \left[f(\sigma_1(\zeta h)) - \dot{\sigma}_1(\zeta h) \right] \mathrm{d}\zeta \\
&= h \int_{0}^{c} \Phi(t_0 + ch, t_0 + \zeta h) \left[\sum_{j \ge 0} P_j(\zeta) \gamma_j(\sigma_1) - \sum_{j=0}^{s-1} P_j(\zeta) \gamma_j(\sigma_1) \right] \mathrm{d}\zeta \\
&= h \sum_{j \ge s} \Big[\int_{0}^{c} P_j(\zeta) \Phi(t_0 + ch, t_0 + \zeta h) \mathrm{d}\zeta \Big] \underbrace{\gamma_j(\sigma_1)}_{= O(h^j)}.
\end{aligned}$$

Consequently, the second part of the statement follows for $c \in (0, 1)$, whereas, when $c = 1$ one deduces, by virtue of Theorem 1:

$$\begin{aligned}
y(t_1) - y_1 &\equiv \hat{\sigma}_1(h) - \sigma_1(h) \\
&= h \sum_{j \ge s} \overbrace{\underbrace{\left[\int_0^1 P_j(\zeta) \underbrace{\Phi(t_1, t_0 + \zeta h)}_{=: G(\zeta h)} \mathrm{d}\zeta \right]}_{= O(h^j)} \gamma_j(\sigma_1)}^{= O(h^{2j})} = O(h^{2s+1}).
\end{aligned}$$

□

For the remaining steps, the following result holds true.

Theorem 4 *With reference to* (19) *and* (20)*, for* $n = 1, \ldots, N$ *one has:*

$$y(t_n) - y_n = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}), \quad \|\delta\sigma_n\| := \max_{c \in [0,1]} |\delta\sigma_n(ch)| = O(h^{s+1}).$$

Proof By induction on n. For $n = 1$ the statement follows from the previous Theorem 3. Assuming it true for $n - 1$, for n one has:

$$\begin{aligned}
\hat{\sigma}_n(ch) - \sigma_n(ch) &= y(t_{n-1} + ch, t_{n-1}, \hat{\sigma}_n(0)) - y(t_{n-1} + ch, t_{n-1} + ch, \sigma_n(ch)) \\
&= \underbrace{y(t_{n-1} + ch, t_{n-1}, \hat{\sigma}_n(0)) - y(t_{n-1} + ch, t_{n-1}, \sigma_n(0))}_{=: E_{n,1}(ch)} \\
&\quad + \underbrace{y(t_{n-1} + ch, t_{n-1}, \sigma_n(0)) - y(t_{n-1} + ch, t_{n-1} + ch, \sigma_n(ch))}_{=: E_{n,2}(ch)}.
\end{aligned}$$

By using similar arguments as those used in the proof of Theorem 3, one deduces that

$$E_{n,2}(ch) = \begin{cases} O(h^{s+1}), & c \in (0,1), \\ O(h^{2s+1}), & c = 1. \end{cases}$$

Moreover, considering that, by the induction hypothesis,

$$\delta\sigma_n(0) = \hat{\sigma}_n(0) - \sigma_n(0) = y(t_{n-1}) - y_{n-1} = (n-1)\, O(h^{2s+1}),$$

from Corollary 2, one has:

$$\begin{aligned} E_{n,1}(ch) &= \underbrace{\Phi(t_{n-1}+ch, t_{n-1})}_{=I+O(ch)} \delta\sigma_n(0) + ch\, O(|\delta\sigma_n(0)|^2) \\ &= y(t_{n-1}) - y_{n-1} + (n-1) O(ch^{2s+2}). \end{aligned}$$

Consequently, for $c = 1$ one obtains the first part of the statement, whereas the second part follows by taking $c \in (0,1)$. □

Remark 1 We observe that the two equivalent equations (see (10), (12), and (14)):

$$\dot{\sigma}_n(ch) = \sum_{j=0}^{s-1} P_j(c) \int_0^1 P_j(\zeta) f(\sigma_n(\zeta h)) \mathrm{d}\zeta, \qquad c \in [0,1], \qquad \sigma_n(0) = y_{n-1},$$

and

$$\sigma_n(ch) = y_{n-1} + h \sum_{j=0}^{s-1} \int_0^c P_j(x) \mathrm{d}x \int_0^1 P_j(\zeta) f(\sigma_n(\zeta h)) \mathrm{d}\zeta, \qquad c \in [0,1], \quad (21)$$

define a so-called HBVM(∞, s) method on the interval $[t_{n-1}, t_n]$ (equation (21) is named *Master Functional Equation* in [12]. See also [9, 10]). Consequently, such method defines an order $2s$ approximation procedure for all $s \geq 1$, which can be also recast as a continuous-stage Runge-Kutta method [1]. In particular, the case $s = 1$ corresponds to the so-called AVF method [40]; the case $s \geq 1$ has been also considered in [26].

An interesting question concerns the difference between the Fourier coefficients of the solution (8)–(10) and those of the polynomial approximation (14) on the interval $[t_{n-1}, t_n]$. The next result clarifies the issue.

Theorem 5 *With reference to* (8), (10), *and* (14), *for all* $n = 1, \dots, N$ *one has:*

$$\delta\gamma_j^n := \gamma_j(\hat{\sigma}_n) - \gamma_j(\sigma_n) = O(h^{2s-j}), \qquad j = 0, \dots, s-1.$$

Proof First of all, from (3), (8), (14), and Theorem 4 we know that:

$$y(t_n) - y_n = y(t_{n-1}) - y_{n-1} + h[\gamma_0(\hat{\sigma}_n) - \gamma_0(\sigma_n)] = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}).$$

Consequently, from the last equality one derives:

$$\delta\gamma_0^n = \gamma_0(\hat{\sigma}_n) - \gamma_0(\sigma_n) = O(h^{2s}).$$

Further, by taking into account (18) and (20), one obtains:

$$O(h^{2s}) = \gamma_0(\hat{\sigma}_n) - \gamma_0(\sigma_n) = \int_0^1 \left[f(\hat{\sigma}_n(\zeta h)) - f(\sigma_n(\zeta h)) \right] \mathrm{d}\zeta$$
$$= \int_0^1 \underbrace{\int_0^1 F(\sigma_n(\zeta h) + c\,\delta\sigma_n(\zeta h))\mathrm{d}c}_{=: G(\zeta h)} \delta\sigma_n(\zeta h)\mathrm{d}\zeta = \int_0^1 G(\zeta h)\delta\sigma_n(\zeta h)\mathrm{d}\zeta.$$

Now, considering that $P_0(x) \equiv 1$ and, for all $\zeta \in [0, 1]$,

$$\int_0^\zeta P_j(x)\mathrm{d}x = \xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta), \qquad j \geq 1, \tag{22}$$
$$with \qquad \xi_j = \left(2\sqrt{4j^2 - 1} \right)^{-1},$$

one has:

$$\delta\sigma_n(\zeta h) = \hat{\sigma}_n(\zeta h) - \sigma_n(\zeta h)$$
$$= y(t_{n-1}) - y_{n-1} + h \sum_{j=0}^{s-1} \int_0^\zeta P_j(x)\mathrm{d}x\, \delta\gamma_j^n + h \sum_{j\geq s} \int_0^\zeta P_j(x)\mathrm{d}x\, \gamma_j(\hat{\sigma}_n)$$
$$= y(t_{n-1}) - y_{n-1} + \zeta h \delta\gamma_0^n + h \sum_{j=1}^{s-1} \left[\xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta) \right] \delta\gamma_j^n$$
$$+ h \sum_{j\geq s} \left[\xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta) \right] \gamma_j(\hat{\sigma}_n).$$

Consequently, from Theorem 1 and Corollary 1, one obtains:

$$O(h^{2s}) = \int_0^1 G(\zeta h)\delta\sigma_n(\zeta h)\,\mathrm{d}\zeta = \underbrace{\int_0^1 G(\zeta h)\mathrm{d}\zeta}_{=O(1)} \underbrace{[y(t_{n-1}) - y_{n-1}]}_{=(n-1)\,O(h^{2s+1})}$$
$$+ h \underbrace{\int_0^1 G(\zeta h)\zeta\mathrm{d}\zeta}_{=O(1)} \underbrace{\delta\gamma_0^n}_{=O(h^{2s})} + h \sum_{j=1}^{s-1} \int_0^1 G(\zeta h) \left[\xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta) \right] \mathrm{d}\zeta\, \delta\gamma_j^n$$
$$+ \underbrace{h \sum_{j\geq s} \overbrace{\int_0^1 G(\zeta h) \left[\xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta) \right] \mathrm{d}\zeta}^{=O(h^{j-1})} \overbrace{\gamma_j(\hat{\sigma}_n)}^{=O(h^j)}}_{=O(h^{2s})},$$

from which,

$$O(h^{2s}) = h \sum_{j=1}^{s-1} \overbrace{\int_0^1 G(\zeta h) \left[\xi_{j+1} P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta) \right] \mathrm{d}\zeta}^{=O(h^{j-1})} \delta\gamma_j^n$$

follows and, therefore, one concludes that $\delta\gamma_j^n = O(h^{2s-j})$, $j = 0 \ldots, s-1$. □

2.3 Conservative/dissipative problems

An interesting case [9, 25, 36] is that when problem (4) is in the form

$$\dot{y}(t) = S\nabla H(y(t)), \qquad t \in [t_0, T], \qquad y(t_0) = y_0 \in \mathbb{R}^m, \tag{23}$$

with $S \in \mathbb{R}^{m\times m}$ either a skew-symmetric matrix, $S^\top = -S$, or a negative semidefinite matrix, $S \leq 0$, whereas ∇H is the gradient of a scalar function usually called the *Hamiltonian*. As is clear:

- when $S^\top = -S$:
$$\frac{\mathrm{d}}{\mathrm{d}t} H(y(t)) = \nabla H(y(t))^\top \dot{y}(t) = \nabla H(y(t))^\top S \nabla H(y(t)) = 0,$$
so that H is a conserved quantity, and the problem is said to be *conservative*;
- when $S \leq 0$:
$$\frac{\mathrm{d}}{\mathrm{d}t} H(y(t)) = \nabla H(y(t))^\top \dot{y}(t) = \nabla H(y(t))^\top S \nabla H(y(t)) \leq 0,$$
and the problem is said to be *dissipative*.

The next result shows that this behavior is preserved by the approximations (12)–(15), upon observing that in this case (10) can be conveniently rewritten as

$$\gamma_j(z) = S\int_0^1 P_j(\zeta)\nabla H(z(\zeta h))\mathrm{d}\zeta =: S\beta_j(z). \tag{24}$$

Theorem 6 *With reference to* (12)–(15) *applied for approximating problem* (23)*, for all* $n = 1, \ldots, N$ *one has:*

- $H(y_n) = H(y_{n-1})$*, when* $S^\top = -S$*;*
- $H(y_n) \leq H(y_{n-1})$*, when* $S \leq 0$*.*

Proof In fact, by considering that $y_n = \sigma_n(h)$, $y_{n-1} = \sigma_n(0)$, and taking into account (24), one has:

$$\begin{aligned}
H(y_n) - H(y_{n-1}) = H(\sigma_n(h)) - H(\sigma_n(0)) &= \int_0^h \frac{\mathrm{d}}{\mathrm{d}t} H(\sigma_n(t))\mathrm{d}t \\
&= h\int_0^1 \nabla H(\sigma_n(ch))^\top \dot{\sigma}_n(ch)\mathrm{d}c = h\int_0^1 \nabla H(\sigma_n(ch))^\top \sum_{j=0}^{s-1} P_j(c)\gamma_j(\sigma_n)\mathrm{d}c \\
&= h\sum_{j=0}^{s-1} \underbrace{\left[\int_0^1 \nabla H(\sigma_n(ch)) P_j(c)\mathrm{d}c\right]^\top}_{=\beta_j(\sigma_n)^\top} S\beta_j(\sigma_n) \\
&= h\sum_{j=0}^{s-1} \beta_j(\sigma_n)^\top S\beta_j(\sigma_n) \;=:\; \Delta H_n.
\end{aligned}$$

Consequently, if $S^\top = -S$, then $\Delta H_n = 0$, whereas $\Delta H_n \leq 0$, when $S \leq 0$. □

Remark 2 According to Remark 1, one then obtains that HBVM(∞, s) methods can preserve the conservative/dissipative feature of problem (23).

2.4 Discretization and Runge-Kutta formulation

Quoting Dahlquist and Björk [22, p. 521] "as is well known, even many relatively simple integrals cannot be expressed in finite terms of elementary functions, and thus must be evaluated by numerical methods." In this context, this quite obvious statement means that the approximation procedure defined by (12) and (10) does not yet provide a "true" numerical method. In fact, the integrals defining the Fourier coefficients,

$$\gamma_j(\sigma_n) = \int_0^1 P_j(\zeta) f(\sigma_n(\zeta h)) \mathrm{d}\zeta, \qquad j = 0, \dots, s-1, \tag{25}$$

need to be numerically approximated by using a quadrature formula. Since we are dealing with a polynomial approximation, it is quite natural to do this by using an interpolatory quadrature with abscissae and weights (c_i, b_i), $i = 1, \dots, k$ (we shall always assume k distinct abscissae):

$$\gamma_j(\sigma_n) = \sum_{i=1}^{k} b_i P_j(c_i) f(\sigma_n(c_i h)) + \Delta_j(h), \tag{26}$$

where $\Delta_j(h)$ is the quadrature error. The following straightforward result holds true.

Theorem 7 *If the quadrature (c_i, b_i), $i = 1, \dots, k$ has order q, i.e., it is exact for polynomial integrands of degree $q-1$, then*

$$\Delta_j(h) = O(h^{q-j}), \qquad j = 0, \dots, s-1.$$

Remark 3 As is well known, since the quadrature (26) is based at k (distinct) abscissae, $q \in \{k, \dots, 2k\}$: the lower limit is obtained by a generic choice of the abscissae, whereas the upper one is achieved by placing them at the zeros of $P_k(c)$.

When using a quadrature, clearly the Fourier coefficients (25) may be not exactly evaluated anymore. This implies that we are actually computing a possibly different piecewise polynomial approximation $u(t)$ such that (compare with (10)–(15)), for all $n = 1, \dots, N$:

$$u_n(ch) \equiv u(t_{n-1} + ch), \qquad c \in [0, 1], \tag{27}$$

$$\dot{u}_n(ch) = \sum_{j=0}^{s-1} P_j(c) \hat{\gamma}_j(u_n), \qquad c \in [0, 1], \qquad u_n(0) = y_{n-1}, \tag{28}$$

with (see (26))

$$\hat{\gamma}_j(u_n) := \sum_{i=1}^{k} b_i P_j(c_i) f(u_n(c_i h)) \equiv \gamma_j(u_n) - \Delta_j(h), \quad j=0,\ldots,s-1, \tag{29}$$

$$u_n(ch) = y_{n-1} + h \sum_{j=0}^{s-1} \int_0^c P_j(x)\mathrm{d}x\, \hat{\gamma}_j(u_n), \qquad c \in [0,1], \tag{30}$$

$$u_n(h) =: y_n \equiv y_{n-1} + h\hat{\gamma}_0(u_n). \tag{31}$$

Actually, (29)–(31) define the nth integration step, by using a timestep h, performed with the k stage Runge-Kutta method having stages:

$$Y_i^n := u_n(c_i h), \qquad i = 1,\ldots,k. \tag{32}$$

In fact, evaluating (30) at the abscissae $c_1,\ldots,c_k$, and substituting in it the s approximate Fourier coefficients (29), one obtains, after rearranging terms,

$$Y_i^n = y_{n-1} + h \sum_{\ell=1}^{k} \underbrace{b_\ell \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, P_j(c_\ell)}_{=: a_{i\ell}} f(Y_\ell^n), \qquad i = 1,\ldots,n, \tag{33}$$

$$y_n = y_{n-1} + h \sum_{i=1}^{k} b_i f(Y_i^n). \tag{34}$$

In other words, we have derived the k-stage Runge-Kutta method with abscissae and weights (c_i, b_i), $i = 1,\ldots,k$, and Butcher matrix $A = \left(a_{i\ell}\right) \in \mathbb{R}^{k\times k}$. Next theorem puts the Butcher tableau in a more compact form [9].

Theorem 8 *The Butcher tableau of the Runge-Kutta method* (33)–(34) *is given by*

$$\begin{array}{c|c} \boldsymbol{c} & \mathcal{I}_s \mathcal{P}_s^\top \Omega \\ \hline & \boldsymbol{b}^\top \end{array} \tag{35}$$

where

$$\boldsymbol{b} = \begin{pmatrix} b_1 & \ldots & b_k \end{pmatrix}^\top, \qquad \boldsymbol{c} = \begin{pmatrix} c_1 & \ldots & c_k \end{pmatrix}^\top, \qquad \Omega = \mathrm{diag}(\boldsymbol{b}),$$

and

$$\mathcal{P}_s = \begin{pmatrix} P_0(c_1) & \ldots & P_{s-1}(c_1) \\ \vdots & & \vdots \\ P_0(c_k) & \ldots & P_{s-1}(c_k) \end{pmatrix}, \qquad \mathcal{I}_s = \begin{pmatrix} \int_0^{c_1} P_0(x)\mathrm{d}x & \ldots & \int_0^{c_1} P_{s-1}(x)\mathrm{d}x \\ \vdots & & \vdots \\ \int_0^{c_k} P_0(x)\mathrm{d}x & \ldots & \int_0^{c_k} P_{s-1}(x)\mathrm{d}x \end{pmatrix}.$$

It is possible to derive an alternative formulation of the Runge-Kutta method (35). In fact, using the relation (22) between the integrals of the Legendre polynomials and the polynomials themselves, and considering that

$$\int_0^c P_0(x)\mathrm{d}x = \xi_1 P_1(c) + \xi_0 P_0(c), \qquad \xi_0 = \frac{1}{2},$$

one has that $\mathcal{I}_s = \mathcal{P}_{s+1}\hat{X}_s$, where

$$\mathcal{P}_{s+1} = \begin{pmatrix} P_0(c_1) & \dots & P_s(c_1) \\ \vdots & & \vdots \\ P_0(c_k) & \dots & P_s(c_k) \end{pmatrix}, \qquad \hat{X}_s = \left(\begin{array}{cccc} \xi_0 & -\xi_1 & & \\ \xi_1 & 0 & \ddots & \\ & \ddots & \ddots & -\xi_{s-1} \\ & & \xi_{s-1} & 0 \\ \hline & & & \xi_s \end{array}\right) =: \left(\begin{array}{c} X_s \\ \hline 0 \ \dots \ 0 \ \xi_s \end{array}\right).$$

Consequently, the Butcher tableau (35) can be rewritten as

$$\begin{array}{c|c} \boldsymbol{c} & \mathcal{P}_{s+1}\hat{X}_s\mathcal{P}_s^\top\Omega \\ \hline & \boldsymbol{b}^\top \end{array}.$$

When the quadrature (26) has order $q \geq 2s$, it is quite straightforward to prove that

$$\mathcal{P}_s^\top\Omega\mathcal{P}_s = I_s, \qquad \mathcal{P}_s^\top\Omega\mathcal{P}_{s+1} = [I_s \ \mathbf{0}] \in \mathbb{R}^{s\times(s+1)},$$

where in general, hereafter, $I_r \in \mathbb{R}^{r\times r}$ is the identity matrix (when the dimension of the identity matrix is not explicitly indicated, it will be easily deducible from the context). Consequently,

$$\mathcal{P}_s^\top\Omega\left[\mathcal{P}_{s+1}\hat{X}_s\mathcal{P}_s^\top\Omega\right]\mathcal{P}_s = X_s,$$

which can be regarded as a generalization of the W-transformation in [28, Theorem 5.6, p. 79]. In addition to this, when $q \geq 2s$ also the following results hold true (for sake of brevity, we do not discuss the case $q < 2s$, since it has no practical interest).

Theorem 9 *With reference to* (4)–(10) *and* (27)–(31)*, and assuming that the quadrature formula* (26) *has order* $q \geq 2s$*, one has:*

$$y(t_1) - y_1 = O(h^{2s+1}), \qquad \max_{c\in[0,1]} |\hat{\sigma}_1(ch) - u_1(ch)| = O(h^{s+1}).$$

Theorem 10 *With reference to* (4)–(10) *and* (27)–(31)*, and assuming that the quadrature formula* (26) *has order* $q \geq 2s$*, for* $n = 1, \dots, N$ *one has:*

$$y(t_n) - y_n = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}), \qquad \max_{c\in[0,1]} |\hat{\sigma}_n(ch) - u_n(ch)| = O(h^{s+1}).$$

Theorem 11 *With reference to* (8)–(10) *and* (29)–(31)*, and assuming that the quadrature formula* (26) *has order* $q \geq 2s$*, for all* $n = 1, \dots, N$ *one has:*

$$\delta\hat{\gamma}_j^n := \gamma_j(\hat{\sigma}_n) - \hat{\gamma}_j(u_n) = O(h^{2s-j}), \qquad j = 0, \dots, s-1.$$

Concerning the case of conservative/dissipative problems in the form (23), the result of Theorem 6 modifies as follows.

Theorem 12 *With reference to* (27)–(31) *applied for approximating problem* (23)*, and assuming that the quadrature formula* (26) *has order* $q \geq 2s$*, for all* $n = 1, \dots, N$ *one has:*

- *if H is a polynomial of degree not larger than q/s, then the result of Theorem 6 continues to hold;*
- *differently,*
  - $H(y_n) = H(y_{n-1}) + O(h^{q+1})$, *when* $S^\top = -S$,
  - $H(y_n) \leq H(y_{n-1}) + O(h^{q+1})$, *when* $S \leq 0$.

We here provide only the proof of Theorem 9 (see also [14, Theorem 4]), since those of Theorems 10, 11, and 12 can be similarly obtained by slightly adapting the corresponding proofs of Theorems 4, 5, and 6, respectively.

Proof (of Theorem 9) By taking into account the result of Theorem 7, one has:

$$
\begin{aligned}
&\hat{\sigma}_1(ch) - u_1(ch) = y(t_0 + ch, t_0, y_0) - y(t_0 + ch, t_0 + ch, u_1(ch)) \\
&= y(t_0 + ch, t_0, u_1(0)) - y(t_0 + ch, t_0 + ch, u_1(ch)) \\
&= \int_{ch}^{0} \frac{\mathrm{d}}{\mathrm{d}t} y(t_0 + ch, t_0 + t, u_1(t)) \mathrm{d}t \\
&= \int_{ch}^{0} \left[\frac{\partial}{\partial \xi} y(t_0 + ch, \xi, u_1(t)) \Big|_{\xi = t_0 + t} + \frac{\partial}{\partial \eta} y(t_0 + ch, t_0 + t, \eta) \Big|_{\eta = u_1(t)} \dot{u}_1(t) \right] \mathrm{d}t \\
&= \int_{0}^{ch} \Phi(t_0 + ch, t_0 + t) \left[f(u_1(t)) - \dot{u}_1(t) \right] \mathrm{d}t \\
&= h \int_{0}^{c} \Phi(t_0 + ch, t_0 + \zeta h) \left[f(u_1(\zeta h)) - \dot{u}_1(\zeta h) \right] \mathrm{d}\zeta \\
&= h \int_{0}^{c} \Phi(t_0 + ch, t_0 + \zeta h) \left[\sum_{j \geq 0} P_j(\zeta) \gamma_j(u_1) - \sum_{j=0}^{s-1} P_j(\zeta) \hat{\gamma}_j(u_1) \right] \mathrm{d}\zeta \\
&= h \int_{0}^{c} \Phi(t_0 + ch, t_0 + \zeta h) \left[\sum_{j \geq 0} P_j(\zeta) \gamma_j(u_1) - \sum_{j=0}^{s-1} P_j(\zeta) \left(\gamma_j(u_1) - \Delta_j(h) \right) \right] \mathrm{d}\zeta \\
&= h \sum_{j \geq s} \left[\int_{0}^{c} P_j(\zeta) \Phi(t_0 + ch, t_0 + \zeta h) \mathrm{d}\zeta \right] \underbrace{\gamma_j(u_1)}_{= O(h^j)} \\
&\quad + h \sum_{j=0}^{s-1} \left[\int_{0}^{c} P_j(\zeta) \Phi(t_0 + ch, t_0 + \zeta h) \mathrm{d}\zeta \right] \underbrace{\Delta_j(h)}_{= O(h^{q-j})} .
\end{aligned}
$$

Consequently, the second part of the statement follows by considering that, for $c \in (0, 1)$, this quantity is

$$
O(h^{s+1}) + O(h^{q-s+2}) = O(h^{s+1}),
$$

since $q \geq 2s$, whereas, when $c = 1$ one deduces, by virtue of Theorem 1, and considering that $t_1 = t_0 + h$:

$$y(t_1) - y_1 \equiv \hat{\sigma}_1(h) - u_1(h) = h \sum_{j \geq s} \overbrace{\underbrace{\left[\int_0^1 P_j(\zeta) \underbrace{\Phi(t_1, t_0 + \zeta h)}_{=: G(\zeta h)} \mathrm{d}\zeta \right]}_{= O(h^j)} \gamma_j(u_1)}^{= O(h^{2j})}$$

$$+ h \sum_{j=0}^{s-1} \overbrace{\underbrace{\left[\int_0^1 P_j(\zeta) \underbrace{\Phi(t_1, t_0 + \zeta h)}_{=: G(\zeta h)} \mathrm{d}\zeta \right]}_{= O(h^j)} \Delta_j(h)}^{= O(h^q)}$$

$$= O(h^{2s+1}) + O(h^{q+1}) = O(h^{2s+1}).$$

□

Remark 4 When the k abscissae are placed at the zeros of $P_k(c)$, and $k \geq s$, one obtains a HBVM(k, s) method, whose order is $2s$ [9, 10, 12]. It is worth mentioning that the HBVM(s, s) method is nothing but the s-stage Gauss-Legendre collocation method. Moreover, the HBVM$(k, 1)$ methods correspond to the second-order Runge-Kutta methods described in [21]. Different choices of the quadrature have been also considered in [15, 31–33].

2.5 Solving the discrete problems

Sometimes, the number of stages k of the Runge-Kutta method (35) can be much larger than the degree s of the underlying polynomial approximation (29)–(31). This is the case, for example, of HBVM(k, s) methods when used as energy-conserving methods [9, 10, 12] (see also Theorem 12 in Section 2.3). In such a case, it is clear that the usual implementation of the Runge-Kutta method leads to the solution of a discrete problem having (block) dimension k. Nevertheless, for sake of completeness we now recall how the discrete problem to be solved can be actually recast so as to have (block) dimension s, independently of k [13]. This clearly allows for relatively large values of k, thus making possible the use of an arbitrarily high-order quadrature (26). Let us then consider the first integration step of the method for solving (4) with timestep h, (thus, we can skip the index n of the step). Setting $\mathbf{1} = \left(1, \ldots, 1\right)^\top \in \mathbb{R}^k$, and Y the stage vector of (block) dimension k, one obtains that the stage equation for (35) is given by:

$$Y = \mathbf{1} \otimes y_0 + h \mathcal{I}_s \mathcal{P}_s^\top \Omega \otimes I_m \, f(Y), \tag{36}$$

with an obvious meaning of $f(Y)$. However, we observe that [13]

$$\mathcal{P}_s^\top \Omega \otimes I_m \, f(Y) =: \hat{\boldsymbol{\gamma}} \equiv \begin{pmatrix} \hat{\gamma}_0(u_1) \\ \vdots \\ \hat{\gamma}_{s-1}(u_1) \end{pmatrix},$$

i.e., the (block) vector with the s coefficients of the polynomial approximation $u_1(ch)$ (see (29)–(30)). Consequently, (36) can be rewritten as

$$Y = \mathbf{1} \otimes y_0 + h\mathcal{I}_s \otimes I_m \, \hat{\boldsymbol{\gamma}}.$$

By combining the last two equations one eventually obtains:

$$\hat{\boldsymbol{\gamma}} = \mathcal{P}_s^{\top} \Omega \otimes I_m \, f\left(\mathbf{1} \otimes y_0 + h\mathcal{I}_s \otimes I_m \, \hat{\boldsymbol{\gamma}}\right), \tag{37}$$

which is a discrete problem, equivalent to (36), having (block) dimension s, *independently* of k. Once this equation has been solved, the new approximation is derived, according to (31), as

$$y_1 = y_0 + h\hat{\gamma}_0(u_1).$$

It is also worth mentioning that very effective nonlinear iterations have been devised for solving (37) [8, 9, 13] (the most effective being that derived from the so-called *blended iteration* introduced in [17], see also [18]).

3 The DDE case

As for the ODE case, also for DDEs we shall consider, without loss of generality, the simpler problem

$$\begin{aligned} \dot{y}(t) &= f(y(t), y(t-\tau)), \qquad t \in [t_0, T], \qquad y(t_0) = y_0, \\ y(t) &= \phi(t), \qquad\qquad\quad t \in [t_0 - \tau, t_0), \end{aligned} \tag{38}$$

in place of (2) where, usually, $y_0 = \phi(t_0)$. Moreover, we shall suppose that both the timestep h defining the discrete mesh (5) and the width of the integration interval, $T - t_0$, are commensurable with the delay:

$$\tau = \nu h, \qquad T - t_0 = K\tau, \qquad K, \nu \in \mathbb{N}, \tag{39}$$

so that the discrete mesh is now given by:

$$t_n = t_0 + nh, \qquad n = -\nu, \ldots, N \equiv K\nu. \tag{40}$$

On one hand, similarly as done in the ODE case, let us denote, for notational purposes, by $\hat{\sigma}(t) \equiv y(t)$ the solution of (38), and

$$\hat{\sigma}_n(ch) := \hat{\sigma}(t_{n-1} + ch), \qquad c \in [0, 1], \qquad n = 1 - \nu, \ldots, N, \tag{41}$$

its restriction to the time interval $[t_{n-1}, t_n]$. Consequently,

$$\hat{\sigma}_n(ch) \equiv \phi(t_{n-1} + ch), \qquad c \in [0, 1], \qquad n = 1 - \nu, \ldots, 0, \tag{42}$$

whereas, for $n = 1, \ldots, N$, one has (compare with (7)–(10)):

$$\dot{\hat{\sigma}}_n(ch) = \sum_{j \geq 0} P_j(c)\gamma_j(\hat{\sigma}_n, \hat{\sigma}_{n-\nu}), \qquad c \in [0, 1], \qquad \hat{\sigma}_n(0) = y(t_{n-1}), \tag{43}$$

so that,

$$\hat{\sigma}_n(ch) = y(t_{n-1}) + h \sum_{j \geq 0} \int_0^c P_j(x)\mathrm{d}x \, \gamma_j(\hat{\sigma}_n, \hat{\sigma}_{n-\nu}), \qquad c \in [0, 1], \tag{44}$$

and

$$y(t_n) = y(t_{n-1}) + h\gamma_0(\hat{\sigma}_n, \hat{\sigma}_{n-\nu}) \equiv \hat{\sigma}(t_n), \tag{45}$$

where, in general, for any suitably regular functions $z, w : [0, h] \to \mathbb{R}^m$,

$$\gamma_j(z, w) := \int_0^1 P_j(\zeta) f(z(\zeta h), w(\zeta h)) \mathrm{d}\zeta. \tag{46}$$

On the other hand, we shall look for a piecewise approximation to $\hat{\sigma}(t)$, i.e., $\sigma(t)$, such that (compare with (11)–(15))

$$\sigma_n(ch) := \sigma(t_{n-1} + ch), \qquad c \in [0, 1], \qquad n = 1 - \nu, \ldots, N, \tag{47}$$

denotes its restriction to the time interval $[t_{n-1}, t_n]$. Consequently, one has:

$$\sigma_n(ch) \equiv \phi(t_{n-1} + ch), \qquad c \in [0, 1], \qquad n = 1 - \nu, \ldots, 0, \tag{48}$$

whereas, for $n = 1, \ldots, N$, $\sigma_n \in \Pi_s$ satisfies the differential equation

$$\dot{\sigma}_n(ch) = \sum_{j=0}^{s-1} P_j(c)\gamma_j(\sigma_n, \sigma_{n-\nu}), \qquad c \in [0, 1], \qquad \sigma_n(0) = y_{n-1}, \tag{49}$$

so that,

$$\sigma_n(ch) = y_{n-1} + h \sum_{j=0}^{s-1} \int_0^c P_j(x)\mathrm{d}x\, \gamma_j(\sigma_n, \sigma_{n-\nu}), \qquad c \in [0, 1], \tag{50}$$

and

$$y_n = y_{n-1} + h\gamma_0(\sigma_n, \sigma_{n-\nu}) =: \sigma_n(h), \tag{51}$$

with $\gamma_j(\sigma_n, \sigma_{n-\nu})$ defined according to (46). In the sequel, we shall discuss the accuracy of the approximations:

$$\begin{aligned} y(t_n) - y_n \;&\equiv\; \hat{\sigma}_n(h) - \sigma_n(h), \\ \delta\sigma_n(ch) \;&:=\; \hat{\sigma}_n(ch) - \sigma_n(ch), \quad c \in (0, 1), \qquad n = 1, \ldots, N. \end{aligned} \tag{52}$$

For this purpose, some preliminary results are given in the next section.

3.1 Preliminary results

We start with the generalization of Corollary 1 to the present setting.

Corollary 3 *With reference to* (46)*, one has:* $\gamma_j(z, w) = O(h^j)$.

Proof Immediate from Theorem 1, by setting $G(\zeta h) := f(z(\zeta h), w(\zeta h))$. □

We also need perturbation results corresponding to those of Theorem 2 for ODEs. For this purpose, it is sufficient to discuss them for a local problem defined on two contiguous time subintervals of width τ: the former containing the *memory*, the latter containing the solution to be computed. Without loss of generality, we shall then fix

the reference interval $[t_0 - \tau, t_0 + \tau]$, where we consider the following problem, defined for a generic $\xi \in [t_0, t_0 + \tau]$:

$$\begin{aligned} \dot{y}(t) &= f(y(t), y(t-\tau)), \qquad t \in [t_0, t_0 + \tau], \qquad y(\xi) = \eta \in \mathbb{R}^m, \\ y(t) &= \phi(t), \qquad\qquad\quad t \in [t_0 - \tau, t_0). \end{aligned} \tag{53}$$

Problem (53) defines a generalization of the localized one associated to (38) (obtained for $\xi = t_0$ and $\eta = y_0$), and we shall denote its solution by

$$y(t) \equiv y(t, \xi, \eta, \phi; t_0), \tag{54}$$

in order to emphasize its dependence on the first four parameters, whereas the last one refers to the time subinterval. We shall also use the following notation:

$$F_1(z, w) = \frac{\partial}{\partial z} f(z, w), \qquad F_2(z, w) = \frac{\partial}{\partial w} f(z, w). \tag{55}$$

Remark 5 It is clear that the function ϕ in (53) represents the *memory term* of the equation, and it is a known function. The same will happen in the subsequent reference interval, $[t_0, t_0 + 2\tau]$, obtained by shifting to the right the previous one by τ, once the solution of (53) has been computed, and so forth.

To begin with, let us state the following straightforward result, whose proof is omitted for brevity.

Theorem 13 *The solution* (54) *of problem* (53) *is defined on the whole time interval* $[t_0, t_0 + \tau]$, *independently of the point* $\xi \in [t_0, t_0 + \tau]$ *where the condition* η *is given.*

The following result then holds true (compare with Theorem 2).

Theorem 14 *With reference to the solution* (54) *of problem* (53), *one has:*

$$\begin{aligned} &a)\ \frac{\partial}{\partial t} y(t) = f(y(t), y(t-\tau)), \quad b)\ \frac{\partial}{\partial \eta} y(t) = \Phi(t, \xi; t_0), \\ &c)\ \frac{\partial}{\partial \xi} y(t) = -\Phi(t, \xi; t_0) f(y(\xi), y(\xi - \tau)), \end{aligned}$$

where $\Phi(t, \xi; t_0)$ *satisfies (see* (55)*):*

$$\begin{aligned} \dot{\Phi}(t, \xi; t_0) &= F_1(y(t), y(t-\tau))\Phi(t, \xi; t_0), \qquad t \in [t_0, t_0 + \tau], \\ \Phi(\xi, \xi; t_0) &= I \in \mathbb{R}^{m \times m}, \\ \Phi(t, \xi; t_0) &= O \in \mathbb{R}^{m \times m}, \qquad t \in [t_0 - \tau, t_0). \end{aligned} \tag{56}$$

Proof The statement *a*) clearly follows from (53). From the same equation one also derives that, for $t \in [t_0, t_0 + \tau]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial}{\partial\eta}y(t)\right) = \frac{\partial}{\partial\eta}\dot{y}(t) = \frac{\partial}{\partial\eta}f(y(t), y(t-\tau)) \tag{57}$$
$$= F_1(y(t), y(t-\tau))\frac{\partial}{\partial\eta}y(t) + F_2(y(t), y(t-\tau))\frac{\partial}{\partial\eta}y(t-\tau).$$

Moreover, at $t = \xi$,

$$\frac{\partial}{\partial\eta}y(\xi) = \frac{\partial}{\partial\eta}\eta = I,$$

and, for $t \in [t_0 - \tau, t_0)$,

$$\frac{\partial}{\partial\eta}y(t) = \frac{\partial}{\partial\eta}\phi(t) = O.$$

This latter equality implies that, for $t \in [t_0, t_0+\tau)$, the term $F_2(y(t), y(t-\tau))\frac{\partial}{\partial\eta}y(t-\tau)$ in (57) vanishes, thus reducing to the first equation in (56), so that *b*) eventually follows. Finally, by virtue of Theorem 13, let t^* be a generic point in the interval $[t_0, t_0 + \tau]$, and denote

$$y^* = y(t^*, \xi, \eta, \phi; t_0).$$

Consequently, since $\xi \in [t_0, t_0 + \tau]$ as well, one has:

$$\eta = y(\xi, t^*, y(t^*, \xi, \eta, \phi; t_0), \phi; t_0),$$

so that we eventually arrive at the identity

$$y^* = y(t^*, \xi, y(\xi, t^*, y^*, \phi; t_0), \phi; t_0).$$

By taking into account the results of the previous points *a*) and *b*), one derives:

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\xi}y(t^*, \xi, y(\xi, t^*, y^*, \phi; t_0), \phi; t_0) \\ &= \frac{\partial}{\partial\xi}y(t^*, \xi, \eta, \phi; t_0) + \frac{\partial}{\partial\eta}y(t^*, \xi, \eta, \phi; t_0)\ \frac{\partial}{\partial t}y(t, t^*, y^*, \phi; t_0)\Big|_{t=\xi} \\ &= \frac{\partial}{\partial\xi}y(t^*, \xi, \eta, \phi; t_0) + \Phi(t^*, \xi; t_0)f(y(\xi), y(\xi - \tau)). \end{aligned}$$

The statement *c*) then follows, by taking into account that t^* is generic. □

One main difference with the ODE case, stems from the fact that now (54) also depends on the *memory term* ϕ, which is a *functional* parameter. Consequently, we now look for a Frechét derivative such that, for any perturbation $\delta\phi \in C([t_0 - \tau, t_0])$ and $t \in [t_0 - \tau, t_0 + \tau]$:

$$\lim_{\varepsilon\to 0}\frac{y(t, \xi, \eta, \phi + \varepsilon\delta\phi; t_0) - y(t, \xi, \eta, \phi; t_0)}{\varepsilon} = \int_{t_0-\tau}^{t_0}\frac{\delta}{\delta\phi(\zeta)}y(t)\delta\phi(\zeta)\mathrm{d}\zeta, \tag{58}$$

where

$$\frac{\delta}{\delta\phi(\zeta)}y(t) : (t, \zeta) \in [t_0 - \tau, t_0 + \tau] \times [t_0 - \tau, t_0) \to \left(\frac{\delta}{\delta\phi_j(\zeta)}y_i(t)\right) \in \mathbb{R}^{m\times m}, \tag{59}$$

is the *functional derivative* of (54) (see, e.g., [24, Appendix A]), with y_i and ϕ_j the respective entries of y and ϕ. For later use, we recall that, for a given $\hat{t} \in [t_0 - \tau, t_0)$ and $i, j = 1, \ldots, m$,

$$\begin{aligned} \frac{\delta}{\delta \phi_j(\hat{t})} y_i(t) &\equiv \lim_{\varepsilon \to 0} \frac{y_i(t, \xi, \eta, \phi + \varepsilon e_j \delta_{\hat{t}}; t_0) - y_i(t, \xi, \eta, \phi; t_0)}{\varepsilon} \\ &= \int_{t_0 - \tau}^{t_0} \frac{\delta}{\delta \phi_j(\zeta)} y_i(t) \delta_{\hat{t}}(\zeta) \mathrm{d}\zeta, \end{aligned} \tag{60}$$

with $e_j \in \mathbb{R}^m$ the jth unit vector and, hereafter, $\delta_{\hat{t}}(t)$ is the Dirac delta function centered at $\hat{t}$. The following result holds true.

Lemma 1 *With reference to* (58) *and* (59)*, for any fixed* $t \in [\xi, t_0 + \tau] \subseteq [t_0, t_0 + \tau]$ *one has:*

$$\frac{\delta}{\delta \phi(\zeta)} y(t) = O \in \mathbb{R}^{m \times m}, \qquad \forall \zeta \in [t_0 - \tau, \xi - \tau) \cup (t - \tau, t_0).$$

Proof Having fixed $t \in [\xi, t_0 + \tau]$, it follows that $\forall \zeta \in [t_0 - \tau, \xi - \tau) \cup (t - \tau, t_0)$, setting as usual δ_ζ the Dirac delta centered at ζ, one has:

$$y(t, \xi, \eta, \phi + \varepsilon \delta_\zeta; t_0) = y(t, \xi, \eta, \phi; t_0), \qquad \forall \varepsilon \in \mathbb{R}.$$

In fact, by virtue of (53), the solution (54) is independent of the values of ϕ outside the interval $[\xi - \tau, t - \tau]$. Consequently, by taking into account (60), it follows that:

$$\frac{\delta}{\delta \phi(\zeta)} y(t) = \lim_{\varepsilon \to 0} \frac{y(t, \xi, \eta, \phi + \varepsilon \delta_\zeta; t_0) - y(t, \xi, \eta, \phi; t_0)}{\varepsilon} = O. \quad \square$$

Taking into account Lemma 1, the following result provides a more practical characterization of the functional derivative (58)–(59). Figure 1 displays the location of the most relevant points and subintervals involved in Theorem 15.

Theorem 15 *With reference to the solution* (54) *of problem* (53)*, for any* $\hat{t} \in (\xi - \tau, t_0)$ *one has:*

$$\frac{\delta}{\delta \phi(\hat{t})} y(t) = \Psi(t, \hat{t}\,; t_0), \tag{61}$$

where $\Psi(t, \hat{t}\,; t_0)$ *satisfies (see* (55)*):*

$$\begin{aligned} \dot{\Psi}(t, \hat{t}\,; t_0) &= F_1(y(t), y(t - \tau)) \Psi(t, \hat{t}\,; t_0), \qquad t \in (\hat{t} + \tau, t_0 + \tau], \\ \Psi(\hat{t} + \tau, \hat{t}\,; t_0) &= F_2(y(\hat{t} + \tau), y(\hat{t}\,)), \\ \Psi(t, \hat{t}; t_0) &= \delta_{\hat{t}}(t) I, \qquad t \in [t_0 - \tau, \hat{t} + \tau), \end{aligned} \tag{62}$$

with $O, I \in \mathbb{R}^{m \times m}$ *and* $\delta_{\hat{t}}(t)$ *the Dirac delta function.*

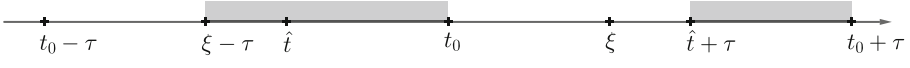

Fig. 1 Relevant time subintervals for Theorem 15

Proof In fact, for $t \in [t_0, t_0 + \tau]$ one has, by virtue of (53):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\delta}{\delta\phi(\hat{t}\,)}y(t,\xi,\eta,\phi;t_0) &= \frac{\delta}{\delta\phi(\hat{t}\,)}\dot{y}(t) = \frac{\delta}{\delta\phi(\hat{t}\,)}f(y(t),y(t-\tau))\\ &= F_1(y(t),y(t-\tau))\frac{\delta}{\delta\phi(\hat{t}\,)}y(t) + F_2(y(t),y(t-\tau))\frac{\delta}{\delta\phi(\hat{t}\,)}y(t-\tau),\end{aligned}$$

i.e., using the notation (61),

$$\begin{aligned}\dot{\Psi}(t,\hat{t}\,;t_0) = {} & F_1(y(t),y(t-\tau))\Psi(t,\hat{t}\,;t_0)\\ & + F_2(y(t),y(t-\tau))\Psi(t-\tau,\hat{t}\,;t_0).\end{aligned} \tag{63}$$

Moreover, at $t = \xi$,

$$\Psi(\xi,\hat{t}\,;t_0) = \frac{\delta}{\delta\phi(\hat{t}\,)}y(\xi) = \frac{\delta}{\delta\phi(\hat{t}\,)}\eta = O, \tag{64}$$

since the condition $y(\xi) = \eta$ is independent of the history ϕ. Further, taking into account (60), for all $i, j = 1, \ldots, m$ and for $t \in [t_0 - \tau, t_0)$, one has:

$$\frac{\delta}{\delta\phi_j(\hat{t}\,)}y_i(t) = \frac{\delta}{\delta\phi_j(\hat{t}\,)}\phi_i(t) = \lim_{\varepsilon\to 0}\frac{[\phi_i(t) + \varepsilon\delta_{ij}\delta_{\hat{t}}(t)] - \phi_i(t)}{\varepsilon} = \delta_{ij}\delta_{\hat{t}}(t),$$

with δ_{ij} the Kronecker delta. Consequently,

$$\Psi(t,\hat{t}\,;t_0) = \frac{\delta}{\delta\phi(\hat{t}\,)}y(t) = \delta_{\hat{t}}(t)I, \qquad t \in [t_0 - \tau, t_0). \tag{65}$$

From (63)–(65), one then derives, considering that (see Fig. 1) $t_0 - \tau \le \xi - \tau < \hat{t}$:

$$\Psi(t,\hat{t}\,;t_0) = \int_{\xi}^{t}\dot{\Psi}(\zeta,\hat{t}\,;t_0)\mathrm{d}\zeta = \begin{cases} O, & t \in [t_0, \hat{t} + \tau),\\ F_2(y(\hat{t} + \tau), y(\hat{t}\,)), & t = \hat{t} + \tau. \end{cases} \tag{66}$$

From (65) and (66) the last two equations in (62) follow. Consequently, from (63), one obtains

$$\dot{\Psi}(t,\hat{t}\,;t_0) = F_1(y(t),y(t-\tau))\Psi(t,\hat{t}\,;t_0), \qquad t \in (\hat{t} + \tau, t_0 + \tau],$$

which completes the proof of (62). □

As a straightforward consequence, the following result holds true, which guarantees the regularity of Ψ w.r.t. its first two arguments (again, for sake of clarity, refer to Fig. 1).

Corollary 4 *With reference to the solution* (54) *of problem* (53)*, and considering* (55)*,* (56)*, and* (62)*, for any* $\hat{t} \in (\xi - \tau, t_0)$ *one has:*

$$\Psi(t,\hat{t}\,;t_0) = \Phi(t,\hat{t} + \tau;t_0)F_2(y(\hat{t} + \tau), y(\hat{t}\,)), \qquad t \in [\hat{t} + \tau, t_0 + \tau]. \tag{67}$$

Finally, the following result holds true (compare with Corollary 2 of the ODE case).

Corollary 5 *With reference to the solution* (54) *of problem* (53)*, and considering* (56) *and* (67)*, for any* $\delta\phi \in C([t_0 - \tau, t_0])$ *one has:*

$$y(t,\xi,\eta+\delta\eta,\phi+\delta\phi;t_0) = y(t,\xi,\eta,\phi;t_0)+\Phi(t,\xi;t_0)\delta\eta + \int_{\xi-\tau}^{t-\tau}\Psi(t,\zeta;t_0)\delta\phi(\zeta)\mathrm{d}\zeta + (t-\xi)\,O(|\delta\eta|+\|\delta\phi\|)^2, \qquad t\in[\xi,t_0+\tau],$$

with $\|\delta\phi\| = \max_{\zeta\in[\xi-\tau,t-\tau]}|\delta\phi(\zeta)|$.

Proof The statement follows from Theorem 14, part *b*), and Theorem 15, by taking into account (58) and the result of Lemma 1. □

3.2 Main results (DDE case)

We are now in the position of discussing the accuracy of the approximations (52). To begin with, the following result holds true.

Theorem 16 *With reference to* (38)–(52)*, for* $n = 1,\dots,\nu$ *one has:*

$$y(t_n)-y_n = y(t_{n-1})-y_{n-1}+O(h^{2s+1}), \qquad \|\delta\sigma_n\| := \max_{c\in[0,1]}|\delta\sigma_n(ch)| = O(h^{s+1}).$$

Proof The statement follows from Theorem 4 by considering that, for $n = 1,\dots,\nu$, $t\in[t_0,t_0+\tau]$ in (38), so that $y(t-\tau)\equiv\phi(t-\tau)$, which is a known function, thus obtaining an ODE. □

This result allows us to state the following one, which generalizes that of Theorem 5 to the present case.

Theorem 17 *With reference to* (44)*,* (46)*,* (50)*, and* (52) *if for* $n\geq 1$ *one has:*

$$y(t_r)-y_r = y(t_{r-1})-y_{r-1}+O(h^{2s+1}), \qquad r=1,\dots,n,$$

then

$$\delta\gamma_j^n := \gamma_j(\hat\sigma_n,\hat\sigma_{n-\nu})-\gamma_j(\sigma_n,\sigma_{n-\nu}) = O(h^{2s-j}), \qquad j=0,\dots,s-1.$$

Proof The proof is by generalized induction. For $n = 1,\dots,\nu$ the statement follows from Theorems 5 and 16 since, in this case,

$$\hat\sigma_{n-\nu}(ch)\equiv\sigma_{n-\nu}(ch)\equiv\phi(t_{n-1}+ch-\tau), \qquad c\in[0,1],$$

so that we are dealing with an ODE. Assume now it true up to $n-1$, and prove for n. By hypothesis, and from (45) and (51), we know that

$$y(t_n)-y_n = y(t_{n-1})-y_{n-1}+O(h^{2s+1}) \equiv y(t_{n-1})-y_{n-1}+h\delta\gamma_0^n,$$

so that $\delta\gamma_0^n = O(h^{2s})$ follows. Then, by taking into account (55), it follows that:

$$
\begin{aligned}
O(h^{2s}) = \delta\gamma_0^n \;&=\; \gamma_0(\hat\sigma_n,\hat\sigma_{n-\nu}) - \gamma_0(\sigma_n,\sigma_{n-\nu})\\
&= \int_0^1 \big[f(\hat\sigma_n(\zeta h),\hat\sigma_{n-\nu}(\zeta h)) - f(\sigma_n(\zeta h),\sigma_{n-\nu}(\zeta h))\big]\,\mathrm{d}\zeta\\
&= \int_0^1\int_0^1 \big[\,F_1(\sigma_n(\zeta h)+c\,\delta\sigma_n(\zeta h),\sigma_{n-\nu}(\zeta h)+c\,\delta\sigma_{n-\nu}(\zeta h))\delta\sigma_n(\zeta h)\\
&\quad + F_2(\sigma_n(\zeta h)+c\,\delta\sigma_n(\zeta h),\sigma_{n-\nu}(\zeta h)+c\,\delta\sigma_{n-\nu}(\zeta h))\delta\sigma_{n-\nu}(\zeta h)\,\big]\,\mathrm{d}c\,\mathrm{d}\zeta\\
&= \int_0^1 \underbrace{\int_0^1 F_1(\sigma_n(\zeta h)+c\,\delta\sigma_n(\zeta h),\sigma_{n-\nu}(\zeta h)+c\,\delta\sigma_{n-\nu}(\zeta h))\mathrm{d}c}_{=:\,G_1(\zeta h)}\;\delta\sigma_n(\zeta h)\,\mathrm{d}\zeta\\
&\quad + \int_0^1 \underbrace{\int_0^1 F_2(\sigma_n(\zeta h)+c\,\delta\sigma_n(\zeta h),\sigma_{n-\nu}(\zeta h)+c\,\delta\sigma_{n-\nu}(\zeta h))\mathrm{d}c}_{=:\,G_2(\zeta h)}\;\delta\sigma_{n-\nu}(\zeta h)\,\mathrm{d}\zeta\\
&= \int_0^1 G_1(\zeta h)\delta\sigma_n(\zeta h)\mathrm{d}\zeta \;+\; \int_0^1 G_2(\zeta h)\delta\sigma_{n-\nu}(\zeta h)\mathrm{d}\zeta.
\end{aligned}
$$

Let us discuss in detail the term

$$\int_0^1 G_1(\zeta h)\delta\sigma_n(\zeta h)\,\mathrm{d}\zeta,$$

since the remaining one,

$$\int_0^1 G_2(\zeta h)\delta\sigma_{n-\nu}(\zeta h)\,\mathrm{d}\zeta = O(h^{2s}),$$

is similarly discussed, by taking into account the induction hypothesis. By virtue of (22), one has:

$$
\begin{aligned}
\delta\sigma_n(\zeta h) &= \hat\sigma_n(\zeta h) - \sigma_n(\zeta h)\\
&= y(t_{n-1}) - y_{n-1} \;+\; h\sum_{j=0}^{s-1}\int_0^\zeta P_j(x)\mathrm{d}x\;\delta\gamma_j^n \;+\; \sum_{j\ge s}\int_0^\zeta P_j(x)\mathrm{d}x\;\gamma_j(\hat\sigma_n,\hat\sigma_{n-\nu})\\
&= y(t_{n-1}) - y_{n-1} \;+\; \zeta h\delta\gamma_0^n + h\sum_{j=1}^{s-1}\big[\xi_{j+1}P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta)\big]\delta\gamma_j^n\\
&\quad + h\sum_{j\ge s}\big[\xi_{j+1}P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta)\big]\gamma_j(\hat\sigma_n,\hat\sigma_{n-\nu}).
\end{aligned}
$$

Consequently, from Theorem 1 and Corollary 3, one obtains:

$$O(h^{2s}) = \int_0^1 G_1(\zeta h)\delta\sigma_n(\zeta h)\,\mathrm{d}\zeta = \underbrace{\int_0^1 G_1(\zeta h)\mathrm{d}\zeta}_{=O(1)}\underbrace{[y(t_{n-1}) - y_{n-1}]}_{=(n-1)\,O(h^{2s+1})}$$
$$+ h\underbrace{\int_0^1 G_1(\zeta h)\zeta\,\mathrm{d}\zeta}_{=O(1)}\underbrace{\delta\gamma_0^n}_{=O(h^{2s})} + h\sum_{j=1}^{s-1}\int_0^1 G_1(\zeta h)\left[\xi_{j+1}P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta)\right]\mathrm{d}\zeta\,\delta\gamma_j^n$$
$$+ \underbrace{h\sum_{j\geq s}\overbrace{\int_0^1 G_1(\zeta h)\left[\xi_{j+1}P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta)\right]\mathrm{d}\zeta}^{=O(h^{j-1})}\overbrace{\gamma_j(\hat{\sigma}_n, \hat{\sigma}_{n-\nu})}^{=O(h^j)}}_{=O(h^{2s})},$$

from which,

$$O(h^{2s}) = h\sum_{j=1}^{s-1}\underbrace{\int_0^1 G_1(\zeta h)\left[\xi_{j+1}P_{j+1}(\zeta) - \xi_j P_{j-1}(\zeta)\right]\mathrm{d}\zeta}_{=O(h^{j-1})}\,\delta\gamma_j^n$$

follows and, therefore, one concludes that $\delta\gamma_j^n = O(h^{2s-j}),\ j = 0\ldots, s-1.$ □

As a consequence, the following result can be stated.

Theorem 18 *With reference to* (38)–(52)*, for* $n = 1, \ldots, N \equiv K\nu$ *one has:*

$$y(t_n) - y_n = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}), \quad \|\delta\sigma_n\| := \max_{c\in[0,1]}|\delta\sigma_n(ch)| = O(h^{s+1}).$$

Proof The proof is done by induction on groups of ν consecutive steps. For the first ν steps, the statement follows from Theorem 16. Assume now, by induction, that it holds true up to $t_{k\nu} = t_0 + k\nu h$, and let us prove for $n = k\nu + 1, \ldots, (k+1)\nu$. For this purpose, for $k = 1, \ldots, K-1$ let us set:

$$\phi_k(t) \equiv \sigma(t), \qquad \hat{\phi}_k(t) \equiv \hat{\sigma}(t) \equiv y(t), \qquad t \in [t_{(k-1)\nu}, t_{k\nu}).$$

Assuming, again, true the statement for $n-1$, and using the notation (54), one has:

$$\begin{aligned}
\delta\sigma_n(ch) &= \hat{\sigma}_n(ch) - \sigma_n(ch)\\
&= y(t_{n-1}+ch, t_{n-1}, y(t_{n-1}), \hat{\phi}_k; t_{k\nu}) - y(t_{n-1}+ch, t_{n-1}+ch, \sigma_n(ch), \phi_k; t_{k\nu})\\
&= \underbrace{y(t_{n-1}+ch, t_{n-1}, \overbrace{\sigma_n(0)}^{=y_{n-1}}, \phi_k; t_{k\nu}) - y(t_{n-1}+ch, t_{n-1}+ch, \sigma_n(ch), \phi_k; t_{k\nu})}_{=: E_{n,1}^{(k)}(ch)}\\
&\quad + \underbrace{y(t_{n-1}+ch, t_{n-1}, y(t_{n-1}), \hat{\phi}_k; t_{k\nu}) - y(t_{n-1}+ch, t_{n-1}, \overbrace{y_{n-1}}^{=\sigma_n(0)}, \phi_k; t_{k\nu})}_{=: E_{n,2}^{(k)}(ch)}.
\end{aligned}$$

From Theorem 16, it follows that

$$E_{n,1}^{(k)}(h) \;=\; \sigma_n(0) - \sigma_n(0) + O(h^{2s+1}) \;=\; O(h^{2s+1}), \tag{68}$$
$$\|E_{n,1}^{(k)}\| := \max_{c\in[0,1]} |E_{n,1}^{(k)}(ch)| = O(h^{s+1}).$$

Moreover, from Corollary 5, and considering that $h\nu = \tau$, one has:

$$\begin{aligned} E_{n,2}^{(k)}(ch) \;=\; & \overbrace{\Phi(t_{n-1}+ch, t_{n-1}; t_{k\nu})}^{=I+O(ch)} \overbrace{\left[y(t_{n-1}) - y_{n-1}\right]}^{=\delta\sigma_n(0)} \\ & + h\int_0^c \Psi(t_{n-1}+ch, t_{n-1}+\zeta h-\tau; t_{k\nu})\delta\sigma_{n-\nu}(\zeta h)\mathrm{d}\zeta \\ & + ch\, O(|\delta\sigma_n(0)| + \|\delta\sigma_{n-\nu}\|)^2. \end{aligned}$$

By considering that

$$|\delta\sigma_n(0)| = (n-1)O(h^{2s+1}), \qquad \|\delta\sigma_{n-\nu}\| = O(h^{s+1}),$$

one eventually derives

$$\|E_{n,2}^{(k)}\| := \max_{c\in[0,1]} |E_{n,2}^{(k)}(ch)| = O(h^{s+2}),$$

from which the second part of the statement follows, by taking into account (68). Moreover, when $c = 1$ then $t_{n-1} + h = t_n$ and, by virtue of Theorem 1 and Theorem 17, one obtains:

$$\begin{aligned} & \int_0^1 \overbrace{\Psi(t_n, t_{n-1}+\zeta h-\tau; t_{k\nu})}^{=:G(\zeta h)} \delta\sigma_{n-\nu}(\zeta h)\mathrm{d}\zeta \\ & = \int_0^1 G(\zeta h)\left[\sum_{j=0}^{s-1} P_j(\zeta)\delta\gamma_j^{n-\nu} + \sum_{j\ge s} P_j(\zeta)\gamma_j(\hat\sigma_{n-\nu}, \hat\sigma_{n-2\nu})\right] \\ & = \sum_{j=0}^{s-1} \underbrace{\int_0^1 P_j(\zeta)G(\zeta h)\mathrm{d}\zeta}_{=O(h^j)} \underbrace{\delta\gamma_j^{n-\nu}}_{=O(h^{2s-j})} + \sum_{j\ge s} \underbrace{\int_0^1 P_j(\zeta)G(\zeta h)\mathrm{d}\zeta}_{=O(h^j)} \underbrace{\gamma_j(\hat\sigma_{n-\nu}, \hat\sigma_{n-2\nu})}_{=O(h^j)} \;=\; O(h^{2s}). \end{aligned}$$

Consequently,

$$E_{n,2}^{(k)}(h) = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}),$$

and also the first part of the statement follows. □

3.3 Discretization

The discretization issue proceeds as in the ODE case. In fact, also in the DDE case, the Fourier coefficients (see (46) and (49)),

$$\gamma_j(\sigma_n, \sigma_{n-\nu}) = \int_0^1 P_j(\zeta) f(\sigma_n(\zeta h), \sigma_{n-\nu}(\zeta h))\mathrm{d}\zeta, \qquad j = 0, \dots, s-1,$$

need to be approximated by using a (interpolatory) quadrature rule of order q, thus providing a possibly different piecewise approximation $u(t)$,

$$u(t) \equiv \phi(t), \quad t < t_0, \qquad u_n(ch) := u(t_{n-1}+ch), \quad c \in [0,1], \quad n = 1-\nu, \ldots, N,$$

such that, for $n \geq 1$:

$$\dot{u}_n(ch) = \sum_{j=0}^{s-1} P_j(c)\hat{\gamma}_j(u_n, u_{n-\nu}), \qquad c \in [0,1], \qquad u_n(0) = y_{n-1}. \tag{69}$$

Consequently,

$$u_n(ch) = y_{n-1} + h\sum_{j=0}^{s-1}\int_0^c P_j(x)\mathrm{d}x\ \hat{\gamma}_j(u_n, u_{n-\nu}), \qquad c \in [0,1], \tag{70}$$

and

$$y_n = y_{n-1} + h\hat{\gamma}_0(u_n, u_{n-\nu}) =: u_n(h), \tag{71}$$

where (see (46)),

$$\hat{\gamma}_j(u_n, u_{n-\nu}) := \sum_{i=1}^{k} b_i P_j(c_i) f(u_n(c_i h), u_{n-\nu}(c_i h)) = \gamma_j(u_n, u_{n-\nu}) - \Delta_j(h), \tag{72}$$

with (c_i, b_i) the abscissae and weights of the quadrature, and $\Delta_j(h) = O(h^{q-j})$ the quadrature error, where q is the order of the quadrature.

Formulae (69) and (46) form a subclass of the so-called natural continuous RK methods for DDEs (see [5, Sec. 6.2]). As a consequence, their convergence properties could be as well derived by more classical approaches such as Bellman's method of steps, which is an analytic procedure specific for DDEs. In the present context, the main goal is to show how the framework based on the perturbation theory applied to the truncated Fourier expansion is easily adapted to cope with DDEs, therefore we will pursue this route of investigation. A further strength of this approach is the possibility of analyzing the convergence properties of the truncated Fourier approximations when these are used as spectral methods in time. In this regard, the analysis for the ODE case has been addressed in [3], while a spectral implementation of the methods for DDEs has been considered in [16].

By using standard arguments (which we omit, as done in the ODE case), we can derive the following results, representing the corresponding counterparts of Theorems 17 and 18, respectively.

Theorem 19 *With reference to* (44)*,* (46)*,* (69)–(71)*, and assuming that the quadrature formula* (72) *has order* $q \geq 2s$*, if for* $n \geq 1$ *one has:*

$$y(t_r) - y_r = y(t_{r-1}) - y_{r-1} + O(h^{2s+1}), \qquad r = 1, \ldots, n,$$

then

$$\delta\hat{\gamma}_j^n := \gamma_j(\hat{\sigma}_n, \hat{\sigma}_{n-\nu}) - \hat{\gamma}_j(u_n, u_{n-\nu}) = O(h^{2s-j}), \qquad j = 0, \ldots, s-1.$$

Theorem 20 *With reference to* (44), (46), (69)–(71)*, and assuming that the quadrature formula* (72) *has order* $q \geq 2s$*, for* $n = 1, \ldots, N \equiv K\nu$ *one has:*

$$y(t_n) - y_n = y(t_{n-1}) - y_{n-1} + O(h^{2s+1}), \qquad \max_{c\in[0,1]} |\hat{\sigma}_n(ch) - u_n(ch)| = O(h^{s+1}).$$

Remark 6 It is worth mentioning that the result of Theorem 20 states that the superconvergence order $2s$ at the mesh-points t_n is obtained, even though possibly different Runge-Kutta methods are used at each integration step, provided that they define a polynomial approximation of degree s. This, in turn, represents a generalization of the results in [4] for collocation methods.

We conclude this section, by recalling that the considerations in Remark 4 continue to hold in the DDE case and by observing that, concerning the implementation of the resulting Runge-Kutta method used for solving problem (38), the arguments in Section 2.5, *mutatis mutandis*, apply as well.

4 Numerical tests

In this section we report a few numerical tests for the DDE case. In fact, in the ODE case, HBVMs have been extensively used as energy-conserving methods for Hamiltonian systems (see, e.g., [2, 9–11, 19]). We show that, under some circumstances, their use can be advantageous also in the DDE case. Hereafter, we consider a class of DDEs defined by a Hamiltonian function

$$H : (q, p) \in \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \tag{73}$$

through the equations

$$\begin{aligned} \dot{q}(t) &= \quad H_p(q(t), p(t)) + \alpha H_p(q(t-\tau), p(t-\tau)), \\ \dot{p}(t) &= -\big[H_q(q(t), p(t)) + \alpha H_q(q(t-\tau), p(t-\tau))\big], \end{aligned} \tag{74}$$

with α a real parameter, $\tau > 0$ the delay, and H_q and H_p the partial derivatives of H w.r.t. q and p, respectively. The problem is completed by the initial conditions

$$q(t) = \phi(t), \qquad p(t) = \psi(t), \qquad t \in [-\tau, 0]. \tag{75}$$

The introduction of such a kind of *delay Hamiltonian system* is partly inspired by the problem of looking for periodic orbits of DDEs, which has been attacked by many authors in the past (see, e.g., [23, 34, 35, 37–39, 41]). In this respect, the first two examples below show an attractive periodic orbit with integer period lying on a level set of the Hamiltonian function (73) which is, therefore, a constant of motion once the periodic orbit has been approached. In the third example we are instead interested in simulating the correct qualitative behavior of a *dissipative Hamiltonian delay problem* in the phase space when the dynamics takes place in a neighborhood of a

separatrix. Taking aside a theoretical discussion of problem (73)–(74) which would go beyond the scopes of the present work, we infer its properties for the three considered instances by preliminarily applying a high-order integrator with very small stepsize, in order to get a very accurate numerical solution that will be taken as a reference trajectory in the phase space.

For all the three problems, we show that a very accurate approximation of the Hamiltonian function allows us to reproduce the correct geometric features of the solution in the discrete setting. To the best of our knowledge, this is the first instance of the use of HBVMs in the context of DDEs displaying geometric properties. For comparison purposes, we also solve the problems with the classical Gauss collocation integrator of the same order. The numerical tests have been implemented in Matlab (R2020b) on a 3 GHz Intel Xeon W10 core computer with 64GB of memory.

4.1 Problem 1

With reference to (73)–(75), the first problem is defined as follows:

$$\begin{aligned} &m = 1, \qquad H(q,p) = \frac{1}{4}\left(q^4 + p^4\right), \qquad \alpha = 10^{-1}, \\ &\tau = 1, \qquad \phi(t) \equiv \sqrt{2}, \qquad \psi(t) \equiv 0. \end{aligned} \tag{76}$$

We solve this problem by using the following methods:

- HBVM(2,2) (i.e., the 2-stage Gauss method),
- HBVM(4,2).

Both methods are fourth-order, according to Theorem 10, with HBVM(4,2) energy-conserving once a periodic orbit of integer period is eventually reached (see Theorem 12). Problem (76) possesses an attracting periodic orbit with period $T = 2\tau = 2$ which suggests using a stepsize h equal to a submultiple of τ, in order to mimic a corresponding discrete periodic solution. As we are going to see, unlike the 2-stage Gauss collocation method, the conservation property of HBVM(4,2) results in a precise resolution of this task. We solve the problem on the interval $[0, 2 \cdot 10^3]$ by using a timestep $h = \tau/5 = 0.2$. Figure 2 summarizes the obtained results.

- In the upper row of the figure are the plots of the numerical Hamiltonian, $H(q_n, p_n)$, from which one deduces that both methods quite soon reach a stationary behavior.
- To better discern the asymptotic behavior of the two numerical solutions, the central pictures show the plots of $|H(q_n, p_n) - H(q_{n-1}, p_{n-1})|$ for the two methods. From these plots one infers that, while the stationary value of the Hamiltonian is constant for the HBVM(4,2) method, it is only approximately constant for the HBVM(2,2) method, with oscillations having amplitude of order 10^{-2}.
- The bottom row contains the plots of the numerical trajectory in the phase plane for both methods, relative to the interval $[2 \cdot 10^2, 2 \cdot 10^3]$ (i.e., after the transient phase). For both methods, the solution seems to repeat every 10 points (i.e., with period $T = 2\tau = 2$). However, the points obtained by the HBVM(2,2)

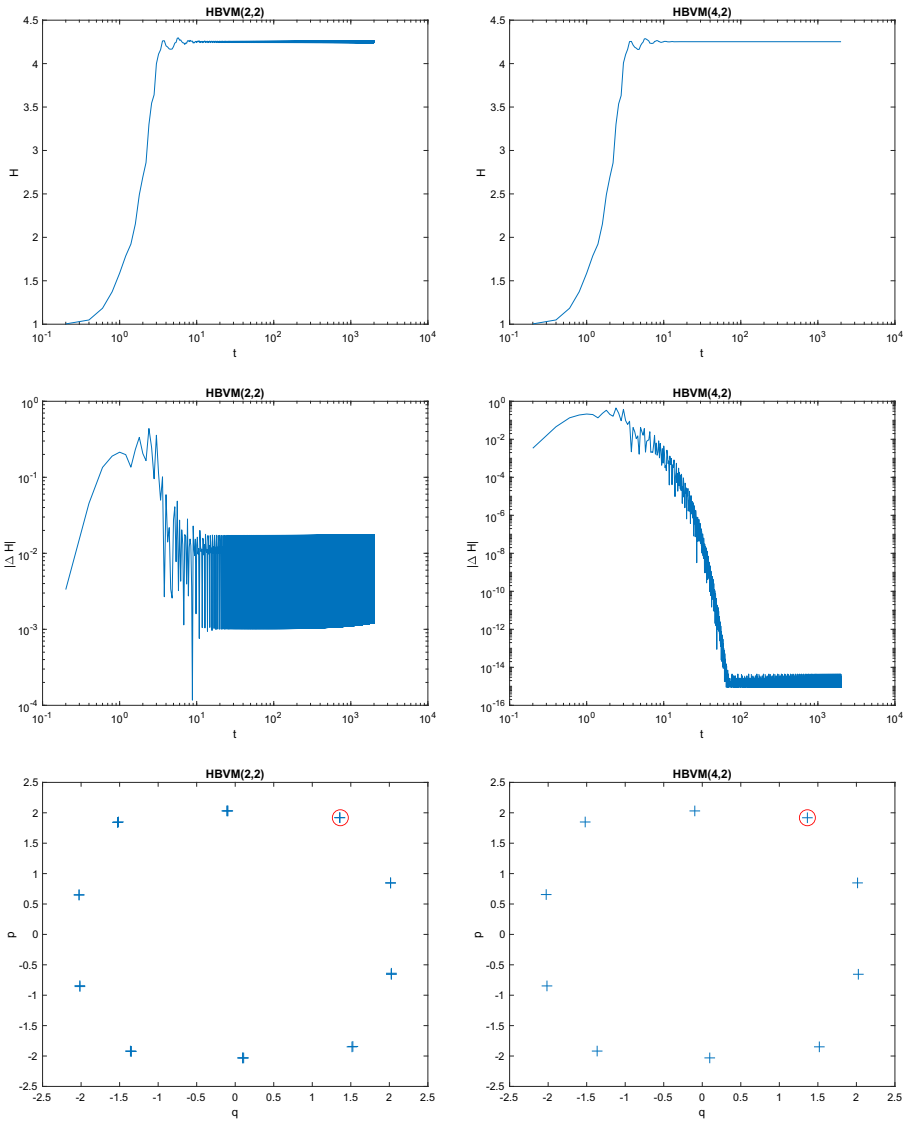

Fig. 2 Numerical results for problem (76) solved by using HBVM(2,2), left plots, and HBVM(4,2), right plots, using a timestep $h = 0.2$ (see the text for details)

method are not actually periodic, whereas they are (within to machine precision and independently of the used stepsize h) for the HBVM(4,2) method. To confirm this, in Table 1 we list the last 20 points of the trajectories computed after each period $T = 10h$ and lying inside the two small circles highlighted in the plots. As one may see, only the first 4 digit of the points of the trajectory computed by the HBVM(2,2) method are retained, whereas the points computed by the HBVM(4,2) method differ at most on the last digit.

Table 1 The last 20 points of the trajectories inside the circles in the plots on the bottom line of Fig. 2

HBVM(2,2)		HBVM(4,2)	
q	p	q	p
1.344913051657652	1.924341608176171	1.364023296679203	1.918490612087558
1.344895222079097	1.924347768593204	1.364023296679201	1.918490612087558
1.344877390115245	1.924353929533504	1.364023296679201	1.918490612087558
1.344859555766308	1.924360090996891	1.364023296679201	1.918490612087558
1.344841719032502	1.924366252983184	1.364023296679200	1.918490612087558
1.344823879914032	1.924372415492206	1.364023296679201	1.918490612087558
1.344806038411116	1.924378578523775	1.364023296679200	1.918490612087558
1.344788194523964	1.924384742077714	1.364023296679201	1.918490612087559
1.344770348252789	1.924390906153841	1.364023296679202	1.918490612087558
1.344752499597800	1.924397070751980	1.364023296679202	1.918490612087558
1.344734648559217	1.924403235871946	1.364023296679200	1.918490612087559
1.344716795137255	1.924409401513561	1.364023296679199	1.918490612087559
1.344698939332121	1.924415567676645	1.364023296679200	1.918490612087559
1.344681081144034	1.924421734361018	1.364023296679199	1.918490612087559
1.344663220573212	1.924427901566499	1.364023296679200	1.918490612087558
1.344645357619866	1.924434069292906	1.364023296679200	1.918490612087558
1.344627492284208	1.924440237540062	1.364023296679200	1.918490612087558
1.344609624566458	1.924446406307785	1.364023296679200	1.918490612087559
1.344591754466830	1.924452575595894	1.364023296679201	1.918490612087558
1.344573881985545	1.924458745404208	1.364023296679203	1.918490612087558

4.2 Problem 2

The second example is similar in nature to the previous one but considers a non-polynomial Hamiltonian function with two degrees of freedom. With reference to (73)–(75), it is defined by:

$$m = 2, \qquad H(q, p) = \frac{1}{4}\left(q_1^4 + q_2^4 + p_1^4 + p_2^4\right) + \frac{\pi}{2}\left(\frac{1}{\|q\|_2^2} + \frac{2}{\|p\|_2^2}\right), \tag{77}$$

$$\alpha = 5 \cdot 10^{-2}, \qquad \tau = 1, \qquad \phi(t) \equiv (0.1,\ 1)^\top, \qquad \psi(t) \equiv (1,\ 0.2)^\top.$$

Again, we have experienced the existence of a periodic orbit with period $T = 2\tau = 2$. We solve this problem on the interval $[0, 10^3]$ with timestep $h = \tau/10 = 0.1$, by using the following methods:

- HBVM(2,2) (i.e., the 2-stage Gauss method),
- HBVM(10,2).

Both methods are fourth-order, the latter being *practically* energy-conserving, for the given timestep, in the event that a periodic orbit is reached.

Also in this case, the conservation property of HBVM(10,2) turns out to be crucial in reproducing a discrete orbit with period precisely equal to 2, while a small phase drift affects the solution yielded by the 2-stage Gauss collocation method. Figure 3, which is similar to Fig. 2, summarizes the obtained results.

- In the upper row of the figure are the plots of the numerical Hamiltonian, namely $H(q_n, p_n)$: for both methods it seems to reach a stationary behavior.

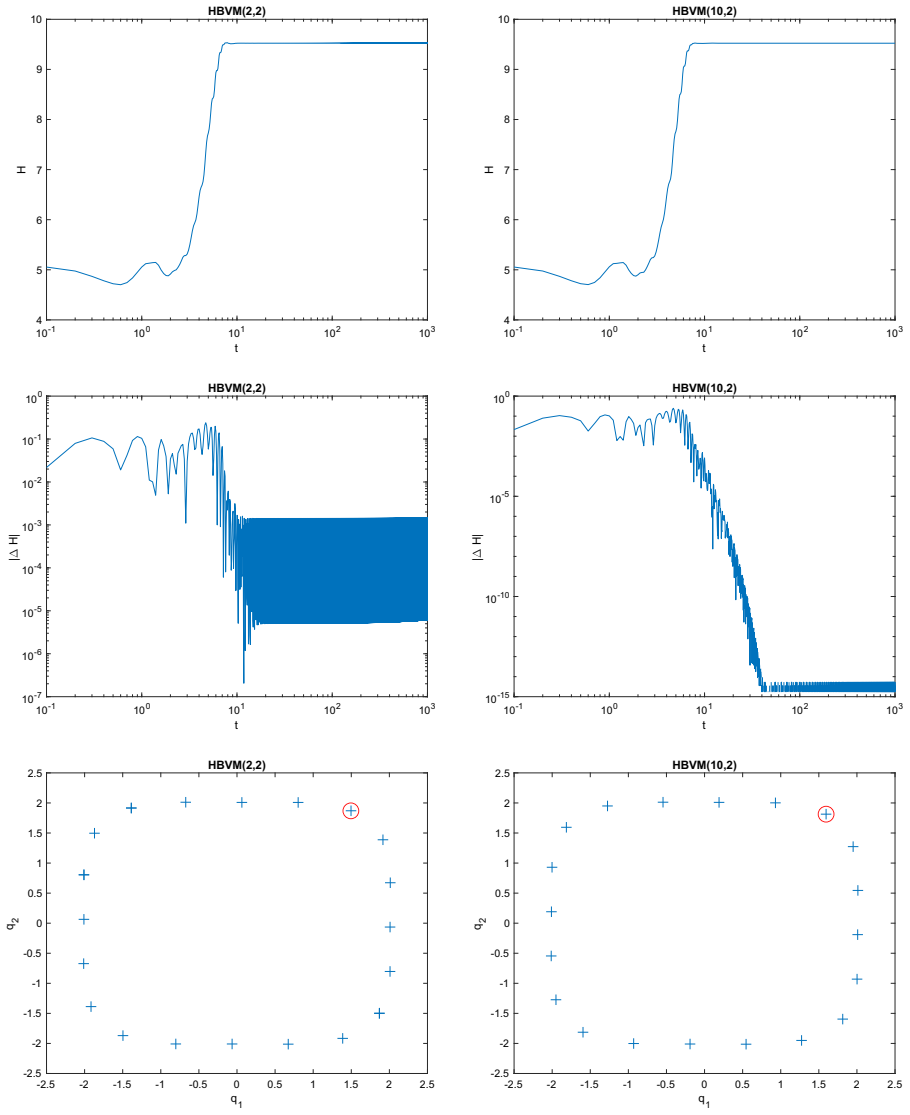

Fig. 3 Numerical results for problem (77) solved by using HBVM(2,2), left plots, and HBVM(10,2), right plots, using a timestep $h = 0.1$ (see the text for details)

- The second row shows the plots of $|H(q_n, p_n) - H(q_{n-1}, p_{n-1})|$ for the two methods. From these plots one infers that, while the stationary value of the Hamiltonian is constant (up to round-off) for the HBVM(10,2) method, it is only approximately constant for the HBVM(2,2) method, with oscillations having amplitude of order 10^{-3}.
- The bottom row contains the plots of the numerical trajectory in the $q_1 - q_2$ plane for both methods, relative to the interval $[10^2, 10^3]$ (i.e., after the transient phase). For both methods, the solution seems to repeat every 20 points (i.e., with period $T = 2\tau = 2$). However, only the points obtained by the HBVM(10,2) method are actually periodic. To confirm this, in Table 2 we list the last 20 points of the trajectories computed after each period $T = 20h$ and lying inside the two small circles displayed in the plots. As one may see, the Gauss collocation method only retain the first 5 digits after each period, whereas the points computed by the HBVM(10,2) method differ at most on the last digit.

Table 2 The last 20 points of the trajectories inside the circles in the plots on the bottom line of Fig. 3

HBVM(2,2)		HBVM(10,2)	
q_1	q_2	q_1	q_2
1.500006047618583	1.868403720200248	1.595245320422993	1.813631211153069
1.500014079966090	1.868399733165645	1.595245320422992	1.813631211153067
1.500022113284264	1.868395745553114	1.595245320422991	1.813631211153069
1.500030147573164	1.868391757362593	1.595245320422994	1.813631211153068
1.500038182832831	1.868387768594017	1.595245320422991	1.813631211153067
1.500046219063319	1.868383779247324	1.595245320422993	1.813631211153069
1.500054256264677	1.868379789322453	1.595245320422994	1.813631211153069
1.500062294436967	1.868375798819335	1.595245320422991	1.813631211153067
1.500070333580236	1.868371807737912	1.595245320422991	1.813631211153068
1.500078373694533	1.868367816078119	1.595245320422993	1.813631211153069
1.500086414779919	1.868363823839889	1.595245320422992	1.813631211153067
1.500094456836429	1.868359831023165	1.595245320422991	1.813631211153069
1.500102499864130	1.868355837627883	1.595245320422994	1.813631211153068
1.500110543863066	1.868351843653977	1.595245320422991	1.813631211153067
1.500118588833287	1.868347849101385	1.595245320422993	1.813631211153069
1.500126634774853	1.868343853970044	1.595245320422994	1.813631211153069
1.500134681687813	1.868339858259893	1.595245320422991	1.813631211153067
1.500142729572213	1.868335861970863	1.595245320422991	1.813631211153068
1.500150778428111	1.868331865102895	1.595245320422993	1.813631211153069
1.500158828255558	1.868327867655927	1.595245320422992	1.813631211153067
1.500166879054607	1.868323869629889	1.595245320422991	1.813631211153069
1.500174930825299	1.868319871024725	1.595245320422994	1.813631211153068

4.3 Problem 3

For the last problem, we are no more interested in periodic trajectories. Instead, we consider a *delay Hamiltonian problem* with *dissipation*. This can be achieved by choosing a negative value of the parameter α in (74). With reference to (73)–(75), the selected parameters are:

$$\begin{aligned} &m = 1, \quad H(q, p) = \frac{1}{2}p^2 - \cos q, \qquad \alpha = -10^{-5}, \\ &\tau = 1, \quad \phi(t) \equiv 0, \quad \psi(t) \equiv 1.99999. \end{aligned} \tag{78}$$

This problem is a dissipative delay-variant of the nonlinear pendulum, with the initial condition chosen close to the separatrix (the level set $H(q, p) = 1$) between the two different regimes of the pendulum: librations around the straight-down stationary position, and rotations. For the given initial conditions, the pendulum should undergo damped oscillations with a decreasing trend of the Hamiltonian function $H(q_n, p_n)$. Consequently, when using relatively large stepsizes, it is fundamental to reproduce the correct dissipation of the Hamiltonian along the numerical trajectory.

We solve this problem on the interval $[0, 500]$, with a timestep $h = \tau/2 = 0.5$, by using the following methods:

- HBVM(2,2) (i.e., the 2-stage Gauss method),
- HBVM(10,2).

Figure 4 summarizes the obtained results.

- In the upper row of the figure are the plots of the numerical Hamiltonian, $H(q_n, p_n)$, from which one deduces that both methods have a dissipation trend of the energy H. Nevertheless, for HBVM(2,2) the values of the Hamiltonian becomes quite larger than 1 in the initial part of the trajectory and undergoes fictitious oscillations which cause the numerical solution to escape the correct region of the phase space where the dynamics should take place, as we are going to see. This is not the case for the HBVM(10,2) method, whose numerical Hamiltonian decreases in the correct way, thus remaining always smaller than 1.
- The central pictures show the numerical solution in the phase space. As one may see, the numerical solution provided by HBVM(2,2) "jumps" twice, before being trapped into an invariant region. This means that the pendulum undergoes two complete rotations until it looses enough energy and begins oscillating around the rest position. On the contrary, the numerical solution obtained by using HBVM(10,2) always remains in the correct region.
- The bottom row contains the plots of the numerical solution w.r.t. time, confirming that the numerical solution provided by the HBVM(2,2) method "jumps" twice, whereas that obtained by the HBVM(10,2) method does not.

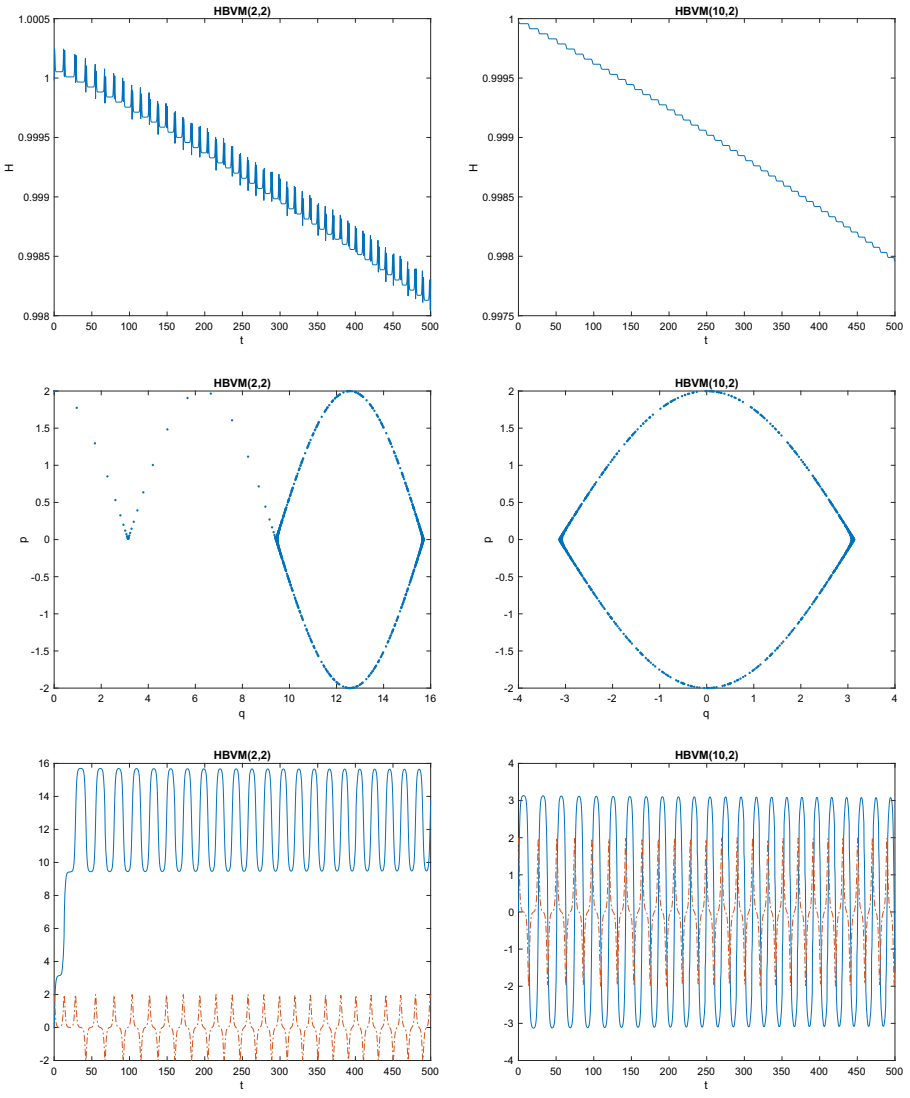

Fig. 4 Numerical results for problem (78) solved by using HBVM(2,2), left plots, and HBVM(10,2), right plots, using a timestep $h = 0.5$ (see the text for details)

5 Conclusions

In this paper we have fully developed a thorough approach for obtaining polynomial approximations to the solution of initial value ODE and DDE problems. It allows us to derive a wide class of Runge-Kutta methods, whose properties are easily discussed within the framework, as well as their actual implementation. Some numerical tests, concerning the numerical simulation of solutions of certain DDE problems of Hamiltonian type, confirm this. The present approach leaves room for generalizations

along several directions: in particular to different kind of problems, besides the ones considered here. Another relevant direction of investigation consists in looking for approximations belonging to functional subspaces different than polynomials: that is, by considering orthonormal functional bases different from (3). Both directions will be the subject of future investigations.

Funding Open access funding provided by Università degli Studi di Bari Aldo Moro within the CRUI-CARE Agreement. The authors received financial support from the *mrSIR* crowdfunding https://www.mrsir.it/en/about-us/.

Declarations

Conflict of interest The authors declare no competing interests

References

1. Amodio, P., Brugnano, L., Iavernaro, F.: A note on the continuous-stage Runge-Kutta-(Nyström) formulation of Hamiltonian Boundary Value Methods (HBVMs). Appl. Math. Comput. **363**, 124634 (2019). https://doi.org/10.1016/j.amc.2019.124634
2. Amodio, P., Brugnano, L., Iavernaro, F.: Continuous-Stage Runge-Kutta Approximation to differential problems. Axioms **11**, 192 (2022). https://doi.org/10.3390/axioms11050192
3. Amodio, P., Brugnano, L., Iavernaro, F.: Analysis of Spectral Hamiltonian Boundary Value Methods (SHBVMs) for the numerical solution of ODE problems. Numer Algorithms **83**, 1489–1508 (2020). https://doi.org/10.1007/s11075-019-00733-7
4. Bellen, A.: One step collocation for delay differential equations. J. Comput. Appl. Math. **10**, 275–283 (1984). https://doi.org/10.1016/0377-0427(84)90039-6
5. Bellen, A., Zennaro, M.: Numerical Methods for Delay Differential Equations. Clarendon Press, Oxford (2003)
6. Betsch, P., Steinmann, P.: Conservation properties of a time FE method. I. Time-stepping schemes for N-body problems. Internat. J. Numer. Methods Engrg. **49**, 599–638 (2000). https://doi.org/10.1002/1097-0207(20001020)49:5<599::AID-NME960>3.0.CO;2-9
7. Bottasso, C.L.: A new look at finite elements in time: a variational interpretation of Runge-Kutta methods. Appl. Numer. Math. **25**, 355–368 (1997). https://doi.org/10.1016/S0168-9274(97)00072-X
8. Brugnano, L., Frasca-Caccia, G., Iavernaro, F.: Efficient implementation of Gauss collocation and Hamiltonian Boundary Value Methods. Numer. Algorithms. **65**, 633–650 (2014). https://doi.org/10.1007/s11075-014-9825-0
9. Brugnano, L., Iavernaro, F.: Line Integral Methods for Conservative Problems. Chapman et Hall/CRC, Boca Raton (2016)
10. Brugnano, L., Iavernaro, F.: Line integral solution of differential problems. Axioms **7**(2), 36 (2018). https://doi.org/10.3390/axioms7020036
11. Brugnano, L., Iavernaro, F., Montijano, J.I., Rández, L.: Spectrally accurate space-time solution of Hamiltonian PDEs. Numer Algorithms **81**, 1183–1202 (2019). https://doi.org/10.1007/s11075-018-0586-z

12. Brugnano, L., Iavernaro, F., Trigiante, D.: Hamiltonian boundary value methods (energy preserving discrete line integral methods). JNAIAM J. Numer. Anal. Ind. Appl. Math. **5**(1-2), 17–37 (2010)
13. Brugnano, L., Iavernaro, F.: D. Trigiante. A note on the efficient implementation of Hamiltonian BVMs. J. Comput. Appl Math. **236**, 375–383 (2011). https://doi.org/10.1016/j.cam.2011.07.022
14. Brugnano, L., Iavernaro, F., Trigiante, D.: A simple framework for the derivation and analysis of effective one-step methods for ODEs. Appl. Math. Comput. **218**, 8475–8485 (2012). https://doi.org/10.1016/j.amc.2012.01.074
15. Brugnano, L., Iavernaro, F., Trigiante, D.: Analisys of Hamiltonian Boundary Value Methods (HBVMs): a class of energy-preserving Runge-Kutta methods for the numerical solution of polynomial Hamiltonian systems. Commun. Nonlinear Sci. Numer. Simul. **20**, 650–667 (2015). https://doi.org/10.1016/j.cnsns.2014.05.030
16. Brugnano, L., Iavernaro, F., Zanzottera, P.: A multiregional extension of the SIR model, with application to the COVID-19 spread in Italy. Math. Meth. Appl Sci. **44**, 4414–4427 (2021). https://doi.org/10.1002/mma.7039
17. Brugnano, L., Magherini, C.: Blended implementation of block implicit methods for ODEs. Appl. Numer Math. **42**, 29–45 (2002). https://doi.org/10.1016/S0168-9274(01)00140-4
18. Brugnano, L., Magherini, C.: Recent advances in linear analysis of convergence for splittings for solving ODE problems. Appl. Numer Math. **59**, 542–557 (2009). https://doi.org/10.1016/j.apnum.2008.03.008
19. Brugnano, L., Montijano, J.I., Rández, L.: On the effectiveness of spectral methods for the numerical solution of multi-frequency highly-oscillatory Hamiltonian problems. Numer. Algorithms **81**, 345–376 (2019). https://doi.org/10.1007/s11075-018-0552-9
20. Brunner, H.: Collocation Methods for Volterra Integral and Related Functional Equations. Cambridge University Press, Cambridge (2004)
21. Celledoni, E., McLachlan, R.I., McLaren, D., Owren, B., Quispel, G.R.W., Wright, W.M.: Energy preserving Runge-Kutta methods. m2AN. Math. Model. Numer. Anal. **43**, 645–649 (2009). https://doi.org/10.1051/m2an/2009020
22. Dahlquist, G., Björk, Å..: Numerical Methods in Scientific Computing, vol. I. SIAM, Philadelphia (2008)
23. Dos Reis, J.G., Baroni, R.L.: On the existence of periodic solutions for autonomous retarded functional-differential equations on R^2. Proc. Roy. Soc. Edinburgh Sect. A **102**, 259–262 (1986). https://doi.org/10.1017/S0308210500026342
24. Engel, E., Dreizler, R.M.: Density Functional Theory, an Advanced Course. Springer, Berlin (2011)
25. Furihata, D., T. Matsuo.: Discrete Variational Derivative Method: A Structure-Preserving Numerical Method for Partial Differential Equations. Chapman and Hall/CRC, Boca Raton (2010)
26. Hairer, E.: Energy-preserving variants of collocation methods. J.AIAM J. Numer. Anal. Ind. Appl. Math. **5**(1-2), 73–84 (2010)
27. Hairer, E., NøRsett, S.P., Wanner, G.: Solving Ordinary Differential Equations I, Nonstiff Problems. Second Revised Edition (3Rd Printing). Springer, Heidelberg (2008)
28. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations I Nonstiff Problems. Second Revised Edition. Springer, Heidelberg (2002)
29. Hulme, B.L.: One-step piecewise polynomial Galerkin methods for initial value problems. Math. Comp. **26**, 415–426 (1972). https://doi.org/10.1090/S0025-5718-1972-0321301-2
30. Hulme, B.L.: Discrete Galerkin and related one-step methods for ordinary differential equations. Math Comp. **26**, 881–891 (1972). https://doi.org/10.1090/S0025-5718-1972-0315899-8
31. Iavernaro, F., Pace, B.: s-Stage trapezoidal methods for the conservation of Hamiltonian functions of polynomial type. AIP Conf. Proc. **936**, 603–606 (2007). https://doi.org/10.1063/1.2790219
32. Iavernaro, F., Pace, B.: Conservative block-boundary value methods for the solution of polynomial Hamiltonian systems. AIP Conf. Proc. **1048**, 888–891 (2008). https://doi.org/10.1063/1.2991075
33. Iavernaro, F., Trigiante, D.: High-order symmetric schemes for the energy conservation of polynomial Hamiltonian problems. JNAIAM J. Numer. Anal. Ind. Appl Math. **4**(1-2), 87–111 (2009)
34. Kaplan, J.L., Yorke, J.A.: Ordinary differential equations which yield periodic solutions of differential delay equations. J. Math. Anal Appl. **48**, 317–324 (1974). https://doi.org/10.1016/0022-247X(74)90162-0
35. Mallet-Paret, J., Nussbaum, R.D.: Stability of periodic solutions of state-dependent delay-differential equations. J Diff. Equ. **250**, 4085–4103 (2011). https://doi.org/10.1016/j.jde.2010.10.023

36. Miyatake, Y., Butcher, J.C.: A characterization of energy-preserving methods and the construction of parallel integrators for Hamiltonian systems. SIAM J. Numer. Anal. **54**, 1993–2013 (2016). https://doi.org/10.1137/15M1020861
37. Nussbaum, R.D.: Periodic solutions of some nonlinear, autonomous functional differential equations. Bull. Amer. Math. Soc. **79**, 811–814 (1973). https://doi.org/10.1016/0022-0396(73)90053-3
38. Nussbaum, R.D.: Periodic solutions of some nonlinear, autonomous functional differential equations. II. J Diff. Equ. **14**, 360–394 (1973). https://doi.org/10.1090/S0002-9904-1973-13330-0
39. Nussbaum, R.D.: Uniqueness and nonuniqueness for periodic solutions of x', $(t) = -g(x(t-1))$. J Diff. Equ. **34**, 25–54 (1979). https://doi.org/10.1016/0022-0396(79)90016-0
40. Quispel, G.R.W., McLaren, D.I.: A new class of energy-preserving numerical integration methods. J. Phys. A **41**, 045206 (2008). https://doi.org/10.1088/1751-8113/41/4/045206
41. Walther, H.-O.: Existence of a non-constant periodic solution of a nonlinear autonomous functional differential equation representing the growth of a single species population. J. Math. Biol. **1**, 227–240 (1975). https://doi.org/10.1007/BF01273745