SIAM J. NUMER. ANAL.
Vol. 28, No. 4, pp. 1081-1096, August 1991


011

# ORDER CONDITIONS FOR CANONICAL RUNGE–KUTTA SCHEMES*

J. M. SANZ-SERNA† AND L. ABIA†

**Abstract.** When numerically integrating Hamiltonian systems of differential equations, it is often advantageous to use canonical methods, i.e., methods that preserve the symplectic structure of the phase space, thus reproducing an important feature of the Hamiltonian flow. An $s$-stage Runge-Kutta (RK) method without redundant stages is canonical if and only if, with a standard notation, $b_i a_{ij} + b_j a_{ji} - b_i b_j = 0$, $1 \leqq i, j \leqq s$. It is shown that for canonical RK methods there are many redundancies in the standard order conditions. For a canonical method to have order $p$ it is sufficient that the $b_i$'s, $a_{ij}$'s satisfy a system of algebraic equations that has, roughly speaking, an equation per *nonrooted* tree of order $\leqq p$. Furthermore, a new methodology is presented for the investigation of the order of canonical integration methods (not necessarily RK methods) when applied to Hamiltonian systems. In the new approach consistency is studied by Taylor expanding a suitable scalar function in terms of so-called canonical elementary differentials.

**Key words.** Runge-Kutta schemes, Hamiltonian systems, trees, order conditions, simplifying assumptions, generating functions for canonical mappings

**AMS(MOS) subject classifications.** primary 65L05; secondary 05C05

## 1. Introduction.

Hamiltonian systems of ordinary differential equations

$$\frac{dp^J}{dt} = -\frac{\partial H}{\partial q^J}, \quad \frac{dq^J}{dt} = \frac{\partial H}{\partial p^J}, \qquad J = 1, \cdots, q, \tag{1.1}$$

arise very frequently in the applications, either in their own right or as the result of the space discretization of Hamiltonian systems of partial differential equations. In (1.1), the integer $g \geqq 1$ is the number of degrees of freedom and $H$ is the Hamiltonian function, a $C^1$ function of the real variables $p^J$ (momenta) and $q^J$ (coordinates). We suppose that $H$ is defined for $(p^1, \cdots, p^g)$ in $R^g$ and $(q^1, \cdots, q^g)$ in an open subset $\Omega$ of $R^g$. The domain $R^g \times \Omega$ is called the phase space of the Hamiltonian system. In (1.1) and later in the paper, capital superscripts refer to components of a vector.

Hamiltonian systems possess many properties (preservation of volume in phase space, Poincaré recurrence, etc.) not shared by more general systems of differential equations. It has become increasingly clear [2] that most of those specific properties follow from the fact that the flow of a Hamiltonian system preserves the symplectic structure of the phase space, i.e., the differential form $\omega = dp_1 \wedge dq_1 + \cdots + dp_g \wedge dq_g$. Mappings that preserve $\omega$ are called *canonical.* When numerically integrating (1.1) with a one-step method of the form $(\mathbf{p}^{*T}, \mathbf{q}^{*T})^T = \mathbf{G}(\mathbf{p}, \mathbf{q}, h; H)$, it may seem desirable to choose the method in such a way that, for each timestep $h$, $\mathbf{G}$ is a canonical mapping. Such a choice guarantees that the dynamics of the numerical solution possesses the specific properties of Hamiltonian flows that have been referred to above [14]. Numerical methods for which the associated mapping $\mathbf{G}$ is canonical for each $h$ and each (smooth) Hamiltonian $H$ are called *canonical or symplectic* [12], [6]. (Here and later, "each $h$" means each $h$ for which $\mathbf{G}$ is defined; arbitrarily long timesteps may, of course, not be feasible with implicit methods.) The practical advantages of using canonical methods have been illustrated in [14] and [4] and will not be discussed in the present paper.

* Received by the editors January 17, 1990; accepted for publication (in revised form) September 10, 1990. This research was supported by "Junta de Castilla y León" (project 1031-89).

† Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain.

Standard numerical schemes, such as explicit Runge-Kutta (RK) methods, are not canonical and a number of special methods have been constructed to achieve canonicity (see the references in [14]). Some of these ad hoc methods require the evaluation of higher derivatives of $H$ and are not very appealing from a practical point of view. Other canonical methods in the literature are only applicable to restricted classes of Hamiltonian systems. These considerations led one of the present authors to the question of whether there are canonical *implicit* RK methods. In [13] it was shown that a sufficient condition for an $s$-stage RK method

$$\begin{pmatrix} A \\ b^T \end{pmatrix} \tag{1.2}$$

to be canonical is

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \qquad 1 \leqq i, j \leqq s. \tag{1.3}$$

This result was discovered later and independently by Lasagni [9] and also by Suris [15], [16]. Furthermore, Lasagni shows that, except for some unimportant methods with redundant stages, condition (1.3) is also necessary for (1.2) to be canonical. In what follows a canonical RK method means an RK method fulfilling (1.3).

The well-known RK schemes based on Gauss-Legendre quadrature are canonical [13]. Semi-implicit methods satisfying (1.3) are easily derived. First note that, in deriving canonical schemes (semi-implicit or not), it may be assumed that all weights are different from zero, because if $b_j = 0$ then (1.3) implies that $b_i a_{ij} = 0$ for all $i$, and hence the method is equivalent to a method with fewer stages. When $b_j \neq 0$, $1 \leqq i \leqq s$, the semi-implicitness requirements $a_{ij} = 0$, $i < j$, and (1.3) lead easily to the following format for the Butcher array:

$$\begin{array}{ccccccc} \beta_1/2 & & & & & \\ \beta_1 & \beta_2/2 & & & & \\ \beta_1 & \beta_2 & \beta_3/2 & & & \\ \vdots & \vdots & \vdots & & & \\ \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_s/2 \\ \hline \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_s \end{array} . \tag{1.4}$$

The family (1.4) has been considered by Suris [16] and Qin and Zhang [11]. In [16] methods are derived with $s = 3$, $p = 3$, while [11] contains a tableau with $s = 4$, $p = 4$. Cooper [5] has discussed the condition (1.3) and constructed a method of the family (1.4) with $s = 3$ and order 3.

When trying to find practical canonical RK methods, either of the general format (1.2)-(1.3) or within a subfamily such as (1.4), the available free parameters would be used, at least in part, to achieve as high an order of consistency as possible. According to standard RK theory [3, Thm. 307B] [7, Thm. 2.13], to obtain order $p \geqq 1$ the coefficients $a_{ij}$, $b_i$ of the tableau (1.2) should satisfy the familiar system of polynomial equations

$$\Phi(\rho\tau) = \frac{1}{\gamma(\rho\tau)} \quad \text{for each rooted } n\text{-tree } \rho\tau, \quad n = 1, \cdots, p, \tag{1.5}$$

where the integer $\gamma(\rho\tau)$ is the density of $\rho\tau$ and $\Phi(\rho\tau)$ denotes the corresponding elementary weight. It is known [3, Thm. 306A] that, if the number of stages and the coefficients $a_{ij}$, $b_i$ are seen as free parameters, each equation in (1.5) is independent of the others. However, for canonical methods, the coefficients $a_{ij}$, $b_i$ are constrained

by the canonicity conditions (1.3) and we may ask whether the equations in (1.5) remain independent. The goal of this paper is to answer such a question. It will be shown that, for canonical methods, there are many redundant equations in (1.5). Roughly speaking, in order that a canonical RK method possess order $p$, it is sufficient to impose as many conditions as the number of (nonrooted) trees with $p$ or less vertices. Thus the canonicity conditions (1.3) operate as simplifying assumptions.

An overview of the organization and the results of the paper is as follows. Section 2 is devoted to some graph theoretical definitions that are necessary to state the main result of the paper. This main result is given in § 3, where we provide conditions for a canonical RK method to have order $p$. Actually, two sets of equivalent conditions are given. In the first set the order equations are homogeneous. In the second set the order equations are simply a subset of (1.5) and hence inhomogeneous. Sections 4 and 5 are devoted to proofs. We emphasize that in §§ 2–5 it is not assumed that the differential system to which the method (1.2) is applied is Hamiltonian; we rather consider a general (smooth) autonomous system

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\mathbf{y}). \tag{1.6}$$

In §§ 6 and 7 we develop an order theory for canonical methods as applied to Hamiltonian systems. In the standard theory RK methods, consistency is investigated by Taylor expanding, with respect to the stepsize $h$, the vector-valued RK map $(\mathbf{p}^{*T}, \mathbf{q}^{*T})^T = \mathbf{G}(\mathbf{p}, \mathbf{q}, h; \mathbf{F})$. In the present approach the function $\Gamma$ to be Taylor expanded is scalar valued. This matches the fact that, while a general differential system (1.6) is specified by a vector field $\mathbf{F}$, a Hamiltonian system is specified by a scalar field $H$. In the standard theory, the expansion involves elementary differentials, i.e., suitable combinations of the partial derivatives of $\mathbf{F}$. Here we find combinations of partial derivatives of $H$ that we call *canonical elementary differentials*. The theory of canonical elementary differentials gives a transparent meaning to the material in § 3: the right-hand sides of the homogeneous order conditions presented there turn out to be the coefficients of the canonical elementary differentials in the Taylor expansion of $\Gamma$.

It should be mentioned that the methodology introduced here for the consistency study or canonical integration methods is not restricted to RK methods. The application to other families of methods will be illustrated in a forthcoming paper [1].

**2. Graph theoretical definitions.** This section is devoted to the presentation of a number of graph theoretical definitions that are needed to formulate the main result of the paper. The material in §§ 2.1–2.3 below is standard and is included here to fix the notation. The material in §§ 2.4–2.5, however, is introduced for the first time here.

**2.1. Preliminaries.** Let $n$ be a positive integer. A *labeled n-graph* is a pair formed by a set $V$ with Card $(V) = n$ and a (possibly empty) set $E$ of unordered pairs $(v, w)$, with $v, w \in V$; $v \neq w$. The elements of $V$ and $E$ are called, respectively, the *vertices* and the *edges* of the labeled graph. Given a labeled $n$-graph, two vertices $v$, $w$ are said to be *adjacent*, if $(v, w) \in E$. A (simple) *path* joining a vertex to a vertex $w$, $(v \neq w)$, is a sequence of pairwise distinct vertices $v = v_0, v_1, \cdots, v_m = w$ with $v_i$ adjacent to $v_{i+1}$, $i = 0, 1, \cdots, m-1$. The integer $m \geqq 1$ is the *length* of the path.

**2.2. Trees.** We employ four kinds of trees, whose definitions are as follows.

(i) A *labeled n-tree* $\lambda\tau$ is a labeled $n$-graph $\{V, E\}$ such that for any pair of distinct vertices $v$ and $w$ there exists a unique path that joins $v$ to $w$.

(ii) Two labeled $n$-trees $\{V_1, E_1\}$, $\{V_2, E_2\}$ are said to be *isomorphic* if a bijection of $V_1$ onto $V_2$ exists that transforms edges in $E_1$ into edges in $E_2$. An *n-tree* $\tau$ is an

equivalence class that comprises a labeled $n$-tree and all labeled $n$-trees isomorphic to it. Each of the labeled $n$-trees that represent $\tau$ is called a *labeling of* $\tau$.

(iii) A *rooted labeled n-tree* $\rho\lambda\tau$ is a labeled $n$-tree in which one of the vertices $r$, called the *root*, has been highlighted. The vertices adjacent to the root are called the *sons* of the root. The sons of the remaining vertices are defined in an obvious recursive way.

(iv) Two rooted labeled $n$-trees $\{V_1, E_1, r_1\}$, $\{V_2, E_2, r_2\}$ are said to be *root-isomorphic* if a bijection of $V_1$ onto $V_2$ exists that transforms edges in $E_1$ into edges in $E_2$ and maps $r_1$ into $r_2$. A *rooted n-tree* $\rho\tau$ is an equivalence class that comprises a rooted labeled $n$-tree and all rooted labeled $n$-trees root-isomorphic to it.

(v) In the remainder of the paper and unless otherwise explicitly stated, *it is assumed that the set of vertices of a labeled n-graph is always* $\{1, 2, \cdots, n\}$.

*Remark* 1. It is standard practice (see, e.g., [3, p. 79]) to use the word tree to mean both the abstract entity defined above as an equivalence class and any of the labeled trees that belong to the class. This practice is particularly natural when trees are handled nonrigorously through pictorial representations rather than through their mathematical definitions. Of course, similar remarks apply to rooted trees. Unfortunately, we shall have to deal simultaneously with different labelings of the same tree and therefore care shall be exercised throughout to distinguish between equivalence classes and individual elements of the classes. To clarify this issue and to illustrate the four kinds of trees, we have considered in Fig. 1 pictorial representations corresponding to the case $n = 3$. There is only one tree, $\tau_{3,1}$, which comprises three different labeled trees denoted by capital Roman numerals. Each labeled tree possesses three rooted labeled trees denoted by lowercase letters. In turn, the rooted labeled trees organize themselves into two rooted trees $\rho\tau_{3,1}$, $\rho\tau_{3,2}$. Finally, the tree $\tau_{3,1}$ can be thought of as the result of the identification of $\rho\tau_{3,1}$ with $\rho\tau_{3,2}$. In general, trees can be considered to be equivalence classes of rooted trees, because a root isomorphism is an isomorphism. Figure 2 depicts all trees and rooted trees for $n$ up to 4.

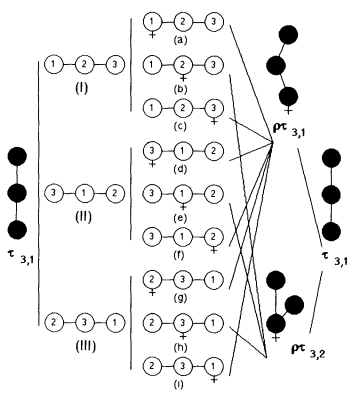


FIG. 1. *The* 3*-tree, labeled* 3*-trees* (I)-(III), *rooted labeled* 3*-trees* (a)-(i), *and rooted* 3*-trees.*

*Remark* 2. The definition of rooted labeled $n$-tree used here differs from that in [7]. The latter only allows so-called *monotonic rooted labelings*, where each vertex is (an integer number $\leqq n$) smaller than all its sons. Thus in Fig. 1, (a) and (e) would be the only rooted labeled 4-trees according to the terminology of [7].

As is standard (see, e.g., [7], [3]), for each rooted tree $\rho\tau$ we denote by $\alpha(\rho\tau)$ the number of its monotonic rooted labeled trees.

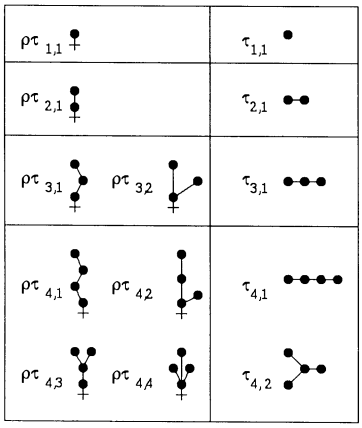


FIG. 2. *Rooted n-trees and n-trees, n = 1, 2, 3, 4.*

**2.3. Rooted trees and RK methods.** As recalled in (1.5), the standard order conditions for RK methods are written in terms of rooted trees. It is perhaps useful to review how to form the elementary weight associated with a given rooted tree. As an illustration, take $\Phi(\rho\tau_{3,2})$. We pick up a rooted labeled representative, say (b). To each vertex $v$ there corresponds a summation index $i_v$ taking values from 1 to the number $s$ of stages; then

$$\Phi(\rho\tau_{3,2}) = \sum_{i_1,i_2,i_3} b_{i_2} a_{i_2 i_1} a_{i_2 i_3}. \tag{2.1}$$

In general, the summation index associated with the root appears as a subscript for the letter b and there are $n-1$ coefficients $a_{ij}$ where $[i,j]$ runs through all [father, son] pairs. Even though a specific rooted labeled tree is used to write down the $\Phi$'s, the result is independent of the chosen representative: a different choice would result in a different notation for the involved summation but not in a different sum. A similar remark applies to any object associated with a tree or rooted tree: a labeled representative must always be used in the construction in such a way that the end result is representative-independent.

**2.4. Superfluous trees.** We now introduce a notion that is required to state the main result of the paper. Let $\tau$ be an $n$-tree and choose one of its labelings $\lambda\tau$. This labeling gives rise to $n$ different rooted labeled trees $\rho\lambda\tau_1, \cdots, \rho\lambda\tau_n$, where $\rho\lambda\tau_i$ has its root at the integer $i$, $1 \leqq i \leqq n$. Suppose that for each edge $(i,j)$ in $\lambda\tau$, it is true that $\rho\lambda\tau_i$ and $\rho\lambda\tau_j$ represent different rooted trees. Then $\tau$ is called *nonsuperfluous*. It is obvious that the definition makes sense, i.e., that the particular labeling $\lambda\tau$ chosen to begin with is immaterial.

As an illustration, consider the only 3-tree $\tau_{3,1}$. When choosing the labeling (I) in Fig. 1, we see that for the edge 1-2 the choice of root 1 leads to $\rho\tau_{3,1}$ and the choice of root 2 leads to $\rho\tau_{3,2}$, while for the edge 2-3 the choice of root 2 leads to $\rho\tau_{3,2}$ and the choice of root 3 leads to $\rho\tau_{3,1}$. Therefore $\tau_{3,1}$ is nonsuperfluous. On the other hand, the 4-tree with labeling 1-2-3-4 is superfluous: changing the root from 2 to the adjacent 3 does not result in a different rooted tree.

Figure 3 depicts all superfluous $n$-trees with $n \leqq 8$. Note that superfluous $n$-trees always have an even number of vertices $n = 2m$, and can be thought of as the result of joining by the roots two copies of the same rooted $m$-tree. The validity of this statement for general $n$ will be established in § 4.

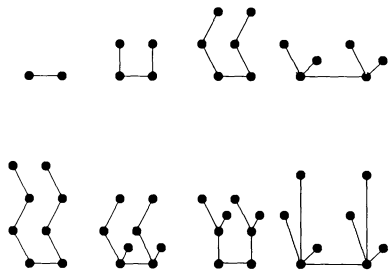

FIG. 3. *Superfluous n-trees,* $n \leqq 8$.

**2.5. Parity.** Let $\rho\tau_1$ and $\rho\tau_2$ be two distinct rooted trees belonging to the same nonsuperfluous $n$-tree $\tau$, $n \geqq 2$. Choose a representative $\rho\lambda\tau_1$ of $\rho\tau_1$ and a representative $\rho\lambda\tau_2$ of $\rho\tau_2$ in such a way that $\rho\lambda\tau_1$ and $\rho\lambda\tau_2$ are the same as labeled graphs and differ only in the location of the root. There is then a well-defined path $P$ that joins the root of $\rho\lambda\tau_1$ to the root of $\rho\lambda\tau_2$. We say that $\rho\tau_1$ and $\rho\tau_2$ are of the same (respectively, different) *parity* if the length $L$ of the path $P$ is an even (respectively, odd) integer. This definition is meaningful because, as shown in § 4, the parity of the integer $L$ does not depend on the specific choices made for $\rho\lambda\tau_1$ and $\rho\lambda\tau_2$. The relation of parity divides each nonsuperfluous $n$-tree with $n \geqq 2$ into two equivalence classes of rooted $n$-trees, called the two *parities* of $\tau$.

An illustration can be made in Fig. 1. The rooted trees $\rho\tau_{3,1}$ and $\rho\tau_{3,2}$ belong to different parities: in the labeling (I) the roots of (a) and (b) are one edge away.

We emphasize that parity only makes sense for nonsuperfluous trees: Consider the superfluous tree $\tau_{4,1}$ in Fig. 2. The rooted trees $\rho\tau_{4,1}$ and $\rho\tau_{4,2}$ would belong to the same parity if the representatives 1-2-3-4 and 1-2-3-4 were chosen (roots two edges away) and different parities with the choices 1-2-3-4 and 1-2-3-4 (roots one edge away).

**3. Order conditions.** We are now in a position to write down the order conditions for canonical methods.

THEOREM 3.1. *Assume that the* RK*-method* (1.2) *is consistent and canonical* (*i.e., satisfies* (1.3)). *Let p be an integer* $p \geqq 2$. *The following conditions are equivalent*:

(i) *The method has order at least p, i.e., produces a local truncation error* $O(h^{p+1})$ *when applied to any smooth system of differential equations* (1.6), *or equivalently, condition* (1.5) *is satisfied.*

(ii) (*Homogeneous form.*) *For each* $n = 2, \cdots, p$,

$$(3.1)\qquad \forall \tau \in \{\textit{nonsuperfluous n-trees}\},\quad \sum_1 \alpha(\rho\tau)\gamma(\rho\tau)\Phi(\rho\tau) - \sum_2 \alpha(\rho\tau)\gamma(\rho\tau)\Phi(\rho\tau) = 0.$$

*In* (3.1) *the first sum is extended to all the rooted trees in a parity of* $\tau$ *and the second sum comprises all the rooted trees in the other parity.*

(iii) (*Inhomogeneous form.*) *For each* $n = 2, \cdots, p$,

$$(3.2)\qquad \forall \tau \in \{\textit{nonsuperfluous n-trees}\},\quad \exists \rho\tau \in \tau \quad \textit{such that } \Phi(\rho\tau) = 1/\gamma(\rho\tau).$$

We postpone the proof of this result until § 5 and turn to some examples of the use of homogeneous and inhomogeneous forms.

For $n = 2$ there is no nonsuperfluous tree. Therefore both (3.1) and (3.2) are satisfied: all consistent canonical methods have order $\geqq 2$.

For $n = 3$, there is one (nonsuperfluous) tree $\tau_{3,1}$. We have seen above that it includes two rooted trees $\rho\tau_{3,1}$ and $\rho\tau_{3,2}$ which belong to different parities. There is

only one monotonic rooted labeled tree for each of $\rho\tau_{3,1}$ and $\rho\tau_{3,2}$, so that $\alpha(\rho\tau_{3,1}) = \alpha(\rho\tau_{3,2}) = 1$ and therefore, with $n = 3$, the homogeneous condition (3.1) reads

$$\gamma(\rho\tau_{3,1})\Phi(\rho\tau_{3,1}) - \gamma(\rho\tau_{3,2})\Phi(\rho\tau_{3,2}) = 0, \tag{3.3}$$

or, in full,

$$6\sum_{ijk} b_i a_{ij} a_{jk} - 3\sum_{ijk} b_i a_{ij} a_{ik} = 0.$$

It is obvious that condition (3.3) is implied by the standard order conditions, that for order 3 demand $\gamma(\rho\tau_{3,1})\Phi(\rho\tau_{3,1}) = 1$ *and* $\gamma(\rho\tau_{3,2})\Phi(\rho\tau_{3,2}) = 1$. If we decide to use the nonhomogeneous form, we should impose, for a canonical method to possess order 3, either $\gamma(\rho\tau_{3,1})\Phi(\rho\tau_{3,1}) = 1$ *or* $\gamma(\rho\tau_{3,2})\Phi(\rho\tau_{3,2}) = 1$.

For $n = 4$ there are two trees $\tau_{4,1}$ and $\tau_{4,2}$. The former is superfluous and does not give rise to order equations. For the latter, there are two rooted trees (belonging to different parities). Then (3.1) becomes

$$\gamma(\rho\tau_{4,3})\Phi(\rho\tau_{4,3}) - \gamma(\rho\tau_{4,4})\Phi(\rho\tau_{4,4}) = 0 \tag{3.4}$$

or

$$12\sum_{ijkl} b_i a_{ij} a_{jk} a_{jl} - 4\sum_{ijkl} b_i a_{ij} a_{ik} a_{il} = 0.$$

Here we have one equation for a method of order $\geqq 3$ to have order $\geqq 4$, instead of the four appearing in the standard theory. When the inhomogeneous form is used, (3.4) should be replaced by either $\gamma(\rho\tau_{4,3})\Phi(\rho\tau_{4,3}) = 1$ or $\gamma(\rho\tau_{4,4})\Phi(\rho\tau_{4,4}) = 1$.

As a final illustration the use of the theorem, we study the homogeneous condition corresponding to the "straight" 5-tree, i.e., o-o-o-o-o. This tree includes three rooted trees +-o-o-o-o, o-o-+-o-o, o-+-o-o-o. The first two belong to the same parity (the roots are two edges away). The values of $\gamma$ are, respectively, 120, 20, and 30; the values of $\alpha$ are 1, 3, and 4 and therefore the condition is

$$\left\{120\sum_{ijklm} b_i a_{ij} a_{jk} a_{kl} a_{lm}\right\} + 3\left\{20\sum_{ijklm} b_i a_{ij} a_{jk} a_{il} a_{lm}\right\} - 4\left\{30\sum_{ijklm} b_i a_{ij} a_{ik} a_{kl} a_{lm}\right\} = 0. \tag{3.5}$$

The inhomogeneous form in Theorem 3.1 is likely to be easier to handle than the homogeneous form. However, the homogeneous form is appealing in that it makes all rooted trees in a given tree play a symmetric role. Either in their homogeneous or in their inhomogeneous form, the new order conditions involve *one equation per nonsuperfluous tree*, rather than an equation per *rooted tree* as is necessary for general methods. This entails an important reduction in the number of equations. It is not difficult to derive a formula for the number of nonsuperfluous $n$-trees. Denote, respectively, by $a_n$, $b_n$, $c_n$, $n = 1, 2, \cdots$, the number of rooted $n$-trees, $n$-trees, and nonsuperfluous $n$-trees and form the generating functions

$$A(z) = a_1 z + a_2 z^2 + \cdots, \quad B(z) = b_1 z + b_2 z^2 + \cdots, \quad C(z) = c_1 z + c_2 z^2 + \cdots.$$

The coefficients of the power series $A$ can be found recursively in various ways (see, e.g., [8, § 2.3.4.4]). Series $B$ can be obtained from $A$, since [8, § 2.3.4.4]

$$B(z) = A(z) - \tfrac{1}{2}[A^2(z) + A(z^2)].$$

We noted above that each superfluous $n$-tree can be seen as a duplication of a rooted $(n/2)$-tree. Thus there are as many superfluous $n$-trees as rooted $(n/2)$-trees, and then

$$C(z) = B(z) - A(z^2),$$

which determines the $c_i$'s once the $a_i$'s are known.

We have used these formulas to compute $a_n$, $b_n$, $c_n$ for $n = 1, 2, \cdots, 10$. The results are presented in Table 1, where the reduction in number of conditions is clearly borne out.

To conclude this section we give a very simple example of use of the new conditions. For the tableau (1.4), for a consistent method to have order 3 the condition (3.3) reads, after some straightforward cancellations,

$$\sum_i \beta_i^3 = 0. \tag{3.6}$$

With two stages, (3.6) is clearly incompatible with the condition for consistency $\sum_i \beta_i = 1$. For three stages we have a one-parameter family of order 3 methods with

$$\beta_1+\beta_2+\beta_3 = 1, \qquad \beta_1^3+\beta_2^3+\beta_3^3 = 0. \tag{3.7}$$

We choose the free parameter so as to have $\beta_1 = \beta_3$. This ensures the symmetry of the method and hence order at least 4 (the order of symmetric methods is even). Subject to symmetry and (3.7) there is a unique method, namely,

$$\beta_1 = \beta_3 = \tfrac{1}{3}(2+\omega+\omega^{-1}), \quad \omega = 2^{1/3}, \quad \beta_2 = 1-2\beta_1. \tag{3.8}$$

Its order is exactly 4; it cannot be 6 since the method has three stages and is not the 3-stage Gauss–Legendre method. Cooper [5] found a method identical to (3.8) except for the ordering of the $\beta_i$. (Cooper's first/second weight is our first/third weight.) Cooper's nonsymmetric method has only order 3. Note that a step of the I-stable, fourth-order method (1.4), (3.8) is essentially a sequence of three steps of the implicit midpoint rule. This makes the method appealing to integrate in time systems of Hamiltonian partial differential equations, e.g., linear wave equations. This issue will be taken up in a future paper. The use of (3.4) reveals that, with three stages, (3.8) is the only choice of coefficients in (1.4) for which order 4 is attained.

**4. Proofs in graph theory.** In this section we deal with some nontrivial statements on superfluous trees and parities that were left unproved in § 2. We also give a graph-theoretic result that will be useful in proving the main theorem.

Recall that the *order* of a labeled graph is the cardinal of the corresponding set of vertices. If $v$ is a vertex of a labeled tree $\lambda\tau$, then the *weight* of $v$ is the maximal order of the subtrees that arise when $v$ is chopped from $\lambda\tau$ (see [8]). A *centroid* of $\lambda\tau$ is a vertex with minimal weight. It can be proved [8] that a labeled tree has at most two centroids. With these preliminaries we can identify all superfluous trees.

THEOREM 4.1. *A superfluous tree possesses an even number of vertices. There is a bijection between the set of rooted m-trees and the set of superfluous* $2m$*-trees,* $m = 1, 2, \cdots$.

TABLE 1

| $N$ | No. rooted trees | No. nonsuperfluous trees |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 0 |
| 3 | 2 | 1 |
| 4 | 4 | 1 |
| 5 | 9 | 3 |
| 6 | 20 | 4 |
| 7 | 48 | 11 |
| 8 | 115 | 19 |
| 9 | 286 | 47 |
| 10 | 719 | 97 |

*Remark.* The proof below shows that, under the bijection of the theorem, the superfluous $2m$-tree that corresponds to a rooted $m$-tree $\rho\tau$ can be conceived of as the result of joining by an edge the roots of two copies of $\rho\tau$ (see Fig. 3).

*Proof.* Let $\tau$ be a superfluous $n$-tree and choose one of its labelings $\lambda\tau$. By definition there are, *at least*, two adjacent vertices $i$ and $j$ in $\lambda\tau$ so that the rooted labeled trees $\rho\lambda\tau_i$, $\rho\lambda\tau_j$ obtained when setting the root at $i$ and $j$, respectively, are root-isomorphic. Denote by $\sigma$ a corresponding isomorphism with $\sigma(i)=j$ and let $V_i$ (respectively, $V_j$) be the set consisting of $i$ (respectively, $j$) and of all vertices that can be joined to $i$ (respectively to $j$) without traveling along the edge $(i,j)$. It is clear from the definition of labeled tree that $\lambda\tau$ contains no edge, other than $(i,j)$, joining a vertex in $V_i$ to an edge in $V_j$. Thus, the deletion of the edge $(i,j)$ defines in an obvious way two rooted labeled trees $\rho\lambda\tau_i$, $\rho\lambda\tau_j$; the former has $V_i$ as a set of vertices and root at $i$ and the latter has $V_j$ as a set of vertices and root at $j$. Actually $\rho\lambda\tau_i$, $\rho\lambda\tau_j$ represent the same rooted tree $\rho\tau_i$, since $\sigma$ maps $i$ into $j$, $V_i$ onto $V_j$, and edges joining vertices in $V_i$ into edges joining vertices in $V_j$. This shows, in particular, that $n=2$ Card $(V_i)$ is even. Also $i$ and $j$ have weight $n/2$ and consequently are the centroids of $\lambda\tau$ so that the unordered pair of vertices $(i,j)$ in the construction above is uniquely defined. It is thus possible to define, in a nonambiguous way, a mapping $T:\tau\to\rho\tau_i$ that associates with each superfluous $n$-tree a rooted $(n/2)$-tree. Clearly, $T$ is a bijection and the result follows. □

We must also show that the notion of parity of a nonsuperfluous tree is well defined. We need two lemmata.

LEMMA 4.2. *Let $\lambda\tau$ be a labeled tree, $v=v_0, v_1, \cdots, v_m=w$ a path in $\lambda\tau$, and $\sigma$ an isomorphism of $\lambda\tau$ onto itself, mapping $v$ into $w$. Then $\sigma(v_i)=v_{m-i}$ for $i=1, 2, \cdots, [m/2]$.*

*Proof.* Clearly it is enough to show that $\sigma(v_1)=v_{m-1}$. Assume that, for a given $\lambda\tau$, there are a path $P$: $v=v_0, v_1, \cdots, v_m=w$ and an isomorphism $\sigma$ with $\sigma(v)=w$, $\sigma(v_1)\neq v_{m-1}$. We shall show that then, in $\lambda\tau$, there are a path $P^*$ of length $2m$ and an isomorphism $\sigma^*$ that maps the first vertex of $P^*$ into the last vertex of $P^*$ but does not map the second vertex of $P^*$ into the penultimate. Therefore $\lambda\tau$ successively contains paths of lengths $m$, $2m$, $4m, \cdots$, which is not possible.

Let us construct $P^*$ and $\sigma^*$. The sets $\{v_0, v_1, \cdots, v_{m-1}\}$, $\{\sigma(v_0), \sigma(v_1), \cdots, \sigma(v_m)\}$ are disjoint. In fact, if they had $v_k=\sigma(v_j)$ in common, then $v_k, v_{k+1}, \cdots, v_m$ and $\sigma(v_j)$, $\sigma(v_{j-1}), \cdots, \sigma(v_0)$ would be two different paths joining $v_k$ to $w$. Then $P^*$: $v_0$, $v_1, \cdots, v_{m-1}$, $\sigma(v_0)$, $\sigma(v_1), \cdots, \sigma(v_{m-1})$, $\sigma(v_m)$, and $\sigma^*=\sigma^2$ serve our purposes, because $\sigma^2(v_1)\neq\sigma(v_{m-1})$. □

LEMMA 4.3. *In the situation of Lemma 4.2, assume that $m$ is odd. Then the rooted labeled trees belonging to $\lambda\tau$ with roots at $v_{[m/2]}$ and $v_{[m/2]+1}$ represent the same rooted tree. As a consequence, the tree represented by $\lambda\tau$ is superfluous.*

*Proof.* By the previous lemma $\sigma(v_{[m/2]})=\sigma(v_{[m/2]+1})$, so that $\sigma$ provides a root isomorphism of the rooted labeled trees. □

The correctness of the definition of parity is a direct corollary of the following result.

THEOREM 4.4. *Let $\rho\lambda\tau_v$, $\rho\lambda\tau_w$, $\rho\lambda\tau_u$ be rooted labeled trees belonging to the same labeled tree $\lambda\tau$. Assume that $\rho\lambda\tau_v$ and $\rho\lambda\tau_w$ are root isomorphic and that the tree defined by $\lambda\tau$ is nonsuperfluous. Then, the lengths of the path joining the root of $\rho\lambda\tau_v$ to the root of $\rho\lambda\tau_u$ possess the same parity as the length of the path joining the root of $\rho\lambda\tau_w$ to the root of $\rho\lambda\tau_u$.*

*Proof.* Denote by $v$, $w$, and $u$ the corresponding roots, $v\neq w$. Let $v=v_0$, $v_1, \cdots, v_{m-1}$, $v_m=u$ and $w=w_0$, $w_1, \cdots, w_{k-1}$, $w_k=u$ be the paths. Let $j$ be the largest integer number so that $j\leqq\min(m,k)$ and $v_m=w_k, v_{m-1}=w_{k-1}, \cdots, v_{m-j}=w_{k-j}$.

Then $\{v_0, v_1, \cdots, v_{m-j-1}\}$ and $\{w_0, w_1, \cdots, w_{k-j-1}, w_{k-j}\}$ are disjoint (proof by contradiction as in Lemma 4.2). Therefore, $v_0, v_1, \cdots, v_{m-j-1}, w_{k-j}, w_{k-j-1}, \cdots, w_0$ is a path joining $v$ to $w$, whose length is $m+k-2j$. If $m+k$ were odd, Lemma 4.3 would imply that we were dealing with a superfluous tree. □

The next result will be crucial in the proof of the equivalence between the homogeneous and inhomogeneous form of the order conditions in Theorem 3.1.

THEOREM 4.5. *With the notation of Theorem* 3.1, *for each nonsuperfluous n-tree* $\tau$, $n \geqq 2$,

$$\sum_1 \alpha(\rho\tau) = \sum_2 \alpha(\rho\tau). \tag{4.1}$$

*Proof.* Clearly each side of (4.1) is the cardinal of the set of monotonic labeled trees belonging to a parity of $\tau$. Denote these sets by $X_1$ and $X_2$. Let $\rho\lambda\tau$ be an element in $X_1$. From this rooted labeled tree we construct a new rooted labeled tree $\rho\lambda\tau^*$ by exchanging the labels 1 and 2. In mathematical terms: (1) The root $\rho\lambda\tau^*$ is at 1. (2) The edges $(i, j)$ with $i$ and $j$ different from 1, 2 in $\rho\lambda\tau^*$ are the same as in $\rho\lambda\tau$. (3) $(1, j)$ is an edge in $\rho\lambda\tau^*$ if and only if $(2, j)$ is an edge in $\rho\lambda\tau$. (4) $(2, j)$ is an edge in $\rho\lambda\tau^*$ if and only if $(1, j)$ is an edge in $\rho\lambda\tau$. It is easy to check that $\rho\lambda\tau^*$ is in fact a rooted monotonically labeled tree. Furthermore, the transformation $T$ that maps $\rho\lambda\tau$ into $\rho\lambda\tau^*$ is one-to-one. Since we are dealing with a nonsuperfluous tree, $\rho\lambda\tau$ and $\rho\lambda\tau^*$ represent rooted trees that belong to different parities, i.e., $\rho\lambda\tau^*$ belongs to $X_2$. Thus a one-to-one mapping from $X_1$ into $X_2$ exists and Card $(X_1) \leqq$ Card $(X_2)$. The result flows after exchanging the roles of $X_1$ and $X_2$. □

**5. Proof of the main result.** Before we prove Theorem 3.1 it is expedient to give a preliminary result in connection with order conditions for RK schemes. Let $\tau$ be an $n$-tree, $n \geqq 2$, and choose one of its labelings $\lambda\tau$. Let $v$ and $w$ be adjacent vertices of $\lambda\tau$ and consider four rooted trees as follows. Denote by $\rho\tau_v$ (respectively, $\rho\tau_w$) the equivalence class of the rooted labeled tree $\rho\lambda\tau_v$ (respectively, $\rho\lambda\tau_w$) obtained by highlighting the vertex $v$ (respectively, $w$) in $\lambda\tau$. Denote by $\rho\tau_V$ and $\rho\tau_W$ the rooted trees that arise when the edge $(v, w)$ is deleted from $\lambda\tau$. A pictorial illustration is given in Fig. 4. The (tedious) mathematical definition of $\rho\tau_V$, $\rho\tau_W$ is not given here, because of its similarity to the construction performed in the proof of Theorem 4.1.

THEOREM 5.1. *With the above notation*:

(i) $$1/\gamma(\rho\tau_v) + 1/\gamma(\rho\tau_w) = (1/\gamma(\rho\tau_V))(1/\gamma(\rho\tau_W)).$$

(ii) *If the* RK *tableau satisfies* (1.3), *then*

$$\Phi(\rho\tau_v) + \Phi(\rho\tau_w) = \Phi(\rho\tau_V)\Phi(\rho\tau_W). \tag{5.1}$$

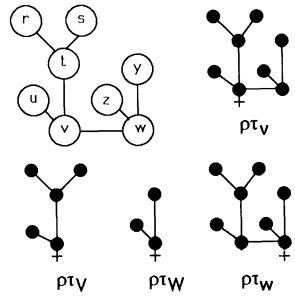


FIG. 4. *The construction for Theorem* 5.1.

*Proof.* The recursive definition of the function $\gamma$ given, e.g., in [3, form. (144a)] or in [7, Chap. II, Definition 2.10] implies that

$$\gamma(\rho\tau_v) = n\gamma(\rho\tau_W)[\gamma(\rho\tau_V)/n(\rho\tau_V)],$$

$$\gamma(\rho\tau_w) = n\gamma(\rho\tau_V)[\gamma(\rho\tau_W)/n(\rho\tau_W)],$$

where $n(\rho\tau_V)$, $n(\rho\tau_W)$ are the orders of $\rho\tau_V$, $\rho\tau_W$, respectively. Part (i) is then a direct consequence of the equality $n = n(\rho\tau_V) + n(\rho\tau_W)$.

Part (ii) is proved as follows. As in (2.1), the left-hand side of (5.1) is

$$\sum_{i_1,i_2,\cdots,i_n} b_{i_v} a_{i_v i_w} \Pi + \sum_{i_1,i_2,\cdots,i_n} b_{i_w} a_{i_w i_v} \Pi, \tag{5.2}$$

where $\Pi$ stands for a product of $n-2$ factors $a_{ij}$. Note that the same product features in both summations, as the pairs [father, son] are the same in $\rho\lambda\tau_v$ and $\rho\lambda\tau_w$ with the exception that $v$ is the father of $w$ in $\rho\lambda\tau_v$ and this relation is reversed in $\rho\lambda\tau_w$. The use of (1.3) in (5.2) yields (5.1). □

The next result needs no proof.

COROLLARY 5.2. *Suppose that the method* (1.2) *is canonical and has order at least* $n-1$, *with* $n \geqq 2$. *Then*

(i) *if* $\rho\tau_v$ *and* $\rho\tau_w$ *are different rooted n-trees, then the standard order condition* $\Phi(\rho\tau_v) = 1/\gamma(\rho\tau_v)$ *holds if and only if* $\Phi(\rho\tau_w) = 1/\gamma(\rho\tau_w)$ *holds.*

(ii) *Assume that* $\rho\tau_v$ *and* $\rho\tau_w$ *are the same rooted tree, which happens if and only if* $\tau$ *is superfluous and* $v$ *and* $w$ *are the two centroids of* $\lambda\tau$. *Then the standard order condition* $\Phi(\rho\tau_v) = 1/\gamma(\rho\tau_v)$ *holds.*

With all the previous preparations, it is possible to give a short proof of Theorem 3.1.

*Proof of Theorem* 3.1. Consider first the inhomogeneous form. It is trivially true that (i) implies (iii). To prove that (iii) is sufficient for (i), use induction in $p$. For $p = 2$ the result is true: the only standard order condition to be considered holds by part (ii) of Corollary 5.2 and consistency. Assume that, for $p-1$, (iii) implies (i). Choose a nonsuperfluous $p$-tree $\tau$. Condition (iii) states that a standard order condition corresponding to a rooted tree $\rho\tau$ in $\tau$ is satisfied. By part (i) of Corollary 5.2, we see that the standard order conditions also hold for the rooted trees in $\tau$ that can be obtained by "moving the root one edge away" in $\rho\tau$. The iteration of this process caters to all rooted trees in $\tau$ since any two vertices can be joined by a path. For a superfluous tree the argument is similar; rather than using (iii) to get a starting point, we use part (ii) of the corollary.

Let us now turn to the homogeneous form. Theorem 4.5 shows that (i) implies (ii). Therefore it is enough to check that (ii) implies (iii). Assume that for a canonical method with order $p-1$, (ii) holds but (iii) does not. Then there is a nonsuperfluous $p$-tree $\tau$ and a rooted $p$-tree $\rho\tau$ belonging to $\tau$ such that $\Phi(\rho\tau) \neq 1/\gamma(\rho\tau)$. If $\gamma(\rho\tau)\Phi(\rho\tau)$ is larger (respectively, smaller) than 1, then, by Theorem 5.1, $\gamma\Phi$ is smaller (respectively, larger) than 1 for each rooted tree that can be obtained by "moving the root one edge away" in $\rho\tau$. These trees belong to the parity of $\tau$ to which $\rho\tau$ does not belong. The argument can be iterated to show that $\gamma\Phi > 1$ for all rooted trees in one parity of $\tau$ and $\gamma\Phi < 1$ for all rooted trees in the other parity of $\tau$. In view of Theorem 4.5, (ii) does not hold, thus contradicting an earlier assumption. □

**6. A canonical order condition theory.** In this section and the next, we concentrate on the integration of (smooth) Hamiltonian systems. It is assumed that $\mathbf{y}^* = \mathbf{G}(\mathbf{y}, h)$ represents a canonical one-step integration method for (1.1), not necessarily an RK

method. Here $\mathbf{y}$ and $\mathbf{y}^*$ denote the vectors with $d = 2g$ components $(\mathbf{p}^T, \mathbf{q}^T)^T$, $(\mathbf{p}^{*T}, \mathbf{q}^{*T})^T$. By definition, $\mathbf{G}(\mathbf{y}, h)$ is, for each fixed $h$, a canonical mapping and, therefore, a well-known result in classical mechanics [2] implies the existence of a (scalar) generating function $S = S(\mathbf{p}, \mathbf{q}^*, h)$ such that the relations

$$p^{*J} = \partial S/\partial q^{*J}, \quad q^J = \partial S/\partial p^J, \qquad J = 1, 2, \cdots, g \tag{6.1}$$

provide $d$ scalar equations that implicitly define $\mathbf{G}$.

Suppose that in the extended phase space of points $(\mathbf{y}^T, t)^T = (\mathbf{p}^T, \mathbf{q}^T, t)^T$ a change of variables is made so as to have $(\mathbf{p}^{*T}, \mathbf{q}^{*T}, t)^T := (\mathbf{G}(\mathbf{y}, -t)^T, t)^T$ as the new coordinate system. Let $(\mathbf{p}(t)^T, \mathbf{q}(t)^T, t)^T$ be an integral curve $C$ of the system (1.1). The theory of generating functions shows that the functions $\mathbf{p}^* = \mathbf{p}^*(t)$, $\mathbf{q}^* = \mathbf{q}^*(t)$ that describe $C$ in the new variables are solutions of the nonautonomous Hamiltonian system with Hamiltonian function

$$H^*(\mathbf{y}^*, t) = H(\mathbf{y}) + \frac{\partial S^-}{\partial t}, \tag{6.2}$$

where $S^-(\mathbf{p}, \mathbf{q}^*, t) = S(\mathbf{p}, \mathbf{q}^*, -t)$. In (6.2) the derivative with respect to $t$ should, of course, be interpreted as derivative with $\mathbf{p}$, $\mathbf{q}^*$ constant; once $\partial S^-/\partial t$ has been found as a function of $\mathbf{p}$, $\mathbf{q}^*$, and $t$, the formulas that define the change of variables must be used to write the right-hand side of (6.2) as a function of $\mathbf{y}^*$ and $t$.

Now, if the method $\mathbf{G}(\mathbf{y}, h)$ were exact for all solutions of (1.1), then, in the new variables, the integral curves would reduce to points, i.e., they would have the form $\mathbf{y}^*(t) = \mathbf{y}^*(0) = \text{constant}$. The substitution of this in the differential system satisfied by $\mathbf{y}^*$ clearly reveals that the corresponding Hamiltonian function $H^*$ would be a constant. This argument can be reversed: if $H^*$ turns out to be a constant, then the method is exact for the solutions of (1.1). In conclusion, the Hamilton–Jacobi equation

$$H(\mathbf{y}) + \frac{\partial S^-}{\partial t} = K = \text{constant}$$

characterizes the generating function of the true solution of the system.

Likewise, order of consistency $p \geqq 1$ is equivalent to the requirement that

$$H^*(\mathbf{y}^*, t) = K + O(t^p), \qquad t \to 0.$$

In other words, $\mathbf{G}$ is consistent if and only if $H^*(\mathbf{y}^*, t) = K + O(t)$, $t \to 0$, and, if $\mathbf{G}$ is consistent, then it has order $p \geqq 2$ for (1.1) if and only if, at $t = 0$,

$$\frac{\partial^{n-1} H^*}{\partial t^{n-1}} = 0, \qquad n = 2, \cdots, p. \tag{6.3}$$

This formula can be written more explicitly as

$$\frac{\partial^{n-1}}{\partial t^{n-1}} \left( H(\mathbf{y}) + \frac{\partial S^-}{\partial t} \right) \Bigg|_{\mathbf{y}^*} = 0, \qquad n = 2, \cdots, p, \quad t = 0, \tag{6.4}$$

where the notation $(\partial^{n-1}/\partial t^{n-1})(\cdot)|_{\mathbf{y}^*}$ means that the differentiation with respect to $t$ is carried out with constant $\mathbf{y}^*$. To sum up, (6.3)–(6.4) give a set of *order conditions* where the function to be differentiated with respect to time is scalar. The idea behind this derivation is essentially due to Ruth [12].

We have found it useful to rewrite (6.3)–(6.4) in an equivalent manner, given in the next theorem.

THEOREM 6.1. *Assume that* $\mathbf{y}^* = \mathbf{G}(\mathbf{y}, h)$ *is a consistent, canonical, one-step method for the integration of a given Hamiltonian system* (1.1) *and let* $S = S(\mathbf{p}, \mathbf{q}^*, h)$ *be the*

*associated generating function. Then* **G** *is consistent of order* $p \geqq 2$ *with* (1.1), *if and only if, at* $h = 0$,

$$\frac{\partial^{n-1}}{\partial h^{n-1}}(\Gamma)\Big|_{\mathbf{y}} = 0, \qquad n = 2, \cdots, p, \tag{6.5}$$

*with*

$$\Gamma = \frac{\partial}{\partial h} S(\mathbf{p}, \mathbf{q}^*, h).$$

*Proof.* For notational simplicity take first the case $n = 2$, for which the chain rule yields

$$\frac{\partial H^*}{\partial t}\bigg|_{\mathbf{y}} = \frac{\partial H^*}{\partial t}\bigg|_{\mathbf{y}^*} + \sum_J \frac{\partial H^*}{\partial y^{*J}}\bigg|_t \frac{\partial y^{*J}}{\partial t}\bigg|_{\mathbf{y}}.$$

If the method is consistent then, at $t = 0$, $H^*(\mathbf{y}^*, 0) = K$ and hence $(\partial H^*/\partial t)\big|_{\mathbf{y}} = 0$ (at $t = 0$) is equivalent to $(\partial H^*/\partial t)\big|_{\mathbf{y}^*} = 0$ (at $t = 0$). A similar argument shows that, for any $n \geqq 2$, $(\partial^{n-1} H^*/\partial t^{n-1})\big|_{\mathbf{y}} = 0$ (at $t = 0$) is equivalent to $(\partial^{n-1} H^*/\partial t^{n-1})\big|_{\mathbf{y}^*} = 0$ (at $t = 0$). Since $S^-(\mathbf{p}, \mathbf{q}^*, t) = S(\mathbf{p}, \mathbf{q}^*, -t)$ the result follows from (6.4) after setting $t = -h$. □

This theorem gives a convenient way of studying the order of canonical methods when applied to Hamiltonian systems. We shall show next that the investigation via (6.5) of the order of a canonical RK scheme results in the homogeneous order conditions (3.1).

**7. Order conditions for canonical RK methods.** In this section we apply Theorem 6.1 to the case of a canonical RK method. Let the RK equations be

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{y} + h \sum_j a_{ij} \mathbf{F}(\mathbf{Y}_j), \qquad 1 \leqq i, \quad j \leqq s, \\ \mathbf{y}^* &= \mathbf{y} + h \sum_i b_i \mathbf{F}(\mathbf{Y}_i), \end{aligned} \tag{7.1}$$

where $\mathbf{y}$ and $\mathbf{y}^*$ denote the vectors $(\mathbf{p}^T, \mathbf{q}^T)^T$, $(\mathbf{p}^{*T}, \mathbf{q}^{*T})^T$ and $\mathbf{F}$ corresponds to the right-hand side of (1.1), i.e.,

$$\mathbf{F} = \Xi \, \mathbf{grad} \, H, \tag{7.2}$$

with $\Xi$ equal to the $2d \times 2d$ skew-symmetric matrix

$$\begin{bmatrix} 0 & -\mathrm{I} \\ \mathrm{I} & 0 \end{bmatrix}.$$

First the generating function can be written in the form [10]

$$\begin{aligned} S(\mathbf{p}, \mathbf{q}^*, h) &= \mathbf{p}^T \mathbf{q}^* - h \sum_i b_i H(\mathbf{Y}_i) - h^2 \sum_{ij} b_i a_{ij} H_{\mathbf{p}}(\mathbf{Y}_i) H_{\mathbf{q}}(\mathbf{Y}_j)^T, \\ H_{\mathbf{p}} &= \left(\frac{\partial H}{\partial p^1}, \cdots, \frac{\partial H}{\partial p^J}\right), \qquad H_{\mathbf{q}} = \left(\frac{\partial H}{\partial q^1}, \cdots, \frac{\partial H}{\partial q^J}\right), \end{aligned} \tag{7.3}$$

where it is assumed that the internal stages $\mathbf{Y}_i$ are functions of $(\mathbf{p}, \mathbf{q}^*, h)$ defined implicitly in (7.1).

Then, the function $\Gamma = (\partial/\partial h)S(\mathbf{p}, \mathbf{q}^*, h)$ to be Taylor expanded is given in the next result.

LEMMA 7.1. *If the* RK *method* (7.1) *is canonical, then*

$$\Gamma = -\sum_i b_i H(\mathbf{Y}_i). \tag{7.4}$$

*Proof.* Differentiate (7.3) with respect to $h$. The result contains derivatives of the stages $\mathbf{Y}_j$ with respect to $h$. These can be expressed, via (7.1), in terms of derivatives of $\mathbf{F}(\mathbf{Y}_j)$ with respect to $h$. The result follows after some manipulation. □

Once $\Gamma$ has been found, standard RK theory (see, e.g., [7, pp. 150–151] can be used to compute the derivatives $(\partial^{n-1}/\partial h^{n-1})(\Gamma)\big|_y$ that feature in the order conditions (6.5). First use Faà di Bruno's formula to express $(\partial^{n-1}/\partial h^{n-1})(\Gamma)\big|_y$ in terms of partial derivatives of $H$ and of derivatives of the internal stages with respect to $h$. Then write the latter derivatives in terms of the corresponding elementary differentials. The result is

$$\frac{\partial^{n-1}\Gamma}{\partial h^{n-1}} = -\frac{1}{n}\sum_{\rho\tau} \alpha(\rho\tau)\gamma(\rho\tau)\Phi(\rho\tau)\theta(\rho\tau), \qquad n \geqq 2, \tag{7.5}$$

where the summation is extended to all rooted $n$-trees $\rho\tau$ and $\theta(\rho\tau)$ is defined as follows. Choose a representative $\rho\lambda\tau$ of $\rho\tau$, and denote by $r$ its root; then

$$\theta(\rho\tau) = \sum_{I_1,\cdots,I_{r-1},I_{r+1},\cdots,I_n} F^{I_1}_{s(1)} \cdots F^{I_{r-1}}_{s(r-1)} H_{s(r)} F^{I_{r+1}}_{s(r+1)} \cdots F^{I_n}_{s(n)}, \tag{7.6}$$

where the indices run from 1 to $d$, subscripts imply differentiation, superscripts denote components, and $s(j)$ refers to the set of all indices $I_k$ such that $k$ is a son of $j$. There is an index $I_k$ for each vertex in $\rho\lambda\tau$ other than the root. We shall say that $\theta(\rho\tau)$ is the *canonical elementary differential* associated with $\rho\tau$.

The right-hand side of (7.5) can be simplified, by using the next lemma, whose proof will be postponed until the end of the section.

LEMMA 7.2. (i) *Canonical elementary differentials associated with the rooted trees of a superfluous tree are zero.*

(ii) $\theta(\rho\tau_1) = \theta(\rho\tau_2)$ *if* $\rho\tau_1$ *and* $\rho\tau_2$ *are rooted trees belonging to the same parity of a nonsuperfluous n-tree,* $n \geqq 2$.

(iii) $\theta(\rho\tau_1) = -\theta(\rho\tau_2)$ *if* $\rho\tau_1$ *and* $\rho\tau_2$ *are rooted trees belonging to different parities of a nonsuperfluous n-tree,* $n \geqq 2$.

As a consequence of the lemma, in the summation in (7.5), the terms corresponding to rooted trees of superfluous trees can be left out and all contributions coming from the same tree can be grouped together. This yields

$$\frac{\partial^{n-1}\Gamma}{\partial h^{n-1}} = -\frac{1}{n}\sum_{\tau}\left\{\sum_1 \alpha(\rho\tau)\gamma(\rho\tau)\Phi(\rho\tau) - \sum_2 \alpha(\rho\tau)\gamma(\rho\tau)\Phi(\rho\tau)\right\}\theta_1(\tau), \qquad n \geqq 2, \tag{7.7}$$

where the summation is extended to all nonsuperfluous $n$-trees, $\sum_1$ and $\sum_2$ are as in (3.1), and $\theta_1(\tau)$ represents the common value of any of the canonical elementary differentials of the rooted trees in the parity of $\tau$ associated with $\sum_1$. When (7.7) is taken into (6.5), it is apparent that the homogeneous order conditions in (3.1) are *sufficient* for a consistent, canonical RK method to have order $p$ when applied to any smooth Hamiltonian system. It is possible to show that the values of $\theta_1(\tau)$ corresponding to different nonsuperfluous trees are independent and hence (3.1) is also *necessary* for order $p$ for Hamiltonian systems. The proof of this independence uses separable Hamiltonians and will be given in a forthcoming paper [1].

It is remarkable that in the "simplifying assumption" derivation of § 4 the order conditions (3.1) are found after manipulation of the $\Phi$'s (which depend on the RK tableau but not on the differential equation), while here we have grouped the $\theta$'s (which are independent of the tableau and depend on the differential equation).

To end the paper we prove Lemma 7.2.

*Proof of Lemma* 7.2. It is clear enough to show that if the rooted trees $\rho\tau_v$ and $\rho\tau_w$ are as at the beginning of § 5, then $\theta(\rho\tau_v) = -\theta(\rho\tau_w)$. Let $w, x1, \cdots, xm$ be the sons of $v$ in $\rho\lambda\tau_v$ and let $v, y1, \cdots, yk$ be the sons of $w$ in $\rho\lambda\tau_w$. According to (7.6), $\theta(\rho\tau_v)$ and $\theta(\rho\tau_w)$ are given by $(n-1)$-fold summations. Namely

$$\theta(\rho\tau_v) = \sum\sum_{I_w} H_{I_w, I_{y1}, \cdots, I_{yk}} F^{I_w}_{I_{x1}, \cdots, I_{xm}} \Pi,$$

$$\theta(\rho\tau_w) = \sum\sum_{I_v} H_{I_v, I_{x1}, \cdots, I_{xm}} F^{I_v}_{I_{y1}, \cdots, I_{yk}} \Pi,$$

where, in both formulae, the outer summation is extended to the indices $I_k$ with $k \neq v$, $w$ and $\Pi$ denotes a product of $n-2$ derivatives of components of $\mathbf{F}$ (the *same* product features in both summations, cf. the proof of Theorem 5.1).

Derivatives of $H$ and components of $\mathbf{F}$ are related in (7.2). Namely,

$$F^I = \sum_J \xi_{IJ} H_J,$$

with $\Xi = (\xi_{IJ})$. Hence

$$\theta(\rho\tau_v) = \sum\sum_{I_w}\sum_{I_v} H_{I_w, I_{y1}, \cdots, I_{yk}} \xi_{I_w, I_v} H_{I_v, I_{x1}, \cdots, I_{xm}} \Pi,$$

$$\theta(\rho\tau_w) = \sum\sum_{I_v}\sum_{I_w} H_{I_v, I_{x1}, \cdots, I_{xm}} \xi_{I_v, I_w} H_{I_w, I_{y1}, \cdots, I_{yk}} \Pi,$$

and the skew-symmetry of $\Xi$ reveals that $\theta(\rho\tau_v) = -\theta(\rho\tau_w)$. □

## REFERENCES


[1] L. ABIA AND J. M. SANZ-SERNA, *Order conditions for partitioned Runge-Kutta schemes for the integration of separable Hamiltonian systems*, in preparation.

[2] V. I. ARNOLD, *Mathematical Methods of Classical Mechanics*, Second ed., Springer-Verlag, New York, 1989.

[3] J. C. BUTCHER, *The Numerical Analysis of Ordinary Differential Equations*, John Wiley, Chichester, U.K., 1987.

[4] P. J. CHANNEL AND C. SCOVEL, *Symplectic integration of Hamiltonian systems*, preprint.

[5] G. J. COOPER, *Stability of Runge-Kutta methods for trajectory problems*, IMA J. Numer. Anal., 7 (1987), pp. 1-13.

[6] K. FENG, *Difference schemes for Hamiltonian formalism and symplectic geometry*, J. Comput. Math., 4 (1986), pp. 279-289.

[7] E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations* I, *Nonstiff Problems*, Springer-Verlag, Berlin, 1987.

[8] D. E. KNUTH, *Fundamental Algorithms*, Second ed., Addison-Wesley, Reading, MA, 1973.

[9] F. LASAGNI, *Canonical Runge-Kutta methods*, Z. Angew. Math. Phys., 39 (1988), pp. 952-953.

[10] ———, *private communication to J. M. Sanz-Serna*, 1989.

[11] M.-Z. QIN AND M.-Q. ZHANG, *Symplectic Runge-Kutta schemes for Hamiltonian systems*, preprint.

[12] R. RUTH, *A canonical integration technique*, IEEE Trans. Nucl. Sci., 30 (1983), pp. 2669-2671.

[13] J. M. SANZ-SERNA, *Runge-Kutta schemes for Hamiltonian Systems*, BIT, 28 (1988), pp. 877-883.

[14] ———, *The numerical integration of Hamiltonian systems*, in Proceedings of the Conference on Computational Differential Equations, Imperial College, London, 1989, to appear.

[15] Y. B. SURIS, *On the preservation of the symplectic structure in the course of the numerical integration of Hamiltonian systems*, in Numerical Solution of Ordinary Differential Equations, S. S. Filippov, ed., M. V. Keldysh Institute of Applied Mathematics, USSR Academy of Sciences, Moscow, 1988, pp. 148–160. (In Russian.)

[16] ———, *Canonical transformations generated by methods of Runge–Kutta type for the numerical integration of the system* $x'' = -\partial U/\partial x$, Zh. Vychisl. Mat. i Mat. Fiz., 29 (1989), pp. 202–211. (In Russian.)