SIAM Review
Vol. 58, No. 1, pp. 3–33



# Symplectic Runge–Kutta Schemes for Adjoint Equations, Automatic Differentiation, Optimal Control, and More*

J. M. Sanz-Serna†

**Abstract.** The study of the sensitivity of the solution of a system of differential equations with respect to changes in the initial conditions leads to the introduction of an adjoint system, whose discretization is related to reverse accumulation in automatic differentiation. Similar adjoint systems arise in optimal control and other areas, including classical mechanics. Adjoint systems are introduced in such a way that they exactly preserve a relevant *quadratic invariant* (more precisely, an inner product). Symplectic Runge–Kutta and partitioned Runge–Kutta methods are defined through the exact conservation of a differential geometric structure, but may be characterized by the fact that they preserve exactly *quadratic invariants* of the system being integrated. Therefore, the symplecticness (or lack of symplecticness) of a Runge–Kutta or partitioned Runge–Kutta integrator should be relevant to understanding its performance when applied to the computation of sensitivities, to optimal control problems, and in other applications requiring the use of adjoint systems. This paper examines the links between symplectic integration and those applications and presents in a new, unified way a number of results currently scattered among or implicit in the literature. In particular, we show how some common procedures, such as the direct method in optimal control theory and the computation of sensitivities via reverse accumulation, imply, probably unbeknownst to the user, "hidden" integrations with symplectic partitioned Runge–Kutta schemes.

**Key words.** Runge–Kutta methods, partitioned Runge–Kutta methods, symplectic integration, Hamiltonian systems, variational equations, adjoint equations, computation of sensitivities, Lagrange multipliers, automatic differentiation, optimal control, Lagrangian mechanics, reflected and transposed Runge–Kutta schemes, differential-algebraic problems, constrained controls

**AMS subject classifications.** 34H05, 49A10, 65L06, 65K10, 65P10, 70H25

**DOI.** 10.1137/151002769

## 1. Introduction.

Symplectic Runge–Kutta (RK) [24], [31], [39] and partitioned Runge–Kutta (PRK) [1], [40] formulas were introduced to integrate Hamiltonian systems over long time intervals. They are defined in terms of a purely geometric property, the conservation of the symplectic structure, and provided the first widely studied instance of what was later termed *geometric integration* [32]. It is well known that symplectic RK methods may be characterized as those that exactly preserve all *quadratic* first integrals (invariants of motion) of the system being integrated. This is a useful property: for instance, the (symplectic) implicit midpoint rule is sometimes

*Received by the editors January 6, 2015; accepted for publication (in revised form) June 17, 2015; published electronically February 4, 2016. This research was supported by projects MTM2010-18246-C03-01 and MTM2013-46553-C3-1-P from Ministerio de Ciencia e Innovación, Spain.
http://www.siam.org/journals/sirev/58-1/100276.html

†Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés (Madrid), Spain (jmsanzserna@gmail.com).

chosen to integrate wave equations because it conserves quadratic invariants. However, quadratic conservation has taken a back seat to the symplectic property itself in the geometric integration literature. The aim of this paper is to emphasize that the conservation of quadratic invariants plays an important role in the computation of numerical sensitivities, in optimal control theory, and in classical mechanics. In all these areas there is an interplay between variational equations and their adjoints, an interplay based *on the conservation of a key quadratic invariant* (see (3.5)). The conservation of this invariant gives relevance to the symplecticness of the integrator. In fact, some widely used procedures, such as the direct method in optimal control theory and the computation of sensitivities via reverse accumulation, imply "hidden" integrations with symplectic PRK schemes; therefore, the theory of symplectic PRK integration should be helpful in understanding such procedures. From a more abstract point of view one might say that the purpose of this article is to clarify the behavior of RK integrators *vis-à-vis* the operation of taking *adjoints*: an RK method is symplectic precisely if it commutes with the formation of adjoints.

The paper presents a coherent treatment of results spread across the literature of various communities together with some new, unifying results. In order to cater to a variety of possible readers, it is written without assuming much background. We hope it will help researchers in optimal control to better understand RK schemes and, similarly, encourage RK experts to consider sensitivities and optimal control problems.

Section 2 provides background on numerical integrators. We introduce the necessary notation and recall a number of properties of symplectic RK and related schemes. In particular, we quote some results (Theorems 2.1 and 2.4) that ensure the exact preservation by the integrator of quadratic conservation laws.

Section 3, the core of the paper, is devoted to the integration of the adjoint variational equations used to perform sensitivity analysis. It is well known that an RK method $\mathcal{M}$ applied to the variational equations of a system $\mathcal{S}$ automatically produces the variational equations for the discretization of $\mathcal{S}$ by means of $\mathcal{M}$ (see Theorem 3.2); in other words, the operation of RK discretization *commutes* with the operation of forming variational equations. The situation for the adjoints is more complicated (cf. [37]), because commutation will only take place if the discretization is carried out so as to *exactly conserve the key quadratic invariant* (3.5) and, in some way, this demands a symplectic integrator. There are three cases of increasing complexity:

- $\mathcal{S}$ is integrated with a *symplectic* RK scheme $\mathcal{M}$. Then the application of $\mathcal{M}$ to the adjoint equations of $\mathcal{S}$ produces the adjoint equations for discretization of $\mathcal{S}$ by means of $\mathcal{M}$ (see Theorem 3.3).
- $\mathcal{S}$ is integrated with a *nonsymplectic* RK scheme $\mathcal{M}$ whose weights do not vanish. Then the adjoint equations for the discretization are obtained by integrating the adjoint equations of $\mathcal{S}$ with a *different* set of RK coefficients, so that the overall procedure is a symplectic PRK method (see Theorem 3.4). The recipe for the adjoint coefficients is given in (3.23) below. The method used for the adjoint equations will in general be of lower order than the RK scheme $\mathcal{M}$ used for the main integration and will also have different stability properties. For these reasons nonsymplectic methods $\mathcal{M}$ should be used with care. The computation of sensitivities of the discrete solution via *automatic differentiation with reverse accumulation* implicitly provides the *symplectic PRK integration* of the adjoint equations with coefficients (3.23) (see Theorem 3.6).

- $\mathcal{S}$ is integrated with a *nonsymplectic* RK scheme $\mathcal{M}$ having one or more null weights. Then, to obtain the adjoint equations of the discretization, the continuous adjoint equations have to be integrated with a fancy integrator outside the RK class (see the appendix). Again an order reduction is likely to take place and again the fancy integration is implicitly performed whenever differentiation with reverse accumulation is used.

Section 4 deals with the Mayer optimal control problem in the case of unconstrained controls. There is again a quadratic conservation law that is of crucial importance and this fact brings symplectic schemes to the foreground. The results there are quite similar to those in the preceding section (the case of vanishing weights is discussed in the appendix):

- For a symplectic RK method, *commutation* [29] takes place: the discretization of the continuous first order conditions necessary for optimality provides the first order necessary conditions for the discrete solution (see Theorem 4.3).
- When the equations for the states are discretized with a *nonsymplectic* RK scheme *with nonvanishing weights*, to achieve commutation the costate equations have to be integrated by means of a clever set of coefficients that does not coincide with the set used for the states (see Theorem 4.3). With this clever set, the overall integration (states+costates) is performed with a symplectic PRK method. In general, an order reduction will take place for states, costates, and controls. As first noted by Hager [17], the required set of coefficients is alternatively defined, not by imposing symplecticness of the integration, but by using the *direct* approach, i.e., by minimizing the cost in the discrete realm with the help of Lagrange multipliers (see Theorem 4.4).

For a *symplectic* RK or PRK integration of the system for states and costates, the direct and indirect approach are mathematically equivalent. When a nonsymplectic PRK is used in the indirect approach, the discrete solution *cannot* be reached via the direct approach, which always implies a symplectic integration of the states+costates system.

Extensions to more general control problems are presented in section 5. Section 6 is devoted to classical mechanics. Hamilton's variational principle may of course be viewed as an optimal control problem: it is a matter of minimizing a functional subject to differential constraints. As is well known, the application of the theory of optimal control to this situation replicates the standard procedure to obtain Hamilton's canonical equations from Hamilton's principle. In the discrete realm, this process provides the variational derivation of symplectic PRK integrators, originally due to Suris [40].

Section 7 relates the preceding material to the notions of reflection and transposition of RK coefficients introduced by Scherer and Türke [35], and section 8 concludes.

The appendix deals with the problem of how to "supplement" a given nonsymplectic RK method with some vanishing weights to produce a symplectic algorithm for partitioned systems.

In order not to clutter the exposition with unwanted details, we shall not be concerned with technical issues such as existence of solutions of implicit integrators, smoothness requirements, and so on. These may be very important in some circumstances (e.g., lack of smoothness poses difficulties if the controls are constrained; see [9]).

To keep the length of this work within reasonable limits some other interesting connections are not discussed. The duality between the Fokker–Planck equations and

the Kolmogorov backward equations in the theory of Markov stochastic processes [12] provides another instance of the occurrence of adjoints; the material in this paper may be easily extended to study that situation. The paper [13] shows how the symplecticness of the integrator may be used to ensure symmetry-preserving simulations of the matrix Riccati equation in the feedback representation of linear/quadratic optimal control problems.

**2. Numerical Integrators.** In this section we review some results on RK and related methods. For more details the reader is referred to [34], [5], [19], [21], [22].

**2.1. Runge–Kutta Schemes.** An RK method with $s$ stages is specified by $s^2+2s$ numbers

$$a_{ij}, \quad i,j=1,\dots,s, \qquad b_i,\ c_i, \quad i=1,\dots,s. \tag{2.1}$$

Given a $D$-dimensional differential system, $F:\mathbb{R}^D\times\mathbb{R}\to\mathbb{R}^D$,

$$\frac{d}{dt}y = F(y,t), \tag{2.2}$$

to be studied in an interval $t_0\le t\le t_0+T$, and an initial condition

$$y(t_0)=A\in\mathbb{R}^D, \tag{2.3}$$

the method (2.1) finds approximations $y_n$ to the values $y(t_n)$, $n=0,1,\dots,N$, of the solution of (2.2)–(2.3), $t_0<t_1<\cdots<t_N=t_0+T$, by setting $y_0=A$ and, recursively,

$$y_{n+1}=y_n+h_n\sum_{i=1}^{s}b_iK_{n,i}, \qquad n=0,1\dots,N-1. \tag{2.4}$$

Here $h_n=t_{n+1}-t_n$ denotes the step-length and $K_{n,i}$, $i=1,\dots,s$, are the "slopes"

$$K_{n,i}=F(Y_{n,i},t_n+c_ih_n) \tag{2.5}$$

at the so-called internal stages $Y_{n,i}$. The vectors $Y_{n,1},\dots,Y_{n,s}$ are in turn defined by the relations

$$Y_{n,i}=y_n+h_n\sum_{j=1}^{s}a_{ij}K_{n,j}, \quad i=1,\dots,s. \tag{2.6}$$

In the particular case where the matrix $(a_{ij})$ is, perhaps after renumbering the stages, strictly lower triangular (explicit RK methods), the stages are computed recursively from (2.5)–(2.6). In the general case, (2.5)–(2.6) provides, for each $n$, a system of coupled equations to be solved for the stages.

The internal stages should not be confused with the values $y_n$ output by the integrator and may be regarded merely as auxiliary variables. Alternatively, the vector $Y_{n,i}$ is sometimes viewed as an approximation to the off-step value $y(t_n+c_ih_n)$. It is important to emphasize that the differences $y(t_n+c_ih_n)-Y_{n,i}$ are typically much larger than the differences $y(t_n)-y_n$.

When the system (2.2) is autonomous, i.e., $F=F(y)$, the $c_i$ play no role. At the other end of the spectrum, if $F$ is independent of $y$, the RK discretization amounts to the use in the interval $t_0\le t\le t_0+T$ of the composite quadrature rule based on the *abscissas* $c_i$ and the *weights* $b_i$.

An RK scheme is said to possess order $\rho$ if, for $t_0 \le t_n \le t_0 + T$ and smooth problems, $|y_n - y(t_n)| = \mathcal{O}(h^\rho)$, where $h = \max_n h_n$. The expansion of the local truncation error in powers of the step-length $h_n$ includes, for each power $h_n^k$, $k = 1, 2, \ldots$, one or several elementary differentials of $F$; an integrator has order $\ge \rho$ if and only if, in that expansion, the coefficients of the elementary differentials of orders $k = 1, \ldots, \rho$ vanish. For instance, the relations (order conditions)

$$\sum_{i=1}^{s} b_i = 1, \quad \sum_{i,j=1}^{s} b_i a_{ij} = \frac{1}{2}, \quad \sum_{i,j,k=1}^{s} b_i a_{ij} a_{jk} = \frac{1}{6}, \quad \sum_{i,j,k=1}^{s} b_i a_{ij} a_{ik} = \frac{1}{3} \tag{2.7}$$

ensure order at least 3 for autonomous problems. They correspond to the elementary differentials $F$ (of order 1), $(\partial_y F)F$ (of order 2), and $(\partial_y F)(\partial_y F)F$, $(\partial_{yy} F)[F, F]$ (both of order 3) ($\partial_y F$ is the Jacobian matrix and $\partial_{yy} F$ the tensor of second derivatives). Since the work of Butcher in the early 1960s, order conditions and elementary differentials have been studied with the help of graphs. To impose order $\ge \rho$ for autonomous problems, there is an independent order condition for each rooted tree with $\rho$ or fewer vertices. Most, but not all useful RK schemes satisfy $c_i = \sum_j a_{ij}$ for each $i$; for them, order $\rho$ for autonomous problems implies order $\rho$ for all problems.

In general RK methods do not conserve exactly the quadratic first integrals of the system being integrated. The simplest illustration is afforded by the familiar Euler's rule ($s = 1$, $b_1 = 1$, $a_{11} = 0$, $c_1 = 0$) applied to the harmonic oscillator ($D = 2$)

$$\frac{d}{dt} y^1 = -y^2, \quad \frac{d}{dt} y^2 = y^1$$

(superscripts denote components). The (quadratic) energy $I = (1/2)((y^1)^2 + (y^1)^2)$ is conserved by the differential system because

$$\frac{d}{dt} I = y^1 \frac{d}{dt} y^1 + y^2 \frac{d}{dt} y^2 = y^1(-y^2) + y^2 y^1 = 0.$$

However, for Euler's rule it is trivial to check that, over one step,

$$I(y_{n+1}^1, y_{n+1}^2) - I(y_n^1, y_n^2) = \frac{h_n}{2}\big((y_n^1)^2 + (y_n^2)^2\big),$$

with an energy increase. This lack of exact preservation takes place for all explicit RK integrators, even when their order $\rho$ is high. On the other hand, it is well known and easy to prove that for the implicit midpoint rule ($s = 1$, $b_1 = 1$, $a_{11} = 1/2$, $c_1 = 1/2$) and the harmonic oscillator $I(y_{n+1}^1, y_{n+1}^2) = I(y_n^1, y_n^2)$.

The present article is based on the following 1987 result of Cooper [8], which ensures that *some* RK methods automatically inherit each quadratic conservation law possessed by the system being integrated.

THEOREM 2.1. *Assume that the system* (2.2) *possesses a quadratic first integral* $I$, *i.e.,* $I(\cdot, \cdot)$ *is a real-valued bilinear mapping in* $\mathbb{R}^D \times \mathbb{R}^D$ *such that, for each* $A$ *and* $t_0$, *the solution* $y(t)$ *of* (2.2)–(2.3) *satisfies* $(d/dt) I(y(t), y(t)) \equiv 0$. *The relations*

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \qquad i, j = 1, \ldots, s, \tag{2.8}$$

*guarantee that, for each RK trajectory* $\{y_n\}$ *satisfying* (2.4)–(2.6), $I(y_n, y_n)$ *is independent of* $n$.

We shall not reproduce the proof of this result here; it is similar to that of Theorem 2.4 below. The relations (2.8) are essentially necessary for an RK scheme to conserve *each* quadratic first integral of *each* differential system [19, Chapter VI, Theorems 7.6 and 7.10].

In many applications the system (2.2) is Hamiltonian. This means that $D$ is even and, after writing $y = [q^{\mathsf{T}}, p^{\mathsf{T}}]^{\mathsf{T}}$, $F = [f^{\mathsf{T}}, g^{\mathsf{T}}]^{\mathsf{T}}$, with $q, p, f, g \in \mathbb{R}^d$, $d = D/2$, there exists a real-valued function $H(p, q, t)$ (the Hamiltonian) such that $f^r = \partial H/\partial p^r$ and $g^r = -\partial H/\partial q^r$, $r = 1, \dots, d$ (superscripts indicate components). Hamiltonian systems are characterized geometrically by the symplectic property of the corresponding solution flow [2]. When $d = 1$, symplecticness means conservation of oriented area; in higher dimensions a similar but more complicated interpretation, based on differential forms, exists; such an interpretation is not required to read this paper. It is often advisable [34], [19], [25] to integrate Hamiltonian problems by means of so-called symplectic algorithms, i.e., algorithms such that the transformation $y_n \mapsto y_{n+1}$ in $\mathbb{R}^{2d}$ is symplectic; those algorithms are particularly advisable in integrations where the interval $t_0 \le t \le t_0 + T$ is long (for a recent reference in that direction, see [11], which is part of a project to integrate the solar system over a 60-million-year interval). Using the method of modified equations [16], each numerical solution may (approximately) be interpreted as a true solution of a nearby differential system called the modified system. For symplectic methods applied to Hamiltonian systems, the modified system is Hamiltonian; for nonsymplectic discretizations, the modified system, while perhaps close to the system being integrated, is not Hamiltonian and this fact is likely to imply a substantial distortion of the long-time dynamics [34], [19].

The first symplectic integrators were constructed in an ad hoc way; it was later discovered (independently by Lasagni [24], Suris [39], and the present author [31]) that the class of RK methods contains many symplectic schemes:

THEOREM 2.2. *Assume that the system* (2.2) *is Hamiltonian. The relations* (2.8) *guarantee that the mapping* $y_n \mapsto y_{n+1}$ *defined in* (2.4)–(2.6) *is symplectic.*

The proof of Theorem 2.2, not included here, is very similar to the proof of Theorem 2.1. Just as for the conservation of quadratic first integrals, it turns out, as seen in [34, section 6.5], that the relations (2.8) are essentially necessary for $y_n \mapsto y_{n+1}$ to be symplectic for each Hamiltonian system.

The set of relations (2.8) thus ensures *two* different properties: quadratic conservation and symplecticness. These two properties are not unrelated: symplecticness may be viewed a consequence of the quadratic conservation because, as noted in [3], the preservation of the symplectic structure by a Hamiltonian solution flow may be interpreted as a bilinear first integral of the solution flow of the associated variational system.

The symplectic character of RK schemes satisfying (2.8) has attracted much attention in view of the importance of Hamiltonian systems in applications. On the other hand, it is fair to say that quadratic conservation has been to some extent played down in the geometric integration literature. For this reason, while schemes satisfying (2.8) could have been called conservative, the following terminology is standard:

DEFINITION 2.3. *The RK scheme* (2.1) *is called* symplectic *(or canonical) if* (2.8) *holds.*

Our focus in this paper is on symplectic schemes in so far as they conserve quadratic invariants, as these are actually crucial in several applications. The discussion of any possible benefits derived from the symplectic character of the map $y_n \mapsto y_{n+1}$, including the existence of modified Hamiltonian systems, is out of our scope here. The paper [7] is, in this sense, complementary to the present work.

It was proved in [33] that the relations (2.8) act as simplifying assumptions *vis-à-vis* the order conditions: once these relations are imposed, the order conditions corresponding to the different elementary differentials/rooted trees are no longer independent. For instance, it is a simple exercise to show that, when (2.8) holds, the second order condition in (2.7) is a consequence of the first and therefore symplectic RK schemes of order $\geq 1$ automatically possess order $\geq 2$. Similarly, the last order condition in (2.7) is a consequence of the first three. In this way, for a general RK method to have order $\geq 3$ for autonomous problems, there are four order conditions; for symplectic methods the number is only two. For a symplectic RK method to have order $\geq \rho$ for autonomous problems there is an order condition for each so-called nonsuperfluous free tree with $\leq \rho$ vertices.

There are many symplectic RK methods [34] including the Gauss methods (of maximal order $2s$ and with positive weights) as first shown in [31]; however, no symplectic RK scheme is explicit. The simplest Gauss method ($s = 1$) is the familiar implicit midpoint rule.

**2.2. Partitioned Runge–Kutta Schemes.** In some applications the components of the vector $y$ in (2.2) appear partitioned into two blocks: $y = [q^{\mathsf{T}}, p^{\mathsf{T}}]^{\mathsf{T}}$, $q \in \mathbb{R}^{D-d}$, $p \in \mathbb{R}^d$. Hamiltonian problems, where $d = D/2$, provide an example, as we have just seen. In those cases it may make sense to use a set of coefficients (2.1) for the integration of the block $q$ and a second set

$$A_{ij}, \quad i,j = 1,\dots,s, \qquad B_i,\, C_i, \quad i = 1,\dots,s, \tag{2.9}$$

for the integration of the block $p$. (There is no loss of generality in assuming that the number of stages $s$ in (2.9) coincides with that in (2.1); see [34, Remark 3.2].) The overall method is called a PRK scheme. A more precise description follows.

Denote by $F = [f^{\mathsf{T}}, g^{\mathsf{T}}]^{\mathsf{T}}$, $f \in \mathbb{R}^{D-d}$, $g \in \mathbb{R}^d$, the partitioning of $F$ induced by the partitioning $[q^{\mathsf{T}}, p^{\mathsf{T}}]^{\mathsf{T}}$ of $y$, so that (2.2) reads

$$\frac{d}{dt}q = f(q,p,t), \qquad \frac{d}{dt}p = g(q,p,t); \tag{2.10}$$

then the equations for the step $n \to n+1$ of the PRK method (2.1), (2.9) are

$$q_{n+1} = q_n + h_n \sum_{i=1}^{s} b_i k_{n,i}, \quad p_{n+1} = p_n + h_n \sum_{i=1}^{s} B_i \ell_{n,i}, \quad n = 0,\dots,N-1, \tag{2.11}$$

where

$$k_{n,i} = f(Q_{n,i}, P_{n,i}, t_n + c_i h_n), \qquad \ell_{n,i} = g(Q_{n,i}, P_{n,i}, t_n + C_i h_n), \tag{2.12}$$

and the internal stages $Q_{n,i}$, $P_{n,i}$, $i = 1,\dots,s$, are defined by the relations

$$Q_{n,i} = q_n + h_n \sum_{i=1}^{s} a_{ij} k_{n,j}, \qquad P_{n,i} = p_n + h_n \sum_{j=1}^{s} A_{ij} \ell_{n,j}. \tag{2.13}$$

PRK methods are not a mathematical nicety: the Verlet algorithm, the method of choice in molecular dynamics [36], is one of them. In its so-called velocity form, the algorithm is written in the molecular dynamics literature as (it is a simple matter

to rewrite the algorithm in the format (2.11)–(2.13))

$$\begin{aligned} p_{n+1/2} &= p_n + \frac{h_n}{2} g(q_n, t_n), \\ q_{n+1} &= q_n + h_n M^{-1} p_{n+1/2}, \\ p_{n+1} &= p_{n+1/2} + \frac{h_n}{2} g(q_{n+1}, t_{n+1}). \end{aligned}$$

Here the vectors $p$, $q$, and $g$ contain, respectively, the momenta, positions, and forces and $M$ is the diagonal matrix of the masses. Note the way the $q$ and $p$ variables are advanced in different ways.

Clearly an RK scheme may be regarded as a particular instance of a PRK method where the two sets (2.1), (2.9) happen to coincide. For PRK methods to possess order $\geq \rho$ for autonomous problems, there is an order condition associated with each bicolor rooted tree with $\rho$ or fewer vertices (see, e.g., [19, Chapter III]). For order $\geq 2$ the order conditions are

$$\sum_i b_i = 1, \quad \sum_i B_i = 1, \tag{2.14}$$

$$\sum_{ij} b_i a_{ij} = \frac{1}{2}, \quad \sum_{ij} b_i A_{ij} = \frac{1}{2}, \quad \sum_{ij} B_i a_{ij} = \frac{1}{2}, \quad \sum_{ij} B_i A_{ij} = \frac{1}{2}, \tag{2.15}$$

which correspond to the elementary differentials $f$, $g$, $(\partial_x f)f$, $(\partial_x f)g$, $(\partial_x g)f$, and $(\partial_x g)g$, respectively. It will be important in what follows to note that, if the PRK (2.1), (2.9) has order $\rho$, then the RK scheme with coefficients (2.1) and the RK scheme with coefficients (2.9) both have order $\rho$. The converse is not true: if (2.1) and (2.9) are the coefficients of two RK schemes of order $\rho$, then the combined PRK scheme may have order $< \rho$. This is plain in (2.15), where the second and third relations are necessary for the PRK to have order $\geq 2$, but are obviously not required for (2.1) and (2.9) to be the coefficients of two different RK schemes of order $\geq 2$.

For PRK methods, the result corresponding to Theorem 2.1 is as follows (cf. [19, Chapter IV, Theorem 2.4], where only the autonomous case is envisaged).

THEOREM 2.4. *Assume that $S(\cdot, \cdot)$ is a real-valued bilinear map in $\mathbb{R}^d \times \mathbb{R}^{D-d}$ such that, for each $t_0$ and $A$, the solution $y(t) = [q(t)^\mathsf{T}, p(t)^\mathsf{T}]^\mathsf{T}$ of (2.3), (2.10) satisfies*

$$\frac{d}{dt} S(q(t), p(t)) \equiv 0.$$

*The relations*

$$b_i = B_i, \quad i = 1, \ldots, s, \quad b_i A_{ij} + B_j a_{ji} - b_i B_j = 0, \quad i, j = 1, \ldots, s, \tag{2.16}$$

*and*

$$c_i = C_i, \quad i = 1, \ldots, s, \tag{2.17}$$

*guarantee that, for each PRK trajectory satisfying (2.11)–(2.13), $S(q_n, p_n)$ is independent of $n$.*

As in the case of RK methods, the condition in the theorem is necessary for conservation to hold for all $S$ and all partitioned differential systems; see [19, Chapter VI, Theorems 7.6 and 7.10]. In the particular case of autonomous problems the abscissas play no role. Thus, to achieve conservation, it is not necessary to impose

the condition (2.17) whenever $f$ and $g$ are independent of $t$. Note that the theorem only applies to a quadratic function of the form $S(q,p)$, which is not the most general possible; for instance, the inner product $q^{\mathsf{T}}q$ is not included in that format.

Before proving the theorem we present a simple algebraic auxiliary result that will be used repeatedly in other contexts in what follows.

LEMMA 2.5. *Let $q_n$, $p_n$, $Q_i$, $P_i$, $k_{n,i}$, $\ell_{n,i}$ be arbitrary vectors satisfying* (2.11) *and* (2.13). *If $S$ is bilinear and* (2.16) *holds, then*

$$S(q_{n+1},p_{n+1}) - S(q_n,p_n) = h_n \sum_i b_i \Big( S(k_{n,i},P_{n,i}) + S(Q_{n,i},\ell_{n,i}) \Big). \tag{2.18}$$

*Proof.* Since $S$ is bilinear, we may write from (2.11)

$$\begin{aligned} S(q_{n+1},p_{n+1}) - S(q_n,p_n) = h_n \sum_i b_i S(k_{n,i},p_n) + h_n \sum_j B_j S(q_n,\ell_{n,j}) \\ + h_n^2 \sum_{ij} b_i B_j S(k_{n,i},\ell_{n,j}). \end{aligned}$$

Now use (2.13) to eliminate $q_n$ and $p_n$ from the right-hand side:

$$\begin{aligned} S(q_{n+1},p_{n+1}) - S(q_n,p_n) = h_n \sum_i b_i S\Big( k_{n,i}, P_{n,i} - h_n \sum_j A_{ij}\ell_{n,j} \Big) \\ + h_n \sum_j B_j S\Big( Q_{n,j} - \sum_i a_{ji} k_{n,i}, \ell_{n,j} \Big) \\ + h_n^2 \sum_{ij} b_i B_j S(k_{n,i},\ell_{n,j}). \end{aligned}$$

In view of the bilinearity and (2.16), the proof is complete. □

*Proof of the theorem.* Conservation of $S$ implies that

$$S(f(q,p,t),p) + S(q,g(q,p,t)) \equiv 0,$$

because, along each solution $q(t)$, $p(t)$,

$$S\left(\frac{d}{dt}q(t),p(t)\right) + S\left(q(t),\frac{d}{dt}p(t)\right) = \frac{d}{dt}S(q(t),p(t)) = 0.$$

Therefore, (2.12) and (2.17) mean that the right-hand side of (2.18) vanishes. □

For the preservation of the symplectic structure, the result (derived in [40] and [1] independently) is as follows.

THEOREM 2.6. *Assume that the system* (2.10) *is Hamiltonian. The relations* (2.16)–(2.17) *guarantee that the mapping $(q_n,p_n) \mapsto (q_{n+1},p_{n+1})$ defined in* (2.11)–(2.13) *is symplectic.*

The conditions (2.16)–(2.17) are essentially necessary for symplecticness [34] and hence the following definition is commonly used.

DEFINITION 2.7. *The PRK scheme* (2.1), (2.9) *is called* symplectic *if* (2.16)–(2.17) *hold.*

If the PRK is symplectic, there is a reduction in the number of independent order conditions; the classes of equivalent order conditions were first described by Hairer [18]. An alternative treatment (see [27]) based on so-called H-trees was given by Murua in his 1995 thesis; cf. [4]. For instance, for a symplectic PRK method to have order $\geq 4$ it is necessary to impose 13 order conditions: for general PRK methods that number is 36.

**3. Variational Systems and Their Adjoints.** We now explore the role of symplectic RK schemes when integrating adjoint variational systems. A comprehensive discussion of the use of adjoints to determine sensitivities is not within the scope of this paper. A general introduction, together with applications to aerodynamics, is provided in [14]. Applications of adjoints to atmospheric models are discussed in [30]. Of course, the idea of an adjoint problem is not restricted to differential equations; see [6] for an early paper describing a very general framework.

**3.1. The Continuous Problem: Quadratic Conservation.** We now present the mathematical foundations of the remainder of the paper. Consider a $d$-dimensional differential system

$$\frac{d}{dt}x = f(x,t) \tag{3.1}$$

and denote by $\alpha \in \mathbb{R}^d$ the corresponding initial value and by $\bar{x}(t)$ the solution that arises from the perturbed initial condition $\bar{x}(t_0) = \alpha + \eta$. Linearization of (3.1) around $x(t)$ shows that as $|\eta| \to 0$, $\bar{x}(t) = x(t) + \delta(t) + o(|\eta|)$, where $\delta$ solves the (linear) *variational system* (see, e.g., [21, section I.14])

$$\frac{d}{dt}\delta = \partial_x f(x(t),t)\,\delta \tag{3.2}$$

($\partial_x f$ is the Jacobian matrix of $f$ with respect to $x$). Thus, when $x(t)$ is known, solving for $\delta(t_0+T)$ the initial value problem given by (3.2) and $\delta(t_0) = \eta$ yields an estimate for the change in solution $\bar{x}(t) - x(t)$; see a simple example in Figure 1.

The *adjoint* system of (3.2) is given by

$$\frac{d}{dt}\lambda = -\partial_x f(x(t),t)^{\mathsf{T}}\,\lambda. \tag{3.3}$$

(To avoid confusion, variables in this paper are always *column vectors*; from a mathematical point of view it would have been better to write sensitivities, Lagrange multipliers, and momenta as row vectors, as they belong to the dual space of the space of states.) The right-hand side in (3.3) has been chosen in such a way that the following proposition is valid. More precisely, it is best to think that *the adjoint is the system for which the conservation property* (3.5) *holds*.

PROPOSITION 3.1. *For each $x$, $\delta$, $\lambda \in \mathbb{R}^d$ and real $t$,*

$$\big(-\partial_x f(x,t)^{\mathsf{T}}\,\lambda\big)^{\mathsf{T}}\delta + \lambda^{\mathsf{T}}\partial_x f(x,t)\delta = 0.$$

*Therefore, if $\delta(t)$ and $\lambda(t)$ are arbitrary solutions of* (3.2), (3.3), *respectively, then*

$$\frac{d}{dt}\lambda(t)^{\mathsf{T}}\delta(t) = \Big(\frac{d}{dt}\lambda(t)\Big)^{\mathsf{T}}\delta(t) + \lambda(t)^{\mathsf{T}}\Big(\frac{d}{dt}\delta(t)\Big) \equiv 0 \tag{3.4}$$

*and, accordingly,*

$$\lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0). \tag{3.5}$$

Why is the adjoint system useful? Regard $\eta$ as a parameter and assume that we are interested in finding $\omega^{\mathsf{T}}\delta(t_0+T)$ for fixed $\omega \in \mathbb{R}^d$, i.e., in estimating, at the final time $t_0+T$, the change along the direction of $\omega$ of the solution of (3.1) induced by the initial perturbation $\alpha \mapsto \alpha + \eta$. (For instance, choosing $\omega$ equal to the $r$th

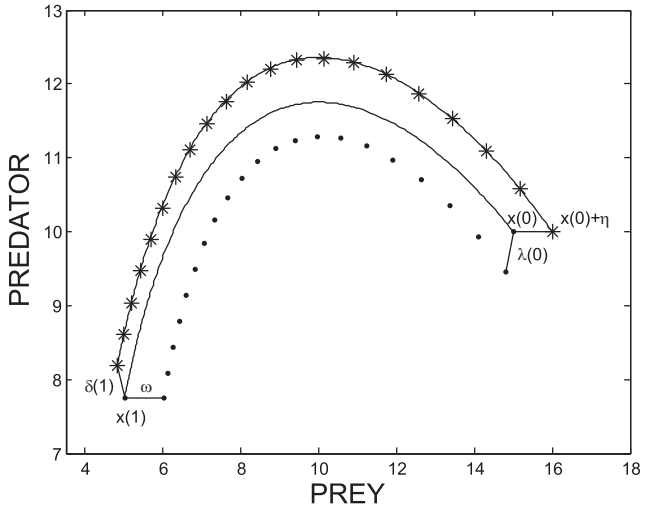


**Fig. 1** *Two-species Lotka–Volterra system $dx^1/dt = x^1 - 0.2x^1x^2$, $dx^2/dt = -2x^2 + 0.2x^1x^2$ (superscripts indicate components of vectors); $x^1$ and $x^2$ represent, in suitable units, numbers of prey and predators, respectively. The solid lines give, for $0 \le t \le 1$, the unperturbed solution $x(t)$ with initial condition $x(0) = (15, 10)$ and a perturbed solution $\bar{x}(t)$ with $\bar{x}(0) = x(0) + \eta = (16, 10)$: an increase in the number of prey at $t = 0$ leads at $t = 1$ to a decrease in the number of prey and to an increase in the number of predators. The stars are the points $x(t) + \delta(t)$, $t = 0, 0.05, 0.10, \ldots$, where $\delta$ solves the variational system; they almost coincide with the corresponding values of the perturbed solution $\bar{x}(t)$. In particular, the change in the number of prey, $\bar{x}^1(1) - x^1(1)$, is very well approximated by $\delta^1(1) = -0.1786\ldots$, i.e., by the inner product $\omega^{\mathsf{T}}\delta(1)$, where $\omega$ denotes the first coordinate vector $(1, 0) = \nabla x^1$. The variational equations move $\eta = \delta(0)$ forward to $\delta(1)$. The dots show how the adjoint equations move $\omega = \lambda(1)$ backward to yield $\lambda(0) = \nabla_{x(0)} x^1(1)$, the gradient of $x^1$ as a function of $x(0)$. The inner product $\omega^{\mathsf{T}}\delta(1)$ exactly coincides with $\lambda(0)^{\mathsf{T}}\eta$. In a Lotka–Volterra system with $d$ species, a single integration of the adjoint system is necessary to find the $d$-dimensional gradient of $x^1(1)$ as a function of $x(0)$.*

coordinate vector would correspond to estimating the change in the $r$th component of the solution.) When $x(t)$ is known, we solve (3.3) with the *final* condition $\lambda(t_0 + T) = \omega$ and note that the quantity we seek coincides with $\lambda(t_0)^{\mathsf{T}}\eta$ because, from the proposition,

$$\omega^{\mathsf{T}}\delta(t_0 + T) = \lambda(t_0 + T)^{\mathsf{T}}\delta(t_0 + T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0) = \lambda(t_0)^{\mathsf{T}}\eta.$$

The advantage of this procedure is that, as $\eta$ varies, the computation of $\lambda(t_0)^{\mathsf{T}}\eta$ requires only *one* integration of (3.3); the computation of $\omega^{\mathsf{T}}\delta(t_0 + T)$ via (3.2) would need a fresh integration for each new choice of $\eta$ (see Figure 1).

As an application, consider the task of computing the gradient, $\nabla_\alpha \mathcal{C}(x(t_0+T))$, of a real-valued function $\mathcal{C}$ with respect to the initial data $\alpha$. We set $\omega = \nabla_x \mathcal{C}(x(t_0+T))$ in the preceding construction and successively let the $r$th coordinate vector, $r = 1, \ldots, d$, play the role of $\eta$ to conclude that the gradient sought has the value $\lambda(t_0)$, where $\lambda(t)$ is the solution of the adjoint system with final condition $\lambda(t_0 + T) = \nabla_x \mathcal{C}(x(t_0 + T))$. Only one integration is required to find $d$ derivatives $\partial/\partial\alpha^r$. The adjoint system (3.3) "pulls back" gradients with respect to $x(t_0 + T)$ into gradients with respect to $x(t_0)$.

**3.2. The Continuous Problem: Lagrange Multipliers.** We shall also need an alternative derivation of the recipe $\nabla_\alpha \mathcal{C}(x(t_0+T)) = \lambda(t_0)$ found above. Since the use of Lagrange multipliers (see, e.g., [14, section 2.5]) in this context (as distinct from their use in minimization) might not be known to some readers, we give full details. Define the Lagrangian functional $\mathcal{L} = \mathcal{L}(\hat\alpha, \hat x, \hat\lambda_0, \hat\lambda)$ as

$$\mathcal{L} = \mathcal{C}(\hat x(t_0+T)) - \hat\lambda_0^{\mathsf{T}}\big(\hat x(t_0) - \hat\alpha\big) - \int_{t_0}^{t_0+T} \hat\lambda(t)^{\mathsf{T}}\Big(\frac{d}{dt}\hat x(t) - f(\hat x(t),t)\Big)\,dt,$$

where $\hat\alpha$, $\hat\lambda_0$ are arbitrary vectors, and $\hat x$, $\hat\lambda$ are arbitrary functions. A key point here is that, whenever $\hat x$ is a solution of (3.1) and $\hat x(t_0) = \hat\alpha$, the value of $\mathcal{L}(\hat\alpha, \hat x, \hat\lambda_0, \hat\lambda)$ coincides with $\mathcal{C}(\hat x(t_0+T))$.

If $\eta$ and $\delta$ are the variations of $\hat\alpha$ and $\hat x$, respectively, the variation $\delta\mathcal{L}$ of the functional is

$$\begin{aligned}\delta\mathcal{L} = \nabla_x \mathcal{C}(\hat x(t_0+T))^{\mathsf{T}}\delta(t_0+T) - \hat\lambda_0^{\mathsf{T}}\big(\delta(t_0) - \eta\big)\\ - \int_{t_0}^{t_0+T} \hat\lambda(t)^{\mathsf{T}}\Big(\frac{d}{dt}\delta(t) - \partial_x f(\hat x(t),t)\delta(t)\Big)\,dt,\end{aligned}$$

so that, after integration by parts,

$$\begin{aligned}\delta\mathcal{L} = \big(\nabla_x \mathcal{C}(\hat x(t_0+T)) - \hat\lambda(t_0+T)\big)^{\mathsf{T}}\delta(t_0+T) + \hat\lambda(t_0)^{\mathsf{T}}\eta\\ + \big(\hat\lambda(t_0) - \hat\lambda_0\big)^{\mathsf{T}}\delta(t_0)\\ + \int_{t_0}^{t_0+T}\Big(\frac{d}{dt}\hat\lambda(t)^T\delta(t) + \hat\lambda(t)^T\partial_x f(\hat x(t),t)\delta(t)\Big)\,dt.\end{aligned}$$

We now make choices $\lambda_0$, $\lambda$ (depending on $\hat\alpha$ and $\hat x$) for the (so far arbitrary) multipliers $\hat\lambda_0$, $\hat\lambda$. We define $\lambda$ as the solution of (3.3) (with $\hat x(t)$ in lieu of $x(t)$) subject to the final condition $\lambda(t_0+T) = \nabla_x \mathcal{C}(\hat x(t_0+T))$ and set $\lambda_0 = \lambda(t_0)$. These choices ensure that, at $\hat\alpha$, $\hat x$, the *intermediate* variation $\delta(t)$ does not contribute to $\delta\mathcal{L}$; we then have (at $\hat\alpha$, $\hat x$) $\delta\mathcal{L} = \lambda(t_0)^{\mathsf{T}}\eta$ or, in other words, $\lambda(t_0)$ is the gradient of $\mathcal{L}$ as a function of $\hat\alpha$. Since, as pointed out above, if $\hat x$ solves (3.1) and $\hat x(t_0) = \hat\alpha$, then $\mathcal{L}(\hat\alpha, \hat x, \hat\lambda_0, \hat\lambda) = \mathcal{C}(\hat x(t_0+T))$, we conclude that $\lambda(t_0) = \nabla_\alpha \mathcal{C}(x(t_0+T))$ as we aimed to prove. The original system (3.1) and the initial condition may also be retrieved from the Lagrangian by setting to zero the variations with respect to $\hat\lambda$ and $\hat\lambda_0$, respectively.

The same approach can also be used if we wish to make things more involved and introduce the velocity $(d/dt)\hat x = \hat k$ as a new argument in the Lagrangian. To simplify the notation we shall hereafter drop all hats, so that the same symbols $\alpha$, $x$, etc., will be used for the arbitrary arguments of the Lagrangian (that previously were written as $\hat\alpha$, $\hat x$, etc.) and for the corresponding values at the solution sought. When the velocity is considered as a new argument, the Lagrangian becomes

$$\begin{aligned}\mathcal{L} = \mathcal{C}(x(t_0+T)) - \lambda_0^{\mathsf{T}}\big(x(t_0) - \alpha\big)\\ - \int_{t_0}^{t_0+T} \lambda(t)^{\mathsf{T}}\Big(\frac{d}{dt}x(t) - k(t)\Big)\,dt\\ - \int_{t_0}^{t_0+T} \Lambda(t)^{\mathsf{T}}\Big(k(t) - f(x(t),t)\Big)\,dt.\end{aligned} \tag{3.6}$$

Taking variations and choosing the multipliers to cancel the undesired contributions to $\delta\mathcal{L}$ leads to the relations $\lambda(t_0) = \nabla_\alpha \mathcal{C}(x(t_0+T))$, $\lambda(t_0+T) = \nabla_x \mathcal{C}(x(t_0+T))$, and $\lambda_0 = \lambda(t_0)$ found above and, additionally, to $\Lambda(t) \equiv \lambda(t)$ (as expected).

**3.3. The Discrete Problem: RK Integration.** Let us suppose that (3.1) has been discretized by means of the RK scheme (2.1) to obtain, $n = 0, \dots, N-1$,

$$x_{n+1} = x_n + h_n \sum_{i=1}^{s} b_i k_{n,i}, \tag{3.7}$$

$$k_{n,i} = f(X_{n,i}, t_n + c_i h_n), \quad i = 1, \dots, s, \tag{3.8}$$

$$X_{n,i} = x_n + h_n \sum_{j=1}^{s} a_{ij} k_{n,j}, \quad i = 1, \dots, s, \tag{3.9}$$

and that, in analogy with the preceding material, we wish to estimate the impact on $x_N$ of a perturbation of the initial condition $x_0 = \alpha$. Linearization of the RK equations (3.7)–(3.9) around $x_n$, $X_{n,i}$ shows that the perturbed RK solution $\bar{x}_n$, $n = 0, \dots, N$, satisfies $\bar{x}_n = x_n + \delta_n + o(|\eta|)$ with

$$\delta_{n+1} = \delta_n + h_n \sum_{i=1}^{s} b_i d_{n,i}, \tag{3.10}$$

$$d_{n,i} = \partial_x f(X_{n,i}, t_n + c_i h_n) \Delta_{n,i}, \quad i = 1, \dots, s, \tag{3.11}$$

$$\Delta_{n,i} = \delta_n + h_n \sum_{j=1}^{s} a_{ij} d_{n,j}, \quad i = 1, \dots, s \tag{3.12}$$

(the vectors $d_{n,i}$ and $\Delta_{n,i}$ are the variations in the slopes $k_{n,i}$ and stages $X_{n,i}$, respectively).

On the other hand, if we regard the given differential equations (3.1) together with the variational equations (3.2) as a $2d$-dimensional system for the vector $y = [x^{\mathsf{T}}, \delta^{\mathsf{T}}]^{\mathsf{T}}$ and apply the RK scheme as in (2.4)–(2.6), we also arrive at (3.7)–(3.12). We have thus proved the following result, as in, say, [19, Chapter VI, Lemma 4.1]:

THEOREM 3.2. *The process of RK discretization commutes with forming variational equations: the RK discretization of the continuous variational equations* (3.1)–(3.2) *yields the variational equations* (3.7)–(3.12) *for the RK discretization.*

The situation for the adjoint equations is not quite as neat (cf. [37]). In order to find the discrete sensitivity $\omega^{\mathsf{T}} \delta_N$ we would like to numerically integrate (3.3) with final condition $\lambda_N = \omega$ in such a way that (cf. (3.5))

$$\lambda_N^{\mathsf{T}} \delta_N = \lambda_0^{\mathsf{T}} \delta_0. \tag{3.13}$$

Although in the actual computation the approximations $\lambda_n$ can be found without using (3.10)–(3.12) for $\delta_n$ (this is the whole point behind the use of adjoints), let us consider for a moment the $3d$-dimensional system (3.1)–(3.3) for the extended vector $y = [x^{\mathsf{T}}, \delta^{\mathsf{T}}, \lambda^{\mathsf{T}}]^{\mathsf{T}}$. Then the condition (3.13) demands that we integrate this large system in such a way as to *exactly* preserve the invariant $I(y(t), y(t)) = \lambda(t)^{\mathsf{T}} \delta(t)$ in (3.4). According to Theorem 2.1, we may achieve this goal by using the RK scheme (2.1) *provided that it is symplectic.* This results in the relations (3.7)–(3.12) in tandem with ($n = 0, \dots, N-1$)

$$\lambda_{n+1} = \lambda_n + h_n \sum_{i=1}^{s} b_i \ell_{n,i}, \tag{3.14}$$

$$\ell_{n,i} = -\partial_x f(X_{n,i}, t_n + c_i h_n)^{\mathsf{T}} \Lambda_{n,i}, \quad i = 1, \dots, s, \tag{3.15}$$

**Table 1** *Euler integration on a uniform grid of the $x$, $\delta$, $\lambda$ equations for the Lotka–Volterra problem in Figure 1. The lack of symplecticness of the integrator results in $\lambda_0^\mathsf{T}\eta$ being different from $\omega^\mathsf{T}\delta_N$: the discretization of the adjoint equations does not provide the adjoint of the discretization. The convergence of the integrator implies that, as the grid is refined, $\lambda_0^\mathsf{T}\eta$ and $\omega^\mathsf{T}\delta_N$ are $\mathcal{O}(h)$ away from their common limit $\lambda(0)^\mathsf{T}\eta = \omega^\mathsf{T}\delta(1) \approx -0.1786$, as borne out by the last two columns. When, alternatively, the $\lambda$ equations are integrated with the Radau method (3.24) the numerical results for $\lambda_0^\mathsf{T}\eta$ coincide with those displayed in the third column of the table.*

| $h$ | $\lambda_0^\mathsf{T}\eta$ | $\omega^\mathsf{T}\delta_N$ | $\lambda_0^\mathsf{T}\eta - \lambda(0)^\mathsf{T}\eta$ | $\omega^\mathsf{T}\delta_N - \omega^\mathsf{T}\delta(1)$ |
|---|---|---|---|---|
| 0.100 | −0.1070 | −0.2497 | 0.0717 | −0.0710 |
| 0.050 | −0.1401 | −0.2135 | 0.0385 | −0.0348 |
| 0.025 | −0.1588 | −0.1959 | 0.0199 | −0.0172 |

$$\Lambda_{n,i} = \lambda_n + h_n \sum_{j=1}^{s} a_{ij}\ell_{n,j}, \quad i = 1, \dots, s. \tag{3.16}$$

Let us summarize the preceding discussion in the following theorem.

THEOREM 3.3. *Assume that the $3d$-dimensional system* (3.1)–(3.3) *is discretized by a* symplectic *RK scheme* (2.1). *Then, for any RK solution,* (3.13) *holds. In particular, for the RK solution specified by the initial condition $x_0 = \alpha$, $\delta_0 = \eta$ together with the final condition $\lambda_N = \omega$,*

$$\omega^\mathsf{T}\delta_N = \lambda_0^\mathsf{T}\eta.$$

For a nonsymplectic RK scheme of order $\rho$, $\omega^\mathsf{T}\delta_N$ and $\lambda_0^\mathsf{T}\eta$ are approximations of order $\rho$ to their continuous counterparts $\omega^\mathsf{T}\delta(t_0 + T)$ and $\lambda(t_0)^\mathsf{T}\eta$, respectively, and therefore $\lambda_0^\mathsf{T}\eta$ will be an $\mathcal{O}(h^\rho)$ approximation to the true sensitivity $\omega^\mathsf{T}\delta_N$ of the discrete solution. See the example in Table 1 where the Euler integrator was chosen to have large errors and show clearly the difference between $\omega^\mathsf{T}\delta_N$ and $\lambda_0^\mathsf{T}\eta$.

In practice, the variational equations (3.2) do *not need to be integrated*. We successively find $x_0$, $x_1$, …, $x_N$ via (3.7)–(3.9) and, once these are available, we set $\lambda_N = \omega$ and compute $\lambda_{N-1}$, …, $\lambda_0$ from (3.14)–(3.16) taken in the order $n = N-1, N-2, \dots, 0$. For this reason, it may be advisable to rewrite (3.14)–(3.16) in the following "reflected" form (see section 7) that emphasizes that the approximation $\lambda_n$ at $t_n$ is to be found from the approximation $\lambda_{n+1}$ at $t_{n+1}$:

$$\lambda_n = \lambda_{n+1} + (-h_n) \sum_{i=1}^{s} b_i \ell_{n,i}, \tag{3.17}$$

$$\ell_{n,i} = -\partial_x f(X_{n,i}, t_{n+1} + (1 - c_i)(-h_n))^\mathsf{T} \Lambda_{n,i}, \quad i = 1, \dots, s, \tag{3.18}$$

$$\Lambda_{n,i} = \lambda_{n+1} + (-h_n) \sum_{j=1}^{s} (b_j - a_{ij})\ell_{n,j}, \quad i = 1, \dots, s. \tag{3.19}$$

In analogy to the continuous case, for a symplectic RK discretization, $\nabla_\alpha \mathcal{C}(x_N)$ may be computed by finding $\lambda_0$ from the recursion (3.14)–(3.16) (or (3.17)–(3.19)) with $\lambda_N = \nabla_x \mathcal{C}(x_N)$.

**3.4. The Discrete Problem: PRK Integration.** Theorem 3.3 may be generalized easily with the help of Theorem 2.4. In what follows it is understood that when using the PRK scheme the $x$, $\delta$ equations are integrated with the set of coefficients

(2.1) (so that the $\delta_n$ are exactly the variations in $x_n$) and the $\lambda$ equations with the set of coefficients (2.9). In other words, the system is partitioned as $q = [x^\mathsf{T}, \delta^\mathsf{T}]^\mathsf{T}$, $p = \lambda$.[1] This approach leads to (3.7)–(3.12) supplemented by the relations obtained by replacing the lowercase coefficients $a_{ij}$, $b_i$, $c_i$ in (3.14)–(3.16) by their uppercase counterparts:

$$\lambda_{n+1} = \lambda_n + h_n \sum_{i=1}^{s} B_i \ell_{n,i}, \tag{3.20}$$

$$\ell_{n,i} = -\partial_x f(X_{n,i}, t_n + C_i h_n)^\mathsf{T} \Lambda_{n,i}, \quad i = 1, \dots, s, \tag{3.21}$$

$$\Lambda_{n,i} = \lambda_n + h_n \sum_{j=1}^{s} A_{ij} \ell_{n,j}, \quad i = 1, \dots, s. \tag{3.22}$$

The generalization of Theorem 3.3 is as follows.

THEOREM 3.4. *Assume that the* $3d$*-dimensional system* (3.1)–(3.3) *is discretized by a* symplectic *PRK scheme* (2.1), (2.9). *Then* (3.13) *holds for any PRK solution. In particular, for the PRK solution specified by the initial condition* $x_0 = \alpha$, $\delta_0 = \eta$ *together with the final condition* $\lambda_N = \omega$,

$$\omega^\mathsf{T} \delta_N = \lambda_0^\mathsf{T} \eta.$$

Once more, for a symplectic PRK discretization, the gradient $\nabla_\alpha \mathcal{C}(x_N)$ coincides with $\lambda_0$ if $\lambda_N = \nabla_x \mathcal{C}(x_N)$. For a nonsymplectic discretization of the adjoint equations, $\lambda_0$ is a only an approximation to $\nabla_\alpha \mathcal{C}(x_N)$. For this reason *nonsymplectic PRK discretizations cannot be implied by the direct differentiation procedure* described in section 3.5.

How do we compute *exactly* (i.e., up to round-off) the sensitivity $\omega^\mathsf{T} \delta_N$ with the help of the adjoint system when the $x$ integration has been performed with a nonsymplectic RK scheme (2.1) and Theorem 3.3 cannot be invoked? Theorem 3.4 suggests a way. For simplicity we only look at the case where in (2.1) none of the weights $b_i$, $i = 1, \dots, s$, vanishes (for the general situation, see the appendix). From the coefficients in (2.1) we compute a new set

$$A_{ji} = b_i - b_i a_{ij}/b_j, \quad i, j = 1, \dots, s, \quad B_i = b_i, \quad C_i = c_i \quad i = 1, \dots, s. \tag{3.23}$$

In view of (2.16)–(2.17), we now have a PRK scheme for the discretization of (3.1)–(3.3) and Theorem 3.4 applies. If (2.1) is explicit, the computations required to descend from $\lambda_N$ to $\lambda_0$ are also explicit. Here is the simplest example. Assume that the $x$ equations are integrated with the explicit Euler rule ($s = 1$, $a_{11} = 0$, $b_1 = 1$, $c_1 = 0$). With that choice, $X_{n,1} = x_n$ and

$$x_{n+1} = x_n + h_n f(x_n, t_n).$$

The trick described above yields $A_{11} = 1$, $B_1 = 1$, $C_1 = 0$. Accordingly, the stage

[1] A variation on this theme is presented in [28, section 6] in the context of optimal control problems. There, the $x$ equations are themselves partitioned and integrated by means of a symplectic PRK.

$\Lambda_{n,1}$ coincides with $\lambda_{n+1}$ and using (2.11) we see that the required $\lambda$ integrator is

$$\lambda_{n+1} = \lambda_n - h_n \partial_x f(x_n, t_n)^{\mathsf{T}} \lambda_{n+1}. \tag{3.24}$$

Obviously this is *not* the explicit Euler rule, because $\lambda$ in the right-hand side appears at time $t_{n+1}$, and, unless the problem is autonomous, it is not the implicit Euler rule either because $t$ is evaluated at the retarded time $t_n$. (For RK enthusiasts only: the coefficients $A_{11} = 1$, $B_1 = 1$, $C_1 = 0$ correspond to the Radau IA method of one stage introduced by Ehle [22, section IV.5].)

In the particular situation where the $x$ integration has been performed by a symplectic RK method (symplectic RK methods possess nonvanishing weights [34, section 8.2]), the recipe (3.23) will lead to $A_{ij} = a_{ij}$ and the resulting PRK method will coincide with the original RK method. In the general case, for (3.13) to hold, *the adjoint equations for $\lambda$ have to be integrated with coefficients different from those used for the original equations for $x$.*

There are hidden difficulties with the use of this recipe. When stability is an issue, as in stiff problems or time-discretizations of partial differential equations, it is necessary to investigate carefully the stability behavior of the $\lambda$ integration [37]. On the other hand, and as noted before, the order of accuracy of the overall PRK, $x$, $\lambda$, integrator may be lower than the order of the RK method (2.1) for the $x$ we started with. When investigating the order of the overall PRK method we have to take into account that the right-hand side of (3.1) is independent of $\lambda$ and the right-hand side of (3.3) is linear in $\lambda$. These features imply that many elementary differentials vanish and that, accordingly, it is not necessary to impose the order conditions associated with them. Furthermore, we have to take into account the reduction in the number of independent order conditions implied by symplecticness.

**3.5. The Discrete Problem: Automatic Differentiation.** According to the preceding discussion, for any RK integration of (3.1) with nonzero weights, it is possible to find the gradient $\nabla_\alpha \mathcal{C}(x_N)$ by means of an integration of the adjoint equations with the coefficients (3.23). It is, however, clear that it is also perfectly possible to compute $\nabla_\alpha \mathcal{C}(x_N)$ by repeatedly using the chain rule in (3.7)–(3.9), as we shall do presently. Since $\mathcal{C}$ is scalar and $\alpha \in \mathbb{R}^d$, where $d$ is possibly large, reverse accumulation [15][2] is preferred, and this can be performed with the help of Lagrange multipliers as in section 3.2.

We shall need the following auxiliary result.

LEMMA 3.5. *Suppose that the mapping $\Omega : \mathbb{R}^{d+d'} \to \mathbb{R}^{d'}$ is such that the Jacobian matrix $\partial_\gamma \Omega$ is invertible at a point $(\alpha_0, \gamma_0) \in \mathbb{R}^d \times \mathbb{R}^{d'}$, so that in the neighborhood of $\alpha_0$, the equation $\Omega(\alpha, \gamma) = 0$ defines $\gamma$ as a function of $\alpha$. Consider a real-valued function in $\mathbb{R}^d$ of the form $\psi(\alpha) = \Psi(\alpha, \gamma(\alpha))$, for some $\Psi : \mathbb{R}^{d+d'} \to \mathbb{R}$. There exists*

[2]Recall that the idea of reverse accumulation is as follows. Imagine an application of the chain rule that leads to a product $J_3 J_2 J_1$, where $J_3$ is the Jacobian matrix $\partial(z)/\partial(y)$ of the final variables $z$ with respect to some intermediate variables $y$ and, similarly, $J_2 = \partial(y)/\partial(x)$, $J_1 = \partial(x)/\partial(w)$ ($w$ are the independent variables). When the dimension of $z$ is much lower than the dimensions of $x$, $y$, and $w$, computing the "short" (few rows) matrices $K = J_3 J_2$ and $K J_1$ (reverse accumulation) is much cheaper than first forming the "tall" (many rows) matrix $L = J_2 J_1$ and then $J_3 L$ (forward accumulation). The *forward* order $J_3(J_2 J_1)$ finds successively the Jacobians $J_1 = \partial(x)/\partial(w)$, $J_2 J_1 = \partial(y)/\partial(w)$, and $J_3 J_2 J_1 = \partial(z)/\partial(w)$. In *reverse* mode, the intermediate Jacobians are $J_3 = \partial(z)/\partial(y)$, $J_3 J_2 = \partial(z)/\partial(x)$, and $J_3 J_2 J_1 = \partial(z)/\partial(w)$. The analogy with the $\delta$ and $\lambda$ equations is manifest.

*a unique vector* $\lambda_0 \in \mathbb{R}^{d'}$ *such that (superscripts denote components)*

$$\nabla_\alpha \psi|_{\alpha_0} = \nabla_\alpha \Psi|_{(\alpha_0,\gamma_0)} + \sum_{r=1}^{d'} \lambda_0^r \nabla_\alpha \Omega^r|_{(\alpha_0,\gamma_0)},$$
$$0 = \nabla_\gamma \Psi|_{(\alpha_0,\gamma_0)} + \sum_{r=1}^{d'} \lambda_0^r \nabla_\gamma \Omega^r|_{(\alpha_0,\gamma_0)}.$$

*Proof.* The second requirement may be rewritten as

$$(\partial_\gamma \Omega)^{\mathsf{T}} \lambda_0 = -\nabla_\gamma \Psi, \tag{3.25}$$

with the matrix and right-hand side evaluated at $\alpha_0$, $\gamma_0$. This is a linear system that uniquely defines $\lambda_0$. To check that the vector $\lambda_0$ we have just found satisfies the first requirement, we use the chain rule

$$\partial_\alpha \psi|_\alpha = \partial_\alpha \Psi|_{(\alpha,\gamma(\alpha))} + \partial_\gamma \Psi|_{(\alpha,\gamma(\alpha))} \partial_\alpha \gamma|_\alpha,$$

differentiate $\Omega(\alpha, \gamma(\alpha)) = 0$ to get

$$\partial_\alpha \Omega|_{(\alpha,\gamma(\alpha))} + \partial_\gamma \Omega|_{(\alpha,\gamma(\alpha))} \partial_\alpha \gamma|_\alpha = 0,$$

evaluate at $\alpha_0$, and eliminate $\partial_\alpha \gamma|_{\alpha_0}$. ∎

It is useful to rephrase the lemma by introducing the Lagrangian

$$\mathcal{L}(\alpha, \gamma, \lambda) = \Psi(\alpha, \gamma) + \lambda^T \Omega(\alpha, \gamma),$$

so that the relation $\Omega(\alpha_0, \gamma_0) = 0$ and the equation (3.25) that defines the multiplier are, respectively,

$$\nabla_\lambda \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0,\gamma_0,\lambda_0)} = 0, \qquad \nabla_\gamma \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0,\gamma_0,\lambda_0)} = 0,$$

while the gradient we seek is computed as

$$\nabla_\alpha \psi|_{\alpha_0} = \nabla_\alpha \mathcal{L}(\alpha, \gamma, \lambda)|_{(\alpha_0,\gamma_0,\lambda_0)}.$$

Note that these developments mimic the material in section 3.2, with $\gamma$ playing the part of $\hat{x}$, $\gamma_0$ the part of $x$, etc.

In numerical differentiation, $\psi$ is the function whose gradient is to be evaluated, the components of $\alpha$ are the independent variables, and the components of $\gamma$ represent intermediate stages toward the computation of $\psi$. (For instance, in the simple case ($d = 1$) where $\psi(\alpha) = \alpha\sqrt{1 + \alpha \exp(\alpha) \cos(\exp(\alpha))}$, we may set the constraints $\Omega^1 = \gamma^1 - \exp(\alpha) = 0$, $\Omega^2 = \gamma^2 - \cos(\gamma^1) = 0$, $\Omega^3 = \gamma^3 - \alpha\gamma^1\gamma^2 = 0$, $\Omega^4 = \gamma^4 - \sqrt{1 + \gamma^3}$, $\psi = \alpha\gamma^4$.) The interpretation of the $\gamma^r$ as successive stages implies that, in practice, $\Omega$ will possess a lower triangular structure: $\Omega^r$ will only involve $\gamma^1, \ldots, \gamma^r$. The evaluation of $\psi$ successively finds the numerical values of $\gamma^1, \ldots, \gamma^{d'}$ in a forward fashion. The numerical values of the components $\lambda_0^r$ are then found by *backward* substitution in the upper-triangular linear system (3.25), and finally the lemma yields the required value of the gradient. If $\Psi$ and $\Omega$ have been judiciously chosen, then the mappings $\nabla_\alpha \Psi$, $\nabla_\gamma \Psi$, $\nabla_\alpha \Omega^r$, $\nabla_\alpha \Omega^r$ required to compute the gradient will have simple analytic expressions, easily derived by a human or by a computer program.

We now apply this technique to find $\nabla_\alpha \mathcal{C}(x_N)$. In (3.7)–(3.9) we let (the components of) $x_n$, $n = 0, \ldots, N$, and $k_{n,i}$, $n = 0, \ldots, N-1$, $i = 1, \ldots, s$, play the role of (the components of) $\gamma$ and introduce the Lagrangian

$$\mathcal{C}(x_N) - \lambda_0^\mathsf{T}(x_0 - \alpha) - \sum_{n=0}^{N-1} h_n \lambda_{n+1}^\mathsf{T}\Big[\frac{1}{h_n}(x_{n+1} - x_n) - \sum_{i=1}^{s} b_i k_{n,i}\Big]$$
$$- \sum_{n=0}^{N-1} h_n \sum_{i=1}^{s} b_i \Lambda_{n,i}^\mathsf{T}\Big[k_{n,i} - f(X_{n,i}, t_n + c_i h_n)\Big], \tag{3.26}$$

where we understand that the stage vectors $X_{n,i}$ are expressed in terms of the $x_n$ and $k_{n,i}$ by means of (3.9). Clearly this discrete Lagrangian is the natural RK approximation to (3.6).

A straightforward application of Lemma 3.5 now directly yields the following result, where we note that the hypothesis $b_i \neq 0$, $i = 1, \ldots, s$, is natural because when, say, $b_1 = 0$, the Lagrangian (3.26) does not incorporate the constraint $k_{n,1} = f(X_{n,1}, t_n + c_1 h_n)$. (The case of zero weights is considered in the appendix.)

THEOREM 3.6. *Consider the RK equations* (3.7)–(3.9)*, with* $b_i \neq 0$*,* $i = 1, \ldots, s$*. The computation of* $\nabla_\alpha \mathcal{C}(x_N)$ *based on the use of Lemma* 3.5 *with Lagrangian* (3.26) *leads to the relations* (3.20)–(3.22)*, with the coefficients* $A_{ij}$*,* $B_i$*,* $C_i$ *given by* (3.23)*, together with* $\nabla_x \mathcal{C}(x_N) = \lambda_N$ *and* $\nabla_\alpha \mathcal{C}(x_N) = \lambda_0$*.*

Note that, in the scenario of the theorem, $\lambda_N$, $\lambda_{N-1}$, $\lambda_{N-2}, \ldots$ successively yield the gradients $\nabla_{x_N} \mathcal{C}(x_N)$, $\nabla_{x_{N-1}} \mathcal{C}(x_N)$, $\nabla_{x_{N-2}} \mathcal{C}(x_N)$, .... It is well known that the reverse mode of differentiation implies an integration of the adjoint equations. The theorem shows additionally that, for an RK computation of $x$, the implied adjoint equation integration is such that the $x$, $\lambda$ system is discretized with a *symplectic* PRK method. Recall that we showed in the preceding subsection that nonsymplectic PRK methods cannot appear in this setting as they do not find $\nabla_\alpha \mathcal{C}(x_N)$ exactly. In a way, the chain rule provided us with symplectic integration *before the latter was invented*.

A further remark: the use of the chain rule with forward accumulation implies an RK integration of the variational equations (3.2) with the original RK coefficients (2.1). In agreement with a previous discussion, the forward mode is more expensive; each partial derivative $\partial/\partial\alpha^r$, $r = 1, \ldots, d$, in the gradient requires a separate integration.

## 4. A Simple Optimal Control Problem.

We explore next the role of symplectic methods when integrating the differential equations that arise in some optimal control problems [38], [41], [42]. In this section we look at the simplest case, where the developments are very similar to those just considered; more general problems are treated in the next section.

### 4.1. The Continuous Problem.

Consider now the $d$-dimensional system

$$\frac{d}{dt}x = f(x, u, t), \tag{4.1}$$

where $x$ is the state vector and $u$ a $\nu$-dimensional vector of controls. Our aim is to find functions $x(t)$ and $u(t)$, subject to (4.1) and the initial condition $x(t_0) = \alpha \in \mathbb{R}^d$, so as to minimize a given cost function $\mathcal{C}(x(t_0 + T))$.

The variational equation is (cf. (3.2))

$$\frac{d}{dt}\delta = \partial_x f(x(t), u(t), t)\, \delta + \partial_u f(x(t), u(t), t)\, \zeta, \tag{4.2}$$

where $\partial_u$ is the Jacobian matrix of $f$ with respect to $u$ and $\zeta$ denotes the variation in $u$; see, e.g., [38, section 2.8], [41, section 5.1]. Now $\delta(t_0) = 0$, as $x(t_0)$ remains nailed down at $\alpha$.

An adjoint system (cf. (3.3))

$$\frac{d}{dt}\lambda = -\partial_x f(x(t), u(t), t)^{\mathsf{T}}\,\lambda \tag{4.3}$$

and constraints

$$\partial_u f(x(t), u(t), t)^{\mathsf{T}}\lambda(t) = 0 \tag{4.4}$$

are introduced; see, e.g., [38, section 9.2]. As was the case with the adjoint in (3.3), the actual form of these equations is chosen to ensure the validity of the conservation property (3.5). More precisely, we have the following result.

PROPOSITION 4.1. *For each choice of vectors $x$, $u$, $\delta$, $\zeta$, $\lambda$ and real $t$,*

$$\Big(-\partial_x f(x,u,t)^{\mathsf{T}}\,\lambda\Big)^{\mathsf{T}}\delta + \lambda^{\mathsf{T}}\Big(\partial_x f(x,u,t)\delta + \partial_u f(x,u)\zeta\Big) = 0. \tag{4.5}$$

*Therefore, if $\delta(t)$, $\lambda(t)$, $\zeta(t)$ satisfy* (4.2)–(4.4)*, then* (3.4)–(3.5) *hold.*

The use of the proposition is as follows. We solve the two-point boundary problem given by the states+costates system (4.1), (4.3)–(4.4) with initial/final conditions

$$x(t_0) = \alpha, \qquad \lambda(t_0 + T) = \nabla\mathcal{C}(x(t_0 + T)). \tag{4.6}$$

Then the variation $\delta(t_0 + T)$ at the end of the interval is orthogonal to the gradient of the cost, since, from (3.5),

$$\nabla\mathcal{C}(x(t_0+T))^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T) = \lambda(t_0)^{\mathsf{T}}\delta(t_0) = 0. \tag{4.7}$$

This of course means that any solution $[x(t)^{\mathsf{T}}, \lambda(t)^{\mathsf{T}}, u(t)^{\mathsf{T}}]^{\mathsf{T}}$ of the boundary value problem satisfies the first order necessary condition for $\mathcal{C}$ to attain a minimum. As in sensitivity analyses, the costates $\lambda$ may be interpreted as *Lagrange multipliers*.

It is customary to introduce the function $H(x, \lambda, u, t) = \lambda^{\mathsf{T}} f(x, u, t)$ (pseudo-Hamiltonian) so that (4.1), (4.3)–(4.4) take the very symmetric form

$$\frac{d}{dt}x = \nabla_\lambda H, \quad \frac{d}{dt}\lambda = -\nabla_x H, \quad \nabla_u H = 0. \tag{4.8}$$

**4.2. The Discrete Problem: Indirect Approach.** In the indirect approach, approximations to the optimal states, costates, and controls are obtained by discretization of the boundary value problem (4.1), (4.3)–(4.4), (4.6). Note that we have to tackle a *differential-algebraic* system [22, Chapter VI.1], with the controls being algebraic variables since $(d/dt)u$ does not feature in any of the equations (4.1), (4.3)–(4.4). Under suitable technical assumptions (invertibility of the second derivative of $H$ with respect to $u$), the system is of *index one*. This means that the constraints (4.4) may be used to express, locally around the solution of interest, the algebraic variables as functions of the differential variables, $u = \Phi(x, \lambda, t)$. (When applying the implicit function theorem, the relevant Jacobian matrix is the Hessian $\partial_{uu}H$ and this will generically be positive definite if Pontryagin's principle [41, section 7.2] holds, so that $H(x, \lambda, \cdot, t)$ is minimized by $\Phi(x, \lambda, t)$.) For a system of index one suppose that the right-hand sides of (4.1) and (4.3) have been written as functions of $x$, $\lambda$, and $t$ by setting $u = \Phi(x, \lambda, t)$, thus transforming the differential-algebraic system into a system

of ordinary differential equations. In fact, the transformed system is the canonical Hamiltonian system with Hamiltonian function $\mathcal{H}(x,\lambda,t) = H(x,\lambda,\Phi(x,\lambda,t),t)$, because the chain rule and $\nabla_u H = 0$ imply that, in (4.8), $\nabla_x H(x,\lambda,u,t) = \nabla_x \mathcal{H}(x,\lambda,t)$ and $\nabla_x H(x,\lambda,u,t) = \nabla_x \mathcal{H}(x,\lambda,t)$. This Hamiltonian system may be discretized with the PRK scheme (2.1), (2.9). (Recall that RK schemes are included as particular cases where both sets of coefficients coincide.) The discrete equations are solved to find the approximations $x_n$ and $\lambda_n$ to $x(t_n)$, $\lambda(t_n)$ and finally the approximations to the controls are retrieved as $u_n = \Phi(x_n,\lambda_n,t_n)$.

The analytic expression of the implicit function $\Phi$ will in general not be available, so it will not be possible to find $\mathcal{H}$ explicitly. This is not a hindrance: the approximations $x_n$, $\lambda_n$, $u_n$ that one would find by a PRK integration of the Hamiltonian system can be found in practice as solutions of the set of equations (4.9)–(4.16) below, obtained by direct discretization of the differential-algebraic format (4.1), (4.3)–(4.4). The equivalence between the two approaches, differential and differential-algebraic, is seen by eliminating the controls from (4.9)–(4.16); see [22, Chapter VI.1].

The discrete equations are ($n = 0,\dots,N-1$)

$$x_{n+1} = x_n + h_n \sum_{i=1}^{s} b_i k_{n,i}, \tag{4.9}$$

$$k_{n,i} = f(X_{n,i}, U_{n,i}, t_n + c_i h_n), \quad i = 1,\dots,s, \tag{4.10}$$

$$X_{n,i} = x_n + h_n \sum_{j=1}^{s} a_{ij} k_{n,j}, \quad i = 1,\dots,s, \tag{4.11}$$

$$\lambda_{n+1} = \lambda_n + h_n \sum_{i=1}^{s} B_i \ell_{n,i}, \tag{4.12}$$

$$\ell_{n,i} = -\partial_x f(X_{n,i}, U_{n,i}, t_n + C_i h_n)^{\mathsf{T}} \Lambda_{n,i}, \quad i = 1,\dots,s, \tag{4.13}$$

$$\Lambda_{n,i} = \lambda_n + h_n \sum_{j=1}^{s} A_{ij} \ell_{n,j}, \quad i = 1,\dots,s, \tag{4.14}$$

$$\partial_u f(X_{n,i}, U_{n,i}, t_n + C_i h_n)^{\mathsf{T}} \Lambda_{n,i} = 0, \quad i = 1,\dots,s, \tag{4.15}$$

together with ($n = 0,\dots,N$)

$$\partial_u f(x_n, u_n, t_n)^{\mathsf{T}} \lambda_n = 0 \tag{4.16}$$

and the boundary conditions $x_0 = \alpha$, $\lambda_N = \nabla C(x_N)$ from (4.6).

What is the accuracy of this technique? We encounter the same difficulty as in the preceding section, relevant here is the order of the overall PRK scheme rather than the (possibly higher) order of the RK coefficients (2.1) used for the state variables. In the preceding section, the approximations $x_n$ were found independently of the $\lambda_n$ and, accordingly, the possible order reduction did not affect them. In the optimal control problem, states and costates are coupled, and any order reduction will harm both of them. This was first noted by Hager, who also provided relevant counterexamples; see [17, Table 3]. Hager [17, Proposition 6.1] also shows that there is no order reduction for explicit fourth order RK schemes with positive weights.

The obvious analogue of Theorem 3.2 holds: the variations $\delta_n$ in the discrete solution $x_n$ satisfy the equations that result from discretizing (4.2) with the coefficients (2.1). These equations are (3.10) and (3.12), where now

$$k_{n,i} = \partial_x f(X_{n,i}, U_{n,i}, t_n + c_i h_n)\, \Delta_{n,i} + \partial_u f(X_{n,i}, U_{n,i}, t_n + c_i h_n)\, Z_{n,i} \tag{4.17}$$

($\Delta_{n,i}$, $Z_{n,i}$ are the stages associated with the variables $\delta$ and $\zeta$).

Assume next that the PRK method is *symplectic*. Recall that symplecticness may be the result of choosing the RK coefficients (2.1) ($b_i \neq 0$, $i = 1, \ldots, s$) for the state variables and retrieving from (3.23) the coefficients (2.9) for the integration of the adjoint system. The symplecticness of the integrator makes it possible to formulate a discrete analogue of Proposition 4.1.

THEOREM 4.2. *Assume that* $x_n$, $\lambda_n$, $u_n$, $n = 0, \ldots, N$, *satisfy* (4.9)–(4.16) *arising from the application of a* symplectic *PRK method and that, furthermore,* $\delta_n$, $n = 0, \ldots, N$, $\delta_0 = 0$ *are the variations in* $x_n$. *Then, for* $n = 0, \ldots, N-1$,

$$\lambda_{n+1}^{\mathsf{T}} \delta_{n+1} = \lambda_n^{\mathsf{T}} \delta_n.$$

*The PRK scheme may be a symplectic RK scheme or the result of choosing freely the RK coefficients* (2.1), $b_i \neq 0$, $i = 1, \ldots, s$, *for the states and then using* (3.23) *to determine the coefficients for the integration of the costates.*

*Proof.* Use Lemma 2.5 with $S(q, p) = \lambda^{\mathsf{T}} \delta$. This results in

$$\lambda_{n+1}^{\mathsf{T}} \delta_{n+1} - \lambda_n^{\mathsf{T}} \delta_n = h_n \sum_i b_i (\Lambda_{n,i}^{\mathsf{T}} k_{n,i} + \ell_{n,i}^{\mathsf{T}} \Delta_{n,i}),$$

where $k_{n,i}$ and $\ell_{n,i}$ come from (4.17) and (4.13), respectively. According to (4.5), each of the terms being summed vanishes. ∎

When the boundary conditions (4.6) are imposed,

$$\nabla \mathcal{C}(x_N)^{\mathsf{T}} \delta_N = \lambda_N^{\mathsf{T}} \delta_N = \lambda_0^{\mathsf{T}} \delta_0 = 0,$$

which means that the discrete solution satisfies the first order necessary conditions for $\mathcal{C}(x_N)$ to achieve a minimum subject to the constraints (4.9)–(4.11) and $x_0 = \alpha$. In this way we have proved that *symplectic discretization* commutes [29] with the process of forming necessary conditions for minimization:

THEOREM 4.3. *Let* $\{x_n\}$, $\{\lambda_n\}$, $\{u_n\}$ *be a solution of* (4.9)–(4.16) *arising from discretizing with a* symplectic *PRK integrator the necessary conditions for the continuous optimal control problem. Then* $\{x_n\}$, $\{\lambda_n\}$, $\{u_n\}$ *satisfy the necessary conditions for* $\mathcal{C}(x_N)$ *to achieve a minimum subject to the discrete constraints* (4.9)–(4.11) *and* $x_0 = \alpha$. *The PRK scheme may be a symplectic RK scheme or the result of choosing freely the RK coefficients* (2.1), $b_i \neq 0$, $i = 1, \ldots, s$, *for the states and then using* (3.23) *to determine the coefficients for the integration of the costates.*

When the states+costates system is integrated by means of a nonsymplectic PRK scheme, $x_N$ will not satisfy the necessary conditions for $\mathcal{C}$ to be minimized subject to the constraints (4.9)–(4.11) and $x_0 = \alpha$. Therefore, nonsymplectic PRK discretizations *cannot* be obtained via the direct approach considered next.

**4.3. The Discrete Problem: Direct Approach.** The direct approach (see, e.g., [41, Chapter 9]) based on RK discretization begins by applying the scheme (2.1) to the differential equation (4.1) to get (4.9)–(4.11). Then these equations and $x_0 = \alpha$ are seen as constraints of a finite-dimensional optimization problem for the minimization of $\mathcal{C}(x_N)$.

We use the standard method of Lagrange multipliers based on the Lagrangian in (3.26), trivially adapted to the present circumstances by letting $f$ depend on the controls. The method leads in a straightforward way to the following result, first proved by Hager [17]; see also [4]. However, [17] does not point out that the relations (3.23) correspond to symplecticness. Furthermore, [17] and [4] do not use a discrete Lagrangian obtained by discretization of the continuous Lagrangian, and they along

with [7] do not point out that the occurrence of symplectic schemes in this context is really due to the conservation property (3.5).

THEOREM 4.4. *The first order necessary conditions for the minimization of* $\mathcal{C}(x_N)$ *subject to* $x_0 = \alpha$ *and* (4.9)–(4.11), $b_i \neq 0$, $i = 1, \ldots, s$, *are* $x_0 = \alpha$ *and* $\nabla\mathcal{C}(x_N) = \lambda_N$ *together with* (4.9)–(4.15), *with the coefficients* $A_{ij}$, $B_i$, $C_i$ *given by* (3.23).

In other words, when the direct approach is used, we arrive at *exactly the same set of equations* for $x_n$, $\lambda_n$, $X_{n,i}$, $\Lambda_{n,i}$, $U_{n,i}$ we obtained, with the help of RK technology, via the indirect approach in Theorem 4.3. Observe that the direct approach does not provide "natural" approximations $u_n$ to $u(t_n)$. Hager [17] suggests defining $u_n$ by locally minimizing $H(x_n, \lambda_n, u, t_n)$, which leads to (4.16). He also notes [17, Table 4] that the order of convergence of the control stages $U_{n,i}$ might be lower than that in $u_n$, which is not at all surprising: typically, internal stages are less accurate than end-of-step approximations. We remark that, in the direct approach and once the RK method for $x$ has been chosen, the minimization of $\mathcal{C}$ implicitly provides the "right" coefficients $A_{ij}$, $B_i$, $C_i$ to be used in the integration of the costates in order to ensure symplecticness of the overall PRK integrator. In the indirect approach, those coefficients have to be determined using the relations (2.16)–(2.17) and Theorem 2.4.

While the direct and indirect approaches can be seen as mathematically equivalent, each has its own appeal. The direct approach suggests solving the discrete PRK equations with the help of optimization techniques, which might be an efficient choice in practice. On the other hand, the direct approach "hides" the PRK integration of the costates, a fact that might lead to the false impression that the order of accuracy of the overall procedure coincides with the order of the RK scheme used to discretize the differential constraint (4.1). This was emphasized in [17], where the order of the PRK method (2.1), (2.9), (3.23) is called the order of the RK method (2.1) *for optimal control problems*. A discussion of the advantages of the direct and indirect approaches is outside our scope here; see, e.g., [41, Chapter 9], [10].

## 5. Some Extensions.

We now consider more general optimal control problems. We need to generalize Theorems 2.1 and 2.4 to the situation where the quantities $I$ or $S$ are not constant along trajectories of the system, but vary in a known manner.

### 5.1. Generalized Conservation.

In this section we give simple generalizations of Theorems 2.1 and 2.4. Only Theorem 5.2 will be proved; the other proof is very similar.

In order to better understand Theorem 5.1, we look at the case where $y$ comprises positions and velocities of a mechanical system and $I$ is the kinetic energy. Conservation of energy demands that the rate of change of $I$ coincides with the rate of change (power) $\varphi$ of the work of the forces. Along each trajectory, the gain in kinetic energy exactly matches the total work exerted by the forces.

THEOREM 5.1. *Assume that, for the differential system* (2.2), *there exist a real-valued bilinear mapping* $I$ *in* $\mathbb{R}^D \times \mathbb{R}^D$ *and a real-valued function* $\varphi$ *in* $\mathbb{R}^D$ *such that, for each solution* $y(t)$,

$$\frac{d}{dt} I(y(t), y(t)) = \varphi(y(t))$$

*and, therefore,*

$$I(y(t_0 + T), y(t_0 + T)) - I(y(t_0), y(t_0)) = \int_{t_0}^{t_0+T} \varphi(y(t))\, dt.$$

*If the system is integrated by means of a symplectic RK scheme as in* (2.4)–(2.6)*, then*

$$I(y_N, y_N) - I(y_0, y_0) = \sum_{n=0}^{N-1} h_n \sum_{i=1}^{s} b_i \, \varphi(Y_{n,i}).$$

Note that the last sum, based on the RK quadrature weights $b_i$ and the approximation $y(t_n + c_i h_n) \approx Y_{n,i}$, is the "natural" RK discretization of the corresponding integral.

THEOREM 5.2. *Assume that, for the partitioned system* (2.10)*, there exist a real-valued bilinear map $S$ in $\mathbb{R}^{D-d} \times \mathbb{R}^d$ and a real-valued function $\varphi$ in $\mathbb{R}^{D-d} \times \mathbb{R}^d$ such that for each solution*

$$\frac{d}{dt} S(q(t), p(t)) = \varphi(q(t), p(t))$$

*and, therefore,*

$$S(q(t_0 + T), p(t_0 + T)) - S(q(t_0), p(t_0)) = \int_{t_0}^{t_0+T} \varphi(q(t), p(t)) \, dt.$$

*If the system is integrated by means of a symplectic PRK scheme as in* (2.11)–(2.13)*, then*

$$S(q_N, p_N) - S(q_0, p_0) = \sum_{n=0}^{N-1} h_n \sum_{i=1}^{s} b_i \, \varphi(Q_{n,i}, P_{n,i}).$$

*Proof.* Use Lemma 2.5 and note that, under the present hypotheses,

$$S(k_{n,i}, P_{n,i}) + S(Q_{n,i}, \ell_{n,i}) = \varphi(Q_{n,i}, P_{n,i}),$$

because $S(f(q,p,t),p) + S(q, g(q,p,t)) \equiv \varphi(q,p)$ (cf. the proof of Theorem 2.4). ∎

**5.2. Other Optimal Control Problems.** Consider first the situation in section 4, but assume that the value $x(t_0)$ is not prescribed. Then $\delta(t_0)$ is free and for (4.7) to hold it is necessary to impose the condition $\lambda(t_0) = 0$. This replaces in (4.6) the initial condition $x(t_0) = \alpha$. The results in section 4 are valid in this setting after the obvious modifications.

We next look at the case where (4.1) and $x(0) = \alpha$ are imposed, but the cost function is given by

$$\mathcal{C}(x(t_0 + T)) + \int_{t_0}^{t_0+T} \mathcal{D}(x(t), u(t), t) \, dt \tag{5.1}$$

(this is often called a Mayer–Lagrange cost [41], as distinct from the Mayer cost $\mathcal{C}(x(t_0 + T))$ envisaged before). The adjoint system and constraints are, respectively,

$$\begin{aligned} &\frac{d}{dt}\lambda = -\partial_x f(x,u,t)^{\mathsf{T}} \, \lambda - \nabla_x \mathcal{D}(x,u,t), \\ &\partial_u f(x,u,t)^{\mathsf{T}} \lambda + \nabla_u \mathcal{D}(x,u,t) = 0. \end{aligned}$$

These take the form given in (4.8) for the pseudo-Hamiltonian $H = \lambda^T f + \mathcal{D}$.

The conservation property (3.5) is replaced by the generalized conservation formula

$$\lambda(t_0+T)^{\mathsf{T}}\delta(t_0+T)-\lambda(t_0)^{\mathsf{T}}\delta(t_0)\\ +\int_{t_0}^{t_0+T}\Big(\nabla_x\mathcal{D}(x(t),u(t),t)^{\mathsf{T}}\delta(t)+\nabla_u\mathcal{D}(x(t),u(t),t)^{\mathsf{T}}\zeta(t)\Big)\,dt=0,$$

which holds for arbitrary $\delta(t)$, $\lambda(t)$ satisfying the variational equations (4.2), the adjoint system, and the constraints. After setting $\delta(t_0)=0$ and $\lambda(t_0+T)=\nabla\mathcal{C}(x(t_0+T))$, the generalized conservation formula expresses that the variation of the cost vanishes, i.e., that the first order necessary conditions for the minimization hold.

For a symplectic PRK discretization of the algebraic-differential system, Lemma 2.5 may be used, as in the proof of Theorem 5.2, to show (the notation should be clear by now) that

$$\lambda_N^{\mathsf{T}}\delta_N-\lambda_0^{\mathsf{T}}\delta_0+\sum_{n=0}^{N-1}h_n\sum_{i=1}^{s}b_i\Big(\nabla_x\mathcal{D}(X_{n,i},U_{n,i},t_n+c_ih_n)^{\mathsf{T}}\Delta_{n,i}\\ +\nabla_u\mathcal{D}(X_{n,i},U_{n,i},t_n+c_ih_n)^{\mathsf{T}}Z_{n,i}\Big)=0.$$

By setting $\lambda_N=\nabla\mathcal{C}(x_N)$ and $\delta_0=0$, this formula expresses the necessary condition (orthogonality between gradient and variation) for the discrete solution to minimize the discretized cost

$$\mathcal{C}(x_N)+\sum_{n=0}^{N-1}h_n\sum_{i=1}^{s}b_i\mathcal{D}(X_{n,i},U_{n,i}).$$

Therefore, in this case as well, results corresponding to Theorems 4.3 and 4.4 hold for a symplectic PRK discretization.

It is of course possible to combine the cost (5.1) with alternative boundary specifications. If $x(t_0)$ is not prescribed, then we have to impose $\lambda(t_0)=0$, as pointed out above. If both $x(t_0)=\alpha$ and $x(t_0+T)=\beta$ are imposed (in which case the term $\mathcal{C}(x(t_0+T))$ may be dropped from the cost), then $\lambda(t_0)$ and $\lambda(t_0+T)$ are both free.

**5.3. Constrained Controls.** Let us go back once more to the problem in section 4 and suppose that the controls $u$ are constrained so that, for each $t$, it is demanded that $u(t)\in U$, where $U$ is a given closed, convex subset of $\mathbb{R}^\nu$. Then (see, e.g., [17]), the constraint (4.4) on $\lambda$ has to be replaced by

$$u(t)\in U,\qquad -\partial_u f(x(t),u(t),t)^{\mathsf{T}}\lambda(t)\in N_U(u(t)),$$

where $N_U(u)$ is the cone of all vectors $w\in\mathbb{R}^\nu$ such that, for each $v\in U$, $w^{\mathsf{T}}(v-u)\le 0$. Proceeding as in Proposition 4.1, we see that now $(d/dt)\lambda(t)^{\mathsf{T}}\delta(t)\ge 0$ and therefore

$$\nabla\mathcal{C}(x(t_0+T))^{\mathsf{T}}\delta(t_0+T)\ge 0,$$

which is the necessary condition for a minimum in the continuous problem. For a PRK discretization of the boundary value for the states+costates system, the relation

$$(d/dt)\lambda(t)^{\mathsf{T}}\delta(t)\ge 0$$

implies

$$k_{n,i}^{\mathsf{T}}\Lambda_{n,i}+\Delta_{n,i}^{\mathsf{T}}\ell_{n,i}\ge 0,$$

and therefore we may use Lemma 2.5 yet again to conclude that for symplectic PRK

methods and if *the weights $b_i$ are positive*,

$$\nabla \mathcal{C}(x_N)^{\mathsf{T}} \delta_N \geq 0.$$

Once more, results similar to Theorems 4.3 and 4.4 hold. See [9] for order reduction results.

**6. Lagrangian Mechanics.** Let us now consider Lagrangian mechanical systems [2]. Denote by $\mathcal{L}(x, u, t)$ the Lagrangian function, where $x \in \mathbb{R}^d$ are the Lagrangian coordinates and $u = (d/dt)x$ the corresponding velocities. According to Hamilton's principle, the trajectories $t \mapsto x(t)$ of the system are characterized by the fact that they render stationary (often minimum) the action integral

$$\int_{t_0}^{t_0+T} \mathcal{L}(x(t), u(t), t)\, dt$$

among all curves $t \mapsto \bar{x}(t)$ with $\bar{x}(t_0) = x(t_0)$ and $\bar{x}(t_0 + T) = x(t_0 + T)$. This may of course be viewed as a control problem that aims to make stationary (or even maximum) the cost (5.1) with $\mathcal{C} \equiv 0$ and $\mathcal{D} = -\mathcal{L}$, subject to the constraint $\dot{x} = u$ with fixed end-values $x(t_0)$ and $x(t_0 + T)$. The theory in section 5 applies and the pseudo-Hamiltonian is $H(x, \lambda, u, t) = \lambda^{\mathsf{T}} u - \mathcal{L}(x, u, t)$. The constraint $\nabla_u H = 0$ reads $\lambda = \nabla_u \mathcal{L}(x, u, t)$; thus, the control costates coincide with the mechanical momenta. The elimination of the controls with the help of Pontryagin's principle would determine $u$ as a function $\Phi(x, \lambda, t)$ by maximizing (recall that we are trying to maximize the cost here!) the function $u \mapsto H(x, \lambda, u, t)$. In mechanics, this corresponds exactly with the theory of the Legendre transformation as presented in [2, section 14]: that theory shows that, if $\mathcal{L}$ is a strictly convex function of $u$, then, at given $x$ and $t$, the velocity vector $u$ that corresponds to a given value of the momentum $\lambda$ is globally uniquely defined and maximizes $\lambda^{\mathsf{T}} u - \mathcal{L}(x, u, t)$. In most mechanical problems $\mathcal{L} = \mathcal{T}(x, u, t) - \mathcal{V}(x, t)$, with $\mathcal{T}$ and $\mathcal{V}$ the kinetic and potential energy, respectively, where $\mathcal{T}$ is quadratic, positive-definite as a function of $u$, thus ensuring the required convexity. In control theory the elimination of the controls $u$ in the pseudo-Hamiltonian $H$ gives rise to the "control" Hamiltonian $\mathcal{H}$; correspondingly, in mechanics the Hamiltonian is defined as the result of expressing in $\lambda^{\mathsf{T}} u - \mathcal{L}(x, u, t)$ the velocities as functions of the momenta (and $x$ and $t$). Finally, the evolution of the states and costates (mechanical coordinates and momenta) obeys Hamilton's canonical equations. Hamiltonian solution flows are symplectic and, in this way, we have traveled all the way from action minimization to symplecticness.

A similar journey may take place in the discrete realm. Choose any RK scheme (2.1) with nonzero weights to discretize the differential constraint $(d/dt)x = u$ and minimize the associated discrete action

$$\sum_{n=0}^{N-1} h_n \sum_{i=1}^{s} b_i\, \mathcal{L}(X_{n,i}, U_{n,i}, t_n + c_i h_n).$$

As we know from Theorem 4.3, this direct approach implies a symplectic PRK integration of the Hamiltonian system for $x$ and $\lambda$, where the $\lambda$ equations are integrated with the coefficients (2.9). This is nothing more than the variational construction of PRK symplectic integrators presented in the early paper [40] by Suris (see [26] for more information on integrators based on the principle of least action; cf. [23]). In this way, Hager's result [17] may be viewed as an extension of Suris's work to general control problems.

**7. What Is the Adjoint of an RK Method? Reflecting and Transposing Coefficients.** In this section we examine the relations between the preceding material and the notion of the adjoint of an RK method.

Scherer and Türke [35] associated with the set of RK coefficients (2.1) two new sets called the reflection and the transposition of the original. The reflected coefficients are given by $(i, j = 1, \ldots, s)$

$$a^r_{ij} = b_j - a_{ij}, \quad b^r_i = b_i, \quad c^r_i = 1 - c_i$$

and the transposed coefficients are defined, only for methods with nonzero weights $b_i$, by

$$a^t_{ij} = b_j a_{ji}/b_i, \quad b^t_i = b_i, \quad c^t_i = 1 - c_i.$$

The operations of reflection and transposition commute: the transposition of the reflection coincides with the reflection of the transposition, as both lead to

$$a^{rt}_{ij} = b_j - b_j a_{ji}/b_i, \quad b^{rt}_i = b_i, \quad c^{rt}_i = c_i.$$

Furthermore, both operations are involutions: each is its own inverse.

The paper [35] introduces the operations of reflection and transposition as algebraic manipulations that make it possible to interrelate important families of RK methods; no attempt is made there to interpret computationally the meaning of integrating with the reflected or transposed coefficients. What do reflection and transposition mean? The interpretation of reflection is well known [34, section 3.6], [21, Chapter II, Theorem 8.3]: a step of length $-h_n$ with the reflected RK method inverts the transformation $y_n \mapsto y_{n+1}$ induced by a step of length $h_n$ with the original method. In this paper we have seen this idea at work when moving from (3.14)–(3.16) to (3.17)–(3.19). The formulas (3.23) give meaning to the idea of transposition: to construct a symplectic PRK method out of a given RK method with nonvanishing weights, the $p$ coefficients are determined by reflecting and transposing the given $q$ coefficients. The transposed of the $q$ coefficients are then those required to integrate backward the $p$ equations in, say, sensitivity analyses.

As a further illustration of these ideas, consider the linear nonautonomous system

$$\frac{d}{dt}q = M(t)q, \quad \frac{d}{dt}p = -M(t)^{\mathsf{T}}p,$$

integrated with the PRK method (2.1), (2.9) (this is a Hamiltonian system). Since $p$ and $q$ are uncoupled, this amounts to an RK integration of the $q$ equations with the coefficients (2.1) together with an RK integration of the $p$ equations with the coefficients (2.9). The system has the invariant $q^{\mathsf{T}}p$; Theorem 2.4 ensures that it will be preserved if the $p$ coefficients are the transposition of the reflection of the $q$ coefficients. Both sets of coefficients only coincide if $q$ itself is integrated symplectically. If we wish to preserve the invariant, a nonsymplectic integration of $q$ is possible, but then one has to compensate by integrating the $p$ equations in an appropriate way and the order and stability of the $p$ integration have to be investigated separately. Again, if the $p$ equations are integrated backward in time, then preservation of $q^{\mathsf{T}}p$ requires that such a backward integration be performed with the transposition of the coefficients used to propagate $q$ forward.

We conclude this section with a remark on terminology. Monographs such as [19] and [34] use the word *adjoint* to refer to the method with reflected coefficients. Section 3 and our last comments suggest that, in order to proceed as in the differential

equation case, it would have been better to keep the word *adjoint* for the reflected and transposed method, and let *reflected* refer to what in [19] or [34] is called *adjoint*. Using that alternative terminology, for RK schemes, symplecticness would simply be *self-adjointness*.

**8. Conclusion.** Symplectic RK and PRK schemes preserve, by definition, the symplectic form in phase space; in addition, they may be characterized as those RK or PRK integrators that exactly preserve each quadratic invariant of the system being integrated. In sensitivity analysis, optimal control, and other areas, adjoint systems are introduced and possess paramount importance; these adjoints are defined so as to preserve the key quadratic invariant (3.5). Therefore, there are tight connections between those areas and the theory of symplectic integration; we hope the present article has helped to explain those connections.

**Appendix. Schemes with Some Vanishing Weights.** If one or more weights $b_i$ in (2.1) vanish, then it is not possible to use the recipe (3.23) to define the coefficients required to create a combined symplectic PRK method (2.1), (2.9). Given the partitioned system (2.10) and the $q$ coefficients (2.1), how do we integrate the $p$ equations so as to have a symplectic scheme? The solution to this problem is rather weird and it is best to begin with the simplest example.

Let us study the second order scheme (due to Runge in his 1895 original paper [21, section II.1]), $s = 2$,

$$a_{11} = a_{21} = a_{22} = 0,\ a_{12} = 1/2,\ b_1 = 1,\ b_2 = 0,\ c_1 = 1/2,\ c_2 = 0. \tag{A.1}$$

While it is customary to label the stages so that the abscissas $c_i$ increase with $i$, we have departed from this practice; if we adopted it, formula (A.6) below would have a rather disordered appearance.

We regularize the zero weight and consider the one-parameter family, $\epsilon \neq 0$,

$$a_{11} = a_{21} = a_{22} = 0,\ a_{12} = 1/2,\ b_1 = 1,\ b_2 = \epsilon,\ c_1 = 1/2,\ c_2 = 0. \tag{A.2}$$

(The regularized scheme is not even consistent, but this does not hinder the argument.) From (3.23), we set

$$A_{11} = 1,\ A_{12} = A_{22} = \epsilon,\ A_{21} = 1 - 1/(2\epsilon),\ B_1 = 1,\ B_2 = \epsilon,\ C_1 = 1/2,\ C_2 = 0. \tag{A.3}$$

Thus, the PRK scheme specified by (A.2)–(A.3) is symplectic for each $\epsilon$. The idea now is to take limits as $\epsilon \to 0$; the limit integrator, *if it exists*, will preserve quadratic invariants and, when applied to Hamiltonian problems, the symplectic structure. The difficulty is that from the equation that defines $P_{n,2}$,

$$P_{n,2} = p_n + h_n \left(1 - \frac{1}{2\epsilon}\right) g(Q_{n,1}, P_{n,1}, t_n + h_n/2) + h_n \epsilon\, g(Q_{n,2}, P_{n,2}, t_n),$$

we may expect that, for fixed $q_n$, $p_n$, the stage vector $P_{n,2}$ grows unboundedly as $\epsilon \to 0$ and that, therefore, a limit integrator cannot be defined. However, the stage $P_{n,2}$ only affects $P_{n,1}$ and $p_{n+1}$ through the *small* coefficients $A_{1,2} = B_2 = \epsilon$, and this makes it possible to prove that the limit scheme exists for some particular differential equations. Specifically, we assume in what follows that in the partitioned differential system (2.10) being integrated, $f$ and $g$ have the special form

$$f = f(q, t), \qquad g = L(q, t) + M(q, t)p \tag{A.4}$$

(with $q = x$, $p = \lambda$; this format includes the system (3.1), (3.3) in section 3). When

(A.4) holds, the $q$ integration with coefficients (A.2) converges, as $\epsilon \to 0$, to the integration with the originally given coefficients (A.1). The system for the $p$ stages $P_1$, $P_2$ (the index $n$ is sometimes dropped to shorten the formulas) may be written as

$$P_1 = p_n + h_n(L_1 + M_1 P_1) + h_n(\epsilon L_2 + h_n M_2 m_2),$$
$$m_2 = \frac{\epsilon}{h_n} p_n + \left(\epsilon - \frac{1}{2}\right)(L_1 + M_1 P_1) + \epsilon(\epsilon L_2 + h_n M_2 m_2),$$

where we have scaled $m_2 = (\epsilon/h_n)P_2$ to avoid blow-up and used the abbreviations

$$L_1 = L(Q_1, t_n + h_n/2), \qquad M_1 = M(Q_1, t_n + h_n/2),$$
$$L_2 = L(Q_2, t_n), \qquad M_2 = M(Q_2, t_n).$$

Now take limits as $\epsilon \to 0$ to find

$$P_1 = p_n + h_n(L_1 + M_1 P_1) + h_n^2 M_2 m_2,$$
$$m_2 = -\frac{1}{2}(L_1 + M_1 P_1).$$

Since $B_1 = A_{11}$ and $B_2 = A_{12}$, the end-of-step approximation is given by $p_{n+1} = P_1$.

We write these equations in a way similar to (2.11)–(2.13):

$$\begin{aligned} p_{n+1} &= p_n + h_n \ell_1 + h_n^2 M_2 m_2, \\ \ell_1 &= g(Q_1, P_1, t_n + h_n/2), \\ M_2 &= M(Q_2, t_n), \\ P_1 &= p_n + h_n \ell_1 + h_n^2 M_2 m_2, \\ m_2 &= -\frac{1}{2}\ell_1. \end{aligned} \tag{A.5}$$

The combination of these formulas for $p$ with the scheme (A.1) for $q$ is a *first order* integrator that conserves quadratic invariants as in Theorem 2.4 and, for Hamiltonian problems, preserves the symplectic structure. Of course, the integrator is *not* a PRK method; since $M = \partial_p g$, the formula (A.5) is reminiscent of RK methods that use higher derivatives of the solution [21, section II.13]. (Such high order derivative methods cannot be symplectic for general problems [20].) Note that, while $\ell_1$ is an approximation to the first derivative $(d/dt)p$, the vector $M_2 m_2$ has the dimensions of the second derivative $(d^2/dt^2)p$.

Let us now turn to the general case. Assume that in (2.1) the first $r$ weights $b_1$, $\ldots$, $b_r$ do not vanish, while $b_{r+1} = \cdots = b_s = 0$. The regularization procedure used for Runge's method leads to the following fancy integrator:

$$p_{n+1} = p_n + h_n \sum_{i=1}^{r} b_i \ell_i + h_n^2 \sum_{\alpha=r+1}^{s} M_\alpha m_\alpha, \tag{A.6}$$

$$P_i = p_n + h_n \sum_{j=1}^{r} \left(b_j - \frac{b_j a_{ji}}{b_i}\right) \ell_j + h_n^2 \sum_{\beta=r+1}^{s} \left(1 - \frac{b_j a_{\beta i}}{b_i}\right) M_\beta m_\beta, \qquad i = 1, \ldots, r, \tag{A.7}$$

$$m_\alpha = -\sum_{j=1}^{r} b_j a_{j\alpha} \ell_j - h_n \sum_{\beta=r+1}^{s} a_{\beta\alpha} M_\beta m_\beta, \qquad \alpha = r+1, \ldots, s. \tag{A.8}$$

Here the $r$ vectors $\ell_i$ are as in (2.12), so that the method uses $r$ slopes and, additionally,

$s - r$ matrices $M_\alpha = M(Q_\alpha, t_n + c_\alpha h_n)$. From the relations (A.8) the $m_\alpha$ may be viewed as functions of the $\ell_i$.

The following result is a consequence of the construction via regularization.

THEOREM A.1. *Consider partitioned systems of the special format* (A.4)*, where the $q$ equations are integrated with the RK scheme* (2.1)*, $b_1 \neq 0, \ldots, b_r \neq 0$, $b_{r+1} = \cdots = b_s = 0$, and the $p$ equations with the formulas in* (A.6)–(A.8)*. If $S(q(t), p(t))$ is a conserved quantity as in Theorem* 2.4*, then $S(q_n, p_n)$ is independent of $n$. If the system is Hamiltonian, then the map $(q_n, p_n) \mapsto (q_{n+1}, p_{n+1})$ is symplectic.*

With the terminology of section 7, for systems of the special form (A.4), the scheme (A.6) may be viewed as the reflected and transposed scheme of (2.1) when it possesses one or more zero weights.

Proofs of Theorem A.1 that do not rely on taking limits as $\epsilon \to 0$ are of course possible. For such an alternative proof of the conservation of $S$, we may note that manipulations (not reproduced here) similar to those used to prove Lemma 2.5 show that for the present method, in lieu of (2.18), we may write

$$S(q_{n+1}, p_{n+1}) - S(q_n, p_n) = h_n \sum_{i=1}^{r} b_i \big(S(k_i, P_i) + S(Q_i, \ell_i)\big) + h_n^2 \sum_{\alpha=r+1}^{s} \big(S(k_\alpha, m_\alpha) + S(Q_\alpha, M_\alpha m_\alpha)\big).$$

This is an algebraic identity that does not require the system integrated to be conservative. When $S$ is conserved, the first sum vanishes as in the proof of Theorem 2.4. For the second sum note that from $S(f(q,t),p) + S(q, L(q,t) + M(q,t)p) \equiv 0$ it follows that $S(f,p) + S(q, Mp) \equiv 0$.

For the adjoint equations in section 3, the conclusion of Theorem 3.4 holds if the $x$ equations are integrated with a (nonsymplectic) RK method with one or more vanishing weights and the $\lambda$ equations are integrated as in (A.6)–(A.8). Similarly, Theorem 3.6 holds for a suitable choice of the Lagrangian (details will not be given, but see below).

What is the situation for the control problem in section 4? Recall that the corresponding system of *differential equations* is given by (4.1), (4.3), where, in the right-hand sides, $u$ has been expressed as $u = \Phi(x, \lambda, t)$. That system of differential equations does *not* possess the format (A.4) for which (A.6) makes sense and, accordingly, we cannot provide analogues to Theorems 4.2 and 4.3.

In order to gain additional insight, let us use the direct approach based on Runge's second order integrator (A.1). We define the Lagrangian (compare with (3.26) and note consistency with (3.6) due to the factor $h_n^2$)

$$\begin{aligned} \mathcal{C}(x_N) - \lambda_0^{\mathsf{T}}(x_0 - \alpha) &- \sum_{n=0}^{N-1} h_n \lambda_{n+1}^{\mathsf{T}} \Big[\frac{1}{h_n}(x_{n+1} - x_n) - k_{n,1}\Big] \\ &- \sum_{n=0}^{N-1} h_n \Lambda_n^{\mathsf{T}} \Big[k_{n,1} - f(X_{n,1}, U_{n,1}, t_n + h_n/2)\Big] \\ &- \sum_{n=0}^{N-1} h_n^2 \mu_n^{\mathsf{T}} \Big[k_{n,2} - f(X_{n,2}, U_{n,2}, t_n)\Big], \end{aligned}$$

where, as on other occasions, the stages $X_{n,1} = x_n + (h_n/2)k_{n,2}$, $X_{n,2} = x_n$ must be seen as known functions of $x_n$ and $k_{n,2}$. Taking gradients with respect to $x_n$, $k_{n,1}$,

$k_{n,2}$ leads to the necessary conditions

$$\begin{aligned}\lambda_{n+1} &= \lambda_n - (\partial_x f(X_{n,1}, U_{n,1}, t_n + h_n/2))^{\mathsf{T}} \Lambda_n \\ &\qquad - h_n^2 (\partial_x f(X_{n,2}, U_{n,2}, t_n))^{\mathsf{T}} \mu_n, \\ \Lambda_n &= \lambda_{n+1}, \\ \mu_n &= \frac{1}{2} (\partial_x f(X_{n,1}, U_{n,1}, t_n + h_n/2))^{\mathsf{T}} \Lambda_n,\end{aligned}$$

which clearly correspond to the integrator (A.5). (By considering the case where $f$ is independent of $u$, this shows that Theorem 3.6 holds in this case.) However, taking gradients with respect to $U_{n,1}$ and $U_{n,2}$ yields

$$(\partial_u f(X_{n,1}, U_{n,1}, t_n + h_n/2))^{\mathsf{T}} \Lambda_n = 0, \qquad (\partial_u f(X_{n,2}, U_{n,2}, t_n))^{\mathsf{T}} \mu_n = 0.$$

The second equation is totally meaningless. It cannot be seen as a discretization of (4.4) because $\mu_n$ is not an approximation to the costate $\lambda$; it does not even possess the right dimensions for that to happen. The values of $U_{n,2}$ retrieved from this constraint will have no relation to the true optimal controls. The paper [17] nicely illustrates this with an example (see also [9]).

Since the trouble arises from the presence of the controls, things may be fixed by tampering with $U_{n,2}$, as pointed out in [17], [9]. However, there is no shortage of RK schemes with nonzero (or even positive) weights, so that, in practice, resorting to such fixes seems ill-advised.

**Acknowledgments.** I am extremely indebted to E. Zuazua for providing me with the initial motivation for this research. He, J. Frank, A. Murua, S. Ober-Blöbaum, and E. Trélat kindly provided useful references. Additionally, A. Murua has to be thanked for sharing some of his many insights, in particular in connection with automatic differentiation. The extremely careful reading of the manuscript made by one of the referees is also acknowledged with gratitude.

## REFERENCES

[1] L. Abia and J. M. Sanz-Serna, *Partitioned Runge-Kutta methods for separable Hamiltonian problems*, Math. Comp., 60 (1993), pp. 617–634.
[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer, New York, 1989.
[3] P. B. Bochev and C. Scovel, *On quadratic invariants and symplectic structure*, BIT, 34 (1994), pp. 337–345.
[4] J. F. Bonnans and J. Laurent-Varin, *Computation of order conditions for symplectic partitioned Runge–Kutta schemes with application to optimal control*, Numer. Math., 103 (2006), pp. 1-10.
[5] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, 2nd ed., J. Wiley, Chichester, UK, 2008.
[6] D. G. Cacuci, *Sensitivity theory for nonlinear systems.* I. *Nonlinear functional analysis approach*, J. Math. Phys., 22 (1981), pp. 2794–2802.
[7] M. Chyba, E. Hairer, and G. Vilmart, *The role of symplectic methods in optimal control*, Optim. Control Appl. Methods, 30 (2009), pp. 367–382.
[8] G. J. Cooper, *Stability of Runge-Kutta methods for trajectory problems*, IMA J. Numer. Anal., 7 (1987), pp. 1–13.
[9] A. L. Dontchev, W. W. Hager, and V. M. Veliov, *Second-order Runge–Kutta approximations in control constrained optimal control*, SIAM J. Numer. Anal., 38 (2000), pp. 202–226.
[10] S. Ervedoza and E. Zuazua, *Numerical Approximation of Exact Control Waves*, Springer, New York, 2013.
[11] A. Farrés, J. Laskar, S. Blanes, F. Casas, J. Makazaga, and A. Murua, *High precision symplectic integrators for the solar system*, Celestial Mechanics and Dynamical Astronomy, 116 (2013), pp. 141–174.

[12] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, J. Wiley, New York, 1971.
[13] J. Frank and S. Zhuk, *Symplectic Möbius integrators for LQ optimal control problems*, in Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, IEEE, 2014, pp. 6377–6382.
[14] M. B. Giles and N. A. Pierce, *An introduction to the adjoint approach to design*, Flow, Turbulence and Combustion, 65 (2000), pp. 393–415.
[15] A. Griewank, *A mathematical view of automatic differentiation*, Acta Numer., 12 (2003), pp. 321–398.
[16] D. F. Griffiths and J. M. Sanz-Serna, *On the scope of the method of modified equations*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 994–1008.
[17] W. W. Hager, *Runge-Kutta methods in optimal control and the transformed adjoint system*, Numer. Math., 87 (2000), pp. 247–282.
[18] E. Hairer, *Backward analysis of numerical integrators and symplectic methods*, Ann. Numer. Math., 1 (1994), pp. 107–132.
[19] E. Hairer, Ch. Lubich, and G. Wanner, *Geometric Numerical Integration*, 2nd ed., Springer, Berlin, 2006.
[20] E. Hairer, A. Murua, and J. M. Sanz-Serna, *The nonexistence of symplectic multiderivative Runge-Kutta methods*, BIT, 34 (1994), pp. 80–87.
[21] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I: *Non-stiff Problems*, 2nd ed., Springer, Berlin, 1993.
[22] E. Hairer, and G. Wanner, *Solving Ordinary Differential Equations* II: *Stiff and Differential-Algebraic Problems*, 2nd ed., Springer, Berlin, 1996.
[23] S. Lall and M. West, *Discrete variational Hamiltonian mechanics*, J. Phys. A, 39 (2006), pp. 5509–5519.
[24] F. M. Lasagni, *Canonical Runge-Kutta methods*, Z. Angew. Math. Phys., 39 (1988), pp. 952–953.
[25] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics,* Cambridge University Press, Cambridge, UK, 2005.
[26] J. E. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numer., 10 (2001), pp. 357–514.
[27] A. Murua, *On order conditions for partitioned symplectic methods*, SIAM J. Numer. Anal., 34 (1997), pp. 2204–2211.
[28] S. Ober-Blöbaum, O. Junge, and J. E. Marsden, *Discrete mechanics and optimal control: An analysis*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 322–352.
[29] I. M. Ross, *A road for optimal control: The right way to commute*, Ann. New York Acad. Sci., 1065 (2006), pp. 210–231.
[30] A. Sandu, D. N. Daescu, G. R. Carmichael, and T. Chai, *Adjoint sensitivity analysis of regional air quality models*, J. Comput. Phys., 204 (2005), pp. 222–252.
[31] J. M. Sanz-Serna, *Runge-Kutta schemes for Hamiltonian systems*, BIT, 28 (1988), pp. 877–883.
[32] J. M. Sanz-Serna, *Geometric integration*, in The State of the Art in Numerical Analysis, I. S. Duff and G. A. Watson, eds., Clarendon Press, Oxford, 1997, pp. 121–143.
[33] J. M. Sanz-Serna and L. Abia, *Order conditions for canonical Runge–Kutta schemes*, SIAM J. Numer. Anal., 28 (1991), pp. 1081–1096.
[34] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall, London, 1994.
[35] R. R. Scherer and H. Türke, *Reflected and transposed Runge-Kutta methods*, BIT, 23 (1983), pp. 262–266.
[36] T. Schlick, *Molecular Modelling and Simulation: An Interdisciplinary Guide*, 2nd ed., Springer, New York, 2010.
[37] Z. Sirkes and E. Tziperman, *Finite difference of adjoint or adjoint of finite difference?*, Monthly Weather Rev., 49 (1997), pp. 5–40.
[38] E. T. Sontag, *Mathematical Control Theory, Deterministic Finite Dimensional Systems*, 2nd ed., Springer, New York, 1998.
[39] Y. B. Suris, *Preservation of symplectic structure in the numerical solution of Hamiltonian systems*, in Numerical Solution of Differential Equations, S. S. Filippov, ed., Akad. Nauk. SSSR, Inst. Prikl. Mat., Moscow, 1988, pp. 138–144 (in Russian).
[40] Y. B. Suris, *Hamiltonian methods of Runge-Kutta type and their variational interpretation*, Math. Model., 2 (1990), pp. 78–87 (in Russian).
[41] E. Trélat, *Contrôle Optimal: Théorie et Applications*, Vuibert, Paris, 2005.
[42] J. Zabczyk, *Mathematical Control Theory: An Introduction*, Birkhäuser, Boston, 1995.