This is the pre-peer reviewed version of the following article: Arbitrarily High-Order Energy-Conserving Methods for Hamilton
with Quadratic Holonomic Constraints, Journal of Computational Mathematics 42 No. 4 (2024) ,
which has been published in final form at DOI:10.4208/jcm.2301-m2022-0065.
This article may be used for non-commercial purposes in accordance with Global Science Press Terms and
Conditions for Use of Self-Archived Versions

arXiv:2203.03275v1 [math.NA] 7 Mar 2022

Arbitrarily high-order energy-conserving methods for Hamiltonian problems with quadratic holonomic constraints

P. Amodio [*] L. Brugnano [†] G. Frasca-Caccia [‡]
F. Iavernaro [*]

March 8, 2022

Abstract

In this paper, we define arbitrarily high-order energy-conserving methods for Hamiltonian systems with quadratic holonomic constraints. The derivation of the methods is made within the so-called *line integral framework*. Numerical tests to illustrate the theoretical findings are presented.

Keywords: constrained Hamiltonian systems; quadratic holonomic constraints; energy-conserving methods; line integral methods; Hamiltonian Boundary Value Methods; HBVMs.

MSC: 65P10, 65L80, 65L06.

1 Introduction

In recent years, much interest has been given to the modeling and/or simulation of tethered systems, where the dynamics of interconnected bodies is studied (see, e.g. [26, 41, 24, 25, 43, 44, 40, 42, 2, 32, 33]). It turns out that the underlying dynamics is often described by a Hamiltonian system, for which the total energy is conserved.

Motivated by this fact, we here investigate the numerical approximation of a constrained Hamiltonian dynamics, described by the separable Hamiltonian

$$H(q,p) = \frac{1}{2}p^\top M^{-1}p - U(q), \qquad q,p \in \mathbb{R}^m, \tag{1}$$

where M is a symmetric and positive-definite (spd) matrix, subject to ν quadratic *holonomic* constraints,

$$g(q) = 0 \in \mathbb{R}^\nu, \qquad \nu \le m, \tag{2}$$

[*]Dipartimento di Matematica, Università di Bari, Italy.
[†]Dipartimento di Matematica e Informatica "U. Dini", Università di Firenze, Italy.
[‡]Dipartimento di Matematica, Università di Salerno, Italy.

i.e., the entries of g are quadratic polynomials. Hereafter, we shall assume all points be regular for the constraints, i.e., $\nabla g(q) \in \mathbb{R}^{m\times\nu}$ has full column rank or, equivalently,

$$\nabla g(q)^\top M^{-1} \nabla g(q) \in \mathbb{R}^{\nu\times\nu} \qquad \text{is spd.} \tag{3}$$

Moreover, we shall assume that its smallest eigenvalue is bounded away from 0, in the domain of interest. Also, for sake of simplicity, in the same domain the potential U will be assumed to be analytic.

It is well-known that the problem defined by (1)–(2) can be cast in Hamiltonian form by defining the augmented Hamiltonian

$$\hat{H}(q,p,\lambda) = H(q,p) + \lambda^\top g(q), \tag{4}$$

where λ is the vector of the Lagrange multipliers. The resulting constrained Hamiltonian system reads:

$$\dot{q} = M^{-1}p, \qquad \dot{p} = \nabla U(q) - \nabla g(q)\lambda, \qquad g(q) = 0, \qquad t \in [0,T], \tag{5}$$

and is subject to consistent initial conditions,

$$q(0) = q_0, \qquad p(0) = p_0, \tag{6}$$

such that

$$g(q_0) = 0, \qquad \nabla g(q_0)^\top M^{-1} p_0 = 0. \tag{7}$$

Clearly, $H(q,p) \equiv \hat{H}(q,p,\lambda)$, provided that the constraints (2) are satisfied, and a straightforward calculation proves that both are conserved along the solution trajectory.

We notice that the condition $g(q_0) = 0$ ensures that q_0 belongs to the manifold

$$\mathcal{M} = \{q \in \mathbb{R}^m : g(q) = 0\}, \tag{8}$$

as required by the constraints, whereas the condition $\nabla g(q_0)^\top M^{-1} p_0$ means that the motion initially stays on the tangent space to $\mathcal{M}$ at q_0. This condition is satisfied by all points on the solution trajectory, since, in order for the constraints to be conserved, the following condition needs to be satisfied as well:

$$\dot{g}(q) = \nabla g(q)^\top \dot{q} = \nabla g(q)^\top M^{-1} p = 0 \in \mathbb{R}^\nu. \tag{9}$$

These latter constraints are usually referred to as *hidden constraints*, and allow the derivation of the vector of the Lagrange multiplier λ. In fact, from (9) and (5)-(6), one obtains

$$\begin{aligned} 0 &= \nabla g(q(t))^\top M^{-1} p(t) \\ &= \nabla g(q(t))^\top M^{-1} \left[p_0 + \int_0^t \nabla U(q(\zeta))\mathrm{d}\zeta - \int_0^t \nabla g(q(\zeta))\lambda(\zeta)\mathrm{d}\zeta \right], \end{aligned} \tag{10}$$

from which one derives the integral equation:

$$\begin{aligned} &\nabla g(q(t))^\top M^{-1} \int_0^t \nabla g(q(\zeta))\lambda(\zeta)\mathrm{d}\zeta \\ &\quad = \nabla g(q(t))^\top M^{-1} \left[p_0 + \int_0^t \nabla U(q(\zeta))\mathrm{d}\zeta \right]. \end{aligned} \tag{11}$$

Remark 1 *From (10)-(11), one deduces that the vector of the Lagrange multipliers depends on the functions q and p. We shall denote this by using the notation:*

$$\lambda(t) := \lambda(q(t), p(t)). \tag{12}$$

We stress that the condition (9) can be conveniently relaxed for a numerical approximation: in fact, we only ask for $\nabla g(q)^\top M^{-1} p$ to be suitably small along the numerical solution. Consequently, when solving the problem on the interval $[0, h]$, we require that the approximations,

$$q_1 \approx q(h), \qquad p_1 \approx p(h), \tag{13}$$

satisfy the conservation of both the Hamiltonian and the constraints,

$$H(q_1, p_1) = H(q_0, p_0), \qquad g(q_1) = 0, \tag{14}$$

whereas the hidden constraints, see (9), become:

$$\nabla g(q(h))^\top M^{-1} p(h) = O(h^{\bar{r}}), \tag{15}$$

for a convenient $\bar{r} \geq 1$.

For sake of completeness, we recall that a formal expression for the vector λ is obtained by further differentiating (9):

$$\ddot{g}(q) = \nabla^2 g(q)(M^{-1}p, M^{-1}p) + \nabla g(q)^\top M^{-1} \left[\nabla U(q) - \nabla g(q)\lambda\right]. \tag{16}$$

In fact, by imposing the vanishing of this derivative yields

$$\left[\nabla g(q)^\top M^{-1} \nabla g(q)\right] \lambda \;=\; \nabla^2 g(q)(M^{-1}p, M^{-1}p) + \nabla g(q)^\top M^{-1} \nabla U(q). \tag{17}$$

Consequently, the following result follows.

Theorem 1 *The vector λ exists and is uniquely determined, provided that (3) holds true. In fact, in such a case, from (17), we obtain*

$$\lambda = \left[\nabla g(q)^\top M^{-1} \nabla g(q)\right]^{-1} \left[\nabla^2 g(q)(M^{-1}p, M^{-1}p) + \nabla g(q)^\top M^{-1} \nabla U(q)\right].$$

The numerical solution of Hamiltonian problems with holonomic constraints has been the subject of several investigations across the years. Among the proposed approaches, we mention: the popular Shake-Rattle method [38, 1], later proved to be symplectic [29]; higher order methods obtained via symplectic PRK methods [23]; composition methods [34, 35]; symmetric LMFs [17]; methods based on discrete derivatives [18]; local parametrizations of the manifold containing the solution [3]; projection techniques [36, 39]. As additional references, we refer to [27, 37, 19, 30, 31] and to the monographs [4, 28, 20, 21], and references therein.

In this paper we pursue a different approach, utilizing the so-called *line integral framework*, which has already been used for deriving the energy-conserving

Runge-Kutta methods, for unconstrained Hamiltonian systems, named *Hamiltonian Boundary Value Methods (HBVMs)* [9, 10, 11, 12, 13], see also the monograph [7] and the review paper [8]. This approach has been already used in [6] for deriving energy-conserving methods for constrained Hamiltonian problems which, however, are in general only second-order accurate. In this paper we propose a class of arbitrarily high-order energy conserving methods for constrained Hamiltonian dynamics, in the relevant case where the constraints are quadratic. As matter of fact, the proposed methods provide approximations (13) satisfying (14) and(15) with $\bar{r}$ arbitrarily large.

With this premise, the structure of the paper is as follows: in Section 2 we define the polynomial approximation defining the basic step of application of the method also studying its conservation properties; in Section 3 we study the order of convergence of the method; in Section 4 we discuss its fully discrete implementation; numerical examples are reported in Section 5; at last, some concluding remarks are given in Section 6.

2 Polynomial approximation

Let us consider the following orthonormal polynomial basis,[1]

$$P_j \in \Pi_j, \qquad \int_0^1 P_i(x)P_j(x)\mathrm{d}x = \delta_{ij}, \qquad \forall i,j = 0,1,\ldots, \tag{18}$$

given by the shifted and scaled Legendre polynomials, and the infinite expansions on the interval $[0,h]$ (see (5)):

$$p(ch) = \sum_{j\geq 0} P_j(c)\gamma_j(p), \quad \nabla U(q(ch)) = \sum_{j\geq 0} P_j(c)\psi_j(q), \quad c\in[0,1], \tag{19}$$

where, for all $j = 0,1,\ldots$,

$$\gamma_j(p) = \int_0^1 P_j(\zeta)p(\zeta h)\mathrm{d}\zeta, \qquad \psi_j(q) = \int_0^1 P_j(\zeta)\nabla U(q(\zeta h))\mathrm{d}\zeta. \tag{20}$$

In order to derive polynomial approximations u and v of degree s,

$$u(ch)\approx q(ch), \qquad v(ch)\approx p(ch), \qquad c\in[0,1],$$

we shall consider the following approximation to (5) on the interval $[0,h]$,

$$\begin{aligned}
\dot{u}(ch) &= M^{-1}\sum_{j=0}^{s-1} P_j(c)\gamma_j(v), \qquad u(0)=q_0, \qquad c\in[0,1],\\
\dot{v}(ch) &= \sum_{j=0}^{s-1} P_j(c)\psi_j(u) - \sum_{j=1}^{s}\ell_j(c)\rho_j, \qquad v(0)=p_0,\\
\rho_j &:= \nabla g(u(c_j h))\lambda_j, \qquad j=1,\ldots,s,
\end{aligned} \tag{21}$$

[1] As is usual, Π_j is the vector space of polynomials of degree j, and δ_{ij} is the Kronecker symbol.

where:

- $\gamma_j(v)$ and $\psi_j(u)$ are defined according to (20), by formally replacing q and p with u and v, respectively;
- $0 < c_1 < \ldots < c_s < 1$ are the zeros of $P_s(c)$, $P_s(c_j) = 0$, $j = 1, \ldots, s$;
- $\ell_j(c)$ are the Lagrange polynomials defined over the previous abscissae;
- finally, by using the notation (12),

$$\lambda_j \approx \bar{\lambda}(c_j h) := \lambda(u(c_j h), v(c_j h)) \tag{22}$$

  are approximation to the Lagrange multipliers, which will be defined in the sequel.

Moreover, let us denote by

$$b_j = \int_0^1 \ell_j(c)\mathrm{d}c, \qquad j = 1, \ldots, s,$$

the weights of the Gauss-Legendre quadrature of order $2s$. The new approximations,

$$q_1 := u(h), \qquad p_1 := v(h), \tag{23}$$

will be then defined such that (14) and (15) are satisfied, for a convenient $\bar{r}$ depending on s.

For later use, we observe that, at the Gauss-Legendre abscissae $\{c_i\}$, (21) reads:

$$\begin{aligned} \dot{u}(c_i h) &= M^{-1} v(c_i h), \\ \dot{v}(c_i h) &= \sum_{j=0}^{s-1} P_j(c_i)\psi_j(u) - \nabla g(u(c_i h))\lambda_i, \qquad i = 1, \ldots, s. \end{aligned} \tag{24}$$

In fact, since $v(ch) \in \Pi_s$, then (see (20))

$$v(ch) = \sum_{j=0}^{s} P_j(c)\gamma_j(v), \qquad c \in [0, 1],$$

however, at the zeros of P_s, one obviously has:

$$v(c_i h) = \sum_{j=0}^{s-1} P_j(c_i)\gamma_j(v), \qquad i = 1, \ldots, s. \tag{25}$$

2.1 Constraints conservation

Let us now study the conservation of the holonomic constraints in (14), i.e. under which conditions $g(q_1) = 0$. One has, by virtue of (24):

$$\begin{aligned} g(q_1) \;\equiv\; g(u(h)) \;&=\; h\int_0^1 \nabla g(u(ch))^\top \dot{u}(ch)\mathrm{d}c \\ &=\; h\sum_{i=1}^{s} b_i \nabla g(u(c_i h))^\top \dot{u}(c_i h) \;=\; h\sum_{i=1}^{s} b_i \nabla g(u(c_i h))^\top M^{-1} v(c_i h), \end{aligned}$$

where the third equality follows from the fact that the integrand has degree $2s-1$, and the last equality follows from (24). Consequently, the following result is proved.

Theorem 2 *$g(q_1) = 0$, provided that:*

$$\nabla g(u(c_i h))^\top M^{-1} v(c_i h) = 0 \in \mathbb{R}^\nu, \qquad i = 1, \ldots, s. \tag{26}$$

2.2 Energy conservation

Let us now study the conservation of the Hamiltonian function (1). By using the usual line integral argument, from (21) one has:

$$\begin{aligned} H(q_1,p_1) - H(q_0,p_0) \;&\equiv\; H(u(h),v(h)) - H(u(0),v(0)) \\ &=\; \int_0^h \frac{\mathrm{d}}{\mathrm{d}t} H(u(t),v(t))\mathrm{d}t \\ &=\; h\int_0^1 \left[H_q(u(ch),v(ch))^\top \dot{u}(ch) + H_p(u(ch),v(ch))^\top \dot{v}(ch)\right]\mathrm{d}c \\ &=\; h\int_0^1 \left[-\nabla U(u(ch))^\top M^{-1}\sum_{j=0}^{s-1} P_j(c)\gamma_j(v) + v(ch)^\top M^{-1}\sum_{j=0}^{s-1} P_j(c)\psi_j(u)\right. \\ &\qquad\qquad \left. -\; v(ch)^\top M^{-1}\sum_{j=1}^{s}\ell_j(c)\rho_j\right]\mathrm{d}c \\ &=\; -h\sum_{j=0}^{s-1}\overbrace{\left[\int_0^1 P_j(c)\nabla U(u(ch))\mathrm{d}c\right]^\top}^{=\psi_j(u)^\top} M^{-1}\gamma_j(v) \\ &\quad + h\sum_{j=0}^{s-1}\overbrace{\left[\int_0^1 P_j(c)v(ch)\mathrm{d}c\right]^\top}^{=\gamma_j(v)^\top} M^{-1}\psi_j(u) - h\int_0^1\sum_{j=1}^{s} v(ch)^\top M^{-1}\ell_j(c)\rho_j\mathrm{d}c \end{aligned}$$

$$= -h\sum_{j=0}^{s-1}\psi_j(u)^\top M^{-1}\gamma_j(v) + h\sum_{j=0}^{s-1}\gamma_j(v)^\top M^{-1}\psi_j(u) - h\sum_{i=1}^{s} b_i v(c_i h)^\top M^{-1}\rho_i$$

$$= -h\sum_{i=1}^{s} b_i v(c_i h)^\top M^{-1}\nabla g(u(c_i h))\lambda_i.$$

In fact, considering that the Gauss-Legendre formula has order $2s$,

$$\int_0^1 \sum_{j=1}^{s} v(ch)^\top M^{-1}\ell_j(c)\rho_j \mathrm{d}c = \sum_{i,j=1}^{s} b_i v(c_i h)^\top M^{-1}\ell_j(c_i)\rho_j$$
$$= \sum_{i=1}^{s} b_i v(c_i h)^\top M^{-1}\rho_i = \sum_{i=1}^{s} b_i v(c_i h)^\top M^{-1}\nabla g(u(c_i h))\lambda_i.$$

Consequently, the following result is proved.

Theorem 3 *Assume that the conditions (26) hold true. Then, $H(q_1, p_1) = H(q_0, p_0)$, i.e., the energy is conserved.*

2.3 Computing the approximate Lagrange multipliers

Before studying the accuracy of the approximations (13) and (23), let us derive the algebraic formulation of the conditions (26), i.e., the discrete analog of (10) at $t = c_i h$, $i = 1, \ldots, s$. For this purpose, let us define the following matrices,

$$\begin{aligned}
\nabla \boldsymbol{g} &= \begin{pmatrix} \nabla g(u(c_1 h)) & & \\ & \ddots & \\ & & \nabla g(u(c_s h)) \end{pmatrix} \in \mathbb{R}^{sm\times s\nu} \\
\mathcal{P}_s &= \begin{pmatrix} P_0(c_1) & \ldots & P_{s-1}(c_1) \\ \vdots & & \vdots \\ P_0(c_s) & \ldots & P_{s-1}(c_s) \end{pmatrix} \in \mathbb{R}^{s\times s}, \qquad (27) \\
\mathcal{I}_s &= \begin{pmatrix} \int_0^{c_1} P_0(x)\mathrm{d}x & \ldots & \int_0^{c_1} P_{s-1}(x)\mathrm{d}x \\ \vdots & & \vdots \\ \int_0^{c_s} P_0(x)\mathrm{d}x & \ldots & \int_0^{c_s} P_{s-1}(x)\mathrm{d}x \end{pmatrix} \in \mathbb{R}^{s\times s}, \\
\Omega &= \mathrm{diag}(b_1, \ldots, b_s) \in \mathbb{R}^{s\times s},
\end{aligned}$$

and vectors,

$$\boldsymbol{\psi} = \begin{pmatrix} \psi_0(u) \\ \vdots \\ \psi_{s-1}(u) \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} v(c_1 h) \\ \vdots \\ v(c_s h) \end{pmatrix} \in \mathbb{R}^{sm},$$

$$\begin{aligned}\boldsymbol{\lambda} &= \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_s \end{pmatrix}, \ \mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}^\top \in \mathbb{R}^{s\nu}, \\ \boldsymbol{e} &= (1, \ldots, 1)^\top \in \mathbb{R}^s. \end{aligned} \tag{28}$$

In so doing, it turns out that (26) can be cast as:

$$\nabla \boldsymbol{g}^\top I_s \otimes M^{-1} \, \boldsymbol{v} = \mathbf{0}.$$

Moreover, from (21) one derives:

$$\boldsymbol{v} = \boldsymbol{e} \otimes p_0 + h\mathcal{I}_s \otimes I_m \boldsymbol{\psi} - h\mathcal{I}_s \mathcal{P}_s^\top \Omega \otimes I_m \nabla \boldsymbol{g} \cdot \boldsymbol{\lambda}.$$

Consequently, combining the two last equations, one formally obtains:

$$\left(\nabla \boldsymbol{g}^\top \, h\mathcal{I}_s \mathcal{P}_s^\top \Omega \otimes M^{-1} \, \nabla \boldsymbol{g}\right) \boldsymbol{\lambda} = \nabla \boldsymbol{g}^\top \left(\boldsymbol{e} \otimes M^{-1} p_0 + h\mathcal{I}_s \otimes M^{-1} \boldsymbol{\psi}\right), \tag{29}$$

which, as observed above, represents the discrete counterpart of (11) at $t = c_i h$, $i = 1, \ldots, s$. Consequently, the following result is proved.

Corollary 1 *Conditions (26) are equivalent to (27)–(29), i.e.,*

$$\begin{aligned} & h\nabla g(u(c_i h))^\top M^{-1} \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x) \mathrm{d}x \sum_{j=1}^{s} b_j P_\ell(c_j) \nabla g(u(c_j h)) \lambda_j \\ \equiv \quad & h\nabla g(u(c_i h))^\top M^{-1} \sum_{j=1}^{s} b_j \overbrace{\sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x) \mathrm{d}x \, P_\ell(c_j)}^{= a_{ij}} \nabla g(u(c_j h)) \lambda_j \qquad (30) \\ = \quad & \nabla g(u(c_i h))^\top M^{-1} \left[p_0 + h \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x) \mathrm{d}x \, \psi_j(u) \right], \qquad i = 1, \ldots, s. \end{aligned}$$

We observe that a_{ij} in (30) is nothing but the $(i,j)th$ entry of the Butcher matrix of the s-stage Gauss method (see, e.g., [7, 8]). Moreover, the problem (29) is well-posed, since $\nabla \boldsymbol{g}$ has full column rank, by hypothesis, and

$$\mathcal{I}_s = \mathcal{P}_s X_s, \qquad \mathcal{P}_s^\top \Omega = \mathcal{P}_s^{-1}, \tag{31}$$

with

$$X_s = \begin{pmatrix} \xi_0 & -\xi_1 & & \\ \xi_1 & 0 & \ddots & \\ & \ddots & \ddots & -\xi_{s-1} \\ & & \xi_{s-1} & 0 \end{pmatrix}, \qquad \xi_i = \left(2\sqrt{|4i^2 - 1|}\right)^{-1}, \tag{32}$$

a nonsingular matrix. Let us now study, for later use, the differences (see (22)):

$$\delta\lambda_j := \bar{\lambda}(c_j h) - \lambda_j, \qquad j = 1, \ldots, s. \tag{33}$$

For this purpose, we need the following lemma.

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

$$\int_0^1 P_j(c)G(ch)\mathrm{d}c = O(h^j), \qquad j = 0, 1, \ldots.$$

Proof See [12, Lemma 1]. □

We also recall that, for any suitably regular function G,

$$\begin{aligned} \int_0^1 P_j(c)G(ch)\mathrm{d}c &= \sum_{i=1}^{s} b_i P_j(c_i)G(c_i h) + \Delta_j(h), \\ &\quad \Delta_j(h) = O(h^{2s-j}), \qquad j = 0, \ldots, s-1, \end{aligned} \tag{34}$$

(c_i, b_i) being, as usual, the abscissae and weights of the Gauss-Legendre quadrature of order $2s$. We are now able to prove the following result.

Theorem 4 *Concerning the differences (33), one has:*

$$\delta\lambda_j \equiv \bar{\lambda}(c_j h) - \lambda_j = O(h^s).$$

Proof In fact, from (11) with q and p replaced by u and v, respectively, and evaluated at $t = c_i h$, one obtains:

$$\nabla g(u(c_i h))^\top M^{-1} v(c_i h) = 0, \qquad i = 1, \ldots, s,$$

i.e.,

$$\begin{aligned} &\nabla g(u(c_i h))^\top M^{-1} h \int_0^{c_i} \nabla g(u(\xi h))\bar{\lambda}(\xi h)\mathrm{d}\xi \\ &\quad = \nabla g(u(c_i h))^\top M^{-1} \left[p_0 + h \int_0^{c_i} \nabla U(u(\xi h))\mathrm{d}\xi \right], \qquad i = 1, \ldots, s. \end{aligned} \tag{35}$$

Considering that $\nabla g(u) \in \Pi_s$, and by virtue of Lemma 1 and (34), one has:

$$\begin{aligned} &\int_0^{c_i} \nabla g(u(\xi h))\bar{\lambda}(\xi h)\mathrm{d}\xi \\ &\quad = \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \int_0^1 P_\ell(\tau)\nabla g(u(\tau h))\bar{\lambda}(\tau h)\mathrm{d}\tau + O(h^s) \\ &\quad = \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \left[\sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u(c_j h))\bar{\lambda}(c_j h) + O(h^{2s-\ell}) \right] + O(h^s) \\ &\quad = \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u(c_j h))\bar{\lambda}(c_j h) + O(h^s). \end{aligned} \tag{36}$$

Similarly, from (19)-(20), one has:

$$\begin{aligned}\int_0^{c_i} \nabla U(u(\xi h))\mathrm{d}\xi &= \sum_{j\geq 0}\int_0^{c_i} P_j(x)\mathrm{d}x\,\psi_j(u) \\ &= \sum_{j=0}^{s-1}\int_0^{c_i} P_j(x)\mathrm{d}x\,\psi_j(u) + O(h^s). \qquad (37)\end{aligned}$$

From (35)–(37) one then obtains:

$$h\nabla g(u(c_i h))^\top M^{-1}\sum_{\ell=0}^{s-1}\int_0^{c_i} P_\ell(x)\mathrm{d}x\sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u(c_j h))\bar{\lambda}(c_j h) \qquad (38)$$

$$= \nabla g(u(c_i h))^\top M^{-1}\left[p_0 + h\sum_{j=0}^{s-1}\int_0^{c_i} P_j(x)\mathrm{d}x\,\psi_j(u) + O(h^{s+1})\right],\ i=1,\ldots,s.$$

Subtracting, side by side, (30) from (38) then gives:

$$\nabla g(u(c_i h))^\top M^{-1}\sum_{\ell=0}^{s-1}\int_0^{c_i} P_\ell(x)\mathrm{d}x\sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u(c_j h))\delta\lambda_j = O(h^s),$$
$$i=1,\ldots,s,$$

from which the statement follows. In fact, the previous equations can be rewritten as (see (29))

$$\left(\nabla \boldsymbol{g}^\top \mathcal{I}_s \mathcal{P}_s^\top \Omega \otimes M^{-1}\nabla \boldsymbol{g}\right)\delta\boldsymbol{\lambda} = O(h^s),$$

having set $\delta\boldsymbol{\lambda} = \left(\begin{array}{ccc}\delta\lambda_1^\top & \ldots & \delta\lambda_1^\top\end{array}\right)^\top$. $\square$

3 Accuracy

Let us now study the accuracy of the approximations (13) and (23). For this purpose, let us denote by [2]

$$y(t) := \begin{pmatrix} q(t) \\ p(t)\end{pmatrix} \equiv y(t,\xi,\eta), \qquad \eta = \begin{pmatrix}\eta_1 \\ \eta_2\end{pmatrix} \in \mathbb{R}^{2m}, \qquad (39)$$

the solution of the problem

$$\dot{y} \equiv \begin{pmatrix}\dot{q} \\ \dot{p}\end{pmatrix} = \begin{pmatrix} M^{-1}p \\ \nabla U(q) - \nabla g(q)\lambda\end{pmatrix} =: f(y), \qquad t>\xi, \qquad y(\xi)=\eta, \quad (40)$$

where we assume that, according to (12) $\lambda(t) = \lambda(q(t),p(t))$ is a known function, given by (11). Accordingly, we have denoted by $\bar{\lambda}(t)$ that defined in (22). We here report standard perturbation results for the solution of (40), w.r.t. its parameters (see, e.g., [22, 7]).

[2] We shall use either one of such notations, depending on the needs.

Lemma 2 *With reference to (39)-(40), one has:*

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

where

$$\Phi(t,\xi,\eta) = \frac{\partial}{\partial \eta}y(t) \in \mathbb{R}^{2m\times 2m}.$$

We also need the expansion, with reference to (22),

$$\begin{aligned} \nabla g(u(ch))\bar{\lambda}(ch) &= \sum_{j\geq 0} P_j(c)\Gamma_j, \qquad c\in[0,1], \\ \Gamma_j &= \int_0^1 P_j(\tau)\nabla g(u(\tau h))\bar{\lambda}(\tau h)\mathrm{d}\tau, \qquad j=0,1,\ldots, \end{aligned} \tag{41}$$

and the following result.

Lemma 3 *The approximation procedure (20)–(23) and (26) is symmetric. Therefore, its convergence order is even.*

Proof The first part of the statement can be proved by using arguments similar to those used in [7, Section 3.3.4], taking into account that the Gauss-Legendre abscissae $\{c_1,\ldots,c_s\}$ are symmetrically distributed in the interval $[0,1]$. The second part then follows from [20, Theorem 3.2]. □

We can now discuss the accuracy of the method, following the framework defined in [12].

Theorem 5 *With reference to (20)–(23) and (26), one has that the approximation procedure has convergence order r, i.e.,*

$$q(h) - q_1 = O(h^{r+1}), \qquad p(h) - p_1 = O(h^{r+1}),$$

where $r = s$ if s is even, or $r = s+1$, when s is odd. In case (see (22) and (33))

$$\delta\lambda_i = 0, \qquad i = 1,\ldots,s, \tag{42}$$

then $r = 2s$.

Proof By using the notation (39)-(40), the results of Theorem 4, Lemmas 1–3, (41), (34), (21), and setting

$$y_i = \begin{pmatrix} q_i \\ p_i \end{pmatrix}, \quad i=0,1, \qquad \sigma(t) = \begin{pmatrix} u(t) \\ v(t) \end{pmatrix}, \quad t\in[0,h], \tag{43}$$

we obtain:

$$
\begin{aligned}
y(h) - y_1 = y(h,0,y_0) - y(h,h,y_1) &= y(h,0,\sigma(0)) - y(h,h,\sigma(h)) \\
&= \int_h^0 \frac{\mathrm{d}}{\mathrm{d}t} y(h,t,\sigma(t))\mathrm{d}t \\
&= \int_h^0 \left[\frac{\partial}{\partial \xi} y(h,\xi,\sigma(t))\Big|_{\xi=t} + \frac{\partial}{\partial \eta} y(h,t,\eta)\Big|_{\eta=\sigma(t)} \dot{\sigma}(t) \right] \mathrm{d}t \\
&= \int_0^h \Phi(h,t,\sigma(t)) \left[f(\sigma(t)) - \dot{\sigma}(t) \right] \mathrm{d}t \\
&= h \int_0^1 \Phi(h,ch,\sigma(ch)) \left[f(\sigma(ch)) - \dot{\sigma}(ch) \right] \mathrm{d}c \\
&= h \int_0^1 \Phi(h,ch,\sigma(ch)) \left[\sum_{j\geq 0} P_j(c) \begin{pmatrix} M^{-1}\gamma_j(v) \\ \psi_j(u) - \Gamma_j \end{pmatrix} \right. \\
&\quad \left. - \sum_{j=0}^{s-1} P_j(c) \begin{pmatrix} M^{-1}\gamma_j(v) \\ \psi_j(u) \end{pmatrix} + \sum_{j=1}^{s} \ell_j(c) \begin{pmatrix} 0 \\ \rho_j \end{pmatrix} \right] \mathrm{d}c \;=:\; (*).
\end{aligned}
$$

Considering that, from the result of Theorem 4,

$$
\begin{aligned}
&\sum_{j=1}^{s} \ell_j(c) \begin{pmatrix} 0 \\ \rho_j \end{pmatrix} = \sum_{j=1}^{s} \ell_j(c) \begin{pmatrix} 0 \\ \nabla g(u(c_j h))\lambda_j \end{pmatrix} \\
&= \sum_{j=0}^{s-1} P_j(c) \begin{pmatrix} 0 \\ \sum_{i=1}^{s} b_i P_j(c_i) \nabla g(u(c_i h))\lambda_i \end{pmatrix} = \sum_{j=0}^{s-1} P_j(c) \begin{pmatrix} 0 \\ \Gamma_j + O(h^s) \end{pmatrix},
\end{aligned}
$$

it follows that

$$
\begin{aligned}
(*) \;=\; & h \underbrace{\sum_{j=0}^{s-1} \overbrace{\int_0^1 P_j(c)\Phi(h,ch,\sigma(ch))\mathrm{d}c}^{=O(h^j)} \begin{pmatrix} 0 \\ O(h^s) \end{pmatrix}}_{=O(h^s)} + \\
& h \overbrace{\sum_{j\geq s} \underbrace{\int_0^1 P_j(c)\Phi(h,ch,\sigma(ch))\mathrm{d}c}_{=O(h^j)} \underbrace{\begin{pmatrix} M^{-1}\gamma_j(v) \\ \psi_j(u) - \Gamma_j \end{pmatrix}}_{=O(h^j)}}^{=O(h^{2s})} \;=\; O(h^{s+1}).
\end{aligned}
$$

Consequently, the statement follows by considering that $r = s$ at least, and, indeed, $r = s+1$ when s is odd, according to Lemma 3.

In the case where (42) holds true, then (see (41)-(34))

$$
\sum_{i=1}^{s} b_i P_j(c_i) \nabla g(u(c_i h))\lambda_i = \Gamma_j + \Delta_j(h), \qquad j = 0, \ldots, s-1,
$$

so that

$$\sum_{j=1}^{s} \ell_j(c) \begin{pmatrix} 0 \\ \rho_j \end{pmatrix} = \sum_{j=0}^{s-1} P_j(c) \begin{pmatrix} 0 \\ \Gamma_j + \Delta_j(h) \end{pmatrix},$$

and, consequently,

$$(*) \;=\; \underbrace{h \sum_{j=0}^{s-1} \overbrace{\int_0^1 P_j(c)\Phi(h, ch, \sigma(ch))\mathrm{d}c}^{=O(h^j)} \overbrace{\begin{pmatrix} 0 \\ \Delta_j(h) \end{pmatrix}}^{=O(h^{2s-j})}}_{=O(h^{2s})} +$$

$$\overbrace{h \sum_{j \geq s} \underbrace{\int_0^1 P_j(c)\Phi(h, ch, \sigma(ch))\mathrm{d}c}_{=O(h^j)} \underbrace{\begin{pmatrix} M^{-1}\gamma_j(v) \\ \psi_j(u) - \Gamma_j \end{pmatrix}}_{=O(h^j)}}^{=O(h^{2s})} \;=\; O(h^{2s+1}). \square$$

Corollary 2 *In the hypothesis of the previous Theorem 5, and using the notation (39) and (43), one has:*[3]

$$\|y - \sigma\| := \max_{\tau \in [0,1]} |y(\tau h) - \sigma(\tau h)| = O(h^{s+1}).$$

Proof In fact, one has:

$$y(\tau h) - \sigma(\tau h) = y(\tau h, 0, \sigma(0)) - y(\tau h, \tau h, \sigma(\tau h)) = \int_{\tau h}^{0} \frac{\mathrm{d}}{\mathrm{d}t} y(\tau h, t, \sigma(t))\mathrm{d}t =: (**).$$

By repeating similar steps as those in the proof of Theorem 5, one then obtains:

$$(**) \;=\; \underbrace{h \sum_{j=0}^{s-1} \overbrace{\int_0^\tau P_j(c)\Phi(\tau h, ch, \sigma(ch))\mathrm{d}c}^{=O(h^0)} \begin{pmatrix} 0 \\ O(h^s) \end{pmatrix}}_{=O(h^s)} +$$

$$\overbrace{h \sum_{j \geq s} \underbrace{\int_0^\tau P_j(c)\Phi(\tau h, ch, \sigma(ch))\mathrm{d}c}_{=O(h^0)} \underbrace{\begin{pmatrix} M^{-1}\gamma_j(v) \\ \psi_j(u) - \Gamma_j \end{pmatrix}}_{=O(h^j)}}^{=O(h^s)} \;=\; O(h^{s+1}). \square$$

3.1 Hidden constraints and Lagrange multipliers

We can now confirm that the value of $\bar{r}$ in (15) has to be at least equal to the order of the method, as stated in Theorem 5. In fact, at the first step one

[3] Hereafter, $|\cdot|$ denotes any convenient vector norm.

obtains:

$$\begin{aligned} 0 &= \nabla g(q(h))^\top M^{-1} p(h) \\ &= \left[\nabla g(u(h)) + O(h^{r+1})\right]^\top M^{-1} \left[v(h) + O(h^{r+1})\right] \\ &= \nabla g(u(h))^\top M^{-1} v(h) + O(h^{r+1}). \end{aligned}$$

As is clear, at the generic nth step, setting $t_n = nh$, and

$$\begin{aligned} u_n(ch) &\approx q_n(ch) := q(t_n + ch), \\ v_n(ch) &\approx p_n(ch) := p(t_n + ch), \qquad c \in [0,1], \end{aligned} \tag{44}$$

the polynomial approximations and the restriction of the solution in the interval $[t_n, t_{n+1}]$, since $u_n(h) = q_n(h) + O(h^r)$ and $v_n(h) = p_n(h) + O(h^r)$, one has:

$$\nabla g(u_n(h))^\top M^{-1} v_n(h) = O(h^r).$$

Similarly, by using the same notation above, we can discuss the accuracy of the approximate Lagrange multipliers,

$$\lambda_i^n \approx \lambda_n(c_i h) := \lambda(t_n + c_i h), \qquad i = 1, \ldots, s, \tag{45}$$

The following result holds true.

Theorem 6 *With reference to (20)–(23) and (26), assume that the procedure has convergence order r, as stated in Theorem 5, in the case where (42) does not hold. Then, the Langrange multipliers are approximated with at least order $r-1$.*

Proof We recall that, in the considered case, the order of the method is $r = s$, if s is even, or $r = s+1$, otherwise. With reference to (44)-(45), the approximate Lagrange multipliers then satisfy:

$$\begin{aligned} & h\nabla g(u_n(c_i h))^\top M^{-1} \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \sum_{j=1}^{s} b_j P_\ell(c_j) \nabla g(u_n(c_j h)) \lambda_j^n \qquad (46) \\ &= \nabla g(u_n(c_i h))^\top M^{-1} \left[v_n(0) + h \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, \psi_j(u_n) \right], \quad i = 1, \ldots, s. \end{aligned}$$

On the other hand, from (11) one deduces that the exact multipliers satisfy:

$$\begin{aligned} & h\nabla g(q_n(c_i h))^\top M^{-1} \sum_{\ell \geq 0} \int_0^{c_i} P_\ell(x)\mathrm{d}x \int_0^1 P_\ell(c) \nabla g(q_n(ch)) \lambda_n(ch)\mathrm{d}c \\ &= \nabla g(q_n(c_i h))^\top M^{-1} \left[p_n(0) + h \sum_{j \geq 0} \int_0^{c_i} P_j(x)\mathrm{d}x\, \psi_j(q_n) \right], \quad i = 1, \ldots, s. \end{aligned}$$

Considering that, by virtue of Corollary 2,

$$\|u_n - q_n\| = \|v_n - p_n\| = O(h^r) + O(h^{s+1}) = O(h^r),$$

from the last equation one derives:

$$\begin{aligned} & h\nabla g(u_n(c_i h))^\top M^{-1} \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \int_0^1 P_\ell(ch)\nabla g(u_n(ch))\lambda_n(ch)\mathrm{d}c \\ = \quad & \nabla g(u_n(c_i h))^\top M^{-1} \left[p_n(0) + h \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, \psi_j(u_n) \right] + O(h^r), \\ & i = 1, \ldots, s, \end{aligned}$$

and, therefore,

$$\begin{aligned} & h\nabla g(u_n(c_i h))^\top M^{-1} \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u_n(c_j h))\lambda_n(c_j h) \\ = \quad & \nabla g(u_n(c_i h))^\top M^{-1} \left[v_n(0) + h \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, \psi_j(u_n) \right] + O(h^r), \\ & i = 1, \ldots, s. \end{aligned} \tag{47}$$

From (46) and (47), one eventually obtains:

$$\begin{aligned} & \nabla g(u_n(c_i h))^\top M^{-1} \sum_{\ell=0}^{s-1} \int_0^{c_i} P_\ell(x)\mathrm{d}x \sum_{j=1}^{s} b_j P_\ell(c_j)\nabla g(u_n(c_j h))(\lambda_n(c_j h) - \lambda_j^n) \\ = \quad & O(h^{r-1}), \qquad i = 1, \ldots, s, \end{aligned}$$

from which the statement follows. □

Remark 2 *Clearly, in the more favourable case where (42) holds true, the accuracy of the approximate Lagrange multipliers could be better.*

4 Discretization

For the actual implementation of the approximation procedure (20)–(23) let us recast it into a more operative form:

$$\begin{aligned} u(c_i h) &= q_0 + hM^{-1} \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, \gamma_j(v), \qquad (48) \\ v(c_i h) &= p_0 + h \sum_{j=0}^{s-1} \int_0^{c_i} P_j(x)\mathrm{d}x\, \left[\psi_j(u) - \zeta_j(u, \boldsymbol{\lambda})\right], \qquad i = 1, \ldots, s, \end{aligned}$$

where, for $j = 0, \ldots, s-1$,

$$\begin{aligned}
\gamma_j(v) &= \int_0^1 P_j(c) v(ch) \mathrm{d}c \equiv \sum_{i=1}^{s} b_i P_j(c_i) v(c_i h), \\
\psi_j(u) &= \int_0^1 P_j(c) \nabla U(u(ch)) \mathrm{d}c, \\
\zeta_j(u, \boldsymbol{\lambda}) &= \int_0^1 P_j(c) \sum_{i=1}^{s} \ell_j(c) \nabla g(u(c_i h)) \lambda_i \equiv \sum_{i=1}^{s} b_i P_j(c_i) \nabla g(u(c_i h)) \lambda_i,
\end{aligned} \tag{49}$$

and the vector $\boldsymbol{\lambda}$, defined in (28), is the solution of (29). Moreover, should U be a quadratic potential, then one could exactly compute

$$\psi_j(u) = \sum_{i=1}^{s} b_i P_j(c_i) \nabla U(u(c_i h)), \qquad j = 0, \ldots, s-1,$$

but, in general, this could be not the case: as matter of fact, many potentials are rational functions. Consequently, we need to use a more accurate quadrature, for approximating the Fourier coefficients $\psi_j(u)$, $j = 0, \ldots, s-1$. For this purpose, following the same approach used for HBVM(k, s) methods (see, e.g., [7, 8]) we can use a Gauss-Legendre quadrature of order $2k \geq 2s$, with weights and abscissae

$$(\hat{c}_i, \hat{b}_i), \qquad i = 1, \ldots, k,$$

thus obtaining:

$$\begin{aligned}
\psi_j(u) &= \sum_{i=1}^{k} \hat{b}_i P_j(\hat{c}_i) \nabla U(u(\hat{c}_i h)) + \hat{\Delta}_j(h) \\
&=: \hat{\psi}_j(u) + \hat{\Delta}_j(h), \qquad j = 0, \ldots, s-1,
\end{aligned} \tag{50}$$

with

$$\hat{\Delta}_j(h) = \begin{cases} 0, & \text{if } U \in \Pi_\mu, \text{ with } \mu \leq 2k/s, \\ O(h^{2k-j}), & \text{otherwise.} \end{cases} \tag{51}$$

Remark 3 *We shall see that we can choose k as large as it will be needed, without increasing the computational cost of the method significantly. Consequently, the quadrature errors can be made arbitrarily small, until they fall within the round-off error level of the used finite precision arithmetic.*

The following result clearly holds true.

Theorem 7 *In case $U \in \Pi_\mu$, with $\mu \leq 2k/s$, then $\psi_j(u) = \hat{\psi}_j(u)$, $j = 0, \ldots, s-1$, and the results of Theorems 2, 3, 5, and 6, as well as Corollary 2, continue to hold.*

Conversely, by continuing to denote by u and v the new polynomial approximations using $\hat{\psi}_j$ defined by (50), instead of ψ_j defined in (20), the statements of the previous results modify as follows, the proofs being similar.

Theorem 8 *Assume that:*

$$\nabla g(u(c_i h))^\top M^{-1} v(c_i h) = 0 \in \mathbb{R}^\nu, \qquad i = 1, \ldots, s. \tag{52}$$

Then, for all $k \geq s$, properties (14) are replaced by

$$\begin{aligned} H(q_1, p_1) &= \begin{cases} H(q_0, p_0), & \text{if } U \in \Pi_\mu, \text{ with } \mu \leq 2k/s, \\ H(q_0, p_0) + O(h^{2k+1}), & \text{otherwise}, \end{cases} \\ g(q_1) &= 0, \end{aligned}$$

whereas for (15) the considerations made in Section 3.1 continue to hold.

Theorem 9 *With reference to (20)–(23) and (52), with $\psi_j(u)$ formally replaced by $\hat{\psi}_j(u)$ as defined in (50), and assuming $k \geq s$, the approximation procedure has convergence order r, i.e.,*

$$q(h) - q_1 = O(h^{r+1}), \qquad p(h) - p_1 = O(h^{r+1}),$$

where $r = s$ if s is even, or $r = s + 1$, when s is odd. In the case where (42) holds true, then $r = 2s$.

Corollary 3 *In the hypothesis of the previous Theorem 9, and using the notation (39) and (43), one has:*

$$\|y - \sigma\| := \max_{\tau \in [0,1]} |y(\tau h) - \sigma(\tau h)| = O(h^{s+1}).$$

Theorem 10 *In the hypothesis of the previous Theorem 9, the result of Theorem 6 continues to hold.*

We now prove Theorem 8, since the other statements can be similarly proved, by slightly modifying the proof of Theorems 5 and 6, and Corollary 2, respectively (the procedure is similar to that used in [12] when discretizing the involved integrals).

Proof (of Theorem 8). We prove only the statement concerning the Hamiltonian, since the remaining ones are straightforward. In the polynomial case, the statement follows from Theorem 7. Conversely, for all $k \geq s$, by taking into account (50)-(51), one has:

$$\begin{aligned} &H(q_1, p_1) - H(q_0, p_0) \equiv H(u(h), v(h)) - H(u(0), v(0)) \\ &= \int_0^h \frac{\mathrm{d}}{\mathrm{d}t} H(u(t), v(t)) \mathrm{d}t \\ &= h \int_0^1 \left[H_q(u(ch), v(ch))^\top \dot{u}(ch) + H_p(u(ch), v(ch))^\top \dot{v}(ch) \right] \mathrm{d}c \\ &= h \int_0^1 \left[-\nabla U(u(ch))^\top M^{-1} \sum_{j=0}^{s-1} P_j(c) \gamma_j(v) + v(ch)^\top M^{-1} \sum_{j=0}^{s-1} P_j(c) \hat{\psi}_j(u) \right. \end{aligned}$$

$$
\begin{aligned}
&\quad - \; v(ch)^\top M^{-1} \sum_{j=1}^{s} \ell_j(c)\rho_j \Bigg] \, \mathrm{d}c \\
&= -h \sum_{j=0}^{s-1} \underbrace{\left[\int_0^1 P_j(c) \nabla U(u(ch)) \mathrm{d}c \right]^\top}_{=\psi_j(u)^\top} M^{-1} \gamma_j(v) \\
&\quad + h \sum_{j=0}^{s-1} \underbrace{\left[\int_0^1 P_j(c) v(ch) \mathrm{d}c \right]^\top}_{=\gamma_j(u)^\top} M^{-1} \hat{\psi}_j(u) - h \int_0^1 \sum_{j=1}^{s} v(ch)^\top M^{-1} \ell_j(c) \rho_j \mathrm{d}c \\
&= -h \sum_{j=0}^{s-1} \psi_j(u)^\top M^{-1} \gamma_j(v) + h \sum_{j=0}^{s-1} \gamma_j(v)^\top M^{-1} \underbrace{[\psi_j(u) \; - \; \overbrace{\hat{\Delta}_j(h)}^{=O(h^{2k-j})} \;]}_{=\hat{\psi}_j(u)} \\
&\quad - h \sum_{j=1}^{s} b_j v(c_j h)^\top M^{-1} \rho_j \\
&= -h \sum_{j=0}^{s-1} \underbrace{\gamma_j(v)^\top M^{-1} \hat{\Delta}_j(h)}_{=O(h^{2k})} - h \sum_{j=1}^{s} b_j \underbrace{v(c_j h)^\top M^{-1} \nabla g(u(c_j h))}_{=0} \lambda_j \\
&= O(h^{2k+1}) . \square
\end{aligned}
$$

4.1 Discrete problem in vector form

Let us now derive a vector formulation of the discrete problem obtained by formally replacing $\psi_j(u)$ in (49) with the quadrature $\hat{\psi}_j(u)$ defined in (50). For this purpose, besides the matrices and vectors defined in (27) and (28), we also need the following ones:

$$
\begin{aligned}
\hat{\mathcal{P}}_s &= \begin{pmatrix} P_0(\hat{c}_1) & \dots & P_{s-1}(\hat{c}_1) \\ \vdots & & \vdots \\ P_0(\hat{c}_k) & \dots & P_{s-1}(\hat{c}_k) \end{pmatrix} \in \mathbb{R}^{k \times s}, \\
\hat{\mathcal{I}}_s &= \begin{pmatrix} \int_0^{\hat{c}_1} P_0(x) \mathrm{d}x & \dots & \int_0^{\hat{c}_1} P_{s-1}(x) \mathrm{d}x \\ \vdots & & \vdots \\ \int_0^{\hat{c}_k} P_0(x) \mathrm{d}x & \dots & \int_0^{\hat{c}_k} P_{s-1}(x) \mathrm{d}x \end{pmatrix} \in \mathbb{R}^{k \times s}, \\
\hat{\Omega} &= \mathrm{diag}(\hat{b}_1, \dots, \hat{b}_k) \in \mathbb{R}^{k \times k}, \qquad (53) \\
\hat{\boldsymbol{\psi}} &= \begin{pmatrix} \hat{\psi}_0(u) \\ \vdots \\ \hat{\psi}_{s-1}(u) \end{pmatrix}, \; \boldsymbol{\gamma} = \begin{pmatrix} \gamma_0(v) \\ \vdots \\ \gamma_{s-1}(v) \end{pmatrix} \in \mathbb{R}^{sm},
\end{aligned}
$$

$$\boldsymbol{\zeta} \;=\; \left(\begin{array}{c} \zeta_0(u,\boldsymbol{\lambda}) \\ \vdots \\ \zeta_{s-1}(u,\boldsymbol{\lambda}) \end{array}\right) \in \mathbb{R}^{sm}, \qquad \hat{\boldsymbol{e}} \;=\; \left(\begin{array}{c} 1 \\ \vdots \\ 1 \end{array}\right) \in \mathbb{R}^{k}.$$

One then obtains, by also taking into account (29), the following discrete problem:

$$\begin{array}{rcl} \nabla \boldsymbol{g} & = & \nabla g(\boldsymbol{e} \otimes q_0 + h\mathcal{I}_s \otimes M^{-1}\boldsymbol{\gamma}) \\ \boldsymbol{\lambda} & = & \left(\nabla \boldsymbol{g}^\top\, h\mathcal{I}_s\mathcal{P}_s^\top \Omega \otimes M^{-1}\, \nabla \boldsymbol{g}\right)^{-1} \nabla \boldsymbol{g}^\top \left[\boldsymbol{e} \otimes M^{-1}p_0 + h\mathcal{I}_s \otimes M^{-1}\hat{\boldsymbol{\psi}}\right] \\ \boldsymbol{\zeta} & = & \mathcal{P}_s^\top \Omega \otimes I_m(\nabla \boldsymbol{g} \cdot \boldsymbol{\lambda}) \\ \boldsymbol{\gamma} & = & \mathcal{P}_s^\top \Omega \otimes I_m \left[\boldsymbol{e} \otimes p_0 + h\mathcal{I}_s \otimes I_m(\hat{\boldsymbol{\psi}} - \boldsymbol{\zeta})\right] \\ \hat{\boldsymbol{\psi}} & = & \hat{\mathcal{P}}_s^\top \hat{\Omega} \otimes I_m\, \nabla U(\hat{\boldsymbol{e}} \otimes q_0 + h\hat{\mathcal{I}}_s \otimes M^{-1}\boldsymbol{\gamma}) \end{array} \tag{54}$$

where, considering that

$$\begin{array}{rcl} \boldsymbol{e} \otimes q_0 + h\mathcal{I}_s \otimes M^{-1}\boldsymbol{\gamma} & \equiv & \left(\begin{array}{c} u(c_1 h) \\ \vdots \\ u(c_s h) \end{array}\right), \\ \hat{\boldsymbol{e}} \otimes q_0 + h\hat{\mathcal{I}}_s \otimes M^{-1}\boldsymbol{\gamma} & \equiv & \left(\begin{array}{c} u(\hat{c}_1 h) \\ \vdots \\ u(\hat{c}_k h) \end{array}\right), \\ \boldsymbol{e} \otimes p_0 + h\mathcal{I}_s \otimes I_m(\hat{\boldsymbol{\psi}} - \boldsymbol{\zeta}) & \equiv & \left(\begin{array}{c} v(c_1 h) \\ \vdots \\ v(c_s h) \end{array}\right), \end{array}$$

it is meant that:

$$\nabla U(\hat{\boldsymbol{e}} \otimes q_0 + h\hat{\mathcal{I}}_s \otimes M^{-1}\boldsymbol{\gamma}) = \left(\begin{array}{c} \nabla U(u(\hat{c}_1 h)) \\ \vdots \\ \nabla U(u(\hat{c}_k h)) \end{array}\right) \in \mathbb{R}^{km},$$

and, similarly, $\nabla g(\boldsymbol{e} \otimes q_0 + h\mathcal{I}_s \otimes M^{-1}\boldsymbol{\gamma})$ computes the $sm \times s\nu$ matrix $\nabla \boldsymbol{g}$ defined in (27).

Remark 4 *In the case where the holonomic constraints are not present, i.e.* $\boldsymbol{\zeta} = \boldsymbol{0}$ *in (54), the first three equations are void. Moreover, the last two equations, combined together, provide the discrete problem of a HBVM*(k,s) *method applied to the corresponding unconstrained problem. For this reason, we shall continue to refer to (54) as a HBVM*(k,s) *method applied for solving the constrained Hamiltonian problem (5).*

We conclude this section by observing that the discrete problem (54), which has to be solved within machine accuracy, in order to gain the conservation

properties, can be numerically solved by suitably modifying the Newton-type iterations studied in [11, 5] (see also [14, 15]). However, for carrying out the numerical tests of Section 5, we shall use the following straightforward fixed-point iteration,

$$\begin{aligned}
\nabla \boldsymbol{g}^{\ell} &= \nabla g(\boldsymbol{e} \otimes q_0 + h\mathcal{I}_s \otimes M^{-1}\boldsymbol{\gamma}^{\ell}) \\
\boldsymbol{\lambda}^{\ell} &= \left((\nabla \boldsymbol{g}^{\ell})^{\top} h\mathcal{I}_s \mathcal{P}_s^{\top}\Omega \otimes M^{-1}\nabla \boldsymbol{g}^{\ell}\right)^{-1} \\
&\qquad (\nabla \boldsymbol{g}^{\ell})^{\top}\left[\boldsymbol{e} \otimes M^{-1}p_0 + h\mathcal{I}_s \otimes M^{-1}\hat{\boldsymbol{\psi}}^{\ell}\right] \\
\boldsymbol{\zeta}^{\ell} &= \mathcal{P}_s^{\top}\Omega \otimes I_m(\nabla \boldsymbol{g}^{\ell} \cdot \boldsymbol{\lambda}^{\ell}) \\
\boldsymbol{\gamma}^{\ell+1} &= \mathcal{P}_s^{\top}\Omega \otimes I_m\left[\boldsymbol{e} \otimes p_0 + h\mathcal{I}_s \otimes I_m(\hat{\boldsymbol{\psi}}^{\ell} - \boldsymbol{\zeta}^{\ell})\right] \\
\hat{\boldsymbol{\psi}}^{\ell+1} &= \hat{\mathcal{P}}_s^{\top}\hat{\Omega} \otimes I_m \nabla U(\hat{\boldsymbol{e}} \otimes q_0 + h\hat{\mathcal{I}}_s \otimes M^{-1}\boldsymbol{\gamma}^{\ell}), \qquad \ell = 0, 1, \ldots,
\end{aligned} \tag{55}$$

which can be conveniently started from

$$\boldsymbol{\gamma}^0 = \hat{\boldsymbol{\psi}}^0 = \mathbf{0} \in \mathbb{R}^{sm}.$$

Assuming a continuous Lipschitz condition for ∇U, and recalling that the constraints are quadratic, the convergence of the previous iteration can be seen to be granted for all sufficiently small timesteps h. Once (55) has been iterated until convergence, the new approximations are given by:

$$q_1 = q_0 + h\gamma_0, \qquad p_1 = p_0 + h(\hat{\psi}_0 - \zeta_0), \tag{56}$$

where we have removed the arguments of the Fourier coefficients, in order to have a more algorithmic description. An approximation of the vector of the Lagrange multipliers can be similarly obtained as

$$\bar{\lambda} = \sum_{i=1}^{s} \ell_i(1)\lambda_i, \tag{57}$$

$\lambda_1, \ldots, \lambda_s$ being, according to (28), the (block) entries of the vector $\boldsymbol{\lambda}$ at convergence.

5 Numerical examples

In this section we report a number of test problems, which we solve by means of HBVM(k, s) methods. The methods have been implemented in Matlab (Rel. 2020b) on a 3GHz Intel Xeon W10 core computer with 64GB RAM. In all numerical tests, the discrete problem (54) is solved according to (55), which is iterated until full machine accuracy is gained. The different problems considered here, are aimed at assessing different facets of the theoretical results.

5.1 The simple pendulum

The first example that we consider is the simple pendulum in cartesian coordinates. We place the reference system, so that the motion is in the plane (q_1, q_2). Assuming to have normalized for the mass and the acceleration of gravity, and also assuming a unit length of the pendulum, the resulting equations are (hereafter, $\|\cdot\|$ will denote the Euclidean norm):

$$\dot{q} = p, \qquad \dot{p} = -e_2, \qquad g(q) := \|q\|^2 - 1 = 0, \tag{58}$$

e_2 being the second unit vector in $\mathbb{R}^2$, with Hamiltonian

$$H(q,p) = \frac{1}{2}\|p\|^2 + e_2^\top q. \tag{59}$$

Having set the origin at the suspension point of the pendulum, a set of consistent initial conditions are:

$$q(0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \qquad p(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{60}$$

Since the Hamiltonian (59) is quadratic, a HBVM(s,s) method (i.e., the s-stage Gauss method) is energy-conserving of order $r = 2\lceil \frac{s}{2} \rceil$, with the approximation of the Lagrange multipliers (57) of order $r-1$, for all $s = 1, 2, \ldots$. Also the hidden constraints are conserved with order r. In Table 1 we list the obtained results, by using the HBVM(s,s) method, $s = 1, 2, 3, 4$, with timestep

$$h_i = 2^{-i}, \qquad i = 0, 1, \ldots, \tag{61}$$

for solving problem (58)–(60) on the interval $[0, 10]$. The error e_y in the solution is numerically estimated. Moreover, we have denoted by e_λ the error in the Lagrange multipliers, e_{hid} the error in the hidden constraints, e_g the error in the constraints, and by e_H the Hamiltonian error (the same notation will be used in the subsequent tests). As one may see, all the obtained results exactly match the expected ones.

As is well known, in such a case, the Lagrange multiplier has the physical meaning of the tension of the suspension rod. Its plot is in Figure 1, together with the phase portrait of the solution.

5.2 A modified pendulum

In this case, the simple pendulum described in Section 5.1 is modified by assuming that the unit-point mass also has an electric charge. Another charged point is placed at q^*, so that the Hamiltonian becomes:

$$H(q,p) = \frac{1}{2}\|p\|^2 + e_2^\top q - \frac{1}{\|q - q^*\|}, \tag{62}$$

subject to the same constraint in (58). We consider the same initial values (60), and choose $q^* = (2,\, 0)^\top$. The corresponding phase portrait and the Lagrange multiplier are depicted in Figure 2, relative to the interval $[0, 20]$.

Table 1: obtained results for problem (58)–(60) by using a HBVM(s,s) method with timestep (61).

$s=1$								
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H
0	5.87e-01	—	5.47e-01	—	6.08e-02	—	5.55e-15	2.22e-15
1	1.66e-01	1.8	1.65e-01	1.7	1.48e-02	2.0	2.22e-15	7.77e-16
2	4.62e-02	1.8	3.58e-02	2.2	3.67e-03	2.0	2.44e-15	1.11e-15
3	1.17e-02	2.0	1.01e-02	1.8	9.16e-04	2.0	1.89e-15	1.33e-15
4	2.95e-03	2.0	3.67e-03	1.5	2.29e-04	2.0	3.33e-15	5.55e-16
5	7.37e-04	2.0	1.56e-03	1.2	5.72e-05	2.0	1.33e-15	8.88e-16
6	1.84e-04	2.0	7.20e-04	1.1	1.43e-05	2.0	3.77e-15	2.22e-15
7	4.61e-05	2.0	3.46e-04	1.1	3.58e-06	2.0	3.11e-15	1.33e-15
8	1.15e-05	2.0	1.69e-04	1.0	8.94e-07	2.0	7.11e-15	4.66e-15
$s=2$								
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H
0	4.61e-02	—	1.14e-02	—	1.98e-02	—	8.88e-16	5.00e-16
1	4.30e-03	3.4	8.41e-04	3.8	4.90e-03	2.0	5.55e-16	5.55e-16
2	5.99e-04	2.8	1.75e-03	-1.1	1.22e-03	2.0	1.55e-15	5.55e-16
3	1.20e-04	2.3	1.31e-03	0.4	3.05e-04	2.0	1.33e-15	2.78e-16
4	2.81e-05	2.1	7.68e-04	0.8	7.63e-05	2.0	8.88e-16	4.44e-16
5	6.91e-06	2.0	4.13e-04	0.9	1.91e-05	2.0	1.33e-15	2.16e-15
6	1.72e-06	2.0	2.14e-04	0.9	4.77e-06	2.0	4.33e-15	1.22e-15
7	4.29e-07	2.0	1.09e-04	1.0	1.19e-06	2.0	3.33e-15	1.72e-15
8	1.07e-07	2.0	5.48e-05	1.0	2.98e-07	2.0	6.66e-15	3.05e-15
$s=3$								
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H
0	2.83e-03	—	5.58e-03	—	1.26e-03	—	4.44e-16	2.22e-16
1	1.86e-04	3.9	2.21e-03	1.3	6.50e-05	4.3	4.44e-16	2.22e-16
2	9.45e-06	4.3	1.94e-04	3.5	3.95e-06	4.0	4.44e-16	1.67e-16
3	5.86e-07	4.0	2.40e-05	3.0	2.46e-07	4.0	8.88e-16	5.55e-16
$s=4$								
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H
0	3.70e-04	—	2.89e-03	—	1.60e-04	—	2.22e-16	1.11e-16
1	2.24e-05	4.0	4.58e-04	2.7	9.63e-06	4.1	2.22e-16	3.33e-16
2	1.44e-06	4.0	6.04e-05	2.9	5.93e-07	4.0	5.55e-16	2.22e-16
3	9.08e-08	4.0	7.32e-06	3.0	3.73e-08	4.0	1.11e-15	5.55e-16

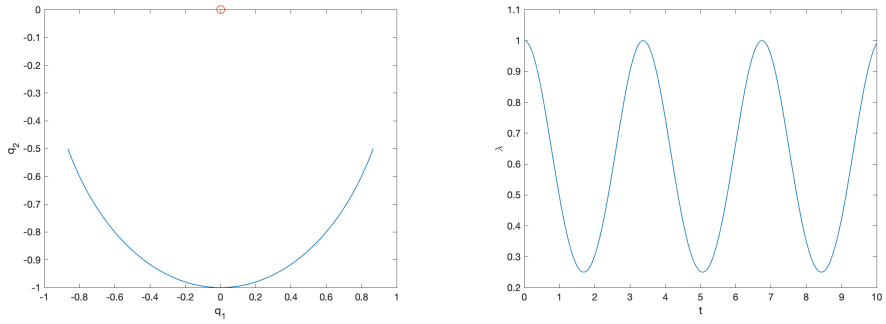

Figure 1: phase portrait (left) and Lagrange multiplier (right) for problem (58)–(60).

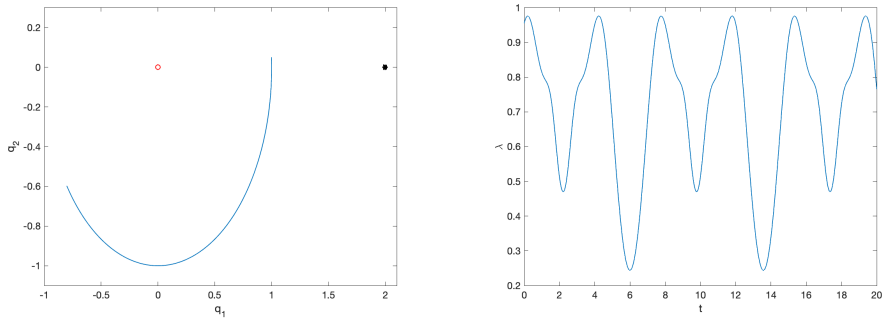

Figure 2: phase portrait (left) and Lagrange multiplier (right) for the problem defined by (62), with $q^* = (2, 0)^\top$, using the initial values (60); in the left-plot, the star denotes the second charged point.

As is clear, the potential is now no more quadratic, so that the HBVM(s, s) method is no more energy-conserving. Nevertheless, by using a HBVM(k, s) method, with $k \geq s$, the Hamiltonian error should behave as $O(h^{2k})$, according to Theorem 8. This is confirmed by the numerical results listed in Table 2, obtained by solving the problem on the interval $[0, 20]$, with timesteps as defined in (61), by means of the HBVM$(k, 1)$ method, $k = 1, 2, 3, 4$. As one may see, the HBVM(4,1) method is (practically) energy-conserving, for the considered values of the timestep.

5.3 A conical pendulum

In this case, the problem is given by:

$$\dot{q} = p, \qquad \dot{p} = -e_3, \qquad g(q) := \|q\|^2 - 1 = 0, \tag{63}$$

Table 2: obtained results for problem (60) and (62) by using HBVM$(k,1)$ methods with timestep (61). The *** mean that the round-off error level is reached.

$k=1$									
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H	rate
3	2.40e-02	—	3.78e-02	—	1.25e-03	—	2.00e-15	9.10e-05	—
4	5.90e-03	2.0	1.52e-02	1.3	3.12e-04	2.0	1.62e-14	2.28e-05	2.0
5	1.47e-03	2.0	6.54e-03	1.2	7.80e-05	2.0	1.78e-15	5.69e-06	2.0
6	3.67e-04	2.0	3.00e-03	1.1	1.95e-05	2.0	2.89e-15	1.42e-06	2.0
7	9.17e-05	2.0	1.43e-03	1.1	4.87e-06	2.0	5.77e-15	3.56e-07	2.0
$k=2$									
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H	rate
3	2.28e-02	—	3.70e-02	—	1.25e-03	—	2.66e-15	8.65e-09	—
4	5.63e-03	2.0	1.50e-02	1.3	3.12e-04	2.0	1.82e-14	5.41e-10	4.0
5	1.40e-03	2.0	6.49e-03	1.2	7.80e-05	2.0	4.44e-15	3.38e-11	4.0
6	3.50e-04	2.0	2.99e-03	1.1	1.95e-05	2.0	4.11e-15	2.11e-12	4.0
7	8.76e-05	2.0	1.43e-03	1.1	4.87e-06	2.0	3.33e-15	1.25e-13	4.1
$k=3$									
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H	rate
3	2.28e-02	—	3.70e-02	—	1.25e-03	—	2.11e-15	7.32e-13	—
4	5.63e-03	2.0	1.50e-02	1.3	3.12e-04	2.0	5.55e-16	9.77e-15	6.2
5	1.40e-03	2.0	6.49e-03	1.2	7.80e-05	2.0	5.55e-16	5.55e-16	***
6	3.50e-04	2.0	2.99e-03	1.1	1.95e-05	2.0	7.77e-16	6.66e-16	***
7	8.76e-05	2.0	1.43e-03	1.1	4.87e-06	2.0	2.89e-15	1.11e-15	***
$k=4$									
i	e_y	rate	e_λ	rate	e_{hid}	rate	e_g	e_H	rate
3	2.28e-02	—	3.70e-02	—	1.25e-03	—	2.22e-16	4.44e-16	—
4	5.63e-03	2.0	1.50e-02	1.3	3.12e-04	2.0	4.44e-16	4.44e-16	***
5	1.40e-03	2.0	6.49e-03	1.2	7.80e-05	2.0	4.44e-16	6.66e-16	***
6	3.50e-04	2.0	2.99e-03	1.1	1.95e-05	2.0	6.66e-16	6.66e-16	***
7	8.76e-05	2.0	1.43e-03	1.1	4.87e-06	2.0	1.55e-15	1.11e-15	***

e_3 being the third unit vector in $\mathbb{R}^3$, with Hamiltonian

$$H(q,p) = \frac{1}{2}\|p\|^2 + e_3^\top q. \tag{64}$$

Considering the set of consistent initial conditions:

$$q(0) = \begin{pmatrix} z_0 \\ 0 \\ -z_0 \end{pmatrix}, \qquad p(0) = \begin{pmatrix} 0 \\ \sqrt{z_0} \\ 0 \end{pmatrix}, \qquad z_0 = \frac{1}{\sqrt{2}}, \tag{65}$$

the motion lies in the plane $q_3 = -z_0$, with period $T = 2^{\frac{3}{4}}\pi$, and the Lagrange multiplier is constant, $\lambda = z_0$. The suspension rod then spans a cone, from which the name of the problem (see Figure 3). For this problem, the Lagrange multiplier is exactly computed, so that condition (42) holds true and HBVM(s,s) methods are expected to be energy-conserving and with order $2s$.

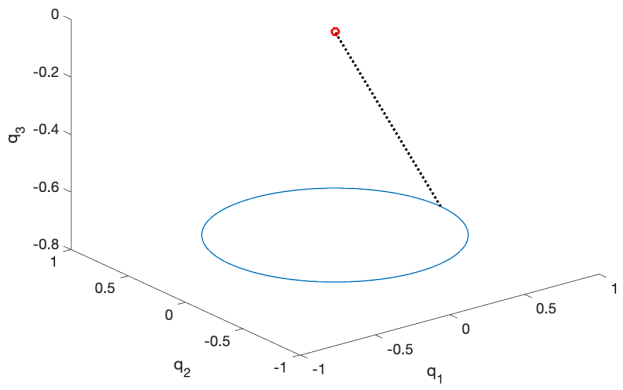

Figure 3: phase portrait the problem (63)–(65).

This is confirmed by the numerical tests listed in Table 3, obtained by using the timestep

$$h_i = 2^{-i}\frac{T}{5}, \qquad i = 0, 1, \ldots, \tag{66}$$

for solving one period. From the results in Table 3, one may see that remarkably enough, also the hidden constraint turns out to be satisfied.

5.4 A tethered satellite system

The last test problem that we consider is a tethered satellite systems of three satellites, forming an equilateral triangle with unit sides, as depicted in Figure 4. Setting

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \qquad p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}, \qquad q_i, p_i \in \mathbb{R}^3, \quad i = 1, 2, 3,$$

we assume the Hamiltonian be given by

$$H(q,p) = \sum_{i=1}^{3} \left(\frac{1}{2}\|p_i\|^2 + \frac{1}{\|q_i\|} + \cos(\|q_i\|) \right), \tag{67}$$

with the constraints:

$$g(q) := \begin{pmatrix} \|q_1 - q_2\|^2 - 1 \\ \|q_2 - q_3\|^2 - 1 \\ \|q_1 - q_3\|^2 - 1 \end{pmatrix} = 0 \in \mathbb{R}^3. \tag{68}$$

Table 3: obtained results for problem (63)–(65) by using HBVM(s,s) methods with timestep (66). The *** mean that the round-off error level is reached.

$s=1$						
i	e_y	rate	e_λ	e_{hid}	e_g	e_H
0	3.61e-01	—	3.03e-14	3.33e-16	7.55e-15	2.22e-15
1	1.20e-01	1.6	3.33e-16	1.28e-15	5.55e-16	1.72e-15
2	3.20e-02	1.9	1.67e-15	4.16e-16	7.77e-16	2.22e-16
3	8.11e-03	2.0	1.60e-14	2.33e-15	4.44e-16	1.67e-16
4	2.03e-03	2.0	1.57e-14	1.33e-15	7.77e-16	4.44e-16
$s=2$						
i	e_y	rate	e_λ	e_{hid}	e_g	e_H
0	1.55e-02	—	1.51e-13	1.11e-15	1.11e-15	6.66e-16
1	1.05e-03	3.9	2.66e-15	1.33e-15	1.33e-15	1.39e-15
2	6.66e-05	4.0	1.07e-14	1.69e-15	2.22e-16	2.78e-16
3	4.18e-06	4.0	1.25e-14	1.44e-15	4.44e-16	3.89e-16
4	2.62e-07	4.0	1.59e-13	4.39e-15	6.66e-16	3.33e-16
$s=3$						
i	e_y	rate	e_λ	e_{hid}	e_g	e_H
0	1.91e-04	—	2.00e-15	9.44e-16	3.33e-16	1.11e-16
1	3.13e-06	5.9	6.22e-15	9.99e-16	3.33e-16	1.11e-16
2	4.94e-08	6.0	1.18e-14	6.66e-16	5.55e-16	2.22e-16
3	7.74e-10	6.0	2.24e-14	8.88e-16	8.88e-16	2.22e-16
4	1.21e-11	6.0	2.24e-13	2.44e-15	8.88e-16	6.66e-16
$s=4$						
i	e_y	rate	e_λ	e_{hid}	e_g	e_H
0	1.23e-06	—	1.02e-14	2.75e-15	2.22e-16	4.44e-16
1	4.97e-09	7.9	1.98e-14	6.66e-16	6.66e-16	5.55e-16
2	1.96e-11	8.0	9.18e-14	2.56e-15	4.44e-16	1.11e-16
3	7.69e-14	8.0	1.38e-13	3.08e-15	4.44e-16	2.78e-16
4	8.80e-16	***	6.98e-14	2.00e-15	4.44e-16	3.89e-16

Consistent initial conditions are chosen as follows:

$$q_1(0) = \begin{pmatrix} 0 \\ \frac{1}{2} \\ z_0 \end{pmatrix}, \quad q_2(0) = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ z_0 \end{pmatrix}, \quad q_3(0) = \begin{pmatrix} 0 \\ 0 \\ z_0 - \frac{\sqrt{3}}{2} \end{pmatrix}, \tag{69}$$

and

$$p_1(0) = p_2(0) = 0 \in \mathbb{R}^3, \qquad p_3(0) = \begin{pmatrix} v_0 \\ 0 \\ 0 \end{pmatrix}, \tag{70}$$

with $z_0 = 20$ and $v_0 > 0$ such that $H(q(0), p(0)) = 0$. Let us numerically solve problem (67)–(70) on the interval $[0, 10^3]$ by using the HBVM(1,1) method with timestep $h = 0.1$. The method is not energy-conserving, and 9.4 sec are required for solving the problem, with a mean number of 15.4 iterations (55) per step. The numerical Hamiltonian is depicted in the left-plot of Figure 5. If we solve the

problem by using HBVM(5,1), the method turns out to be practically energy-conserving, and requires 11.6 sec, with the same mean number of iterations (55) per step. The corresponding numerical Hamiltonian is shown in the right-plot of Figure 5. As one may see, the energy-conserving method has a much more appropriate behavior, with a very moderate increase of the computational time.

Similar results are obtained by solving the problem, with the same timestep $h = 0.1$, by using the HBVM(3,3) and HBVM(5,3) methods, the latter method being (practically) energy-conserving. Both methods need a mean of 14 iterations (55) per step, with execution times of 14.2 sec and 15.4 sec, respectively. The corresponding numerical Hamiltonians are depicted in Figure 6. Also in this case, the qualitative results obtained by the energy-conserving method are more favorable, with only a very limited increase of the overall computational cost.

6 Conclusions

In this paper we have introduced a class of energy-conserving methods for Hamiltonian problems with quadratic holonomic constraints. The methods reduce to the class of Hamiltonian Boundary Value Methods (HBVMs), when the constraints are not given. Arbitrarily high-order methods are obtained within this class. A detailed analysis of the methods has been presented, along with their practical implementation and solution of the generated discrete problems. A number of numerical tests duly confirm the theoretical results, also proving the usefulness of the approach. Future directions of investigation will include the extension of the methods for numerically solving Hamiltonian problems with general holonomic constraints.

Conflict of interests. The authors declare no conflict of interests.

Acknowledgements. The authors acknowledge the *mrSIR project* [45] for the financial support.

References

[1] H.C. Andersen. Rattle: a "velocity" version of the Shake algorithm for molecular dynamics calculations. *J. Comp. Phys.* **52** (1983) 24–34.

[2] M.F. Baião, T.J. Stuchi. Dynamics of tethered satellites in the vicinity of the Lagrangian point L_2 of the Earth-Moon system. *Astrophys. Space Sci.* **362** (2017) 134

[3] G. Benettin, A.M. Cherubini, F. Fassò. A changing chart symplectic algorithm for rigid bodies and other Hamiltonian systems on manifolds. *SIAM J. Sci. Comput.* **23**, No. 4 (2001) 1189–1203.

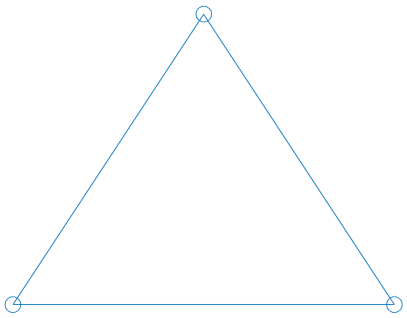

Figure 4: tethered satellite system described by problem (67)–(70).

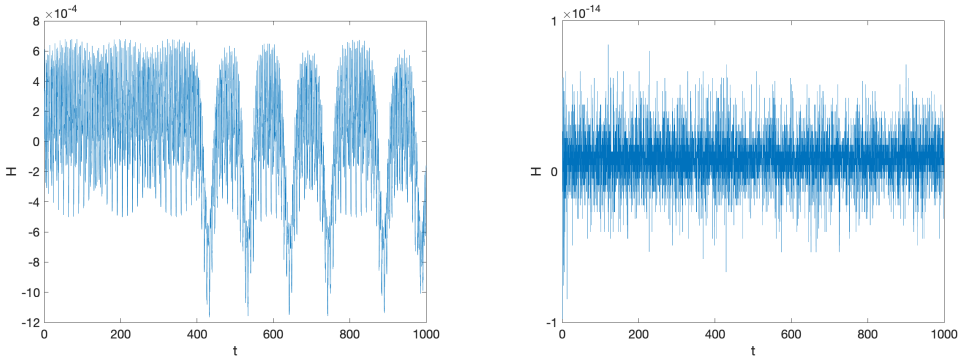

Figure 5: numerical Hamiltonian when solving problem (67)–(70) with timestep $h = 0.1$ by using HBVM(1,1) (left plot) and HBVM(5,1) (right plot).

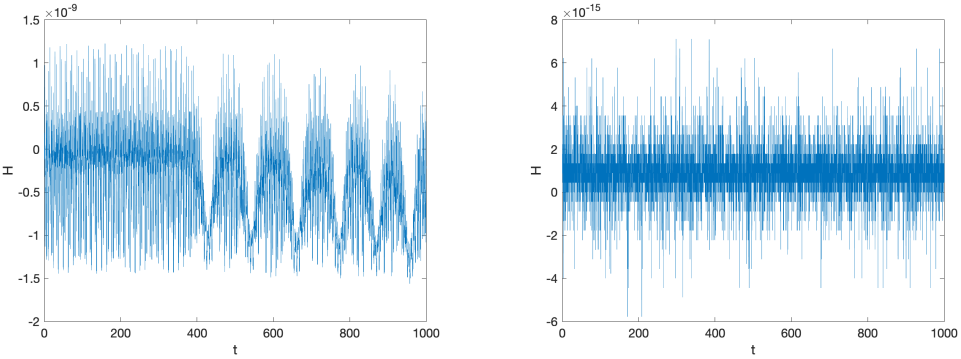

Figure 6: numerical Hamiltonian when solving problem (67)–(70) with timestep $h = 0.1$ by using HBVM(3,3) (left plot) and HBVM(5,3) (right plot).

[4] K.E. Brenan, S.L. Campbell, L.R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*. SIAM, Philadelphia, PA, 1996.

[5] L. Brugnano, G. Frasca-Caccia, F. Iavernaro. Efficient implementation of Gauss collocation and Hamiltonian Boundary Value Methods. *Numer. Algorithms* **65** (2014) 633–650.

[6] L. Brugnano, G. Gurioli, F. Iavernaro, E. Weinmueller. Line Integral Solution of Hamiltonian Systems with Holonomic Constraints. *Appl. Numer. Math.* **127** (2018) 56–77.

[7] L. Brugnano, F. Iavernaro. *Line Integral Methods for Conservative Problems*. Chapman et Hall/CRC, Boca Raton, FL, 2016.

[8] L. Brugnano, F. Iavernaro. Line Integral Solution of Differential Problems. *Axioms* **7**(2) (2018) 36.

[9] L. Brugnano, F. Iavernaro, D. Trigiante. Hamiltonian BVMs (HBVMs): a family of "drift-free" methods for integrating polynomial Hamiltonian systems. *AIP Conf. Proc.* **1168** (2009) 715–718.

[10] L. Brugnano, F. Iavernaro, D. Trigiante. Hamiltonian Boundary Value Methods (Energy Preserving Discrete Line Integral Methods). *JNAIAM. J. Numer. Anal. Ind. Appl. Math.* **5**, No. 1-2 (2010) 17–37.

[11] L. Brugnano, F. Iavernaro, D. Trigiante. A note on the efficient implementation of Hamiltonian BVMs. *J. Comput. Appl. Math.* **236** (2011) 375–383.

[12] L. Brugnano, F. Iavernaro, D. Trigiante. A simple framework for the derivation and analysis of effective one-step methods for ODEs. *Appl. Math.Comput.* **218** (2012) 8475–8485.

[13] L. Brugnano, F. Iavernaro, D. Trigiante. 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** (2015) 650–667.

[14] L. Brugnano, C. Magherini. Blended Implementation of Block Implicit Methods for ODEs. *Appl. Numer. Math.* **42** (2002) 29–45.

[15] L. Brugnano, C. Magherini. Recent Advances in Linear Analysis of Convergence for Splittings for Solving ODE problems. *Appl. Numer. Math.* **59** (2009) 542–557.

[16] Z.Q. Cai, X.F. Li, X. F., H. Zhou. Nonlinear dynamics of a rotating triangular tethered satellite formation near libration points. *Aerospace Science and Technology* **42** (2015) 384–391.

[17] P. Console, E. Hairer, C. Lubich. Symmetric multistep methods for constrained Hamiltonian systems. *Numer. Math.* **124** (2013) 517–539.

[18] O. Gonzalez. Mechanical systems subject to holonomic constraints: Differential–algebraic formulations and conservative integration. *Physica D* **132** (1999) 165–174.

[19] E. Hairer. Global modified Hamiltonian for constrained symplectic integrators. *Numer. Math.* **95** (2003) 325–336.

[20] E. Hairer, C. Lubich, G. Wanner. Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations, Second Edition. Springer-Verlag, Berlin, 2006.

[21] E. Hairer, G. Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems, Second Revised Edition.* Springer-Verlag, Berlin, 2010.

[22] E. Hairer, S.P. Nørsett, G. Wanner. *Solving Ordinary Differential Equations I, nonstiff problems. Second revised edition (3rd printing).* Springer, Heidelberg, 2008.

[23] L. Jay. Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *SIAM J. Numer Anal.* **33**, No. 1 (1996) 368–387.

[24] W. Jung, A.P. Mazzoleni, J. Chung. Dynamic analysis of a tethered satellite system with a moving mass. *Nonlinear Dyn.* **75** (2014) 267–281.

[25] W. Jung, A.P. Mazzoleni, J. Chung. Nonlinear dynamic analysis of a three-body tethered satellite system with deployment/retrieval. *Nonlinear Dyn.* **82** (2015) 1127–1144.

[26] D.T. Kawano. On modeling and simulation of tethered systems with sharp, rough contacts *Commun. Nonlinear Sci. Numer. Simul.* **16** (2011) 550–565.

[27] B. Leimkhuler, S. Reich. Symplectic integration of constrained Hamiltonian systems. *Math. Comp.* **63**, No. 208 (1994) 589–605.

[28] B. Leimkhuler, S. Reich. *Simulating Hamiltonian Dynamics.* Cambridge University Press, 2004.

[29] B. Leimkhuler, R.D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.* **112** (1994) 117–125.

[30] S. Leyendecker, P. Betsch, P. Steinmann. Energy-conserving integration of constrained Hamiltonian systems – a comparison of approaches. *Computational Mechanics* **33** (2004) 174–185.

[31] L. Li, D. Wang. Energy and quadratic invariants preserving methods for Hamiltonian systems with holonomic constraints. `arXiv:2007.06338 [math.NA]`.

[32] X. Lian, J. Liu, J. Zhang, C. Wang. Chaotic motion and control of a tethered-sailcraft system orbiting an asteroid. *Commun. Nonlinear Sci. Numer. Simul.* **77** (2019) 203–224.

[33] C. Luo, H. Wen, D. Jin, S. Xu. Dynamics of a flexible multi-tethered satellite formation in a Halo orbit with uncertain parameters. *Commun. Nonlinear Sci. Numer. Simul.* **99** (2021) 105828.

[34] S. Reich. Symplectic integration of constrained Hamiltonian systems by composition methods. *SIAM J. Numer. Anal.* **33** No. 2 (1996) 475–491.

[35] S. Reich. On higher-order semi-explicit symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *Numer. Math.* **76** (1997) 231–247.

[36] W.M. Seiler. Position versus momentum projections for constrained Hamiltonian systems. *Numer. Algorithms* **19** (1998) 223–234.

[37] W.M. Seiler. Numerical integration of constrained Hamiltonian systems using Dirac brackets. *Math. Comp.* **68**, No. 226 (1999) 661–681.

[38] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23** (1977) 327–341.

[39] Y. Wei, Z. Deng, Q. Li, B. Wang. Projected Runge-Kutta methods for constrained Hamiltonian systems. *Appl. Math. Mech. – Engl. Ed.* **37**, No. 8 (2016) 1077–1094.

[40] C.Y. Wang, C.M. Wang, R. Freund. Vibration of Heavy String Tethered to Mass-Spring System. *Int. J. Struct. Stab. Dyn.* **17** (2017) 1771002, 12 pp.

[41] M. Xu, J.-M. Zhu, T. Tan, S.-J. Xu. Equilibrium configurations of the tethered three-body formation system and their nonlinear dynamics. *Acta Mech. Sin.* **28** (2012) 1668–1677.

[42] B. Yu, D. Jin, H. Wen. Nonlinear dynamics of flexible tethered satellite system subject to space environment. *Appl. Math. Mech. -Engl. Ed.* **37**(4) (2016) 485–500.

[43] Y.M. Zabolotnov. Control of the deployment of a tethered orbital system with a small load into a vertical position. *J. Appl. Math. Mech.* **79**, no. 1 (2015) 28–34.

[44] A.E. Zakrzhevskii. Method of deployment of a tethered space system along the local vertical. *Internat. Appl. Mech.* **51** (2015) 670–681.

[45] `https://www.mrsir.it/en/`