Applied Mathematics and Computation 219 (2012) 2158–2179

Contents lists available at SciVerse ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Time finite element methods: A unified framework for numerical discretizations of ODEs

Wensheng Tang, Yajuan Sun *

LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Keywords:
Time finite element methods
Collocation methods
(Partitioned) Runge–Kutta methods
Hamiltonian systems
Energy-preserving methods
Symplectic methods
Symmetric

ABSTRACT

We present a unified framework for the numerical discretization of ODEs based on time finite element methods. We relate time finite element methods to Runge–Kutta methods with infinitely many stages. By means of the corresponding numerical quadrature, we establish the relation between time finite element methods and (partitioned) Runge–Kutta methods. We also provide order estimates and superconvergence of the corresponding numerical methods in use of the simplifying assumptions. We apply time finite methods to Hamiltonian systems and investigate the conservation of energy and symplectic structure for the resulting numerical discretizations. For Hamiltonian systems, we also construct new classes of symplectic integrators by combining different time finite element methods and verify the results by performing some numerical experiments.

1. Introduction

In this paper, we investigate finite element methods applied to ODEs with initial values. This approach is called as the time finite element (TFE) method. The pioneer works can date back to early 1960s. In 1969, the discrete variational formulation has been introduced in [1,21] for the integration of time-dependent differential equations. Furthermore, some researchers have presented the variational formulation by allowing the TFE solution to be discontinuous at the end of each time element interval. For example, the value of the approximation solution at t_n (one end of an interval) is given by $\alpha \boldsymbol{u}(t_n^-) + (1-\alpha)\boldsymbol{u}(t_n^+)$ in [18]. Different time discretizations can be derived from variational formulations with different discontinuities. Here, our interests are focused on introducing a unified TFE framework which can cover four classes of TFE methods: continuous time finite element (C-TFE) methods, left-discontinuous time finite element (LD-TFE) methods, right-discontinuous time finite element (RD-TFE) methods and bi-discontinuous time finite element (BD-TFE) methods. With suitable quadratures, we can not only relate TFE methods to classical time discretizations e.g. Runge–Kutta (RK) methods, but also construct new numerical methods.

In recent decades, geometric numerical integration has been widely developed and studied by many researchers for solving differential equations with certain structures, and most people are interested in the numerical discretization of Hamiltonian systems. It is known that the analytic solution of Hamiltonian systems can preserve both the symplectic structure and the Hamiltonian function (namely the energy). Though it would be nice to have a numerical method preserving both the symplectic structure and the energy, unfortunately it has been shown in [22] that for general Hamiltonian systems such numerical method does not exist. One is therefore confined to consider methods which can preserve one of these properties only. Symplectic methods have been well developed in the past decades (see [25] and references therein). Recently energy-preserving methods have also been studied fruitfully. A systematic way to construct energy-preserving methods is

* Corresponding author.
E-mail addresses: tangws@lsec.cc.ac.cn (W. Tang), sunyj@lsec.cc.ac.cn (Y. Sun).

0096-3003/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2012.08.062

the so-called discrete gradient approach [31]. These methods usually have only low order, and thus to get higher-order methods one may use a bootstrapping technique [30]. However, this technique is not always practically useful. To overcome the disadvantage of the discrete gradient approach, recently the Averaged Vector Field (AVF) method [32], the Hamiltonian boundary value methods (HBVMs) [5,6,9] and the energy-preserving collocation methods [24] have been presented.

We are interested in the preservation of energy and symplecticity for numerical discretizations based on TFE methods. For a given linear Hamiltonian system, it has been shown in [14] that C-TFE methods are symplectic and preserve the quadratic Hamiltonian function. For nonlinear Hamiltonian systems, C-TFE methods are not symplectic, but are proved to be energy-preserving [20,2,14] (up to a certain order if quadrature formulae are used). The effect of preservation of the energy in the computational practice mainly depends on the order of the numerical quadrature formula used for the integration of nonlinear terms. Theoretically, the higher the order of the numerical quadrature formula, the better the preservation of the energy for the numerical solution. Practically, the optimal order of the numerical quadrature to get the better energy preservation of the numerical discretization can be found in numerical experiments. In the TFE framework, we will show that the AVF method, the HBVMs and the energy-preserving collocation methods can be retrieved by the numerical discretization from the C-TFE method with suitable quadratures. It is also known that for general Hamiltonian systems the resulting discretization based on TFE methods is not symplectic. However, we can construct symplectic methods by combining different TFE methods, e.g. combining the LD-TFE and RD-TFE methods gives a class of symplectic PRK methods. These newly obtained symplectic methods can cover some classical schemes, e.g. implicit midpoint rule, symplectic Euler methods, Gauss methods, Lobatto IIIA-IIIB, Lobatto IIIC-IIIC, Radau IA-IĀ, Radau IIA-IIĀ, Gauss IA-IĀ, Radau IB and Radau IIB (see [33,34]) etc.

The outline of this paper is as follows. In the next section, we introduce a unified framework of TFE methods for solving ODEs with initial values. In Section 3, we first provide the definition of continuous (discontinuous) collocation methods. Then, we derive the relation between TFE methods and collocation methods. In Section 4, we study the energy preservation of TFE methods. By using various numerical quadrature formulae the numerical discretizations based on C-TFE methods can be equivalent to some existing energy-preserving numerical methods. Section 5 is devoted to the construction of new symplectic methods via combining different TFE methods. We present numerical experiments in Section 6. As last, we conclude this paper.

2. Time finite element formulation

In this section, we first introduce a unified framework for continuous time finite element (C-TFE) methods and discontinuous time finite element (D-TFE) methods presented in literature (see [2,4,12,14,18,20,23,28]). Then, we present their reformulation by using a numerical quadrature formula.

Consider the initial value problem for ODEs

$$\begin{cases} \dot{\boldsymbol{z}} = \boldsymbol{f}(t, \boldsymbol{z}), \quad t \in I, \quad \boldsymbol{z} \in \mathbf{R}^d, \\ \boldsymbol{z}(0) = \boldsymbol{z}_0, \end{cases} \tag{2.1}$$

where $\boldsymbol{f}$ is Lipschitz continuous, and $I = [0, T]$.

We partition I as $I = \bigcup_{n=0}^{N-1} [t_n, t_{n+1}]$. Let $I_n = (t_n, t_{n+1}), \bar{I}_n = [t_n, t_{n+1}]$ and $h_n = t_{n+1} - t_n$. Define $(\mathbb{P}^k(I_n))^d = \mathbb{P}^k(I_n) \times \cdots \times \mathbb{P}^k(I_n)$, where $\mathbb{P}^k(I_n)$ is the set of polynomials of degree k defined on I_n. The Galerkin variational problem is:

Find $\boldsymbol{u}_n(t) \in \left(\mathbb{P}^k(I_n)\right)^d$ and $\boldsymbol{u}_{n+1} \in \mathbf{R}^d$ such that

$$\begin{cases} \int_{t_n}^{t_{n+1}} \left(\dot{\boldsymbol{u}}_n(t) - \boldsymbol{f}(t, \boldsymbol{u}_n(t))\right) \cdot \boldsymbol{v}(t)\, \mathrm{d}t = \left(\boldsymbol{u}_n(t_{n+1}^-) - \boldsymbol{u}_{n+1}\right) \cdot \boldsymbol{v}(t_{n+1}) - \left(\boldsymbol{u}_n(t_n^+) - \boldsymbol{u}_n\right) \cdot \boldsymbol{v}(t_n) \\ \boldsymbol{u}_0 = \boldsymbol{z}_0 \end{cases} \tag{2.2}$$

or, equivalently,

$$\begin{cases} \int_{t_n}^{t_{n+1}} \left(\boldsymbol{u}_n(t) \cdot \dot{\boldsymbol{v}}(t) + \boldsymbol{f}(t, \boldsymbol{u}_n(t)) \cdot \boldsymbol{v}(t)\right) \mathrm{d}t = \boldsymbol{u}_{n+1} \cdot \boldsymbol{v}(t_{n+1}) - \boldsymbol{u}_n \cdot \boldsymbol{v}(t_n) \\ \boldsymbol{u}_0 = \boldsymbol{z}_0 \end{cases} \tag{2.3}$$

holds for arbitrary $\boldsymbol{v}(t) \in \left(\mathbb{P}^m(\bar{I}_n)\right)^d$ and $n = 0, 1, \ldots, N-1$.

Eqs. (2.2) and (2.3) are called k-degree time finite element (k-TFE) methods. By choosing the appropriate degree m, we derive four classes of TFE methods:

(a) When $m = k-1, \boldsymbol{u}_n(t_{n+1}^-) = \boldsymbol{u}_{n+1}$ and $\boldsymbol{u}_n(t_n^+) = \boldsymbol{u}_n$, this is the k-degree continuous time finite element (k-C-TFE) method (see [2,14,20,28]).
(b) When $m = k$ and $\boldsymbol{u}_n(t_{n+1}^-) = \boldsymbol{u}_{n+1}$, this is the k-degree left-discontinuous TFE (k-LD-TFE) method.
(c) When $m = k$ and $\boldsymbol{u}_n(t_n^+) = \boldsymbol{u}_n$, this is the k-degree right-discontinuous TFE (k-RD-TFE) method.
(d) When $m = k+1, \boldsymbol{u}_n(t_{n+1}^-) \neq \boldsymbol{u}_{n+1}$ and $\boldsymbol{u}_n(t_n^+) \neq \boldsymbol{u}_n$, this is the k-degree bi-discontinuous time finite element (k-BD-TFE) method (see [4]).

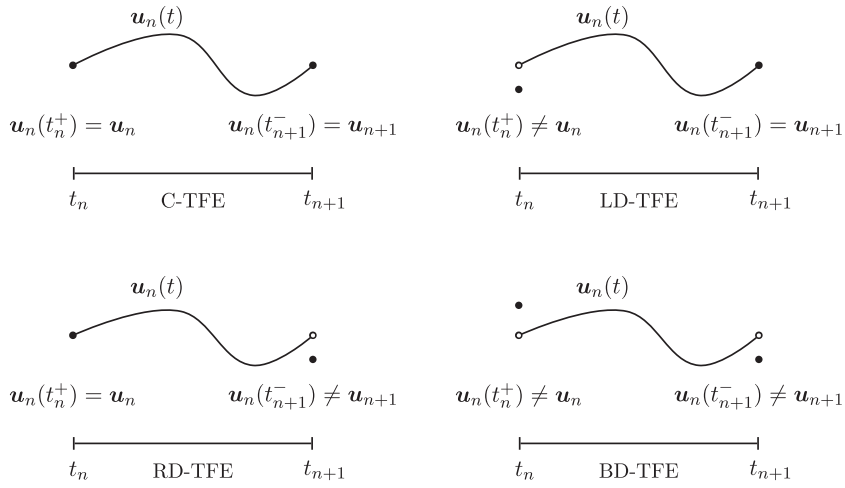

Fig. 2.1. Plots for various TFE methods in the nth element I_n (solid dots stand for $\boldsymbol{u}_n$ and $\boldsymbol{u}_{n+1}$).

The methods (b) and (c) are also known by a joint name, the k-degree singly discontinuous time finite element (k-SD-TFE) methods [4,18]. Fig. 2.1 shows the plots of the above TFE solutions. Obviously, it is convenient to define a one-step method via establishing a mapping from $\boldsymbol{u}_n$ to $\boldsymbol{u}_{n+1}$ in the nth element I_n and $\boldsymbol{u}_n$ can be naturally regarded as the approximation of $\boldsymbol{z}(t_n)$ for $n \geqslant 1$.

For the TFE methods, it is only necessary to consider the nth element I_n since the variational problem can be solved element by element. To simplify, hereafter we replace the notations $\boldsymbol{u}_n(t_n^+), \boldsymbol{u}_n(t_{n+1}^-)$ and $\boldsymbol{u}_n(t)$ with $\boldsymbol{u}(t_n), \boldsymbol{u}(t_{n+1})$ and $\boldsymbol{u}(t)$, respectively. When $\boldsymbol{f}(t,\boldsymbol{u})$ is a polynomial, the integral term on the left side of (2.2) or (2.3) can be calculated exactly. In most cases $\boldsymbol{f}(t,\boldsymbol{u})$ is nonlinear and more complicated, so we need to compute the integration by numerical quadrature formulae. Introduce a q-point quadrature formula as

$$\int_0^1 F(\tau)\,\mathrm{d}\tau \approx \sum_{i=1}^{q} w_i F(c_i), \tag{2.4}$$

where $c_1,\ldots,c_q \in [0,1]$ are distinct abscissae and $w_i \neq 0, i = 1,\ldots,q$ are the corresponding weights. Applying (2.4) to (2.2) and (2.3), we derive the following equalities

$$\sum_{i=1}^{q} w_i(\boldsymbol{u}'[i] - h\boldsymbol{f}[i])\phi_j(c_i) = (\boldsymbol{u}(t_{n+1}) - \boldsymbol{u}_{n+1})\phi_j(1) - (\boldsymbol{u}(t_n) - \boldsymbol{u}_n)\phi_j(0), \tag{2.5}$$

$$\sum_{i=1}^{q} w_i\Big(\boldsymbol{u}[i]\phi_j'(c_i) + h\boldsymbol{f}[i]\phi_j(c_i)\Big) = \boldsymbol{u}_{n+1}\phi_j(1) - \boldsymbol{u}_n\phi_j(0) \tag{2.6}$$

for $j = 1,\ldots,m+1$, where $\{\phi_j(\tau)\}_{j=1}^{m+1}$ is a basis of $\mathbb{P}^m([0,1])$, $\boldsymbol{u}'[i] := \frac{d}{d\tau}(\boldsymbol{u}(t_n+\tau h))|_{\tau=c_i}$ and $\boldsymbol{f}[i] := \boldsymbol{f}(t_n + c_i h, \boldsymbol{u}(t_n + c_i h))$. For convenience, we call (2.5) and (2.6) the k-degree time finite element methods with quadrature (k-QTFE methods).

Remark 2.1. The k-QTFE methods in the form (2.5) or (2.6) are independent of the choice of the basis $\{\phi_j(\tau)\}_{j=1}^{m+1}$.

3. Time finite element methods and collocation methods

In this section, we show that the QTFE methods presented in the previous section can be related to continuous (discontinuous) collocation methods. We first introduce two definitions.

Definition 3.1 [25]. Let $c_1,\ldots,c_s \in [0,1]$ be distinct real numbers. The collocation polynomial $\boldsymbol{u}(t)$ is a polynomial of degree s satisfying

$$\begin{cases} \boldsymbol{u}(t_0) = \boldsymbol{z}_0, \\ \dot{\boldsymbol{u}}(t_0 + c_i h) = \boldsymbol{f}(t_0 + c_i h, \boldsymbol{u}(t_0 + c_i h)), \qquad i = 1,\ldots,s. \end{cases} \tag{3.1}$$

The numerical solution of the *s-stage* (*continuous*) *collocation method* is defined by $\boldsymbol{z}_1 = \boldsymbol{u}(t_0 + h)$.

Definition 3.2 [25]. Let $c_2,\ldots,c_{s-1} \in [0,1]$ be distinct real numbers, and let $b^{(1)}, b^{(2)}$ be two arbitrary real numbers. The corresponding *s-stage discontinuous collocation method* is then defined via a polynomial of degree $s-2$ satisfying

$$\begin{cases} \boldsymbol{u}(t_0) = \boldsymbol{z}_0 - hb^{(1)}(\dot{\boldsymbol{u}}(t_0) - \boldsymbol{f}(t_0, \boldsymbol{u}(t_0)), \\ \dot{\boldsymbol{u}}(t_0 + c_i h) = \boldsymbol{f}(t_0 + c_i h, \boldsymbol{u}(t_0 + c_i h)), \qquad i = 2, \ldots, s-1, \\ \boldsymbol{z}_1 = \boldsymbol{u}(t_1) - hb^{(2)}(\dot{\boldsymbol{u}}(t_1) - \boldsymbol{f}(t_1, \boldsymbol{u}(t_1)). \end{cases} \tag{3.2}$$

It is well known that s-stage continuous (discontinuous) collocation methods are equivalent to s-stage RK methods with superconvergence order (see [25]). The s-stage discontinuous collocation method is also called as the singly discontinuous collocation method for $b^{(1)} = 0$ or $b^{(2)} = 0$. When $b^{(1)} = b^{(2)} \equiv 0$, s-stage discontinuous collocation methods can be reduced to $(s-2)$-stage continuous collocation methods.

Now we show that, by choosing the proper quadrature formula the QTFE methods can cover all the variations of these collocation methods as described above.

Theorem 3.1. *Take $q = m + 1$ and assume the order of quadrature formula (2.4) is p (i.e. exact for polynomials of degree $p - 1$), then*

(i) *the QTFE method (2.5) based on k-C-TFE method is equivalent to the k-stage collocation method, and the order is p ($\leqslant 2k$);*
(ii) *when $c_1 = 0$, the QTFE method (2.5) based on k-LD-TFE method is equivalent to the $(k+2)$-stage discontinuous collocation method with $(b^{(1)}, b^{(2)}) = (w_1, 0)$. The order of the corresponding QTFE method is p ($\leqslant 2k+1$);*
(iii) *when $c_q = 1$, the QTFE method (2.5) based on k-RD-TFE method is equivalent to the $(k+2)$-stage discontinuous collocation method with $(b^{(1)}, b^{(2)}) = (0, w_q)$. The order of the corresponding QTFE method is p ($\leqslant 2k+1$);*
(iv) *when $c_1 = 0$ and $c_q = 1$, the QTFE method (2.5) based on k-BD-TFE method is equivalent to the $(k+2)$-stage discontinuous collocation method with $(b^{(1)}, b^{(2)}) = (w_1, w_q)$. The order of the corresponding QTFE method is p ($\leqslant 2k+2$).*

Proof. (i) For the QTFE method (2.5) based on k-C-TFE method, $m = k - 1$ and thus $q = k$. By Remark 2.1 we can take $\phi_j(\tau)$ as the Lagrange interpolation polynomials $l_j(\tau)$ (w.r.t. $c_1, \ldots, c_q$), then

$$\begin{cases} \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \\ \sum_{i=1}^{k} w_i(\boldsymbol{u}'[i] - h\boldsymbol{f}[i]) l_j(c_i) = 0, \qquad j = 1, \ldots, k \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}). \end{cases} \tag{3.3}$$

Since $l_j(c_i) = \delta_{ij}$ and $\boldsymbol{u}'[i] = h\dot{\boldsymbol{u}}(t_n + c_i h)$, (3.3) is reduced to the collocation method of k-stage. The order of the QTFE method can be derived from [25]. (ii)–(iv) can be discussed in a similar way. □

By considering different values of q, we derive the following results.

Theorem 3.2. *Take $q = m$ or $q = m - 1$ and assume the order of (2.4) is p, then*

(i) *when $q = m$ and $c_i \in (0, 1]$ (respectively $c_i \in [0, 1)$), the QTFE method (2.5) based on k-LD-TFE (respectively k-RD-TFE) method is equivalent to the k-stage collocation method. The order is p ($\leqslant 2k$);*
(ii) *the QTFE method (2.5) based on k-BD-TFE method is equivalent to*
- *the k-stage collocation method when $q = m - 1$ and $c_i \in (0, 1)$. The order is p ($\leqslant 2k$);*
- *the $(k+2)$-stage discontinuous collocation method with $(b^{(1)}, b^{(2)}) = (w_1, 0)$ (respectively $(b^{(1)}, b^{(2)}) = (0, w_2)$), when $q = m$ and $c_1 = 0$, $c_i \in (0, 1)$ for $i \geqslant 2$ (respectively $c_q = 1, c_i \in (0, 1)$ for $i \leqslant q - 1$). The order is p ($\leqslant 2k+1$);*
- *the $(k+1)$-stage collocation method when $q = m$ and $c_i \in (0, 1)$. The order is p ($\leqslant 2k+2$).*

Remark 3.1. In integrals $\int_0^1 \boldsymbol{u}'\phi_j \, d\tau$ and $\int_0^1 \boldsymbol{u}'\phi_j \, d\tau, \boldsymbol{u}$ and ϕ_j for $1 \leqslant j \leqslant m+1$ are polynomials. Therefore, by quadrature formula (2.4) with high enough order p the integrals can be calculated exactly. For instance, in k-C-TFE case we should require $p \geqslant 2k - 1$. In such a case, the QTFE method (2.6) is equivalent to (2.5), and thus we can also relate (2.6) to the continuous (discontinuous) collocation method.

From Theorem 3.1 it follows that the k-C-TFE method with k-point Gaussian quadrature is equivalent to Gauss collocation method. For linear Hamiltonian systems, since the integral can be calculated exactly by the k-point Gaussian quadrature, the one-step time discretization based on k-C-TFE method is symplectic and energy-preserving. This conclusion was also presented in [13] by relating k-C-TFE method to $2k$-order diagonal Padé approximation. Similarly, the k-BD-TFE method applied to linear Hamiltonian systems also provides a symplectic numerical flow.

4. TFE methods, RK methods and energy-preserving integrators

In [4], Bottasso presented a variational interpretation for some RK methods, and investigated the k-QTFE method (2.6) based on k-BD-TFE and k-LD-TFE methods with a $(k+1)$-point quadrature formula. In this section, we study more general cases. We first interpret TFE methods as RK methods with continuous stage [24], then the TFE methods can be related to the

usual RK methods by using the corresponding quadrature formulae. We also provide a variational interpretation for some existing energy-preserving integrators.

4.1. Time finite element methods and Runge–Kutta methods

Denote $\xi_k = \sqrt{2k+1}$, the k-degree normalized shifted Legendre polynomial is defined as

$$p_k(x) = \frac{\xi_k}{k!}\frac{d^k}{dx^k}[(x^2-x)^k].$$

The roots of the polynomials $p_k(x)$, $p_k(x)/\xi_k + p_{k-1}(x)/\xi_{k-1}$, $p_k(x)/\xi_k - p_{k-1}(x)/\xi_{k-1}$ and $\int_0^x p_{k-1}(\sigma)\,d\sigma$ are called k-order Gaussian, Radau-left, Radau-right and Lobatto points, respectively.

We list some properties of $p_k(x)$ as follows:

$$\left.\begin{aligned}
&p_k(1-x) = (-1)^k p_k(x), \quad p_k(0) = (-1)^k\xi_k, \quad p_k(1) = \xi_k, \qquad k = 0,1,2,\dots\\
&\int_0^\alpha p_k(x)\,dx = \begin{cases} \frac{p_{k+1}(\alpha)}{2\xi_k\xi_{k+1}} - \frac{p_{k-1}(\alpha)}{2\xi_k\xi_{k-1}}, & k = 1,2,3,\dots\\ \alpha, & k = 0\end{cases}\\
&\sum_{k=0}^n \xi_k\int_0^\alpha p_k(x)\,dx = \frac{p_{n+1}(\alpha)}{2\xi_{n+1}} + \frac{p_n(\alpha)}{2\xi_n}, \qquad n = 0,1,2,\dots\\
&\sum_{k=0}^n (-1)^k\xi_k\int_0^\alpha p_k(x)\,dx = \frac{(-1)^n p_{n+1}(\alpha)}{2\xi_{n+1}} - \frac{(-1)^n p_n(\alpha)}{2\xi_n} + 1, \qquad n = 0,1,2,\dots\\
&\sum_{k=0}^n p_k(\alpha)\int_0^\beta p_k(x)\,dx = \sum_{k=1}^n \frac{p_k(\alpha)p_{k+1}(\beta)}{2\xi_k\xi_{k+1}} - \sum_{k=0}^{n-1}\frac{p_k(\beta)p_{k+1}(\alpha)}{2\xi_k\xi_{k+1}} + \beta, \qquad n = 0,1,2,\dots,
\end{aligned}\right\} \tag{4.1}$$

where $\alpha, \beta \in \mathbf{R}$.

It is clear that the normalized shifted Legendre polynomials $p_j(\tau), j = 0,\dots,k$ form a basis of $\mathbb{P}^k([0,1])$. For $\boldsymbol{u}(t_n+\tau h) \in (\mathbb{P}^k([0,1]))^d$ thus $\boldsymbol{u}'(t_n+\tau h) \in (\mathbb{P}^{k-1}([0,1]))^d$, we have the following expansions (The same notations are presented in [6,8])

$$\boldsymbol{u}(t_n+\tau h) = \sum_{l=0}^k \boldsymbol{\lambda}_l p_l(\tau), \qquad \boldsymbol{u}'(t_n+\tau h) = \sum_{l=0}^{k-1}\boldsymbol{\gamma}_l p_l(\tau). \tag{4.2}$$

Integrating the second equality of (4.2) leads to another expansion of $\boldsymbol{u}(t_n+\tau h)$

$$\boldsymbol{u}(t_n+\tau h) = \boldsymbol{u}(t_n) + \sum_{l=0}^{k-1}\boldsymbol{\gamma}_l\int_0^\tau p_l(x)\,dx. \tag{4.3}$$

In what follows, we interpret TFE methods as RK methods with continuous stage (namely infinitely many stages) which can be reduced to the usual RK methods by using a quadrature formula. For convenience, hereafter sometimes we use the simplified notation $\boldsymbol{v}(\tau) := \boldsymbol{v}(t_n+\tau h)$.

- k-C-TFE case ($m = k-1$).

  By taking[1] $\boldsymbol{v} = p_j(\tau), j = 0,\dots,k-1$, it follows from (2.2) that

$$\begin{cases} (\boldsymbol{u}', p_j) - (h\boldsymbol{f}, p_j) = 0, \qquad j = 0,\dots,k-1,\\ \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \quad \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}),\end{cases} \tag{4.4}$$

where $(\cdot,\cdot)$ is the L^2 inner product on $[0,1]$.

Due to the orthogonality of $\{p_j\}_{j=0}^{k-1}$, substituting the second equality of (4.2) into (4.4) leads to $\boldsymbol{\gamma}_j = (h\boldsymbol{f}, p_j)$. Therefore, by means of (4.1), from (4.3) and (4.4) we obtain

$$\begin{aligned}
&\boldsymbol{u}(t_n+\tau h) = \boldsymbol{u}_n + h\int_0^1\left(\sum_{l=0}^{k-1}\int_0^\tau p_l(x)\,dx\,p_l(\sigma)\right)\boldsymbol{f}(t_n+\sigma h, \boldsymbol{u}(t_n+\sigma h))\,d\sigma,\\
&\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + h\int_0^1 \boldsymbol{f}(t_n+\sigma h, \boldsymbol{u}(t_n+\sigma h))\,d\sigma.
\end{aligned} \tag{4.5}$$

Set $\boldsymbol{U}_\tau = \boldsymbol{u}(t_n+\tau h)$, then (4.5) can be viewed as a RK method with continuous stage $\tau \in [0,1]$

$$\begin{aligned}
&\boldsymbol{U}_\tau = \boldsymbol{u}_n + h\int_0^1 A_{\tau,\sigma}\boldsymbol{f}(t_n + C_\sigma h, \boldsymbol{U}_\sigma)\,d\sigma,\\
&\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + h\int_0^1 B_\tau\boldsymbol{f}(t_n + C_\tau h, \boldsymbol{U}_\tau)\,d\tau,
\end{aligned} \tag{4.6}$$

[1] It is more convenient to replace the vector test functions by scalar ones, which will not violate the equivalence.

where

$$A_{\tau,\sigma}=\sum_{\iota=0}^{k-1}\int_0^\tau p_\iota(x)\,\mathrm{d}x\,p_\iota(\sigma),\quad B_\tau=1,\quad C_\tau=\tau. \tag{4.7}$$

With abuse of notation $\boldsymbol{U}_i=\boldsymbol{u}(t_n+c_ih)$, integrating (4.6) by the quadrature formula (2.4) gives a q-stage RK method

$$\begin{aligned}&\boldsymbol{U}_i=\boldsymbol{u}_n+h\sum_{j=1}^{q}a_{ij}\boldsymbol{f}(t_n+c_jh,\boldsymbol{U}_j),\qquad i=1,\dots,q,\\&\boldsymbol{u}_{n+1}=\boldsymbol{u}_n+h\sum_{i=1}^{q}b_i\boldsymbol{f}(t_n+c_ih,\boldsymbol{U}_i)\end{aligned} \tag{4.8}$$

with

$$a_{ij}=w_j\sum_{\iota=0}^{k-1}\int_0^{c_i}p_\iota(x)\,\mathrm{d}x\,p_\iota(c_j),\quad b_i=w_i. \tag{4.9}$$

In [5–9], (4.5) and (4.8)–(4.9) are called ∞-HBVMs and HBVM (q,k), respectively, which possess energy-preserving property for solving Hamiltonian systems.

- k-LD-TFE case $(m=k)$.

In this case $\boldsymbol{u}_{n+1}=\boldsymbol{u}(t_{n+1})$. Using the first expansion of (4.2) and taking $\boldsymbol{v}=1,\int_0^\tau p_0(x)\mathrm{d}x,\dots,\int_0^\tau p_{k-1}(x)\mathrm{d}x$, from (2.3) we derive a RK method in the form (4.6) with

$$A_{\tau,\sigma}=1+\sum_{\iota=0}^{k-1}\int_0^\sigma p_\iota(x)\,\mathrm{d}x\left(\frac{\sqrt{2\iota+1}}{\sqrt{2k+1}}p_k(\tau)-p_\iota(\tau)\right),\quad B_\tau=1,\quad C_\tau=\tau. \tag{4.10}$$

Inserting the quadrature formula (2.4) yields a q-stage RK method in the form (4.8) with

$$a_{ij}=w_j\left(1+\sum_{\iota=0}^{k-1}\int_0^{c_j}p_\iota(x)\,\mathrm{d}x\left(\frac{\sqrt{2\iota+1}}{\sqrt{2k+1}}p_k(c_i)-p_\iota(c_i)\right)\right),\quad b_i=w_i. \tag{4.11}$$

In particular, when $k=0$ the LD-TFE method is implicit Euler method (assuming the ODEs is autonomous).

- k-RD-TFE case $(m=k)$.

In this case $\boldsymbol{u}(t_n)=\boldsymbol{u}_n$. Analogously to the k-C-TFE case, from (2.2) we deduce a RK method in the form (4.6) with

$$A_{\tau,\sigma}=\sum_{\iota=0}^{k-1}\int_0^\tau p_\iota(x)\,\mathrm{d}x\left(p_\iota(\sigma)-\frac{\sqrt{2\iota+1}}{\sqrt{2k+1}}p_k(\sigma)\right),\quad B_\tau=1,\quad C_\tau=\tau, \tag{4.12}$$

and a q-stage RK method in the form (4.8) with

$$a_{ij}=w_j\sum_{\iota=0}^{k-1}\int_0^{c_i}p_\iota(x)\,\mathrm{d}x\left(p_\iota(c_j)-\frac{\sqrt{2\iota+1}}{\sqrt{2k+1}}p_k(c_j)\right),\quad b_i=w_i. \tag{4.13}$$

In particular, when $k=0$ the RD-TFE method is explicit Euler method (assuming the ODEs is autonomous).

- k-BD-TFE case $(m=k+1)$.

Analogously to the k-LD-TFE case, from (2.3) we derive a RK method in the form (4.6) with

$$A_{\tau,\sigma}=1-\sum_{\iota=0}^{k}p_\iota(\tau)\int_0^\sigma p_\iota(x)\,\mathrm{d}x,\quad B_\tau=1,\quad C_\tau=\tau, \tag{4.14}$$

and a q-stage RK method in the form (4.8) with

$$a_{ij}=w_j\left(1-\sum_{\iota=0}^{k}p_\iota(c_i)\int_0^{c_j}p_\iota(x)\,\mathrm{d}x\right),\quad b_i=w_i. \tag{4.15}$$

When $k=0$, (4.15) becomes $a_{ij}=w_j(1-c_j),b_i=w_i$ for $i,j=1,\dots,q$. In particular, when $q=2,(c_1,c_2)=(0,1)$ and $(w_1,w_2)=(1/2,1/2)$, we get the 2-stage Lobatto IIIB method.

Remark 4.1. If we calculate the second integral term (i.e. the nonlinear term) on the left side of (2.2) or (2.3) by a quadrature formula, then the resulting variational method is equivalent to the q-stage RK method (4.8) with coefficients (4.9), (4.11), (4.13) and (4.15), respectively. The RK method (4.8) can also be related to the continuous (discontinuous) collocation method (see Remark 3.1).

Remark 4.2. When an interpolatory quadrature formula corresponding to the abscissae $\{c_i\}_{i=1}^q$ satisfying $c_{q+1-i} = 1 - c_i$, $i = 1, \ldots, q$ is used, then the q-stage RK method with coefficients (4.9) ((4.15) respectively) is symmetric (see Theorem 2.3 in [25]). Here, we point out that the RK methods with continuous stage based on C-TFE method (BD-TFE method respectively) are also symmetric (one only need to verify $A_{\tau,\sigma} + A_{1-\tau,1-\sigma} = 1$).

From Remark 4.1, we know that the k-C-TFE method can be related to Gauss and Radau IIA collocation methods by integrating the nonlinear term with the k-point Gaussian and Radau-right quadrature formulae, respectively. Besides, by using the proper quadrature formulae for the k-SD-TFE and k-BD-TFE methods, we can also obtain Radau IA, Lobatto IIIC, Lobatto IIIB methods and the methods of Butcher [11] etc.

To study the order of the RK methods with continuous stage, we introduce the following simplifying assumptions presented in [24]

$$\overline{B}(\xi): \quad \int_0^1 B_\tau C_\tau^{\kappa-1}\,\mathrm{d}\tau = \frac{1}{\kappa}, \quad \kappa = 1, \ldots, \xi,$$

$$\overline{C}(\eta): \quad \int_0^1 A_{\tau,\sigma} C_\sigma^{\kappa-1}\,\mathrm{d}\sigma = \frac{1}{\kappa} C_\tau^{\kappa}, \quad \kappa = 1, \ldots, \eta,$$

$$\overline{D}(\zeta): \quad \int_0^1 B_\tau C_\tau^{\kappa-1} A_{\tau,\sigma}\,\mathrm{d}\tau = \frac{1}{\kappa} B_\sigma (1 - C_\sigma^{\kappa}), \quad \kappa = 1, \ldots, \zeta.$$

Applying the method (4.6) to (2.1), we have the following lemma which is similar to the corresponding result for the usual RK methods (see [26]).

Lemma 4.1. *Assume $\boldsymbol{f}$ is Lipschitz continuous with constant L, and the partition of the interval $[0, T]$ is uniform, i.e. $h_n = h = T/N$. If*

$$h < \frac{1}{L \max_{\tau\in[0,1]} \int_0^1 |A_{\tau,\sigma}|\,\mathrm{d}\sigma},$$

then there exists a unique solution of (4.6).

Theorem 4.1. *The TFE methods have the following error estimates:*

(i) *At the end node t_{n+1}, the k-TFE methods have the local truncation error estimate (superconvergence): $\boldsymbol{u}_{n+1} - \boldsymbol{z}(t_{n+1}) = \mathcal{O}(h^{2k+l})$, provided that $\boldsymbol{u}_n \equiv \boldsymbol{z}(t_n)$. Specifically, (a) for the k-C-TFE method, $l = 1$; (b) for the k-SD-TFE method, $l = 2$; (c) for the k-BD-TFE method, $l = 3$.*

(ii) *Fix $I = [0, T]$, then for each element, the k-TFE methods have the global error estimate: $\boldsymbol{u}_n(t_n + \tau h) - \boldsymbol{z}(t_n + \tau h) = \mathcal{O}(h^{k+1})$, uniformly for $\tau \in [0, 1]$ and $n = 0, 1, \ldots, N - 1$.*

(iii) *At special nodes $\{\tau_i\}_{i=1}^{k+1}$ of each element, the k-TFE methods have the following global error estimate (superconvergence): $\boldsymbol{u}_n(t_n + \tau_i h) - \boldsymbol{z}(t_n + \tau_i h) = \mathcal{O}(h^{k+2})$ uniformly for $n = 0, 1, \ldots, N - 1$, where $\{\tau_i\}_{i=1}^{k+1}$ is a set of (a) $(k+1)$-order Lobatto points for the k-C-TFE method when $k \geqslant 2$; (b) $(k+1)$-order Radau-right points for the k-LD-TFE method when $k \geqslant 1$; (c) $(k+1)$-order Radau-left points for the k-RD-TFE method when $k \geqslant 1$; (d) $(k+1)$-order Gaussian points for the k-BD-TFE method when $k \geqslant 0$.*

Proof. (i) By the orthogonality of $\{p_j\}_{j=0}^{k-1}$, we know τ^{s-1} can be expanded as

$$\tau^{s-1} = \sum_{l=0}^{k-1} \int_0^1 \sigma^{s-1} p_l(\sigma)\,\mathrm{d}\sigma\, p_l(\tau), \quad s = 1, \ldots, k,$$

then

$$\int_0^1 \sigma^r p_s(\sigma)\,\mathrm{d}\sigma = 0, \quad 0 \leqslant r < s \in \mathbb{Z}.$$

By (4.1) and the above equalities we can deduce that (4.7) satisfies $\overline{B}(\infty), \overline{C}(k)$ and $\overline{D}(k-1)$; both (4.10) and (4.12) satisfy $\overline{B}(\infty), \overline{C}(k)$ and $\overline{D}(k)$; (4.14) satisfies $\overline{B}(\infty), \overline{C}(k)$ and $\overline{D}(k+1)$. Similarly to the classical result presented by Butcher (see [10]), we can get the order of the k-TFE methods.

(ii) The proof of this part is very similar to that of Lemma 7.5 in [26] and it motivates us to obtain the proof of (iii) by considering a simple modification.

(iii) Denote the coefficient w.r.t. the k-TFE methods by a new notation $A_{\tau,\sigma}^{(k)}$. It follows from the proof of (i) that $A_{\tau,\sigma}^{(k+1)}$ satisfies $\overline{C}(k+1)$, and by (4.1) $A_{\tau_i,\sigma}^{(k+1)} = A_{\tau_i,\sigma}^{(k)}$ for $i = 1, \ldots, k+1$, which implies that $A_{\tau_i,\sigma}^{(k)}$ also satisfies $\overline{C}(k+1)$. Therefore, for the exact solution it gives

$$\boldsymbol{z}(t_n+\tau_i h)=\boldsymbol{z}(t_n)+h\int_0^1 A^{(k)}_{\tau_i,\sigma}\dot{\boldsymbol{z}}(t_n+\sigma h)\,\mathrm{d}\sigma+\mathcal{O}(h^{k+2}). \tag{4.16}$$

It is known that the numerical solution derived by the k-TFE methods satisfies

$$\boldsymbol{u}_n(t_n+\tau_i h)=\boldsymbol{u}_n+h\int_0^1 A^{(k)}_{\tau_i,\sigma}\boldsymbol{f}(t_n+\sigma h,\boldsymbol{u}_n(t_n+\sigma h))\,\mathrm{d}\sigma. \tag{4.17}$$

Subtracting (4.16) from (4.17) gives

$$\boldsymbol{u}_n(t_n+\tau_i h)-\boldsymbol{z}(t_n+\tau_i h)=\boldsymbol{u}_n-\boldsymbol{z}(t_n)+h\int_0^1 A^{(k)}_{\tau_i,\sigma}(\boldsymbol{f}(t_n+\sigma h,\boldsymbol{u}_n(t_n+\sigma h))-\boldsymbol{f}(t_n+\sigma h,\boldsymbol{z}(t_n+\sigma h)))\,\mathrm{d}\sigma+\mathcal{O}(h^{k+2}), \tag{4.18}$$

where $\boldsymbol{f}$ is Lipschitz continuous. Then the result follows from (i) and (ii). □

Corollary 4.1. *Applying the quadrature formula (2.4) of order p to the integration of nonlinear term for the TFE methods gives that (a) when $p \geqslant 2k$, the corresponding q-stage RK method based on k-C-TFE method is of order $2k$; (b) when $p \geqslant 2k+1$, the q-stage RK method based on k-SD-TFE method is of order $2k+1$; (c) when $p \geqslant 2k+2$, the q-stage RK method based on k-BD-TFE method is of order $2k+2$. Similarly to the results shown in (ii) and (iii) of Theorem 4.1, we can also get the global error estimates and superconvergence at special nodes for the resulting numerical discretizations.*

Proof. By the corresponding quadrature formula, the simplifying assumptions $\overline{B}(\xi),\overline{C}(\eta)$ and $\overline{D}(\zeta)$ can be related to the order conditions for the usual RK methods. Using Theorem 7 in [10] and Theorem 4.1 (iii) gives the order estimates of the associated numerical discretizations. □

From Corollary 4.1, we know that the quadrature formula of high enough order needs to be chosen to reach the optimal order of TFE methods. Generally, when the optimal order has been reached the use of quadrature formula with more quadrature points will not improve the order. However, when TFE methods e.g. C-TFE methods are applied to Hamiltonian systems, it is shown in [2,6] that using quadrature formula with more quadrature points is helpful for the sake of energy preservation.

Remark 4.3. The superconvergence of C-TFE and LD-TFE methods shown in Theorem 4.1 and Corollary 4.1 have also been proven in virtue of the theory of finite element methods (see [13,19] and references therein).

Corollary 4.2. *For the k-C-TFE method, applying $(k+1)$-point Lobatto quadrature gives $(k+1)$-stage Lobatto IIIA method; For the k-LD-TFE method, applying $(k+1)$-point Radau-right quadrature gives $(k+1)$-stage Radau IIA method; For the k-RD-TFE method, applying $(k+1)$-point Radau-left quadrature gives $(k+1)$-stage Radau IĀ method [33,34]; For the k-BD-TFE method, applying $(k+1)$-point Gaussian quadrature gives $(k+1)$-stage Gauss method.*

Proof. By corresponding quadrature formula, it follows from the proof of (iii) in Theorem 4.1 that (4.9), (4.11), (4.13) and (4.15) all satisfy $C(k+1)$. This implies that the resulting methods are collocation methods (see Theorem 7.8 in [26]). □

The relation between TFE methods and collocation methods have been studied in Theorems 3.1, 3.2 and Remark 3.1. Furthermore, Corollary 4.2 implies that the $(k+1)$-stage continuous collocation method can be determined by a polynomial of degree k instead of $k+1$. In the case of Lobatto IIIA method, such a property has been observed in [5].

Assume $\theta\in(-\infty,\infty)$, let $A_{\tau,\sigma}$ and $\overline{A}_{\tau,\sigma}$ be two groups of coefficients for two RK methods with continuous stage. Define $A^*_{\tau,\sigma}=\theta A_{\tau,\sigma}+(1-\theta)\overline{A}_{\tau,\sigma}, B_\tau=1$ and $C_\tau=\tau$, then we get a new class of methods in the form (4.6). If $A_{\tau,\sigma}$ and $\overline{A}_{\tau,\sigma}$ satisfy $\overline{C}(\eta_1)$ and $\overline{C}(\eta_2)$, respectively, then $A^*_{\tau,\sigma}$ satisfies $\overline{C}(\eta)$ with $\eta=\min(\eta_1,\eta_2)$. A similar analysis can be made for $\overline{D}(\eta)$. By means of the simplifying assumptions it is easy to obtain the order estimates for the newly defined method. For example, if $A_{\tau,\sigma}$ and $\overline{A}_{\tau,\sigma}$ are derived from the k-LD-TFE and the k-RD-TFE method, respectively, then the newly defined method has order $2k+1$; If $A_{\tau,\sigma}$ and $\overline{A}_{\tau,\sigma}$ are derived from the k-C-TFE and the $(k-1)$-BD-TFE method, respectively, then the newly defined method has order $2k-1$. By the quadrature formula (2.4), the corresponding RK methods can be obtained. In Table 4.1, we list the RK methods of order 5 which are derived from the 2-LD-TFE and 2-RD-TFE methods with 4-point Lobatto quadrature. If the 3-point Gaussian quadrature is used, then we obtain a class of 3-stage RK methods of order 5 (not listed) which are equivalent to Gauss IA methods [34].

Assume the quadrature formula (2.4) is of order p. One can check that the coefficients of the RK methods (4.8) satisfy the order conditions $B(\xi),C(\eta)$ and $D(\zeta)$ with ξ,η and ζ listed below:

(i) $\xi=p$, $\eta=\min(k,p-k+1)$ and $\zeta=\min(k-1,p-k)$ for (4.9);
(ii) $\xi=p$, $\eta=\min(k,p-k)$ and $\zeta=\min(k,p-k)$ for (4.11) and (4.13);
(iii) $\xi=p,\eta=\min(k,p-k-1)$ and $\zeta=\min(k+1,p-k)$ for (4.15).

Table 4.1
New RK method of order 5 with parameter $\theta \in (-\infty, \infty)$.

0	$\frac{1}{12}\theta$	$-\frac{1}{12}\theta$	$-\frac{1}{12}\theta$	$\frac{1}{12}\theta$
$\frac{5-\sqrt{5}}{10}$	$\frac{1}{12}\theta + (1-\theta)\frac{1}{10}$	$\frac{13}{60}\theta + (1-\theta)\frac{1}{5}$	$\frac{13-6\sqrt{5}}{60}\theta + (1-\theta)\frac{2-\sqrt{5}}{10}$	$-\frac{1}{60}\theta$
$\frac{5+\sqrt{5}}{10}$	$\frac{1}{12}\theta + (1-\theta)\frac{1}{10}$	$\frac{13+6\sqrt{5}}{60}\theta + (1-\theta)\frac{2+\sqrt{5}}{10}$	$\frac{13}{60}\theta + (1-\theta)\frac{1}{5}$	$-\frac{1}{60}\theta$
1	$\frac{1}{12}\theta$	$\frac{5}{12}\theta + (1-\theta)\frac{1}{2}$	$\frac{5}{12}\theta + (1-\theta)\frac{1}{2}$	$\frac{1}{12}\theta$
	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

With these order conditions, we can get the error estimates for the q-stage Runge–Kutta methods (4.8) derived from the TFE methods.

Theorem 4.2. *The q-stage Runge–Kutta method (4.8) with coefficients (4.9), (4.11), (4.13) and (4.15) respectively is of order at least* $\min(p, 2\eta+2, \eta+\zeta+1)$, *where η and ζ are as above.*

The results presented in Theorem 4.1 can be understood as the limit case of Theorem 4.2 when $p = \infty$.

4.2. Energy-preserving integrators and their variational interpretation

Consider the autonomous Hamiltonian system

$$\begin{cases} \dot{\boldsymbol{z}} = J^{-1}\nabla H(\boldsymbol{z}), \quad t \in I, \quad \boldsymbol{z} \in \mathbf{R}^{2d}, \\ \boldsymbol{z}(0) = \boldsymbol{z}_0. \end{cases} \tag{4.19}$$

Applying the k-C-TFE method to (4.19) gives the variational formulation

$$\begin{cases} \int_{t_n}^{t_{n+1}} \left(\dot{\boldsymbol{u}}(t) - J^{-1}\nabla H(\boldsymbol{u}(t))\right) \cdot \boldsymbol{v}(t)\,\mathrm{d}t = 0, \\ \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \quad \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}) \end{cases} \tag{4.20}$$

for $\boldsymbol{u} \in \left(\mathbb{P}^k(I_n)\right)^{2d}$ and arbitrary $\boldsymbol{v} \in \left(\mathbb{P}^{k-1}(\bar{I}_n)\right)^{2d}$. It follows from (4.20) by taking $\boldsymbol{v} = J^{-1}\dot{\boldsymbol{u}}$,

$$0 = \int_{t_n}^{t_{n+1}} \dot{\boldsymbol{u}}(t)^T J^{-1}\dot{\boldsymbol{u}}(t)\,\mathrm{d}t = \int_{t_n}^{t_{n+1}} \nabla H(\boldsymbol{u})^T \dot{\boldsymbol{u}}(t)\,\mathrm{d}t = H(\boldsymbol{u}_{n+1}) - H(\boldsymbol{u}_n). \tag{4.21}$$

This implies that the C-TFE method can preserve the Hamiltonian H exactly (see [20,2,14]).

After integrating the first term of (4.20) by the quadrature formula (2.4), we have

$$\begin{cases} \sum_{\iota=1}^{q} w_\iota \boldsymbol{u}'(t_n + c_\iota h) \cdot \boldsymbol{v}(c_\iota) = h\int_0^1 J^{-1}\nabla H(\boldsymbol{u}(t_n + \tau h)) \cdot \boldsymbol{v}(\tau)\,\mathrm{d}\tau, \\ \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \quad \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}) \end{cases} \tag{4.22}$$

for arbitrary $\boldsymbol{v}(\tau) \in \left(\mathbb{P}^{k-1}([0,1])\right)^{2d}$, where we introduce the simplified notations $\boldsymbol{v}(c_\iota) := \boldsymbol{v}(t_n + c_\iota h)$ as well as $\boldsymbol{v}(\tau) := \boldsymbol{v}(t_n + \tau h)$. If the order of the q-point quadrature formula is p ($p \geqslant 2k-1$), (4.22) becomes (4.20). Denote $\boldsymbol{f}(\boldsymbol{u}) := J^{-1}\nabla H(\boldsymbol{u})$. Substituting (4.3) into (4.22) and taking $\boldsymbol{v} = p_i(\tau)$ for $0 \leqslant i \leqslant k-1$, gives

$$\begin{cases} \sum_{j=0}^{k-1}\sum_{\iota=1}^{q} w_\iota p_i(c_\iota) p_j(c_\iota)\gamma_j = h\int_0^1 \boldsymbol{f}\left(\boldsymbol{u}_n + \sum_{j=0}^{k-1}\gamma_j \int_0^\tau p_j(x)\,\mathrm{d}x\right) \cdot p_i(\tau)\,\mathrm{d}\tau, \ i = 0, \ldots, k-1, \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \boldsymbol{\gamma}_0. \end{cases} \tag{4.23}$$

Let $M = (M_{ij})$ be the $k \times k$ matrix with $M_{ij} = \sum_{\iota=1}^{q} w_\iota p_i(c_\iota) p_j(c_\iota)$. If M is non-degenerate, then (4.23) can still be recast as a RK method with continuous stage, the order of which can be analyzed by the corresponding simplifying assumptions. It is easy to check that both (4.22) and (4.23) can preserve the Hamiltonian. In practice, we need to use a quadrature formula for the integration of the nonlinear term on the right side of (4.23). This leads to that the resulting discretization can only preserve the Hamiltonian up to a certain order.

In what follows, we relate C-TFE methods to some existing energy-preserving integrators.

- Averaged Vector Field method (AVF method) [32].
  Consider the 1-C-TFE method, in which $\boldsymbol{u}(t) = \boldsymbol{u}(t_n + \tau h) = (1-\tau)\boldsymbol{u}_n + \tau\boldsymbol{u}_{n+1}$ and the test function $\boldsymbol{v}(t)$ is constant. It follows from (4.20) that

$$\int_{t_n}^{t_{n+1}} (\dot{\boldsymbol{u}}(t) - \boldsymbol{f}(\boldsymbol{u}(t))) \cdot 1\,\mathrm{d}t = 0,$$

which is the AVF method

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + h\int_0^1 \boldsymbol{f}((1-\tau)\boldsymbol{u}_n + \tau\boldsymbol{u}_{n+1})\,\mathrm{d}\tau.$$

- Hamiltonian boundary value methods (HBVMs) [5,6].
  When the Hamiltonian in (4.20) is a polynomial, the integral can be calculated exactly by an appropriate quadrature formula. As already noticed in Section 4.1, HBVMs are the C-TFE methods with corresponding quadrature. The ∞-HBVMs [6] are nothing but another equivalent form of the C-TFE method. Besides, the s-stage trapezoidal methods presented in [29] can also be related to the 1-C-TFE method.
- Energy-preserving collocation methods [24].
  Take $q = k$ in (4.22) and notice $\boldsymbol{u}'(t_n + c_\iota h) = h\dot{\boldsymbol{u}}(t_n + c_\iota h)$, then

$$\begin{cases} \sum_{\iota=1}^{k} w_\iota \dot{\boldsymbol{u}}(t_n + c_\iota h)\cdot \boldsymbol{v}(c_\iota) = \int_0^1 \boldsymbol{f}(\boldsymbol{u}(t_n+\tau h))\cdot \boldsymbol{v}(\tau)\,\mathrm{d}\tau, \\ \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \quad \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}) \end{cases} \tag{4.24}$$

holds for arbitrary $\boldsymbol{v}(\tau) \in (\mathbb{P}^{k-1}([0,1]))^{2d}$. Choose $\boldsymbol{v} = l_i(\tau)$ (the Lagrange polynomial) for $1 \leqslant i \leqslant k$, then it follows (note that $l_i(c_\iota) = \delta_{i\iota}$)

$$\begin{cases} \dot{\boldsymbol{u}}(t_n + c_i h) = \frac{1}{w_i}\int_0^1 l_i(\tau)\boldsymbol{f}(\boldsymbol{u}(t_n+\tau h))\,\mathrm{d}\tau, \quad i = 1,\ldots,k, \\ \boldsymbol{u}(t_n) = \boldsymbol{u}_n, \quad \boldsymbol{u}_{n+1} = \boldsymbol{u}(t_{n+1}). \end{cases} \tag{4.25}$$

This is the energy-preserving collocation method presented by Hairer [24]. When $\{c_i\}_{i=1}^k$ are chosen as Gaussian (or Radau) points, (4.25) is equivalent to the k-C-TFE method. In [24], it has been proven that the method (4.25) with optimal order $2k$, i.e. the k-C-TFE method, is conjugate–symplectic up to at least order $2k+2$. Therefore, the C-TFE method should have the symplectic-like property.

Example 4.1. When $k = 1$ and $\sum_{\iota=1}^{q} w_\iota = 1$, (4.23) is the AVF method.

Example 4.2. Denote $\widetilde{\gamma}_j = \sqrt{2j+1}\gamma_j$ for $0 \leqslant j \leqslant k-1$. When $k = 2$, using the compound trapezoidal formula with $(c_1, c_2, c_3) = (0, 1/2, 1)$ and $(w_1, w_2, w_3) = (1/4, 1/2, 1/4)$, from (4.23) we have

$$\begin{cases} \widetilde{\gamma}_0 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}_n + \widetilde{\gamma}_0\tau + (\tau^2-\tau)\widetilde{\gamma}_1)\,\mathrm{d}\tau, \\ \frac{1}{2}\widetilde{\gamma}_1 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}_n + \widetilde{\gamma}_0\tau + (\tau^2-\tau)\widetilde{\gamma}_1)\cdot(2\tau-1)\,\mathrm{d}\tau, \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \widetilde{\gamma}_0 \end{cases} \tag{4.26}$$

which is of order 2. Using the (uncompounded) trapezoidal formula with $(c_1, c_2) = (0, 1)$ and $(w_1, w_2) = (1/2, 1/2)$, from (4.23) we have

$$\begin{cases} \widetilde{\gamma}_0 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}_n + \widetilde{\gamma}_0\tau + (\tau^2-\tau)\widetilde{\gamma}_1)\,\mathrm{d}\tau, \\ \widetilde{\gamma}_1 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}_n + \widetilde{\gamma}_0\tau + (\tau^2-\tau)\widetilde{\gamma}_1)\cdot(2\tau-1)\,\mathrm{d}\tau, \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \widetilde{\gamma}_0 \end{cases} \tag{4.27}$$

which is just the energy-preserving collocation method of order 2 [24]. By 2-point Gaussian (Radau) quadrature, (4.23) with $k = 2$ is exactly the 2-C-TFE method, namely HBVM $(\infty, 2)$ defined in [6].

Example 4.3. When $k = 3$, using the compound Lobatto formula with $(c_1, c_2, c_3, c_4, c_5) = (0, 1/4, 1/2, 3/4, 1)$ and $(w_1, w_2, w_3, w_4, w_5) = (1/12, 1/3, 1/6, 1/3, 1/12)$, from (4.23) we derive

$$\begin{cases} \widetilde{\gamma}_0 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}(t_n+\tau h))\,\mathrm{d}\tau, \\ \frac{1}{3}\widetilde{\gamma}_1 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}(t_n+\tau h))\cdot(2\tau-1)\,\mathrm{d}\tau, \\ \frac{7}{32}\widetilde{\gamma}_2 = h\int_0^1 \boldsymbol{f}(\boldsymbol{u}(t_n+\tau h))\cdot(6\tau^2-6\tau+1)\,\mathrm{d}\tau, \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \widetilde{\gamma}_0, \end{cases} \tag{4.28}$$

where $\boldsymbol{u}(t_n+\tau h) = \boldsymbol{u}_n + \widetilde{\gamma}_0\tau + (\tau^2-\tau)\widetilde{\gamma}_1 + (2\tau^3 - 3\tau^2 + \tau)\widetilde{\gamma}_2$ and the notation $\widetilde{\gamma}_j$ is defined as in the previous example. One can check that the method (4.28) is of order 4. Similarly, by 3-point Gaussian (Radau) quadrature, we get the 3-C-TFE method, namely HBVM $(\infty, 3)$.

It is well known that for general Hamiltonian systems, there exists no numerical integrator which can conserve the energy and symplecticity of the system simultaneously [25]. However, in TFE framework the two properties are closely linked with each other by C-TFE methods with quadratures of various order: Gauss collocation methods (symplectic) are derived from the k-C-TFE method with Gaussian quadrature of k-point; The quasi-energy-preserving methods are derived from the k-C-TFE method with Gaussian quadrature of higher-order. The similar results were presented in [6,15].

5. PRK schemes obtained by combining different TFE methods

Let us start this section by considering the following ODEs in partitioned form

$$\begin{cases} \dot{\boldsymbol{p}} = \boldsymbol{f}(t,\boldsymbol{p},\boldsymbol{q}), \quad \boldsymbol{p}(0) = \boldsymbol{p}_0, \quad t \in I, \quad \boldsymbol{p},\boldsymbol{q} \in \mathbf{R}^d, \\ \dot{\boldsymbol{q}} = \boldsymbol{g}(t,\boldsymbol{p},\boldsymbol{q}), \quad \boldsymbol{q}(0) = \boldsymbol{q}_0. \end{cases} \tag{5.1}$$

Analogously to the results for TFE methods, approximating $\boldsymbol{p}$ and $\boldsymbol{q}$ in (5.1) with different TFE methods will lead to the PRK method with continuous stage $\tau \in [0,1]$ described as

$$\begin{aligned} &\boldsymbol{Y}_\tau = \boldsymbol{y}_n + h\int_0^1 A_{\tau,\sigma}\boldsymbol{f}(t_n + C_\sigma h, \boldsymbol{Y}_\sigma, \boldsymbol{Z}_\sigma)\,\mathrm{d}\sigma, \quad \boldsymbol{Z}_\tau = \boldsymbol{z}_n + h\int_0^1 \widehat{A}_{\tau,\sigma}\boldsymbol{g}(t_n + C_\sigma h, \boldsymbol{Y}_\sigma, \boldsymbol{Z}_\sigma)\,\mathrm{d}\sigma, \\ &\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\int_0^1 B_\tau \boldsymbol{f}(t_n + C_\tau h, \boldsymbol{Y}_\tau, \boldsymbol{Z}_\tau)\,\mathrm{d}\tau, \quad \boldsymbol{z}_{n+1} = \boldsymbol{z}_n + h\int_0^1 \widehat{B}_\tau \boldsymbol{g}(t_n + C_\tau h, \boldsymbol{Y}_\tau, \boldsymbol{Z}_\tau)\,\mathrm{d}\tau, \end{aligned} \tag{5.2}$$

where $B_\tau = \widehat{B}_\tau = 1$ and $C_\tau = \tau$. When quadrature formulae are used to calculate the integrals, the methods (5.2) are reduced to the corresponding PRK schemes. For simplicity, in the following we will only consider the autonomous system with the time variable t suppressed in (5.1).

- PRK schemes obtained by combining the k-C-TFE and s-C-TFE methods.
  The Galerkin variational problem based on combining the k-C-TFE and s-C-TFE methods is to find $(\boldsymbol{y},\boldsymbol{z}) \in (\mathbb{P}^k(I_n))^d \times (\mathbb{P}^s(I_n))^d$ and $(\mathbf{y}_{n+1},\mathbf{z}_{n+1}) \in \mathbf{R}^d \times \mathbf{R}^d$ such that

$$\begin{cases} \int_{t_n}^{t_{n+1}} (\dot{\mathbf{y}} - \mathbf{f}(\mathbf{y},\mathbf{z}))\cdot \mathbf{v}\,\mathrm{d}t = 0, \quad \mathbf{y}(t_n) = \mathbf{y}_n, \\ \int_{t_n}^{t_{n+1}} (\dot{\mathbf{z}} - \mathbf{g}(\mathbf{y},\mathbf{z}))\cdot \widetilde{\mathbf{v}}\,\mathrm{d}t = 0, \quad \mathbf{z}(t_n) = \mathbf{z}_n, \\ \mathbf{y}_{n+1} = \mathbf{y}(t_{n+1}), \quad \mathbf{z}_{n+1} = \mathbf{z}(t_{n+1}) \end{cases} \tag{5.3}$$

holds for arbitrary $(\mathbf{v},\widetilde{\mathbf{v}}) \in (\mathbb{P}^{k-1}(\bar{I}_n))^d \times (\mathbb{P}^{s-1}(\bar{I}_n))^d$. Applying the quadrature formula (2.4) to the integration of nonlinear terms, from (5.3) we derive a q-stage PRK scheme

$$\begin{aligned} &\mathbf{Y}_i = \mathbf{y}_n + h\sum_{j=1}^{q} a_{ij}\mathbf{f}(\mathbf{Y}_j,\mathbf{Z}_j), \quad \mathbf{Z}_i = \mathbf{z}_n + h\sum_{j=1}^{q} \widehat{a}_{ij}\mathbf{g}(\mathbf{Y}_j,\mathbf{Z}_j), \quad i = 1,\ldots,q, \\ &\mathbf{y}_{n+1} = \mathbf{y}_n + h\sum_{i=1}^{q} b_i\mathbf{f}(\mathbf{Y}_i,\mathbf{Z}_i), \quad \mathbf{z}_{n+1} = \mathbf{z}_n + h\sum_{i=1}^{q} \widehat{b}_i\mathbf{g}(\mathbf{Y}_i,\mathbf{Z}_i), \end{aligned} \tag{5.4}$$

where

$$a_{ij} = w_j\sum_{\iota=0}^{k-1}\int_0^{c_i} p_\iota(x)\,\mathrm{d}x\,p_\iota(c_j), \quad b_i = w_i, \quad \widehat{a}_{ij} = w_j\sum_{\iota=0}^{s-1}\int_0^{c_i} p_\iota(x)\,\mathrm{d}x\,p_\iota(c_j), \quad \widehat{b}_i = w_i. \tag{5.5}$$

- PRK schemes obtained by combining the k-LD-TFE and s-RD-TFE methods.
  The Galerkin variational problem based on combining the k-LD-TFE and s-RD-TFE methods is to find $(\mathbf{y},\mathbf{z}) \in (\mathbb{P}^k(I_n))^d \times (\mathbb{P}^s(I_n))^d$ and $(\mathbf{y}_{n+1},\mathbf{z}_{n+1}) \in \mathbf{R}^d \times \mathbf{R}^d$ such that

$$\begin{cases} \int_{t_n}^{t_{n+1}} (\dot{\mathbf{y}} - \mathbf{f}(\mathbf{y},\mathbf{z}))\cdot \mathbf{v}\,\mathrm{d}t = -(\mathbf{y}(t_n) - \mathbf{y}_n)\cdot \mathbf{v}(t_n), \\ \int_{t_n}^{t_{n+1}} (\dot{\mathbf{z}} - \mathbf{g}(\mathbf{y},\mathbf{z}))\cdot \widetilde{\mathbf{v}}\,\mathrm{d}t = (\mathbf{z}(t_{n+1}) - \mathbf{z}_{n+1})\cdot \widetilde{\mathbf{v}}(t_{n+1}), \\ \mathbf{y}_{n+1} = \mathbf{y}(t_{n+1}), \quad \mathbf{z}(t_n) = \mathbf{z}_n \end{cases} \tag{5.6}$$

holds for arbitrary $(\mathbf{v},\widetilde{\mathbf{v}}) \in (\mathbb{P}^k(\bar{I}_n))^d \times (\mathbb{P}^s(\bar{I}_n))^d$. In particular, when $s = k = 0$, (5.6) is the symplectic Euler method. Applying the quadrature formula (2.4) to the integration of nonlinear terms, we derive a q-stage PRK scheme in the form (5.4) with

$$\begin{aligned} &a_{ij} = w_j\left(1 + \sum_{\iota=0}^{k-1}\int_0^{c_j} p_\iota(x)\,\mathrm{d}x\left(\tfrac{\sqrt{2\iota+1}}{\sqrt{2k+1}}p_k(c_i) - p_\iota(c_i)\right)\right), \quad b_i = w_i, \\ &\widehat{a}_{ij} = w_j\sum_{\iota=0}^{s-1}\int_0^{c_i} p_\iota(x)\,\mathrm{d}x\left(p_\iota(c_j) - \tfrac{\sqrt{2\iota+1}}{\sqrt{2s+1}}p_s(c_j)\right), \quad \widehat{b}_i = w_i. \end{aligned} \tag{5.7}$$

- PRK schemes obtained by combining the k-C-TFE and s-BD-TFE methods.
  The Galerkin variational problem based on combining the k-C-TFE and s-BD-TFE methods is to find $(\mathbf{y},\mathbf{z})\in(\mathbb{P}^k(I_n))^d\times(\mathbb{P}^s(I_n))^d$ and $(\mathbf{y}_{n+1},\mathbf{z}_{n+1})\in\mathbf{R}^d\times\mathbf{R}^d$ such that

$$\begin{cases}\int_{t_n}^{t_{n+1}}(\dot{\mathbf{y}}-\mathbf{f}(\mathbf{y},\mathbf{z}))\cdot\mathbf{v}\,\mathrm{d}t=0,\\ \int_{t_n}^{t_{n+1}}(\dot{\mathbf{z}}-\mathbf{g}(\mathbf{y},\mathbf{z}))\cdot\widetilde{\mathbf{v}}\,\mathrm{d}t=[\mathbf{z}(t_{n+1})-\mathbf{z}_{n+1}]\cdot\widetilde{\mathbf{v}}(t_{n+1})-[\mathbf{z}(t_n)-\mathbf{z}_n]\cdot\widetilde{\mathbf{v}}(t_n),\\ \mathbf{y}(t_n)=\mathbf{y}_n,\quad \mathbf{y}_{n+1}=\mathbf{y}(t_{n+1})\end{cases}\tag{5.8}$$

holds for arbitrary $(\mathbf{v},\widetilde{\mathbf{v}})\in\left(\mathbb{P}^{k-1}(\bar{I}_n)\right)^d\times\left(\mathbb{P}^{s+1}(\bar{I}_n)\right)^d$. Applying the quadrature formula (2.4) to the integration of nonlinear terms, we derive a q-stage PRK scheme in the form (5.4) with

$$a_{ij}=w_j\sum_{\iota=0}^{k-1}\int_0^{c_i}p_\iota(x)\,\mathrm{d}x\,p_\iota(c_j),\quad b_i=w_i,\quad \widehat{a}_{ij}=w_j\left(1-\sum_{\iota=0}^{s}p_\iota(c_i)\int_0^{c_j}p_\iota(x)\,\mathrm{d}x\right),\quad \widehat{b}_i=w_i.\tag{5.9}$$

To discuss the order of the q-stage PRK schemes $(a_{ij},b_i,\hat{a}_{ij},\hat{b}_i)$ derived as above, we introduce the following simplifying assumptions presented in [17]

$$\mathcal{B}(\xi):\ \sum_{i=1}^{q}b_ic_i^{\kappa-1}\widehat{c}_i^{\,\iota}=\tfrac{1}{\kappa+\iota},\quad 1\leqslant\kappa+\iota\leqslant\xi,\quad \mathcal{C}(\eta):\ \sum_{j=1}^{q}a_{ij}c_j^{\kappa-1}\widehat{c}_j^{\,\iota}=\tfrac{c_i^{\kappa+\iota}}{\kappa+\iota},\quad 1\leqslant\kappa+\iota\leqslant\eta,$$

$$\widehat{\mathcal{C}}(\widehat{\eta}):\ \sum_{j=1}^{q}\widehat{a}_{ij}c_j^{\kappa-1}\widehat{c}_j^{\,\iota}=\tfrac{\widehat{c}_i^{\kappa+\iota}}{\kappa+\iota},\quad 1\leqslant\kappa+\iota\leqslant\widehat{\eta},\quad \mathcal{D}(\zeta):\ \sum_{i=1}^{q}b_ic_i^{\kappa-1}\widehat{c}_i^{\,\iota}a_{ij}=\tfrac{b_j(1-\widehat{c}_j^{\kappa+\iota})}{\kappa+\iota},\quad 1\leqslant\kappa+\iota\leqslant\zeta,$$

$$\widehat{\mathcal{D}}(\widehat{\zeta}):\ \sum_{i=1}^{q}b_ic_i^{\kappa-1}\widehat{c}_i^{\,\iota}\,\widehat{a}_{ij}=\tfrac{b_j(1-\widehat{c}_j^{\kappa+\iota})}{\kappa+\iota},\ 1\leqslant\kappa+\iota\leqslant\widehat{\zeta}.$$

Lemma 5.1 [17]. *If a q-stage PRK method with coefficients satisfying $\widehat{b}_i=b_i, c_i=\sum_{j=1}^{q}a_{ij}$ and $\widehat{c}_i=\sum_{j=1}^{q}\widehat{a}_{ij}$ for all i meets the conditions $\mathcal{B}(\xi),\mathcal{C}(\eta),\widehat{\mathcal{C}}(\eta),\mathcal{D}(\zeta),\widehat{\mathcal{D}}(\zeta)$, then it is of order at least* $\min(\xi,2\eta+2,\eta+\zeta+1)$.

Assume the quadrature formula (2.4) has order p. One can check that the coefficients of PRK methods (5.4) satisfy the order conditions $\mathcal{B}(\xi),\mathcal{C}(\eta),\widehat{\mathcal{C}}(\widehat{\eta}),\mathcal{D}(\zeta),\widehat{\mathcal{D}}(\widehat{\zeta})$ with $(\xi,\eta,\zeta,\widehat{\eta},\widehat{\zeta})$ as below (see Theorem 4.2):

(i) $\xi=p,\ \eta=\min(k,p-k+1),\ \widehat{\eta}=\min(s,p-s+1),\ \zeta=\min(k-1,p-k)$ and $\widehat{\zeta}=\min(s-1,p-s)$ for (5.5);
(ii) $\xi=p,\ \eta=\min(k,p-k),\ \widehat{\eta}=\min(s,p-s),\ \zeta=\min(k,p-k)$ and $\widehat{\zeta}=\min(s,p-s)$ for (5.7);
(iii) $\xi=p,\ \eta=\min(k,p-k+1),\ \widehat{\eta}=\min(s,p-s-1),\ \zeta=\min(k-1,p-k)$ and $\widehat{\zeta}=\min(s+1,p-s)$ for (5.9).

By Lemma 5.1, we have the following corollary.

Corollary 5.1. *The q-stage PRK method (5.4) with coefficients (5.5), (5.7) and (5.9) respectively is of order at least* $\min(p,2\widetilde{\eta}+2,\widetilde{\eta}+\widetilde{\zeta}+1)$, *where* $\widetilde{\eta}:=\min(\eta,\widehat{\eta}),\ \widetilde{\zeta}:=\min(\zeta,\widehat{\zeta})$.

As we have pointed out at the beginning of this section, all the combination methods based on different TFE methods are equivalent to PRK methods with continuous stage in the form (5.2). With the corresponding simplifying assumptions, we present the following order estimates.

Theorem 5.1 (*The combination TFE methods have the following superconvergence order*[2]):

(i) *The method (5.3) obtained by combining the k-C-TFE and s-C-TFE methods, is of order at least* $\min(2k,2s)$;
(ii) *The method (5.6) obtained by combining the k-LD-TFE and s-RD-TFE methods, is of order at least* $\min(2k+1,2s+1)$;
(iii) *The method (5.8) obtained by combining the k-C-TFE and s-BD-TFE methods, is of order at least* $\min(2k,k+s,2s+2)$.

For the numerical discretization derived by combining the C-TFE and BD-TFE methods, we provide the further results. By means of the Lobatto quadrature, the resulting discretization can be related to the Lobatto IIIA-IIIB method which is a class of symplectic PRK methods. The results are shown in the following theorem.

Theorem 5.2. *Combining the k-C-TFE and s-BD-TFE methods gives*

(i) *the $(k+1)$-stage Lobatto IIIA-IIIB method of order $2k$ by means of the $(k+1)$-point Lobatto quadrature when $s=k-1$ for $k\geqslant 1$ (or $s=k$ for $k>0$);*
(ii) *the k-stage Lobatto IIIA-IIIB method of order $2k-2$ by means of the k-point Lobatto quadrature when $s=k-2$ for $k\geqslant 2$;*
(iii) *a symmetric PRK method when we calculate the integrals with an interpolatory quadrature formula corresponding to the nodes $\{c_i\}_{i=1}^{q}$ satisfying $c_{q+1-i}=1-c_i,\ i=1,\ldots,q$.*

[2] Here, we present the superconvergence only for three kinds of combination methods, one can deduce the similar results for the other cases.

Proof. (i) When $s = k - 1$, by Corollary 4.2, Theorem 3.1 and Remark 4.1, it is easy to know that the PRK method with coefficients (5.9) is the pair of $(k+1)$-stage Lobatto IIIA and Lobatto IIIB methods.

When $s = k$, with the $(k+1)$-order Lobatto points $\{c_j\}_{j=1}^{k+1}$ (notice that $\int_0^{c_j} p_k(x)\,dx = 0$), the third equality of (5.9) becomes

$$\widehat{a}_{ij} = w_j\left(1 - \sum_{l=0}^{k} p_l(c_i)\int_0^{c_j} p_l(x)\,dx\right) = w_j\left(1 - \sum_{l=0}^{k-1} p_l(c_i)\int_0^{c_j} p_l(x)\,dx\right),$$

which is reduced to the case for $s = k - 1$.

(ii) This result can be proved similarly as above.

(iii) The result follows from Remark 4.2 and Theorem 2.5 in [25].

Remark 5.1. When $s = k - 1$, Theorem 5.1 shows that the method obtained by combining the k-C-TFE and s-BD-TFE methods is of order at least $2k - 1$. In fact, it is easy to prove that this combination method (even for arbitrary s) is symmetric, which implies that the order is $2k$. Theorem 5.2 shows that the optimal order $2k$ can be reached by using a high-order quadrature formula (e.g. the $(k+1)$-point Lobatto quadrature).

Next, by combining different TFE methods, we present new classes of symplectic (P) RK schemes. Recall that a q-stage PRK method is symplectic if the coefficients $(a_{ij}, b_i, \widehat{a}_{ij}, \widehat{b}_i)$ satisfy [25]

$$b_i\widehat{a}_{ij} + \widehat{b}_j a_{ji} = b_i\widehat{b}_j, \quad i,j = 1,\ldots,q, \tag{5.10}$$

$$b_i = \widehat{b}_i, \qquad i = 1,\ldots,q. \tag{5.11}$$

If we assume b_i and $\widehat{b}_i$ are nonzero, then (5.10) implies that the two groups of RK coefficients can be determined by each other. Besides, the sufficient condition for a q-stage RK method (a_{ij}, b_i) to be symplectic, is

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \qquad i,j = 1,\ldots,q. \tag{5.12}$$

It is well known that a RK method can be regarded as a special PRK method with $a_{ij} = \widehat{a}_{ij}$, thus (5.12) is included in (5.10,5.11). When the (P) RK method is irreducible [27], the conditions (5.10,5.11) and (5.12) are also necessary [25] for symplecticity. Similarly, for the (P) RK methods with continuous stage, the symplecticity condition is

$$B_\tau\widehat{A}_{\tau,\sigma} + \widehat{B}_\sigma A_{\sigma,\tau} = B_\tau\widehat{B}_\sigma, \quad \tau,\sigma \in [0,1],$$

$$B_\tau = \widehat{B}_\tau, \quad \tau \in [0,1].$$

It is shown in [34] that new symplectic RK method can be constructed by taking the arithmetic mean of the coefficients of a given symplectic PRK method, i.e. via $(a^*_{ij}, b^*_i) = ((a_{ij} + \widehat{a}_{ij})/2,\ b_i)$. The same technique can be used for constructing new symplectic RK method with continuous stage.

- Symplectic schemes obtained by combining the k-C-TFE and s-C-TFE methods.
  Substituting (5.5) into (5.10) yields

$$w_i w_j\sum_{l=0}^{s-1} p_l(c_j)\int_0^{c_i} p_l(x)\,dx + w_j w_i\sum_{l=0}^{k-1} p_l(c_i)\int_0^{c_j} p_l(x)\,dx = w_i w_j.$$

Denote $\widetilde{p}_j(\tau) = p_j(\tau)/\sqrt{2j+1}$. Eliminating the nonzero factor $w_i w_j$ and using (4.1) provides

$$\left(\sum_{l=1}^{s-1} - \sum_{l=0}^{k-2}\right)\widetilde{p}_l(c_j)\widetilde{p}_{l+1}(c_i) + \left(\sum_{l=1}^{k-1} - \sum_{l=0}^{s-2}\right)\widetilde{p}_l(c_i)\widetilde{p}_{l+1}(c_j) = 2(1 - c_i - c_j),\ i,j = 1,\ldots,q. \tag{5.13}$$

Without loss of generality, we assume $s \geqslant k \geqslant 1$:

(I) When $s = k$, (5.4–5.5) becomes the RK method derived by the k-C-TFE method. From (5.13) we have

$$\widetilde{p}_{k-1}(c_j)\widetilde{p}_k(c_i) + \widetilde{p}_{k-1}(c_i)\widetilde{p}_k(c_j) = 0, \quad i,j = 1,\ldots,q.$$

When $j = i$, it is

$$\widetilde{p}_{k-1}(c_i)\widetilde{p}_k(c_i) = 0, \qquad i = 1,\ldots,q.$$

This implies that $q \leqslant k$ and $\{c_i\}_{i=1}^q$ are the roots of $\widetilde{p}_{k-1}(\tau)$ or $\widetilde{p}_k(\tau)$.

Remark 5.2. Analogously to the above result, it is shown that for the k-C-TFE method to be symplectic the number of quadrature points must be less than or equal to k. When the number of the quadrature points increases, the C-TFE method is not symplectic anymore, though it will preserve the Hamiltonian up to a certain order.

Table 5.1
Two classes of symplectic RK methods derived from the 3-C-TFE method.

$$\begin{array}{c|cc} \frac{1}{2}-\frac{\sqrt{3}}{6} & \frac{1}{2}w_1 & (\frac{1}{2}-\frac{\sqrt{3}}{3})w_2 \\ \frac{1}{2}+\frac{\sqrt{3}}{6} & (\frac{1}{2}+\frac{\sqrt{3}}{3})w_1 & \frac{1}{2}w_2 \\ \hline & w_1 & w_2 \end{array}$$

$$\begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{1}{2}w_1 & (\frac{1}{2}-\frac{3\sqrt{15}}{20})w_2 & (\frac{1}{2}-\frac{3\sqrt{15}}{25})w_3 \\ \frac{1}{2} & (\frac{1}{2}+\frac{3\sqrt{15}}{20})w_1 & \frac{1}{2}w_2 & (\frac{1}{2}-\frac{3\sqrt{15}}{20})w_3 \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & (\frac{1}{2}+\frac{3\sqrt{15}}{25})w_1 & (\frac{1}{2}+\frac{3\sqrt{15}}{20})w_2 & \frac{1}{2}w_3 \\ \hline & w_1 & w_2 & w_3 \end{array}$$

For the 3-C-TFE method ($s=k=3$), the roots of $\widetilde{p}_2(\tau)$ are $1/2\pm\sqrt{3}/6$ and the roots of $\widetilde{p}_3(\tau)$ are $1/2$, $1/2\pm\sqrt{15}/10$. Two classes of new symplectic RK methods are shown in Table 5.1, where $\{w_i\}$ are the parameters. By taking $(w_1,w_2)=(1/2,1/2)$ or $(w_1,w_2,w_3)=(5/18,4/9,5/18)$, we derive the classical 2-stage and 3-stage Gauss collocation methods.

Theorem 3.1 shows that k-C-TFE methods can be related to k-stage Gauss collocation methods by means of k-point Gaussian quadrature. Furthermore, we have the following theorem.

Theorem 5.3. *By means of $(k-1)$-point Gaussian (respectively k-point Lobatto) quadrature, the k-C-TFE method can be related to the $(k-1)$-stage Gauss (respectively k-stage Lobatto IIIA) collocation method.*

Proof. It is known that $p_{k-1}(c_i)=0$ (respectively $\int_0^{c_i}p_{k-1}(x)\mathrm{d}x=0$) at the Gaussian nodes $\{c_i\}_{i=1}^{k-1}$ (Lobatto nodes $\{c_i\}_{i=1}^{k}$, respectively). Therefore, a_{ij} in (5.5) is reduced to

$$a_{ij}=w_j\sum_{l=0}^{k-1}p_l(c_j)\int_0^{c_i}p_l(x)\,\mathrm{d}x=w_j\sum_{l=0}^{k-2}p_l(c_j)\int_0^{c_i}p_l(x)\,\mathrm{d}x.$$

The results follow directly from Theorem 3.1 and Remark 3.1 (Corollary 4.2, respectively). □

(II) When $s=k+1$, it follows from (5.13) that

$$\widetilde{p}_{k-1}(c_j)\widetilde{p}_k(c_i)+\widetilde{p}_k(c_j)\widetilde{p}_{k+1}(c_i)=0,\qquad i,j=1,\ldots,q.$$

Setting $j=i$ gives

$$\widetilde{p}_k(c_i)(\widetilde{p}_{k-1}(c_i)+\widetilde{p}_{k+1}(c_i))=0,\qquad i=1,\ldots,q$$

which holds for the roots of $\widetilde{p}_k(\tau)$.

(III) When $s>k+1$, it is not easy to solve (5.13) except for the trivial case, e.g. $q=1$.

- Symplectic schemes obtained by combining the k-LD-TFE and s-RD-TFE methods.
  When $k=s=0$ and $\sum_i w_i=1$, the PRK method based on combining the k-LD-TFE and s-RD-TFE methods is the symplectic Euler method. In the following, we assume $k,s\geqslant 1$.

(I) When $s=k$, the symplecticity condition (5.10, 5.11) holds for all the coefficients in (5.7). Therefore, we have the following theorem.

Theorem 5.4. *By means of the quadrature with order p, combining the k-LD-TFE and k-RD-TFE methods gives the symplectic PRK methods of order at least* $\min(p,2k+1,2p-2k+1)$. *When $p=\infty$, namely, the corresponding PRK method with continuous stage is also symplectic and of order $2k+1$.*

Proof. The order can be gained directly from Corollary 5.1 and Theorem 5.1. □

Table 5.2
Symplectic PRK methods obtained by combining the LD-TFE and RD-TFE methods.

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array} \quad \text{and} \quad \begin{array}{c|cc} \frac{1}{3} & \frac{1}{3} & 0 \\ 1 & 1 & 0 \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

$$\begin{array}{c|ccc} 0 & \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{6} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array} \quad \text{and} \quad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 1 & 0 & 1 & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

$$\begin{array}{c|cccc} 0 & \frac{1}{12} & -\frac{1}{12} & -\frac{1}{12} & \frac{1}{12} \\ \frac{1}{2}-\frac{\sqrt{5}}{10} & \frac{1}{12} & \frac{13}{60} & \frac{13}{60}-\frac{\sqrt{5}}{10} & -\frac{1}{60} \\ \frac{1}{2}+\frac{\sqrt{5}}{10} & \frac{1}{12} & \frac{13}{60}+\frac{\sqrt{5}}{10} & \frac{13}{60} & -\frac{1}{60} \\ 1 & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \\ \hline & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array} \quad \text{and} \quad \begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2}-\frac{\sqrt{5}}{10} & \frac{1}{10} & \frac{1}{5} & \frac{1}{5}-\frac{\sqrt{5}}{10} & 0 \\ \frac{1}{2}-\frac{\sqrt{5}}{10} & \frac{1}{10} & \frac{1}{5}+\frac{\sqrt{5}}{10} & \frac{1}{5} & 0 \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array}$$

$$\begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{29}{180} & \frac{8}{45}-\frac{\sqrt{15}}{15} & \frac{29}{180}-\frac{\sqrt{15}}{30} \\ \frac{1}{2} & \frac{1}{9}+\frac{\sqrt{15}}{24} & \frac{5}{18} & \frac{1}{9}-\frac{\sqrt{15}}{24} \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{29}{180}+\frac{\sqrt{15}}{30} & \frac{8}{45}+\frac{\sqrt{15}}{15} & \frac{29}{180} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array} \quad \text{and} \quad \begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{7}{60} & \frac{4}{15}-\frac{\sqrt{15}}{15} & \frac{7}{60}-\frac{\sqrt{15}}{30} \\ \frac{1}{2} & \frac{1}{6}+\frac{\sqrt{15}}{24} & \frac{1}{6} & \frac{1}{6}-\frac{\sqrt{15}}{24} \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{7}{60}+\frac{\sqrt{15}}{30} & \frac{4}{15}+\frac{\sqrt{15}}{15} & \frac{7}{60} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array}$$

Table 5.3
Symplectic PRK methods obtained by combining the 2-C-TFE and 1-BD-TFE methods.

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \quad \text{and} \quad \begin{array}{c|cc} 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

$$\begin{array}{c|ccc} 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array} \quad \text{and} \quad \begin{array}{c|cc} 0 & \frac{1}{4} & -\frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & \frac{5}{12} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$$

$$\begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{5}{36}-\frac{\sqrt{15}}{90} & \frac{2}{9}-\frac{2\sqrt{15}}{45} & \frac{5}{36}-\frac{2\sqrt{15}}{45} \\ \frac{1}{2} & \frac{5}{36}+\frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36}-\frac{\sqrt{15}}{24} \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{5}{36}+\frac{2\sqrt{15}}{45} & \frac{2}{9}+\frac{2\sqrt{15}}{45} & \frac{5}{36}+\frac{\sqrt{15}}{90} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array} \quad \text{and} \quad \begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{5}{36}+\frac{\sqrt{15}}{90} & \frac{2}{9}-\frac{\sqrt{15}}{15} & \frac{5}{36}-\frac{2\sqrt{15}}{45} \\ \frac{1}{2} & \frac{5}{36}+\frac{\sqrt{15}}{36} & \frac{2}{9} & \frac{5}{36}-\frac{\sqrt{15}}{36} \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{5}{36}+\frac{2\sqrt{15}}{45} & \frac{2}{9}+\frac{\sqrt{15}}{15} & \frac{5}{36}-\frac{\sqrt{15}}{90} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array}$$

In Table 5.2, we show four symplectic PRK schemes: The first (Radau IIA-II$\bar{\text{A}}$) is derived by combining the 1-LD-TFE and 1-RD-TFE methods with Radau-right quadrature; The second (Lobatto IIIC-IIIC) and the third are derived by combining the 2-LD-TFE and 2-RD-TFE methods with 3-point and 4-point Lobatto quadratures, respectively; The last (Gauss IA-I$\overline{A}$) is derived by combining the 2-LD-TFE and 2-RD-TFE methods with 3-point Gaussian quadrature. The orders of these numerical methods are 3, 4, 5 and 5, respectively.

Table 5.4
Symplectic PRK methods derived from (iv-vii).

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array} \quad \text{and} \quad \begin{array}{c|cc} 0 & \frac{1}{4} & -\frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & \frac{5}{12} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$$

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array} \quad \text{and} \quad \begin{array}{c|cc} \frac{1}{3} & \frac{1}{3} & 0 \\ 1 & 1 & 0 \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

$$\begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{4+\sqrt{6}}{10} & \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ 1 & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \\ \hline & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array} \quad \text{and} \quad \begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{1}{5}-\frac{\sqrt{6}}{120} & \frac{1}{5}-\frac{11\sqrt{6}}{120} & 0 \\ \frac{4+\sqrt{6}}{10} & \frac{1}{5}+\frac{11\sqrt{6}}{120} & \frac{1}{5}+\frac{\sqrt{6}}{120} & 0 \\ 1 & \frac{1}{2}-\frac{\sqrt{6}}{12} & \frac{1}{2}+\frac{\sqrt{6}}{12} & 0 \\ \hline & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array}$$

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 \\ \frac{6-\sqrt{6}}{10} & \frac{9+\sqrt{6}}{75} & \frac{24+\sqrt{6}}{120} & \frac{168-73\sqrt{6}}{600} \\ \frac{6+\sqrt{6}}{10} & \frac{9-\sqrt{6}}{75} & \frac{168+73\sqrt{6}}{600} & \frac{24-\sqrt{6}}{120} \\ \hline & \frac{1}{9} & \frac{16+\sqrt{6}}{36} & \frac{16-\sqrt{6}}{36} \end{array} \quad \text{and} \quad \begin{array}{c|ccc} 0 & \frac{1}{9} & \frac{-1-\sqrt{6}}{18} & \frac{-1+\sqrt{6}}{18} \\ \frac{6-\sqrt{6}}{10} & \frac{1}{9} & \frac{11}{45}+\frac{7\sqrt{6}}{360} & \frac{11}{45}-\frac{43\sqrt{6}}{360} \\ \frac{6+\sqrt{6}}{10} & \frac{1}{9} & \frac{11}{45}+\frac{43\sqrt{6}}{360} & \frac{11}{45}-\frac{7\sqrt{6}}{360} \\ \hline & \frac{1}{9} & \frac{16+\sqrt{6}}{36} & \frac{16-\sqrt{6}}{36} \end{array}$$

In fact, by means of the $(k+1)$-point Gaussian, Radau-left, Radau-right and Lobatto quadrature the combination method based on k-LD-TFE and k-RD-TFE methods will lead to the Gauss IA-IĀ, Radau IA-IĀ, Radau IIA-IIĀ and Lobatto IIIC-IIIC* methods, respectively.

(II) When $s \neq k$, it is not easy to construct symplectic PRK schemes.

- Symplectic schemes obtained by combining the k-C-TFE and s-BD-TFE methods.

Now we consider the PRK schemes derived by combining the k-C-TFE and s-BD-TFE methods.

(I) When $s = k-2$, substituting (5.9) into (5.10) leads to

$$p_{k-1}(c_i)\int_0^{c_j} p_{k-1}(x)\,dx = 0, \quad i,j = 1,\ldots,q.$$

This implies that the PRK scheme is symplectic iff $\{c_i\}$ are the roots of $p_{k-1}(\tau)$ or $\int_0^\tau p_{k-1}(x)\,dx$, i.e. the Gaussian or Lobatto points.

(II) When $s = k-1$, we have the following theorem.

Theorem 5.5. *By means of the quadrature with order p, combining the k-C-TFE and $(k-1)$-BD-TFE methods gives the symplectic PRK methods of order at least* $\min(p, 2k, 2p-2k+2)$. *When $p = \infty$, namely, the corresponding PRK method with continuous stage is also symplectic and of order 2k.*

Proof. The order estimate follows directly from the result presented in [34] and Remark 5.1. □

Remark 5.3. By taking the arithmetic mean of the coefficients of the PRK methods described in Theorems 5.4 and 5.5 we can get lots of symplectic RK methods.

Remark 5.4. Since the PRK schemes described in Theorems 5.4 and 5.5 are symplectic, the two groups of corresponding RK coefficients are linked with each other through (5.10). By virtue of this point we could prove Corollary 4.2 in another way.

In Table 5.3, we show three symplectic PRK schemes derived by combining the 2-C-TFE and 1-BD-TFE methods with Lobatto, Radau-left and Gaussian quadrature, respectively. The first two schemes are Lobatto IIIA-IIIB and Radau IA-IĀ, respectively.

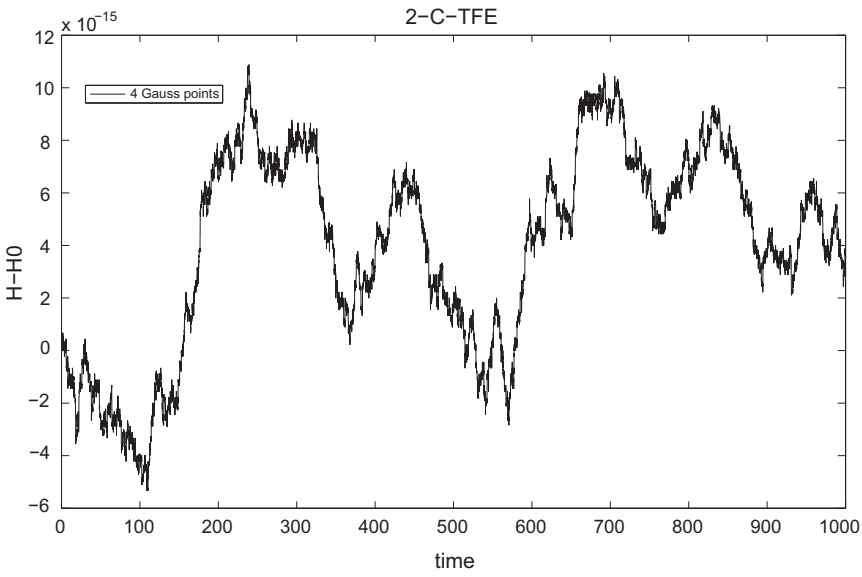

Fig. 6.1. 2-C-TFE method with 4-point Gaussian quadrature for the Huygens system, step size $h = 0.05$.

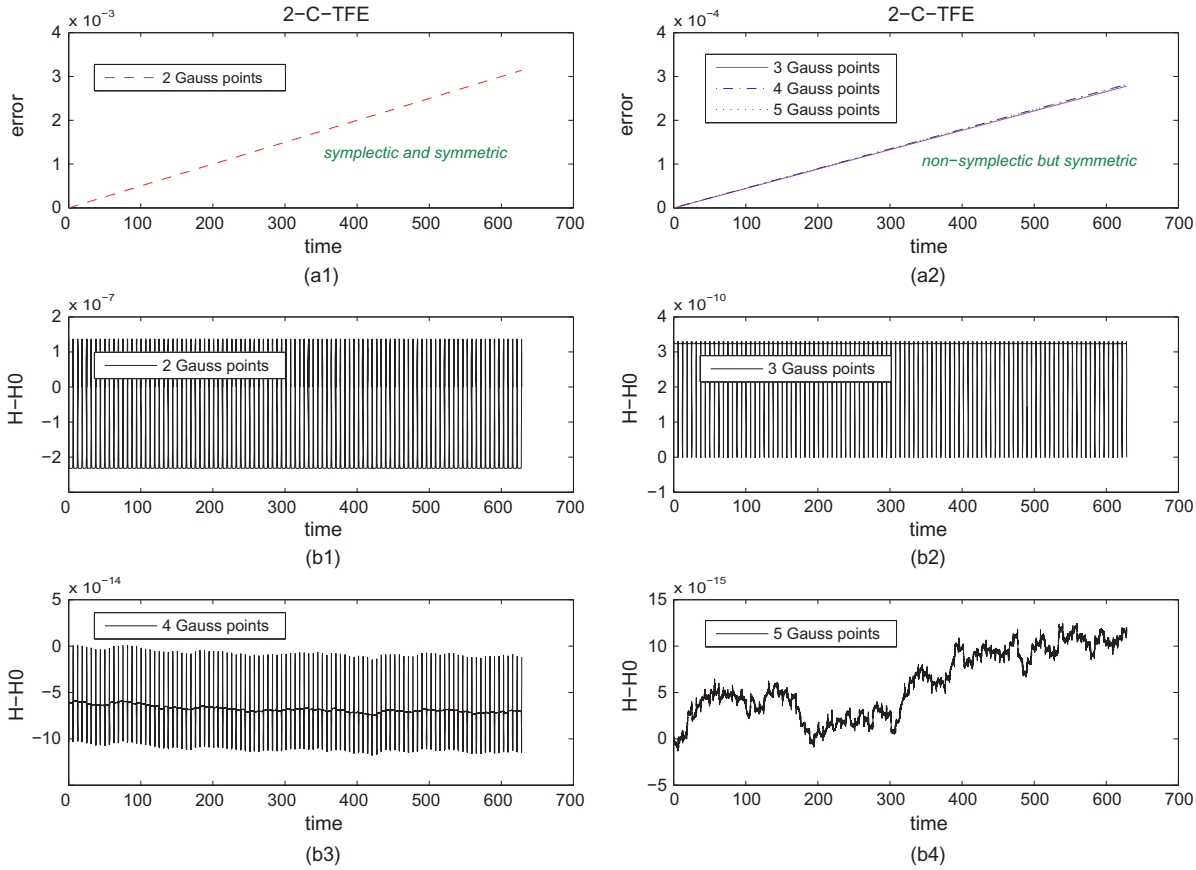

Fig. 6.2. 2-C-TFE method with Gaussian quadrature of different points for the Kepler problem, step size $h = 2\pi/256$. (a1-a2): Solution errors; (b1-b4): Energy errors.

Furthermore, by means of the k-point Radau-right and Lobatto quadrature the combination method based on the k-C-TFE and $(k-1)$-BD-TFE methods will lead to the k-stage Radau IIA-IIĀ and Lobatto IIIA-IIIB methods, respectively.

(III) When $s = k$, substituting (5.9) into (5.10) gives

$$p_k(c_i)\int_0^{c_j} p_k(x)\,\mathrm{d}x = 0, \quad i,j = 1,\ldots,q.$$

Thus, the corresponding PRK method is symplectic iff the abscissae $\{c_i\}$ are the Gaussian or Lobatto points.

(IV) When $s \neq k, k-1, k-2$, it is not easy to construct the corresponding symplectic PRK schemes.

In what follows, we list the conditions for symplecticicty in other cases.

(i) When $s = k, \widetilde{p}_k(c_i)\widetilde{p}_k(c_j) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-LD-TFE and s-LD-TFE methods.
(ii) When $s = k, \widetilde{p}_k(c_i)\widetilde{p}_k(c_j) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-RD-TFE and s-RD-TFE methods.
(iii) When $s = k, \widetilde{p}_k(c_j)\widetilde{p}_{k+1}(c_i) + \widetilde{p}_k(c_i)\widetilde{p}_{k+1}(c_j) = 0$ or when $s = k+1, \widetilde{p}_k(c_j)\widetilde{p}_{k+1}(c_i) + \widetilde{p}_{k+1}(c_j)\widetilde{p}_{k+2}(c_i) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-BD-TFE and s-BD-TFE methods.
(iv) When $s = k-1, \widetilde{p}_{k-1}(c_i)(\widetilde{p}_k(c_j) + \widetilde{p}_{k-1}(c_j)) = 0$ or when $s = k, \widetilde{p}_k(c_i)(\widetilde{p}_k(c_j) + \widetilde{p}_{k-1}(c_j)) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-C-TFE and s-LD-TFE methods.
(v) When $s = k-1, \widetilde{p}_{k-1}(c_i)(\widetilde{p}_k(c_j) - \widetilde{p}_{k-1}(c_j)) = 0$ or when $s = k, \widetilde{p}_k(c_i)(\widetilde{p}_k(c_j) - \widetilde{p}_{k-1}(c_j)) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-C-TFE and s-RD-TFE methods.
(vi) When $s = k-1, \widetilde{p}_k(c_i)(\widetilde{p}_k(c_j) - \widetilde{p}_{k-1}(c_j)) = 0$ or when $s = k, \widetilde{p}_k(c_i)(\widetilde{p}_{k+1}(c_j) - \widetilde{p}_k(c_j)) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-LD-TFE and s-BD-TFE methods.
(vii) When $s = k-1, \widetilde{p}_k(c_i)(\widetilde{p}_k(c_j) + \widetilde{p}_{k-1}(c_j)) = 0$ or when $s = k, \widetilde{p}_k(c_i)(\widetilde{p}_{k+1}(c_j) + \widetilde{p}_k(c_j)) = 0, i,j = 1,\ldots,q$ for the combination methods based on the k-RD-TFE and s-BD-TFE methods.

In Table 5.4, we show the symplectic PRK methods derived by combining different 2-TFE methods from cases (iv-vii), which are the class of Radau IA-IĀ or Radau IIA-IIĀ pair.

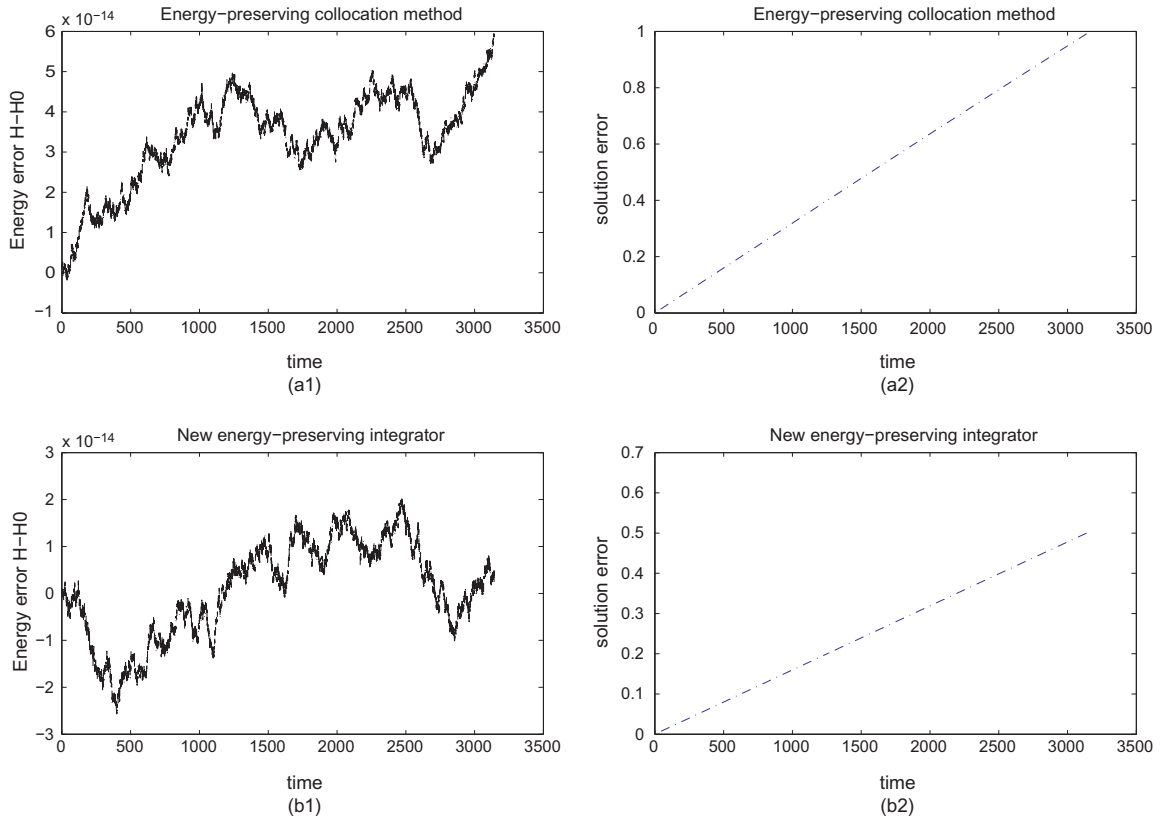

Fig. 6.3. Energy and solution errors for the Kepler problem by two schemes with 5-point Gaussian quadrature, step size $h = 2\pi/512$. Above: Hairer's method (4.27); Below: Our method (4.26).

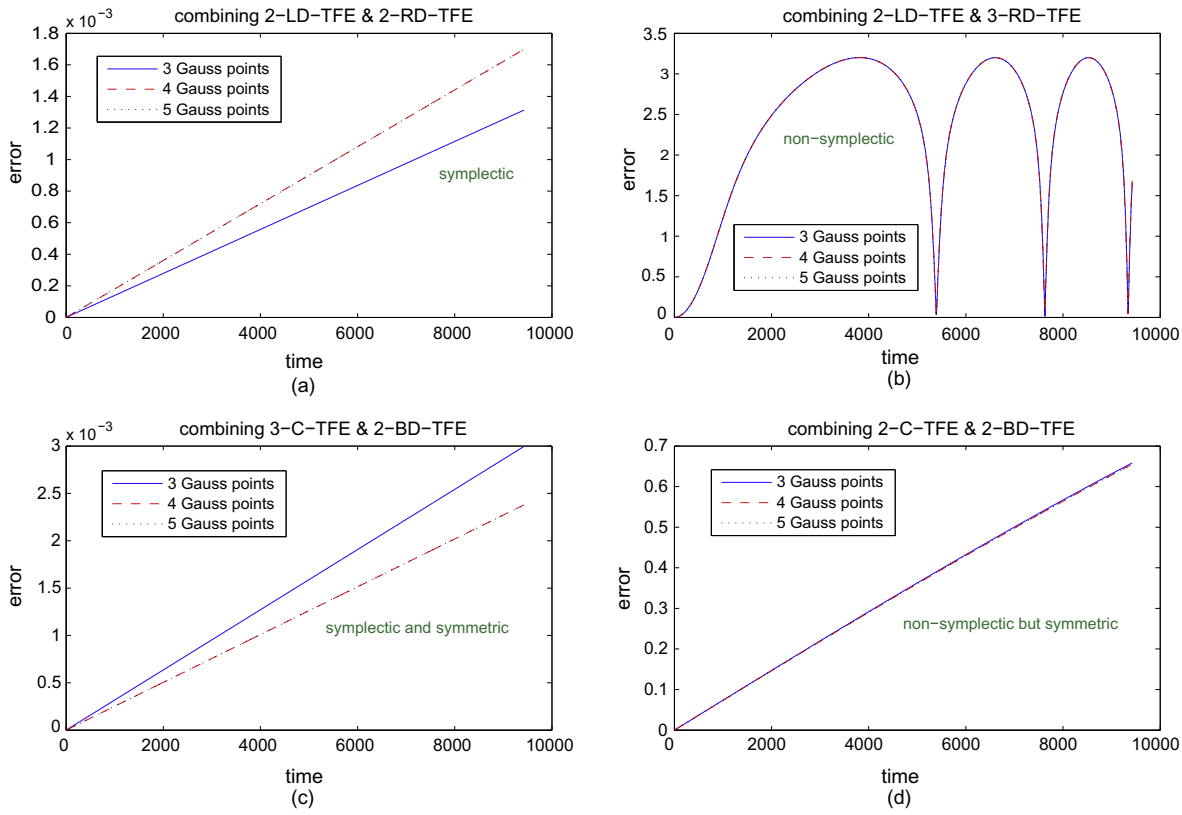

Fig. 6.4. Solution errors for the Kepler problem by combining various TFE methods, step size $h = 2\pi/128$ for 192,000 steps.

6. Numerical results

In this section, we present the numerical experiments for the Hamiltonian system

$$\begin{cases} \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}, \mathbf{q}), \\ \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}). \end{cases} \tag{6.1}$$

Example 6.1. When $H(p,q) = p^2 - q^2 + q^4$, this is the 2-D nonlinear Huygens system. We take the initial conditions as $p(0) = 0$, $q(0) = 1.1$.

Example 6.2. For the well-known Kepler problem, which describes the movement of two celestial bodies, the related Hamiltonian function is

$$H(p_1, p_2, q_1, q_2) = \frac{1}{2}(p_1^2 + p_2^2) - \frac{1}{\sqrt{q_1^2 + q_2^2}}.$$

In the numerical tests, the initial conditions $p_1(0) = 0$, $p_2(0) = 2$, $q_1(0) = 0.4$, $q_2(0) = 0$ are used. Note that in this case the true solution is periodic and the corresponding period is 2π [25].

We first compute the two systems by the 2-C-TFE method. For the Huygens system, the Hamiltonian H is a polynomial, therefore the integral involving H can be evaluated exactly e.g. by 4-point Gaussian quadrature. As expected, Fig. 6.1 shows that the C-TFE method can preserve the Hamiltonian very well. For the Kepler problem, we calculate the numerical solutions by using Gaussian quadrature formula with different points. In Fig. 6.2 (a1-a2), at the end of each period we compute the numerical solution errors in Euclidean norm. When the quadrature formula with more than two points is used, the symplecticity of the numerical discretization disappears. However, the numerical errors still grow linearly w.r.t. time probably due to the symmetric property inherited by the numerical method. When the optimal order has been reached, the use of the quadrature rule with more points will not improve the accuracy of the numerical solution but it does give the numerical solution the benefit of preserving the energy of the system. In Fig. 6.1 (b1-b4), we show

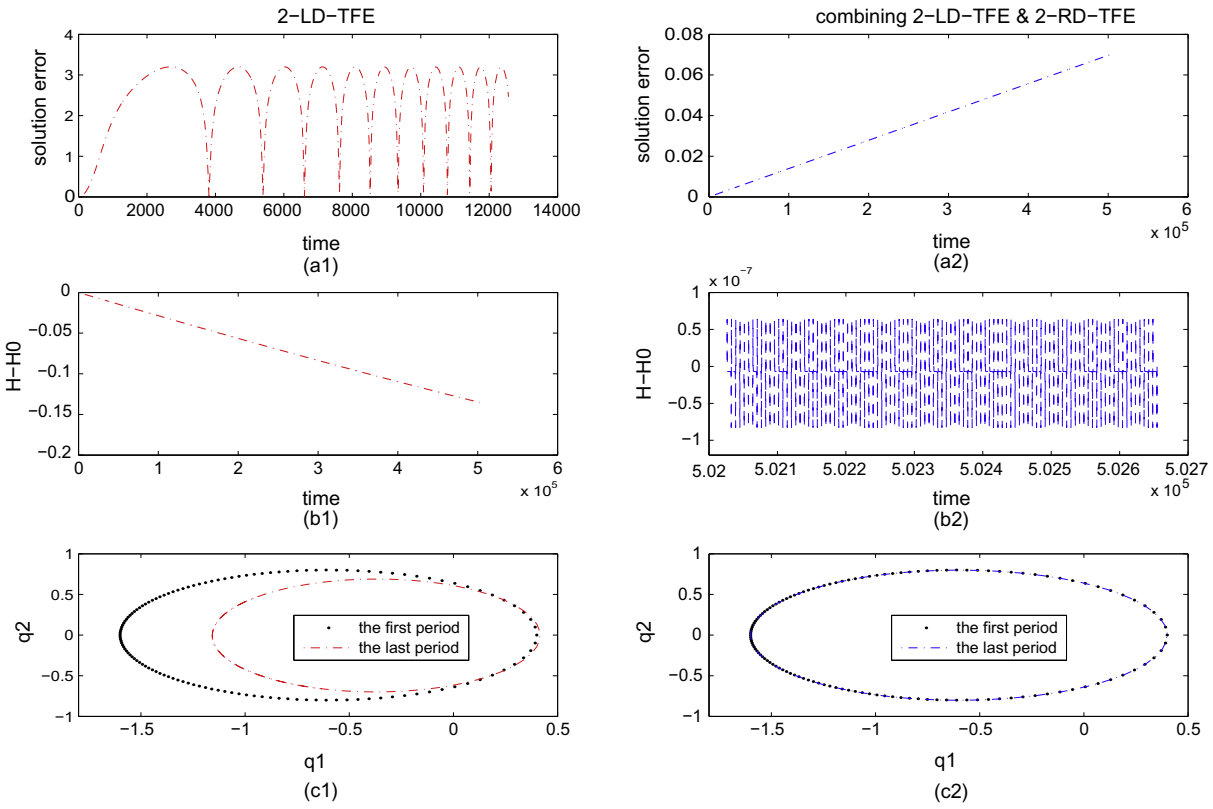

Fig. 6.5. Solution errors, energy errors and numerical orbits for the Kepler problem by two methods with 3-point Gaussian quadrature, step size $h = 2\pi/128$ for 10,240,000 steps. Left: LD-TFE method; Right: LD-TFE and RD-TFE pair.

the energy error by Gaussian quadrature with up to five points. We observe that to get a better conservation of the Hamiltonian the quadrature formula with more points, e.g. 4-point quadrature rule, needs to be used. We then compare the numerical method (4.26) presented by us and Hairer's method (4.27). In Fig. 6.3, the numerical solutions computed by our method show a little better result. It is also shown in Fig. 6.2 that the errors of numerical solutions computed by the two methods both grow linearly w.r.t. time.

For the Kepler problem, in Fig. 6.4 we compare four kinds of numerical methods obtained by combining different pairs of TFE methods: the 2-LD-TFE and 2-RD-TFE pair, the 2-LD-TFE and 3-RD-TFE pair, the 3-C-TFE and 2-BD-TFE pair, and the 2-C-TFE and 2-BD-TFE pair. The Gaussian quadrature is used to calculate the integrals. For the 2-LD-TFE and 2-RD-TFE pair (see Fig. 6.3 (a)), it is surprising to observe that the use of a quadrature formula based upon more points leads to a poorer accuracy. Except for the 2-LD-TFE and 3-RD-TFE pair which is non-symplectic and non-symmetric, we have observed that the numerical solution errors of the other three methods show a linearly growth w.r.t. time. The numerical solution error obtained by the 2-LD-TFE and 3-RD-TFE pair shows a nonlinear growth.

In Fig. 6.5, we compare the 2-LD-TFE method and the pair of 2-LD-TFE and 2-RD-TFE with 3-point Gaussian quadrature. The computation is done for a relatively long period of time (10,240,000 steps). Since the former is non-symplectic and the latter is symplectic, we observe that the solution errors show a nonlinear growth and a linear growth, respectively. For the two numerical methods, we also compare the energy errors. As the 2-LD-TFE and 2-RD-TFE pair is symplectic, the numerical discretization shows a better energy conservation than the 2-LD-TFE method. Furthermore, the solution orbit is simulated quite well by the combination method for a very long time. For the LD-TFE method, the numerical orbit spirals inwards and gives a completely wrong behavior.

In Fig. 6.6, we discretize **p** and **q** in (6.1) by different combinations of TFE methods: the 2-C-TFE and 1-BD-TFE pair, and the 1-BD-TFE and 2-C-TFE pair. When the Gaussian quadrature is used, both numerical discretizations are symplectic and symmetric. Therefore, the numerical results for both methods show a linear error growth. However, a slight difference in the numerical solution and energy errors is observed. The numerical solution error shown in (a) is larger than that shown in (b) by a factor 4.699 for a long time (see (c)). The energy error shown in (d) is also larger than that in (e). This suggests that we may suitably choose the combination of TFE methods by considering the different TFE discretizations for the **q**-variables and **p**-variables to get better results.

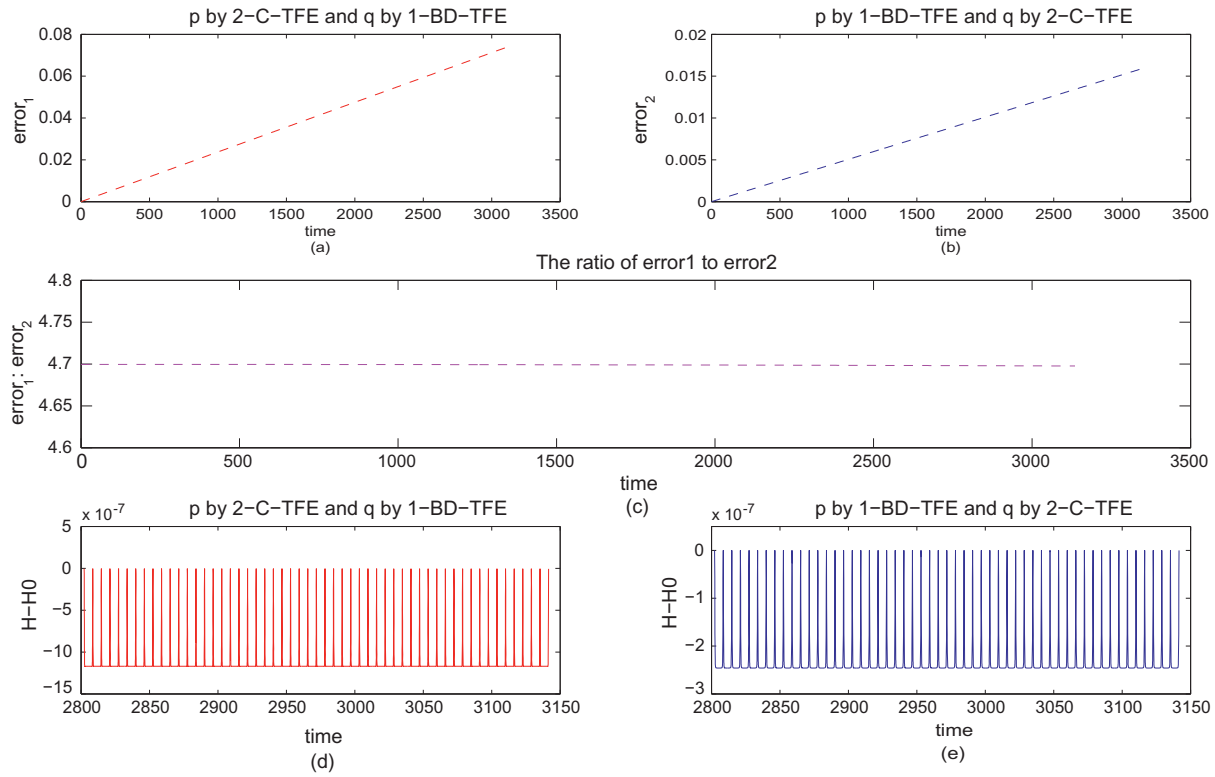

Fig. 6.6. Solution and energy errors by two schemes with 3-point Gaussian quadrature for the Kepler system, step size $h = 2\pi/256$ for 128,000 steps. (a), (d): 2-C-TFE and 1-BD-TFE pair; (b), (e): 1-BD-TFE and 2-C-TFE pair; (c) shows the ratio of the two solution errors.

7. Conclusions

In this paper we present a unified framework for the numerical discretization of ODEs with initial values based on time finite element (TFE) methods. Differently from the two-point boundary value problems for ODEs, the finite element basis functions of the initial value problems for ODEs are defined on every element $[t_n, t_{n+1}]$. This implies that TFE methods are local, the idea of TFE methods is similar to that of discontinuous finite element methods [16]. Therefore, the jumps $\mathbf{u}_n(t_{n+1}^+) - \mathbf{u}_{n+1}$ and $\mathbf{u}_n(t_n^+) - \mathbf{u}_n$ in (2.2) can be understood as the "discrete flux" in fluid dynamics. By suitably choosing the jumps, various variational formulations can be derived [3]. Here, we focus on four kinds of TFE methods and their combinations. Applying TFE methods to ODEs with initial values gives one-step time discretizations which are equivalent to RK methods with infinitely many stages. To calculate the integrals appearing in the TFE variational formulations, usually the suitable numerical quadrature formulae have to be used. In such a way, we can relate TFE methods to classical RK methods including continuous (discontinuous) collocation methods. We have seen that the use of numerical quadrature with enough high-order accuracy is needed for the TFE solutions to reach the theoretic optimal order of the TFE methods. When the optimal order has been reached, using the numerical quadrature formulae of higher order can not affect the accuracy of the numerical solution anymore. However, for the C-TFE methods, the error of energy preservation of the resulting discretization will be improved when the methods are applied to Hamiltonian systems. We have also presented a variational interpretation for some existing energy-preserving integrators including the AVF methods, the HBVMs and the energy-preserving collocation methods. For Hamiltonian systems, generally, the numerical discretizations derived from the TFE methods are not symplectic except for special cases, e.g. the C-TFE or BD-TFE method applied to linear Hamiltonian systems. However, by combining different TFE methods (i.e. approximating the different components of the unknown vector function with different TFE methods respectively) we are able to recover lots of existing symplectic methods, and find new ones. The order and superconvergence of the corresponding numerical methods can be derived by means of the simplifying assumptions, which seems easier to be used than the common analysis techniques in the theory of finite element methods.

Acknowledgments

We would like to express our sincere gratitude to the referees for their constructive comments which help us to improve the presentation of this paper. This work was supported by the Foundation of the NNSFC (60931002), the Foundation for Innovative Research Groups of the NNSFC (11021101) and the National Basic Research Program of China (2010CB832702).

References

[1] J.H. Argyris, D.W. Scharpf, Finite element in time and space, Aer. J. Royal Aer. Soc. 73 (1969) 1041–1044.
[2] P. Betsch, P. Steinmann, Inherently energy conserving time finite elements for classical mechanics, J. Comput. Phys. 160 (2000) 88–116.
[3] M. Borri, C. Bottasso, A general framework for interpreting time finite element formulations, Comput. Mech. 13 (1993) 133–142.
[4] C.L. Bottasso, A new look at finite elements in time: a variational interpretation of Runge–Kutta methods, Appl. Numer. Math. 25 (1997) 355–368.
[5] L. Brugnano, F. Iavernaro, D. Trigiante, Analysis of Hamiltonian Boundary Value Methods (HBVMs) for the numerical solution of polynomial Hamiltonian dynamical systems, Preprint 2009, arXiv:0909.5659.
[6] L. Brugnano, F. Iavernaro, D. Trigiante, Hamiltonian boundary value methods: energy preserving discrete line integral methods, J. Numer. Anal., Indust. Appl. Math. 5 (1–2) (2010) 17–37.
[7] L. Brugnano, F. Iavernaro, D. Trigiante, A note on the efficient implementation of Hamiltonian BVMs, Jour. Comput. Appl. Math. 236 (2011) 375–383.
[8] L. Brugnano, F. Iavernaro, D. Trigiante, The lack of continuity and the role of infinite and infinitesimal in numerical methods for ODEs: the case of symplecticity, Appl. Math. Comp. 218 (2012) 8053–8063.
[9] L. Brugnano, F. Iavernaro, D. Trigiante, A simple framework for the derivation and analysis of effective one-step methods for ODEs, Appl. Math. Comp. 218 (2012) 8475–8485.
[10] J.C. Butcher, Implicit Runge–Kutta processes, Math. Comput. 18 (1964) 50–64.
[11] J.C. Butcher, Integration processes based on Radau quadrature formulas, Math. Comput. 18 (1964) 233–244.
[12] C. Chen, Structure theory of superconvergence of finite elements (in Chinese), Hunan Science and Technology Press, Changsha, 2001.
[13] C. Chen, Introduction to scientific computation (in Chinese), China Science Press, Beijing, 2007.
[14] C. Chen, Q. Tang, Continuous finite element methods for Hamiltonian systems, Appl. Math. Mech. 28 (8) (2007) 1071–1080.
[15] C. Chen, Q. Tang, S. Hu, Long-time behaviour of the finite element method for Hamiltonian systems: Look up to the Kang Feng's Conjecture (in chinese), Research report in Institute of Computational Mathematics, AMSS, Beijing, July 2010.
[16] B. Cockburn, G. Karniadakis, C.-W. Shu (Eds.), Discontinuous Galerkin Methods: Theory, Computation and Applications, Lecture Notes in Computational Science and Engineering, bol. 11, Springer, 2000.
[17] D. Cohen, E. Hairer, Linear energy-preserving integrators for Poisson systems, BIT. Numer. Math. 51 (2011) 91–101.
[18] M. Delfour, W. Hager, F. Trochu, Discontinuous Galerkin methods for ordinary differential equations, Math. Comp. 36 (1981) 455–473.
[19] K. Deng, Z. Xiong, Superconvergence of a discontinuous finite element method for a nonlinear ordinary differential equation, Appl. Math. Comp. 217 (2010) 3511–3515.
[20] D.A. French, J.W. Schaeffer, Continuous finite element methods which preserve energy properties for nonlinear problems, Appl. Math. Comp. 39 (3) (1990) 271–295.
[21] I. Fried, Finite element analysis of time dependent phenomena, AIAA J. 7 (1969) 1170–1173.
[22] Z. Ge, J.E. Marsden, Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators, Phys. Lett. A 133 (3) (1988) 134–139.
[23] S. Gottlieb, G.W. Wei, S. Zhao, A unified discontinuous Galerkin framework for time integration, preprint, 2010.
[24] E. Hairer, Energy-preserving variant of collocation methods, JNAIAM, J. Numer. Anal. Indust. Appl. Math. 5 (2010) 73–84.
[25] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration: Structure-Preserving Algorithms For Ordinary Differential Equations, second ed., Springer Series in Computational Mathematics, vol. 31, Springer-Verlag, Berlin, 2006.
[26] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I: Nonstiff Problems, Springer Series in Computational Mathematics, vol. 8, Springer-Verlag, Berlin, 1993.
[27] E. Hairer, G. Wanner, Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems, second ed., Springer Series in Computational Mathematics, vol. 14, Springer-Verlag, Berlin, 1996.
[28] B.L. Hulme, Discrete Galerkin and related one-step methods for ordinary differential equations, Math. Comp. 26 (1972) 881–891.
[29] F. Iavernaro, B. Pace, s-stage trapezoidal methods for the conservation of Hamiltonian functions of polynomial type, AIP Conf. Proc. 936 (2007) 603–606.
[30] D.I. McLaren, G.R.W. Quispel, Bootstrapping discrete-gradient integral-preserving integrators to fourth order, BIT Numer. Math. 43 (1) (2003) 001–018.
[31] G.R.W. Quispel, G. Turner, Discrete gradient methods for solving ODE's numerically while preserving a first integral, J. Phys. A 29 (1996) 341–349.
[32] G.R.W. Quispel, D.I. McLaren, A new class of energy-preserving numerical integration methods, J. Phys. A: Math. Theor. 41 (2008) 045206.
[33] G. Sun, Construction of high order symplectic PRK methods, J. Comput. Math. 13 (1) (1995) 40–50.
[34] G. Sun, A simple way constructing symplectic Runge–Kutta methods, J. Comput. Math. 18 (1) (2000) 61–68.