Formal series and numerical integrators: some history and some new techniques

Jesús María Sanz-Serna and Ander Murua

Abstract. This paper provides a brief history of B-series and the associated Butcher group and presents the new theory of word series and extended word series. B-series (Hairer and Wanner 1976) are formal series of functions parameterized by rooted trees. They greatly simplify the study of Runge-Kutta schemes and other numerical integrators. We examine the problems that led to the introduction of B-series and survey a number of more recent developments, including applications outside numerical mathematics. Word series (series of functions parameterized by words from an alphabet) provide in some cases a very convenient alternative to B-series. Associated with word series is a group $\mathcal{G}$ of coefficients with a composition rule simpler than the corresponding rule in the Butcher group. From a more mathematical point of view, integrators, like Runge-Kutta schemes, that are affine equivariant are represented by elements of the Butcher group, integrators that are equivariant with respect to arbitrary changes of variables are represented by elements of the word group $\mathcal{G}$..

Mathematics Subject Classification (2000). Primary 65L05; Secondary 34C29, 70H05, 16T05.

Keywords. B-series, word series, extended word series, Lie groups, Lie algebras, Hamiltonian problems, averaging, normal forms, splitting algorithms, processing numerical methods, modified systems, oscillatory problems, resonances.

1. Introduction

This contribution presents a survey of the use of formal series in the analysis of numerical integrators. The first part of the paper (Section 2) reviews a number of developments (the study of the order of consistency of Runge-Kutta methods, Butcher's notions of composition of Runge-Kutta schemes and effective order, etc.) that underlie the introduction by Hairer and Wanner [28] of the concept of *B-series* in 1976. B-series are formal series parameterized by rooted trees. We also show how B-series and the associated Butcher group have gained prominence in view not only of their relevance to the analysis of geometric integrators but also of their applications to several mathematical theories not directly related to numerical analysis. In the second part of the paper (Sections 3–5) we summarize the recent theory of word series and extended word series developed in [37]. These series are parameterized by words from an alphabet, rather than by rooted trees, and may

Proceedings of the ICIAM,
Beijing, China, 2015

be used advantageously to analyze some numerical integrators, including splitting algorithms. *Word series* appeared implicitly [17] and explicitly in [18], [19]. *Extended word series* were introduced in [37]. While series of *differential operators* parameterized by words (Chen-Fliess series) are very common in control theory and dynamical systems (see e.g. [30], [21]) and have also been used in numerical analysis (see, among others, [34], [35], [20]), word series, as defined here, are, like B-series, series of *functions*.

Word series and extended word series are, when applicable, more convenient than B-series. Associated with word series is a group $\mathcal{G}$ of coefficients with a composition rule simpler than the corresponding rule in the Butcher group. From a more mathematical point of view, integrators, like Runge-Kutta schemes, that are affine equivariant are represented by elements of the Butcher group, integrators that are equivariant with respect to arbitrary changes of variables are represented by elements of the word group $\mathcal{G}$.

It should also be mentioned that B-series and word series have been recently applied to reduce efficiently dynamical systems to normal forms, to perform high-order averaging and to find formal conserved quantities [16], [17], [18], [19], [37]. The approach via B- and word series also makes it possible to simplify the derivation of the corresponding error estimates. These developments provide applications outside numerical mathematics of tools originally devised to analyze numerical methods.

2. Series based on rooted trees

This section contains an overview of the use B-series in the analysis of numerical integrators.

2.1. Finding the order of a Runge-Kutta method. A Runge-Kutta (RK) method with s stages is specified by $s+s^2$ real coefficients b_i, a_{ij}, $i,j=1,\dots,s$. When the method is applied with stepsize h to the initial value problem[1]

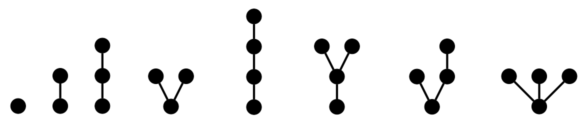
$$\frac{d}{dt}x = f(x), \qquad x(0) = x_0 \in \mathbb{R}^D, \tag{1}$$

and the approximation x_n to the exact solution value $x(nh)$, $n=0,1,\dots$, has been computed, the formulas to find the next approximation x_{n+1} are

$$\begin{aligned} X_{n,i} &= x_n + h\sum_{j=1}^{s} a_{ij} f(X_{n,j}), \qquad 1 \le i \le s,\\ x_{n+1} &= x_n + h\sum_{i=1}^{s} b_i f(X_{n,i}) \end{aligned} \tag{2}$$

[1]Our attention is restricted to deterministic differential equations in Euclidean spaces. We do not consider the extension to stochastic differential equations, differential equations on Lie groups, differential-algebraic equations, etc.

Table 1. Rooted trees with ≤ 4 vertices. For each tree the root is the bottom vertex

$\lvert u \rvert$	1	2	3	3	4	4	4	4
$\sigma(u)$	1	1	1	2	1	2	1	6
$u!$	1	2	6	3	24	12	8	4
								

(the $X_{n,i}$ are the auxiliary internal stages). If f is well behaved and $|h|$ is small, the relations (2) define x_{n+1} as a function of x_n, i.e. $x_{n+1} = \psi_h(x_n)$. The f-dependent mapping ψ_h is an approximation to the solution flow ϕ_h of (1) for which $x((n+1)h) = \phi_h(x(nh))$; for smooth f and an RK method of order ν, $\psi_h(x) - \phi_h(x) = \mathcal{O}(h^{\nu+1})$ as $h \to 0$. To determine the value of ν for a given method it is therefore necessary to write down the expansion of $\psi_h(x)$ in powers of h, a job trickier than one may think: it took mathematicians more than fifty years [6] to come up with an easy, systematic way of carrying it out. After the work of Butcher [3] —following on earlier developments by Gill and Merson— the Taylor series for ψ_h is written, with the help of rooted trees, in the form

$$\psi_h(x) = x + \sum_{n=1}^{\infty} h^n \sum_{u \in \mathcal{T}_n} \frac{1}{\sigma(u)} c_u \, \mathcal{F}_u(x). \tag{3}$$

Here:

- $\mathcal{T}_n$ denotes the set of all rooted trees with n vertices (see Table 1) and, for each rooted tree u, $\sigma(u)$ is the cardinal of the group of symmetries of u.

- $\mathcal{F}_u$ (the elementary differential associated with u) is an $\mathbb{R}^D$-valued function of x that depends on (1) but does not change with the RK coefficients b_i, a_{ij}. Using the rooted tree  as an example,

  $\mathcal{F}$$(x) = \partial_{xx} f(x)\Big[f(x), \partial_x f(x)\big[f(x)\big]\Big],$

  where $\partial_{xx} f(x)[\cdot,\cdot]$ and $\partial_x f(x)[\cdot]$ respectively denote the second and first Frechet derivatives of f evaluated at x. The key point is to observe how the structure of the elementary differential mimics that of the corresponding rooted tree.

- c_u (the elementary weight associated with u) is a a real number that changes with the coefficients b_i, a_{ij} and is independent of the system (1) being integrated. Taking again [tree] as an example,

$$c_{[tree]} = \sum_{i=1}^{s} b_i \Big(\sum_{j=1}^{s} a_{ij} \sum_{k=1}^{s} a_{ik} \big(\sum_{\ell=1}^{s} a_{k\ell} \big) \Big);$$

the structure of the rooted tree is reflected in the summations.

For the solution flow ϕ_h of (1), the expansion is

$$\phi_h(x) = x + \sum_{n=1}^{\infty} h^n \sum_{u \in \mathcal{T}_n} \frac{1}{\sigma(u)} \frac{1}{u!} \mathcal{F}_u(x). \tag{4}$$

where $u!$ (the density of u) is an integer that is easily computed recursively. By comparing (3) and (4) we conclude that an RK method has order $\geq \nu$ if and only if $c_u = 1/u!$ for each $u \in \mathcal{T}_n$, $n \leq \nu$. As an illustration, we note that for order ≥ 3 the coefficients have to satisfy the following set of *order conditions*

$$\sum_i b_i = 1, \quad \sum_{ij} b_i a_{ij} = \frac{1}{2}, \quad \sum_{ijk} b_i a_{ij} a_{jk} = \frac{1}{6}, \quad \sum_{ijk} b_i a_{ij} a_{ik} = \frac{1}{3}.$$

These equations may be shown to be mutually independent [7].

2.2. Composing Runge-Kutta methods. Processing. Butcher [5] developed an algebraic theory of RK methods. If $\psi_h^{(1)}$ and $\psi_h^{(2)}$ represent two RK schemes with s_1 and s_2 stages respectively, the composition of mappings $\psi_h^{(2)} \circ \psi_h^{(1)}$ corresponds to a single step of size $2h$ of a third RK scheme (with s_1+s_2 stages). As distinct from the f-dependent mapping $\psi_h^{(2)} \circ \psi_h^{(1)}$, this new scheme is completely independent of the system (1) under consideration and, accordingly, it is possible to define an operation of composition between the RK schemes themselves. The study of this binary operation is facilitated by considering *classes* of equivalent RK schemes rather than the schemes themselves, where a class consists of a scheme and all others that generate the same ψ_h (note that, for instance, reordering the stages of the method changes the value of the a_{ij} and b_i but does not lead to an essentially different computation, see [7]). Furthermore, if $y = \psi_h(x)$ is an RK map, then x may be recovered from y by taking a step of size $-h$ from y with *another* RK scheme, the so-called adjoint of the original scheme [43]. This should be compared with the situation for the flow, where $y = \phi_h(x)$ implies $x = \phi_{-h}(x)$.

In order to see that these developments are of relevance to practical computation, we discuss briefly the idea of processing also introduced by Butcher [4]. If χ_h is a near-identity mapping in $\mathbb{R}^D$ and ψ_h represents any one-step integrator for (1), the mapping

$$\widehat{\psi}_h = \chi_h^{-1} \circ \psi_h \circ \chi_h$$

defines a *processed* numerical integrator (this is easily interpreted in terms of a change of variables, see [32]). For $m \geq 1$,

$$\widehat{\psi}_h^m = \big(\chi_h^{-1} \circ \psi_h \circ \chi_h\big)^m = \chi_h^{-1} \circ \psi_h^m \circ \chi_h;$$

therefore to advance m steps with the method $\widehat{\psi}_h$ one preprocesses the initial condition to find $\chi_h(x_0)$, advances m steps with the original method and then postprocess the numerical solution by applying χ_h^{-1}. Postprocessing is only performed when output is desired, not at every time step. If both ψ_h and χ_h correspond to RK methods, then so does $\widehat{\psi}_h$ as discussed above. The idea of processing is useful in different scenarios. If χ_h may be chosen in such a way that $\widehat{\psi}_h$ is more accurate than the original ψ_h, one obtains extra accuracy at the (hopefully small) price of having to carry out the processing; ψ_h is said to have *effective order* $\widehat{\nu}$ [8] when $\widehat{\psi}_h$ has order $\widehat{\nu}$ larger than the order ν of ψ_h. As a second possibility, the processed integrator may possess, when ψ_h does not, some of the valuable geometric properties presented below.

2.3. B-series. Butcher series (B-series for short) were introduced in [28] as a means to systematize the derivation and use of expansions like (3) or (4). It is convenient to introduce an empty rooted tree $\emptyset$ with $\sigma(\emptyset) = 1$, $\emptyset! = 1$, $\mathcal{F}_\emptyset(x) = x$ and, for each RK method, $c_\emptyset = 1$. If $\mathcal{T}$ represents the set of all rooted trees (including $\emptyset$), then (3) and (4) become respectively:

$$\psi_h(x) = \sum_{u\in\mathcal{T}} h^{|u|} \frac{1}{\sigma(u)}\, c_u\, \mathcal{F}_u(x), \qquad \phi_h(x) = \sum_{u\in\mathcal{T}} h^{|u|} \frac{1}{\sigma(u)} \frac{1}{u!}\, \mathcal{F}_u(x)$$

($|u|$ is the number of vertices of u; $|\emptyset| = 0$). The right-hand sides of these equalities provide examples of *B-series*: if $\delta \in \mathbb{R}^{\mathcal{T}}$ (i.e. δ is a mapping that associates with each $u \in \mathcal{T}$ a real number δ_u), the corresponding B-series is, by definition, the formal series[2]

$$B_\delta(x) = \sum_{u\in\mathcal{T}} h^{|u|} \frac{1}{\sigma(u)}\, \delta_u\, \mathcal{F}_u(x). \tag{5}$$

Note that B-series are relative to the system (1) being studied because the elementary differentials change with f.

Obviously the set of all B-series is a vector space. A more important algebraic feature is that, if $\gamma \in \mathbb{R}^{\mathcal{T}}$ is a family of coefficients with $\gamma_\emptyset = 1$ and $\delta \in \mathbb{R}^{\mathcal{T}}$, then the *composition* $B_\delta\big(B_\gamma(x)\big)$ is again a B-series $B_\zeta(x)$; furthermore, the elements ζ_u of the family ζ are functions of the elements of δ and γ and do not vary with h or f. For instance for the rooted tree $\dot{\vee}\!:$,

$$\zeta_{\dot{\vee}:} = \delta_\emptyset\, \gamma_{\dot{\vee}:} + \delta_{\bullet}\, \gamma_{\bullet}\, \gamma_{:} + \delta_{:}\, \gamma_{:} + \delta_{:}\, \gamma_{\bullet}\, \gamma_{\bullet} + \delta_{\vdots}\, \gamma_{\bullet} + \delta_{\vee}\, \gamma_{\bullet} + \delta_{\dot{\vee}:}\, \gamma_\emptyset. \tag{6}$$

The rooted tree in the left-hand side has been 'pruned' in all possible ways (including complete uprooting $\dot{\vee}\!: \to \emptyset$ and no pruning $\dot{\vee}\!: \to \dot{\vee}\!:$); in the right-hand side δ is evaluated at the rooted tree that remains after pruning and γ is evaluated at the pieces that have been removed.

As a first example of the use of B-series, we outline the derivation of the RK expansion (3). We begin by assuming that the RK solution x_{n+1} and each internal stage $X_{n,i}$ are expressed as B-series (evaluated at x_n) with undetermined

[2]Rather than using the normalizing factor $1/\sigma(u)$, the original paper [28] uses an alternative factor. The present normalization simplifies many formulas.

coefficients $c \in \mathbb{R}^{\mathcal{T}}$, $C^i \in \mathbb{R}^{\mathcal{T}}$. Those series are then substituted in (2) and the composition rule we have just described is used to find the B-series for $f(X_{n,i})$; this procedure yields equations that determine the elementary weights c_u. The expansion (4) may be derived analogously.

2.4. The Butcher group. If $\gamma_\emptyset = 1$ and $B_\delta\big(B_\gamma(x)\big) = B_\zeta(x)$ as above, we may write $\zeta = \delta \star \gamma$ (recall ζ does not change with f or h). The set of all $\gamma \in \mathbb{R}^{\mathcal{T}}$ with $\gamma_\emptyset = 1$ is a group $\mathcal{G}_B$ for the operation $\star$: the so-called *Butcher group*. The (Taylor expansion of the) true solution flow ϕ_h and (the Taylor expansion of) the RK map ψ_h correspond to the elements of $\mathcal{G}_B$ $\{1/u!\}$ and $\{c_u\}$ respectively; equivalent RK schemes give rise to the same element of $\mathcal{G}_B$. For points in $\mathcal{G}_B$ associated with RK schemes the operation $\star$ obviously corresponds to the composition of RK schemes defined by Butcher.

Assume that the system (1) undergoes an affine change of variables $x = M\overline{x} + c$ and becomes $(d/dt)\overline{x} = \overline{f}(\overline{x})$, with $\overline{f}(\overline{x}) = M^{-1}f(Mx + c)$. If $\gamma \in \mathcal{G}_B$, $B_\gamma(x) = M\overline{B}_\gamma(\overline{x}) + c$, where the notation $\overline{B}$ indicates that the elementary differentials in the series are based on $\overline{f}$; in this way B_γ is affine equivariant. In general,[3] B_γ is not equivariant with respect to more general changes $x = \chi(\overline{x})$.

Taylor series methods, multiderivative RK methods and other one-step integrators $x_{n+1} = \psi_h(x_n)$ not of the form (2) are also represented by elements of $\mathcal{G}_B$. The characterization of the family of integrators that correspond to elements of $\mathcal{G}_B$ has been obtained only recently [33]; note that those integrators necessarily must be affine equivariant.

A forest is a formal juxtaposition of nonempty rooted trees, such as {•, ∶, •, •} or {∨̇, ⁝, ∶}; it is understood that here the order is irrelevant (e.g. {•,, ∶, •, •}, and {•,, •, •, ∶}, are the same forest). Also considered is an empty forest that contains no rooted trees. Thus the set of forests $\mathcal{F}$ is the commutative monoid generated by the set of nonempty rooted trees; by taking all formal linear combinations with real coefficients of forests we obtain a commutative, associative algebra $\mathcal{A}$. The decomposition of each rooted tree that is necessary to compute $B_\delta(B_\gamma(x))$ (for instance from (6) ∨̇ is decomposed as

∅ ⊗ ∨̇ + • ⊗ •∶ + ∶ ⊗ ∶ + ∶ ⊗ •• + ⁝ ⊗ • + ∨ ⊗ • + ∨̇ ⊗ ∅)

defines a coproduct that turns $\mathcal{A}$ into a Hopf algebra: the Connes-Kreimer algebra that appears in renormalization in quantum field theory and elsewhere. Since $\mathcal{F}$ is a basis for the vector space $\mathcal{A}$, each linear form on $\mathcal{A}$ may be identified with a mapping $\in \mathbb{R}^{\mathcal{F}}$. Each element $\gamma \in \mathcal{G}_B \subset \mathbb{R}^{\mathcal{T}}$ may be extended to a mapping $\in \mathbb{R}^{\mathcal{F}}$ by defining γ at a forest as $\prod_u \gamma_u$, where the product is extended to all the trees in the forest. In this way the elements of $\mathcal{G}_B$ may be seen as the characters (multiplicative morphisms) of the Hopf algebra $\mathcal{A}$. The operation $\star$ defined above is the restriction to $\mathcal{G}_B$ of the convolution product in the dual of $\mathcal{A}$. In this way Butcher's work anticipated many results that were to be later used in different subfields of mathematics [2].

[3] Of course the B-series for ϕ_h is equivariant with respect to all changes of variables.

2.5. Modified equations. Since the properties of maps (discrete dynamical systems) are often more difficult to investigate than those of differential equations (continuous dynamical systems), it may make sense, given a one-step integrator ψ_h, to seek a differential system $(d/dt)x = \widetilde{f}_h(x)$ whose h flow $\widetilde{\phi}_h$ coincides with the map ψ_h. While it is well known [38] that it is not possible in general to find such $\widetilde{f}_h$, for typical integrators and smooth f, one may construct a formal series

$$\widetilde{f}_h(x) = \widetilde{f}^{(0)}(x) + h\widetilde{f}^{(1)}(x) + h^2\widetilde{f}^{(2)}(x) + \cdots,$$

whose formal h-flow exactly matches the Taylor expansion of $\psi_h(x)$. For instance, for Euler's rule with $\psi_h(x) = x + hf(x)$, $\widetilde{f}_h(x)$ is found to be

$$f(x) - \frac{h}{2}\partial_x f(x)[f(x)] + \frac{h^2}{3}\partial_x f(x)[\partial_x f(x)[f(x)]] + \frac{h^2}{12}\partial_{xx} f(x)[f(x), f(x)] + \cdots. \quad (7)$$

The fact that $\widetilde{f}_h(x) - f(x) = \mathcal{O}(h)$ indicates that the integrator has order 1; for an integrator of order ν, $\widetilde{f}_h(x) - f(x) = \mathcal{O}(h^\nu)$. When the terms of order h^μ, $\mu = 1, 2, \ldots$, and higher are suppressed from (7), one obtains a vector field $\widetilde{f}_h^{[\mu]}(x)$ whose h-flow differs from ψ_h in $\mathcal{O}(h^{\mu+1})$. Thus Euler's rule applied to (1) is an approximation of order μ to the *modified* differential system $(d/dt)x = \widetilde{f}_h^{[\mu]}(x)$; for μ large and $|h|$ small the modified system may be expected to provide a very accurate description of the behaviour of the numerical solution. The idea of using modified systems is very old, see e.g. the references in [24], and, as we shall see later, has gained prominence with the growing interest in geometric integration.

Note that, for $\widetilde{f}_h(x)$ in (7), $h\widetilde{f}_h(x)$ is a B-series. In fact [25], [10], each $B_\gamma(x)$ with $\gamma \in \mathcal{G}_B$ coincides with the h-flow of a uniquely defined vector field $\widetilde{f}_h(x)$ where $h\widetilde{f}_h(x)$ is a B-series $B_\beta(x)$ with $\beta_\emptyset = 0$. If we denote by $\mathfrak{g}_B$ the set of such β's, the mapping $\gamma \mapsto \beta$ is a bijection from $\mathcal{G}_B$ onto $\mathfrak{g}_B$. Thus B-series integrators may be handled either by using the coefficients $\{\gamma_u\}$ of the B-series for $\psi_h(x)$ or by the coefficients $\{\beta_u\}$ for the corresponding $h\widetilde{f}_h(x)$. The second option is often advantageous because, while $\mathfrak{g}$ is a linear space, $\mathcal{G}_B$ is not; this implies that in many situations properties that are nonlinear when expressed in terms of the γ_u become linear for the corresponding β_u. Formally $\mathcal{G}_B$ is a Lie group and $\mathfrak{g}_B$ is its Lie algebra. The Lie bracket in $\mathfrak{g}_B$ maybe expressed in terms of the convolution product $\star$ in the dual of $\mathcal{A}$ (similar developments for the case of words will be presented in the next section).

2.6. The Hamiltonian case. Geometric integration. In many applications, the system (1) is Hamiltonian, i.e. the dimension D is even and the vector field $f(x)$ is of the form $J^{-1}\nabla H(x)$, where H is the (real-valued) Hamiltonian function and J is the matrix

$$J = \begin{bmatrix} O & I \\ -I & O \end{bmatrix}$$

(the four blocks are of size $D/2 \times D/2$). Hamiltonian systems are characterized by the property that their solution flow ϕ_h is, for each h, a *canonical* or *symplectic* transformation, i.e. (at each x)

$$\big(\partial_x \phi(x)\big)^T J\, \partial_x \phi(x) = J.$$

It is useful, particularly in the context of long-time simulations, to consider one-step integrators $x_{n+1} = \psi_h(x_n)$ that when applied to Hamiltonian system originate a transformation ψ_h that is likewise canonical. These integrators are called symplectic [45], [26], [22] and their importance was first highlighted by Feng Kang. It was proved independently in [31], [40], [46] that the RK method (2) is symplectic if

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \qquad 1 \leq i, j \leq s.$$

In the spirit of the material above, it is sometimes useful to check symplecticness by looking at the B-series of the method rather than by examining the method coefficients. In [11] it was proved that, for Hamiltonian systems (1), the B-series (5) with $\delta_\emptyset = 1$ (not necessarily associated with an RK integrator) is symplectic if and only if, for each pair of nonempty rooted trees u, v,

$$\delta_{u \circ v} + \delta_{v \circ u} = \delta_u \delta_v. \tag{8}$$

Here $u \circ v$ denotes the so-called Butcher product of u and v, i.e. the rooted tree obtained by grafting the root of v into the root of u (e.g. $\bullet \circ$ ⁚ $=$ ⁝, ⁚ $\circ \bullet =$ ∨). In [27] the characterization (8) was used as a stepping stone to establish the nonexistence of symplectic multiderivative RK schemes.

In terms of modified vector fields, symplectic integrators ψ_h may be characterized as methods that when applied to a Hamiltonian problem (1) result in a Hamiltonian $\widetilde{f}_h(x)$. Thus symplectic integrators may be alternatively seen as those integrators that, when applied to a Hamiltonian system, generate a map ψ_h that formally coincides with the flow of *another* Hamiltonian system, hopefully close to the true Hamiltonian. For nonsymplectic schemes ψ_h will coincide with the flow of a vector field perhaps close to f but not within the Hamiltonian class. This interpretation is crucial in symplectic integration [45], [26].

The set of all B-series $h\widetilde{f}_h(x)$ that are Hamiltonian when f is Hamiltonian provides a Lie subalgebra $\mathfrak{g}_0$ of $\mathfrak{g}_B$; the associated Lie subgroup of $\mathcal{G}_B$ corresponds to of all symplectic ψ_h. An element $\beta \in \mathfrak{g}_B$ belongs to $\mathfrak{g}_0$ if and only if, for nonempty u and v,

$$\beta_{u \circ v} + \beta_{v \circ u} = 0.$$

Note that this relation for the Lie algebra is linear, as distinct from the corresponding relation (8) for the group. For $\beta \in \mathfrak{g}_0$ the Hamiltonian function $\widetilde{H}_h(x)$ for the Hamiltonian vector field $\widetilde{f}_h(x)$ is given by a formal series somehow similar to (5) but based on so-called (scalar) elementary Hamiltonians [25] rather than on (vector-valued) elementary differentials. While there is an elementary differential per rooted tree, there are 'fewer' elementary Hamiltonians (one per so-called nonsuperfluous free tree [44]); this explains that for symplectic RK schemes the order conditions corresponding to different rooted trees are not independent.

The study of symplectic integrators for Hamiltonian problems was the first step in what was termed in [41] *geometric integration*: the integration of differential equations by schemes that preserve important geometric features of the system being integrated [26]. In this way, the literature has envisaged volume preserving integrators to integrate divergence free differential equations, integrators that preserve relevant invariants, etc. Formal series have been a key element in these studies, see e.g. [12], [13], [15].

2.7. Extensions. There are several useful extensions of the Runge-Kutta format (2). In some problems the components of the vector x come to us divided into two or more different groups (for instance in mechanical problems we may have positions and velocities). In those circumstances we may use different coefficients b_i, a_{ij} for different groups of components; the result is a *partitioned* RK scheme. An important particular case arises when a second-order system $(d^2/dt^2)y = F(y,(d/dt)y)$ is rewritten as a first-order system for $x = (y,(d/dt)y)$; this is the realm of *Runge-Kutta-Nyström* methods. In other instances the vector field in (1) may be decomposed as a sum of N parts

$$f(x) = f_1(x) + \cdots + f_N(x)$$

and we may resort to additive RK methods [1] based on evaluations of the individual parts f_i rather than on evaluations of f.

The material outlined in previous paragraphs may be adapted to cover all those extensions. A unifying technique has been given in [34]. Let us take the additive case as an example. When Taylor expanding the map ψ_h, we find elementary differentials like $\partial_x f_1(x)[f_2(x)]$ or $\partial_{xx} f_2(x)[f_1(x), f_2(x)]$. The structure of these is captured by *coloured* rooted trees, i.e. rooted trees where each vertex has been marked with one of the symbols (colours) 1, …, N (for the elementary differential $\partial_{xx} f_2(x)[f_1(x), f_2(x)]$, the root of $\because$ is coloured as 2 and the two terminal vertices as 1 and 2). In the definition (5) one has to replace $\mathcal{T}$ by the set of all coloured rooted trees; after that, all the developments above are easily adapted to the additive case [1].

In addition to the composition law for B-series that has been the key element above, a *substitution law* has been introduced [14], [9].

The use of formal series based on rooted trees goes beyond integrators based on the RK idea of repeated evaluations of the vector field. Thus the paper [36] studies order conditions for splitting and composition methods.

To conclude this section we mention the possibility of using the rooted tree machinery to perform efficiently high-order averaging in periodically or quasiperiodically forced dynamical systems [16], [17], [18], [19]. The idea is to expand the solution as a B-series with oscillatory coeffients and show that those coefficients may be interpolated by nonoscillatory functions of t. As a byproduct one may obtain in some circumstances formal conserved quantities. These are other applications to nonnumerical mathematics of the series methodologies developed to analyze numerical integrators.

3. Word series

While the material above dealt with series paratemerized by rooted trees or coloured rooted trees, we now turn our attention to series parameterized by words from an alphabet.

3.1. Definition. In many situations (see [35], [37]) the problem to be integrated is of the form

$$\frac{d}{dt}x = \sum_{a\in A} \lambda_a(t) f_a(x), \qquad x(0) = x_0, \tag{9}$$

where A is a finite or infinite countable set of indices and, for each $a \in A$, λ_a is a scalar-valued function and f_a a D-vector-valued map. The simplest example is furnished by the autonomous system

$$\frac{d}{dt}x = f_a(x) + f_b(x), \tag{10}$$

where $A = \{a, b\}$ and $\lambda_a(t) = \lambda_b(t) = 1$. A more complicated nonautonomous example will appear in the next section.

The solution of (9) has the formal expansion [37]:

$$x(t) = x_0 + \sum_{n=1}^{\infty} \sum_{a_1,\dots,a_n\in A} \alpha_{a_1\cdots a_n}(t) f_{a_1\cdots a_n}(x_0), \tag{11}$$

where the mappings $f_{a_1\cdots a_n}(x)$ and the scalar functions $\alpha_{a_1\cdots a_n}$ are defined by the recursions

$$f_{a_1\cdots a_n}(x) = \partial_x f_{a_2\cdots a_n}(x)\, f_{a_1}(x), \quad n > 1, \tag{12}$$

and

$$\begin{aligned} \alpha_{a_1}(t) &= \int_0^t \lambda_{a_1}(t_1)\, dt_1, \\ \alpha_{a_1\cdots a_n}(t) &= \int_0^t \lambda_{a_n}(t_n)\, \lambda_{a_1\cdots a_{n-1}}(t_n)\, dt_n, \quad n > 1. \end{aligned} \tag{13}$$

For the particular case (10), the inner sum in (11) comprises 2^n terms and each of them has a coefficient $t^n/n!$: the expansion (11) is the Taylor series for $x(t)$ as a function of t written in terms of the pieces f_a and f_b rather than in terms of f. If the flows ϕ_h^a and ϕ_h^b of the split systems $(d/dt)x = f_a(x)$ and $(d/dt)x = f_b(x)$ are available analytically (or may be easily approximated numerically), it is often advantageous to consider integrating (10) by means of *splitting* integrators of the form

$$\psi_h = \phi^b_{d_s h} \circ \phi^a_{c_s h} \circ \cdots \circ \phi^b_{d_1 h} \circ \phi^a_{c_1 h} \tag{14}$$

with c_i, d_i, $1 \le i \le s$, constants that specify the method. The Lie-Trotter $\phi_h^b \circ \phi_h^a$ and Strang $\phi^a_{h/2} \circ \phi^b_h \circ \phi^a_{h/2}$ splittings provide the simplest examples. The Taylor expansion of ψ_h involves the individual pieces f_a and f_b and it then makes sense to write the Taylor expansion of the solution flow in the form (11) we have just considered. Of course splitting integrators that use the flow of the vector fields f_a and f_b are not to be confused with additive RK schemes that just avail themselves of the capability of evaluating the fields f_a, f_b.

The notation in (11) may be made slightly more compact by considering A as an *alphabet* and the strings $a_1 \cdots a_n$ as *words*. Then, if $\mathcal{W}_n$ represents the set of all words with n letters, (11) reads

$$x(t) = x_0 + \sum_{n=1}^{\infty} \sum_{w \in \mathcal{W}_n} \alpha_w(t)\, f_w(x_0).$$

If we furthermore introduce the empty word $\emptyset$ and set $\mathcal{W}_0 = \{\emptyset\}$, $\alpha_\emptyset = 1$, $f_\emptyset(x) = x$, then the last expansion becomes

$$x(t) = \sum_{n=0}^{\infty} \sum_{w \in \mathcal{W}_n} \alpha_w(t)\, f_w(x_0) = \sum_{w \in \mathcal{W}} \alpha_w(t)\, f_w(x_0), \tag{15}$$

where $\mathcal{W}$ represents the set of all words. This suggests the following definition: If δ maps $\mathcal{W}$ into $\mathbb{C}$ (i.e. $\delta \in \mathbb{C}^{\mathcal{W}}$),[4] then its corresponding *word series* is the formal series

$$W_\delta(x) = \sum_{w \in \mathcal{W}} \delta_w f_w(x). \tag{16}$$

The scalars δ_w and the functions f_w will be called the *coefficients* of the series and the *word-basis functions* respectively. Thus, for each fixed t, (15) is the word series with coefficients $\alpha_w(t)$. As we shall see below, in the particular case (10), the mapping ψ_h that represents the splitting method (14) corresponds, for each fixed h, to a word series.

The definition of word series is clearly patterned after that of B-series. Note however that in (16) the coefficients δ_w play the role that in (5) is played by $h^{|u|}\delta_u$. As we have pointed out already, the expansion (11) corresponds to a *family* of word series: one for each fixed value of t. The reason for this small lack of parallelism between the definition of B- and word series is that while (3) or (4) depend on h through powers h^j which may be made to feature in the definition, the time dependence of (15) and related expansions changes with the functions $\lambda_a(t)$.

Each word-basis function f_w, $w \neq \emptyset$, is build up from partial derivatives of the f_a, $a \in A$, e.g., if $a, b, c \in A$,

$$\begin{aligned} f_{ba}(x) &= \partial_x f_a(x)\, f_b(x), \\ f_{cba}(x) &= \partial_x f_{ba}(x) f_c(x) = \partial_{xx} f_a(x)[f_b(x), f_c(x)] + \partial_x f_a(x)\, \partial_x f_b(x)\, f_c(x). \end{aligned}$$

Clearly, $\partial_x f_a(x)\, f_b(x)$, $\partial_{xx} f_a(x)[f_b(x), f_c(x)]$, $\partial_x f_a(x)\, \partial_x f_b(x)\, f_c(x)$ are elementary differentials based on rooted trees coloured by the letters of A. Thus by expanding

[4] While in the case of B-series we only considered real-valued families of coefficients δ, it will be convenient later to consider word series with complex coefficients.

each word-basis function in terms of elementary differentials it would be possible in principle to avoid the introduction of words and work with coloured rooted trees. However such a move would not be always be advisable because word series are more *compact* than B-series (there are 'fewer' word basis functions than elementary differentials) and, additionally, the composition rule for word series is simpler than its counterpart for B-series.

3.2. The word series group $\mathcal{G}$. Given $\delta, \delta' \in \mathbb{C}^{\mathcal{W}}$, we associate with them its *convolution product* $\delta \star \delta' \in \mathbb{C}^{\mathcal{W}}$ defined by

$$(\delta \star \delta')_{a_1 \cdots a_n} = \delta_\emptyset \delta'_{a_1 \cdots a_n} + \sum_{j=1}^{n-1} \delta_{a_1 \cdots a_j} \delta'_{a_{j+1} \cdots a_n} + \delta_{a_1 \cdots a_n} \delta'_\emptyset \tag{17}$$

(here it is understood that $(\delta \star \delta')_\emptyset = \delta_\emptyset \delta'_\emptyset$). The convolution product is not commutative, but it is associative and has a unit (the element $1\!\!1 \in \mathbb{C}^{\mathcal{W}}$ with $1\!\!1_\emptyset = 1$ and $1\!\!1_w = 0$ for $w \neq \emptyset$).

If $w \in \mathcal{W}_m$ and $w' \in \mathcal{W}_n$ are words, $m, n \geq 1$, its *shuffle product* $w ⧢ w'$ [39] is the formal sum of the $(m+n)!/(m!n!)$ words with $m+n$ letters that may be obtained by interleaving the letters of w and w' while preserving the order in which the letters appear in each word. In addition $\emptyset ⧢ w = w ⧢ \emptyset = w$ for each $w \in \mathcal{W}$. The operation $⧢$ is commutative and associative and has the word $\emptyset$ as a unit. We denote by $\mathcal{G}$ the set of those $\gamma \in \mathbb{C}^{\mathcal{W}}$ that satisfy the so-called *shuffle relations:* $\gamma_\emptyset = 1$ and, for each $w, w' \in \mathcal{W}$,

$$\gamma_w \gamma_{w'} = \sum_{j=1}^{N} \gamma_{w_j} \qquad \text{if} \qquad w ⧢ w' = \sum_{j=1}^{N} w_j.$$

The set $\mathcal{G}$ with the operation $\star$ is a (non-commutative) formal Lie group, which plays here the role played by $\mathcal{G}_B$ in the preceding section. For each fixed t, the family of coefficients defined by (13) and $\alpha_\emptyset(t) = 1$ is an element of the group $\mathcal{G}$, [39, Corollary 3.5]. As we shall see below, some numerical integrators for (9) are also associated with families of elements of $\mathcal{G}$ parameterized by the stepsize.

For $\gamma \in \mathcal{G}$, the word series $W_\gamma(x)$ is equivariant with respect to *arbitrary* (not necessarily affine) changes of variables $x = \chi(\overline{x})$ [17, Proposition 3.1]. In fact, if word series are rewritten as B-series (with colours from A), then the family of word series $W_\gamma(x)$ with $\gamma \in \mathcal{G}$ exactly corresponds to the family of B-series that are equivariant with respect to arbitrary changes of variables. By implication only integrators that are equivariant with respect to arbitrary changes of variables are candidates to possess a word-series expansion with coefficients in $\mathcal{G}$. Thus $\mathcal{G}$ may be regarded as a (small) subgroup of $\mathcal{G}_B$.

In analogy with the composition of B-series, for $\gamma \in \mathcal{G}$, $W_\gamma(x)$ may be substituted in an arbitrary word series $W_\delta(x)$, $\delta \in \mathbb{C}^{\mathcal{W}}$, to get a new word series; more precisely

$$W_\delta\big(W_\gamma(x)\big) = W_{\gamma \star \delta}(x), \tag{18}$$

i.e. the coefficients of the word series resulting from the substitution are given by the convolution product $\gamma \star \delta$. It follows immediately that, in the particular case

(10), the map ψ_h defined in (14) corresponds, for each fixed h, to a word series with coefficients in $\mathcal{G}$.

The Lie algebra $\mathfrak{g}$ of the group $\mathcal{G}$ consists of the elements $\beta \in \mathbb{C}^{\mathcal{W}}$ such that $\beta_\emptyset = 0$ and, for each pair of nonempty words w, w',

$$\sum_{j=1}^{N} \beta_{w_j} = 0 \qquad \text{if} \qquad w \sqcup\!\sqcup\, w' = \sum_{j=1}^{N} w_j.$$

The bracket operation in $\mathfrak{g}$ is

$$[\beta, \beta'] = \beta \star \beta' - \beta' \star \beta$$

and corresponds to the Jacobi bracket (commutator) of the associated word series, i.e. for $\beta, \beta' \in \mathfrak{g}$:

$$\big(\partial_x W_{\beta'}(x)\big) W_\beta(x) - \big(\partial_x W_\beta(x)\big) W_{\beta'}(x) = W_{[\beta,\beta']}(x).$$

Since for $\beta \in \mathfrak{g}$, the word series $W_\beta(x)$ belongs to the Lie algebra (for the Jacobi bracket) generated by the f_a, the Dynkin-Specht-Wever formula [29] may be used to rewrite the word series in terms of iterated commutators of these mappings:

$$W_\beta(x) = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{a_1,\dots,a_n \in A} \beta_{a_1 \cdots a_n} [[\cdots[[f_{a_1}, f_{a_2}], f_{a_3}]\cdots], f_{a_n}](x).$$

(For $n = 1$ the terms in the inner sum are of the form $\beta_{a_1} f_{a_1}(x)$.)

In the particular case where each f_a is a Hamiltonian vector field, the iterated commutators are well known to be Hamiltonian and therefore so is $W_\beta(x)$. Furthermore the Hamiltonian function of the vector field $W_\beta(x)$ is

$$\mathcal{H}_\beta(x) = \sum_{w \in \mathcal{W},\, w \neq \emptyset} \beta_w H_w(x),$$

where, for each nonempty word $w = a_1 \cdots a_n$,

$$H_w(x) = \frac{1}{n} \{\{\cdots\{\{H_{a_1}, H_{a_2}\}, H_{a_3}\}\cdots\}, H_{a_n}\}(x). \tag{19}$$

Here $\{\cdot,\cdot\}$ is the Poisson bracket defined by $\{A, B\}(x) = \nabla A(x)^T J^{-1} B(x)$.

We conclude this subsection with an interpretation of the material above in terms of Hopf algebras. The product $\sqcup\!\sqcup$ may be extended in a bilinear way from words to linear combinations of words. After such an extension, the vector space $\mathbb{C}\langle A\rangle$ of all such linear combinations is a unital, commutative, associative algebra, *the shuffle algebra*, denoted by $\mathrm{sh}(A)$ (see [39], [35]). Deconcatenation defines a coproduct and turns $\mathrm{sh}(A)$ into a (commutative, connected, graded) *Hopf algebra* [2]. The sets $\mathcal{G}$ and $\mathfrak{g}$ are then respectively the group of characters and the Lie algebra of infinitesimal characters of the Hopf algebra $\mathrm{sh}(A)$.

4. Extended word series

Extended word series, introduced in [37], are a generalization of word series to cope with perturbed integrable problems and its discretizations.

4.1. Definition.

We now consider systems of the form

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + f(y,\theta), \tag{20}$$

where $y \in \mathbb{R}^{D-d}$, $0 < d \leq D$, $\omega \in \mathbb{R}^d$ is a vector of frequencies $\omega_j > 0$, $j = 1, \dots, d$, and θ comprises d angles, so that $f(y,\theta)$ is 2π-periodic in each component of θ with Fourier expansion

$$f(y,\theta) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \exp(i\mathbf{k}\cdot\theta)\, \hat{f}_{\mathbf{k}}(y)$$

($\hat{f}_{\mathbf{k}}(y)$ and $\hat{f}_{-\mathbf{k}}(y)$ are mutually conjugate, so as to have a real problem). Systems of this form appear in many applications, perhaps after a change of variables. For instance, any system $(d/dt)z = Mz + F(z)$, where M is a skew-symmetric constant matrix, may be brought to the format (20); other examples are discussed in [37]. When $f \equiv 0$ the system is integrable (the angles rotate with uniform angular velocity and y remains constant) and accordingly we refer to problems of the form (20) as perturbed integrable problems and to f as the perturbation (some readers may prefer to substitute ϵf for f).

After introducing the functions

$$f_{\mathbf{k}}(y,\theta) = \exp(i\mathbf{k}\cdot\theta)\, \hat{f}_{\mathbf{k}}(y), \qquad y \in \mathbb{R}^{D-d},\ \theta \in \mathbb{R}^d, \tag{21}$$

we have

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + f(y,\theta) = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + \sum_{\mathbf{k}\in\mathbb{Z}^d} f_{\mathbf{k}}(y,\theta). \tag{22}$$

To find the solution with initial conditions

$$y(0) = y_0, \qquad \theta(0) = \theta_0, \tag{23}$$

we perform the time-dependent change of variables $\theta = \eta + t\omega$ to get

$$\frac{d}{dt}\begin{bmatrix} y \\ \eta \end{bmatrix} = \sum_{\mathbf{k}\in\mathbb{Z}^d} \exp(i\mathbf{k}\cdot\omega t)\, f_{\mathbf{k}}(y,\eta),$$

a particular instance of (9). The formula (11) yields

$$\begin{bmatrix} y(t) \\ \eta(t) \end{bmatrix} = \begin{bmatrix} y(0) \\ \eta(0) \end{bmatrix} + \sum_{n=1}^{\infty} \sum_{\mathbf{k}_1,\dots,\mathbf{k}_n} \alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t)\, f_{\mathbf{k}_1\cdots\mathbf{k}_n}(y(0),\eta(0)),$$

where the coefficients α are still given by (13) (with $\lambda_{\mathbf{k}}(t) = \exp(i\mathbf{k}\cdot\omega t)$) and the word basis functions are defined by (21) and (12) (the Jacobian in (12) is of course taken with respect to the D-dimensional variable (y,θ)). We conclude that, in the original variables, the solution flow of (22), has the formal expansion

$$\phi_t(y_0,\theta_0) = \begin{bmatrix} y(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix} + \begin{bmatrix} 0 \\ t\omega \end{bmatrix} + \sum_{n=1}^{\infty} \sum_{\mathbf{k}_1,\dots,\mathbf{k}_n} \alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(t)\, f_{\mathbf{k}_1\cdots\mathbf{k}_n}(y_0,\theta_0). \quad (24)$$

Note that the word basis functions are *independent of the frequencies* ω and the coefficients α are *independent of* f.

With the notation of the preceding section, we write (24) in the following form (here and later $x=(y,\theta)$):

$$x(t) = \begin{bmatrix} 0 \\ t\omega \end{bmatrix} + W_{\alpha(t)}(x_0).$$

We introduce the vector space $\mathcal{C} = \mathbb{C}^d \times \mathbb{C}^{\mathcal{W}}$ and, if $(v,\delta) \in \mathcal{C}$, define its corresponding *extended word series* to be the formal series

$$\overline{W}_{(v,\delta)}(x) = \begin{bmatrix} 0 \\ v \end{bmatrix} + \sum_{w\in\mathcal{W}} \delta_w f_w(x).$$

Then the solution (24) of (22)–(23) has the expansion

$$x(t) = \overline{W}_{(t\omega,\alpha(t))}(x_0), \qquad (t\omega,\alpha(t)) \in \mathcal{C},$$

with $\alpha(t) \in \mathcal{G} \subset \mathbb{C}^{\mathcal{W}}$ as defined before.

4.2. The extended word series group $\overline{\mathcal{G}}$.

The symbol $\overline{\mathcal{G}}$ denotes the subset of $\mathcal{C}$ comprising the elements (u,γ) with $u \in \mathbb{C}^d$ and $\gamma \in \mathcal{G}$. For each t, the solution coefficients $(t\omega,\alpha(t)) \in \mathcal{C}$ found above provide an example of element of $\overline{\mathcal{G}}$. Some numerical integrators for (20) will be shown below to have, for each value of the stepsize h, an expansion in extended word series with coefficients in $\overline{\mathcal{G}}$.

We introduce linear opertors Ξ_v, ξ_v as follows. If v is a d-vector, Ξ_v is the linear operator in $\mathbb{C}^{\mathcal{W}}$ that maps each $\delta \in \mathbb{C}^{\mathcal{W}}$ into the element of $\mathbb{C}^{\mathcal{W}}$ defined by $(\Xi_v\delta)_\emptyset = \delta_\emptyset$ and

$$(\Xi_v\delta)_w = \exp(i(\mathbf{k}_1+\cdots+\mathbf{k}_n)\cdot v)\,\delta_w,$$

for $w=\mathbf{k}_1\dots\mathbf{k}_n$. For the linear operator ξ_v on $\mathbb{C}^{\mathcal{W}}$, $(\xi_v\delta)_\emptyset = 0$, and for each word $w=\mathbf{k}_1\cdots\mathbf{k}_n$,

$$(\xi_v\delta)_w = i(\mathbf{k}_1+\cdots+\mathbf{k}_n)\cdot v\,\delta_w.$$

With the help of Ξ_u we define a binary operation $\bigstar$: if $(u,\gamma)\in\overline{\mathcal{G}}$ and $(v,\delta)\in\mathcal{C}$, then

$$(u,\gamma)\bigstar(v,\delta) = (\gamma_\emptyset v + \delta_\emptyset u, \gamma \star (\Xi_u\delta)) \in \mathcal{C}.$$

By using (18), it is a simple exercise to check that $\overline{\mathcal{G}}$ acts by *composition* on extended word series as follows:

$$\overline{W}_{(v,\delta)}\big(\overline{W}_{(u,\gamma)}(x)\big) = \overline{W}_{(u,\gamma)\bigstar(v,\delta)}(x), \qquad \gamma\in\mathcal{G}.$$

In fact we have defined the operation $\bigstar$ so as to ensure this property. The set $\overline{\mathcal{G}}$ is a group for the product $\bigstar$ and $\mathbb{C}^d$ and $\mathcal{G}$ are subgroups of $\overline{\mathcal{G}}$.[5] The unit of $\overline{\mathcal{G}}$ is the element $\overline{\mathbb{1}} = (0, \mathbb{1})$.

As a set, the Lie algebra $\overline{\mathfrak{g}}$ of the group $\overline{\mathcal{G}}$ consists of the elements $(v,\delta) \in \mathcal{C}$ with $\delta \in \mathfrak{g}$. For $(v,\delta),(u,\eta) \in \mathbb{C}^d \times \mathfrak{g}$, the Jacobi bracket of the vector fields $\overline{W}_{(v,\delta)}$, $\overline{W}_{(u,\eta)}$ may be shown to be given by [37]

$$[\overline{W}_{(v,\delta)}, \overline{W}_{(u,\eta)}] = \overline{W}_{(0,\xi_v\eta-\xi_u\delta+\delta*\eta-\eta*\delta)};$$

accordingly the bracket of $\overline{\mathfrak{g}}$ has the expression

$$[(v,\delta),(u,\eta))] = (0, \xi_v\eta - \xi_u\delta + \delta*\eta - \eta*\delta).$$

The 0 in the right-hand side reflects the fact that $\mathbb{C}^d$ is an Abelian subgroup of $\overline{\mathcal{G}}$.

Assume that in (22) the dimension D is even with $D/2 - d = m \geq 0$ and that the vector of variables takes the form

$$x = (y,\theta) = (p^1,\dots,p^m; q^1,\dots,q^m; a^1,\dots,a^d; \theta^1,\dots,\theta^d),$$

where p^j is the momentum canonically conjugate to the co-ordinate q^j and a^j is the momentum (action) canonically conjugate to the co-ordinate (angle) θ^j. If each $f_{\mathbf{k}}(x)$ in (21) is a Hamiltonian vector field with Hamiltonian function $H_{\mathbf{k}}(x)$, then the system (22) is itself Hamiltonian for the Hamiltonian function

$$\sum_{j=1}^{d} \omega_j a^j + \sum_{\mathbf{k}\in\mathbb{Z}^d} H_{\mathbf{k}}(x).$$

For each $(\omega,\beta) \in \overline{\mathfrak{g}}$, the extended word series $\overline{W}_{(\omega,\beta)}(x)$ is a Hamiltonian formal vector field, with Hamiltonian function

$$\sum_{j=1}^{d} \omega_j a^j + \sum_{w\in\mathcal{W},\, w\neq\emptyset} \beta_w H_w,$$

with $H_w(x)$ as in (19).

The paper [37] shows how to use the algebraic rules we have just described to reduce (20) to *normal* form where all oscillatory components are removed from the solution by means of a suitable change of variables.

5. Analyzing splitting methods for perturbed integrable problems

Splitting algorithms are natural candidates to integrate (22). Given real coefficients, a_j and b_j, $j = 1,\dots,r$, we study the splitting integrator

$$\psi_h = \phi^{(P)}_{b_r h} \circ \phi^{(U)}_{a_r h} \circ \cdots \circ \phi^{(P)}_{b_1 h} \circ \phi^{(U)}_{a_1 h}, \tag{25}$$

[5] Consider the group homomorphism from the additive group $\mathbb{C}^d$ to the group of automorphisms of $\mathcal{G}$ that maps each $\mu \in \mathbb{C}^d$ into Ξ_μ. Then $\overline{\mathcal{G}}$ is the (outer) semidirect product of $\mathcal{G}$ and the additive group $\mathbb{C}^d$ with respect to this homomorphism.

where $\phi_t^{(U)}$ and $\phi_t^{(P)}$ denote respectively the t-flows of the split systems corresponding to the unperturbed dynamics

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix},$$

and the perturbation

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = f(y,\theta).$$

Since the unperturbed dynamics with frequencies ω_j is reproduced exactly by (25), one would naively hope that the accuracy of the integrator would be dictated for the size of f uniformly in ω. It is well known that such an expectation is unjustified, see e.g. [23], [42], since the oscillatory character of the solution leads to a very complex behaviour of the numerical solution. The algebraic machinery of extended word series has been used in [37] to provide a very detailed description of the performance of the integrators; we shall quote below some of the results in that paper.

5.1. The truncation error. Clearly, the mapping $\phi_t^{(U)}$ has an expansion in extended word series

$$\phi_t^{(U)}(x) = \overline{W}_{(t\omega,\, \mathbb{1}\,)}(x), \qquad (t\omega,\, \mathbb{1}\,) \in \overline{\mathcal{G}};$$

furthermore,

$$\phi_t^{(P)}(x) = \overline{W}_{(0,\tau(t))}(x), \qquad (0,\tau(t)) \in \overline{\mathcal{G}},$$

where $\tau(t) \in \mathcal{G}$ comprises the Taylor coefficients, i.e. $\tau_w(t) = t^n/n!$ if $w \in \mathcal{W}_n$. It follows that ψ_h also possesses an expansion as an extended word series with coefficients in $\overline{\mathcal{G}}$. A simple computation using the definition of $\bigstar$ shows that:

$$\psi_h(x) = \overline{W}_{(ha\omega,\widetilde{\alpha}(h))}(x),$$

where $a = \sum_i a_i$, and $\widetilde{\alpha}(h) \in \mathcal{G}$ is specified by $\widetilde{\alpha}_\emptyset(h) = 1$ and, for $n = 1, 2, \dots,$

$$\widetilde{\alpha}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = h^n \sum_{1\le j_1\le\cdots\le j_n\le r} \frac{b_{j_1}\cdots b_{j_n}}{\sigma_{j_1\cdots j_n}} \exp(i\,(c_{j_1}\mathbf{k}_1 + \cdots + c_{j_n}\mathbf{k}_n)\cdot\omega h).$$

Here,

$$c_j = a_1 + \cdots + a_j, \qquad 1 \le j \le r,$$

and,

$$\sigma_{j_1\cdots j_n} = \frac{1}{n!} \quad \text{if} \quad j_1 = \cdots = j_n,$$
$$\sigma_{j_1\cdots j_n} = \frac{1}{\ell!}\,\sigma_{j_{\ell+1}\cdots j_n} \quad \text{if} \quad \ell < n, \quad j_1 = \cdots = j_\ell < j_{\ell+1} \le \cdots \le j_n.$$

By subtracting the extended word series of the true solution and of the integrator, we obtain an expansion of the local error that may be used to investigate the order of consistency [37]. However such an investigation throws light on the behaviour of the numerical method only when $|h|$ is small with respect to the periods of the oscillations in the problem. A much better description of the numerical solution may be obtained by means of the modified systems that we describe next.

5.2. Modified systems. For $n = 1, 2, \ldots,$ we look for a system

$$\frac{d}{dt}x = \overline{W}_{(\omega,\widetilde{\beta}(h))}(x), \qquad \widetilde{\beta}(h) \in \mathfrak{g},$$

where $\beta_w(h) = 0$ for words with more than n letters, such that, for words with $\leq n$ letters, the extended word series expansion of its flow matches that of the integrator. Such a system may be constructed [37] provided that there is no numerical resonance of order $\leq n$, i.e. there is no set $\mathbf{k}_1, \ldots, \mathbf{k}_r$ with $(\mathbf{k}_1+\cdots+\mathbf{k}_r)\cdot\omega h = 2\pi j$, $r \leq n$, $j \neq 0$. Furthermore, in the Hamiltonian case, the modified systems will also be Hamiltonian.

In the limit where n increases indefinitely one obtains, if there is no numerical resonance of any order, a modified system whose formal h-flow exactly reproduces the extended word series expansion of $\widetilde{\phi}_h$. As detailed in [37] such modified systems provide a very accurate description of the behaviour of the computed solutions. Among other things, it makes it possible to prove $\mathcal{O}(h)$ error bounds for values of h away from first order numerical resonances but not necessarily small with respect to the periods of the fastest oscillations in the problem.

Acknowledgements

A. Murua and J.M. Sanz-Serna have been supported by projects MTM2013-46553-C3-2-P and MTM2013-46553-C3-1-P from Ministerio de Economía y Comercio, Spain. Additionally A. Murua has been partially supported by the Basque Government (Consolidated Research Group IT649-13).

References

[1] Araujo, A. L., Murua, A., Sanz-Serna, J. M., Symplectic methods based on decompositions, *SIAM J. Numer. Anal.* **34** (1997), 1926-1947.

[2] Brouder, Ch., Trees, renormalization and differential equations, *BIT Numerical Mathematics* **44** (2004), 425–438.

[3] Butcher, J. C., Coefficients for the study of Runge-Kutta integration processes, *J. Austral. Math. Soc.* **3** (1963), 185–201.

[4] Butcher, J. C., The effective order of Runge-Kutta methods. In *Conference on the numerical solution of differential equations* (ed. by J. Ll. Morris). Lecture Notes in Math. Vol. 109, Springer, Berlin, 1969, 133–139.

[5] Butcher, J. C., An algebraic theory of integration methods, *Math. Comp.* **19** (1972), 79–106.

[6] Butcher, J. C., A history of Runge-Kutta methods, *Appl. Numer. Math.* **20** (1996), 247–260.

[7] Butcher, J. C., *Numerical Methods for Ordinary Differential Equations, 2nd ed.* Wiley, Chichester, 2008.

[8] Butcher, J. C., Sanz-Serna, J. M., The number of conditions for a Runge-Kutta method to have effective order p, *Appl. Numer. Math.* **22** (1996), 103–111.

[9] Calaque, D., Ebrahimi-Fard, K., Manchon, D., Two interacting Hopf algebras of trees: A Hopf-algebraic approach to composition and substitution of B-series, *Adv. Appl. Math.* **47** (2011), 282–308.

[10] Calvo, M. P., Murua, A., and Sanz-Serna, J. M., Modified equations for ODEs. In Chaotic Numerics (ed. by P. E. Kloeden and K. J. Palmer). Contemporary Mathematics, Vol. 172, American Mathematical Society, Providence, 1944, 63–74.

[11] Calvo, M. P., Sanz-Serna, J. M., Canonical B-series, *Numer. Math.* **67** (1994), 161–175.

[12] Celledoni, E., McLachlan, R. I., Owren, B., and Quispel, G. R. W., Energy-preserving integrators and the structure of B-series, *Found. Comput. Math.* **10** (2010), 673–693.

[13] Chartier, P., Faou, E., Murua, A., An algebraic approach to invariant preserving integrators: The case of quadratic and Hamiltonian invariants, *Numer. Math.*, **103** (2006), 575–590.

[14] Chartier, P., Hairer, E., Vilmart, G., Algebraic structures of B-series, *Found. Comput. Math.***10** (2010) 407-427.

[15] Chartier, P., Murua, A., Preserving first integrals and volume forms of additively split systems, *IMA J. Numer. Anal.* **27** (2007), 381–405.

[16] Chartier, P., Murua, A., Sanz-Serna, J. M., Higher-Order averaging, formal series and numerical integration I: B-series, *Found. Comput. Math.* **10** (2010), 695–727.

[17] Chartier, P., Murua, A., Sanz-Serna, J. M., Higher-Order averaging, formal series and numerical integration II: the quasi-periodic case, *Found. Comput. Math.* **12** (2012), 471–508.

[18] Chartier, P., Murua, A., Sanz-Serna, J. M., A formal series approach to averaging: exponentially small error estimates, *DCDS A* **32** (2012), 3009–3027.

[19] Chartier, P., Murua, A., Sanz-Serna, J. M., Higher-Order averaging, formal series and numerical integration III, *Found. Comput. Math.* (2013) DOI 10.1007/s10208-013-9175-7.

[20] Ebrahimi-Fard, K., Lundervold, A., Malham, S. J. A., Munte-Kaas, H., Wiese, A., Algebraic structure of stochastic expansions and efficient simulation, *Proc. R. Soc. A* **468** (2012), 2361–2382.

[21] Fauvet, F., and F. Menous, M., Ecalle's arborification-coarborification transforms and Connes-Kreimer Hopf algebra, arXiv; 1212.4740v2.

[22] Feng, H., Qin, M., *Symplectic Geometric Algorithms for Hamiltonian Systems.* Springer, Berlin, 2010.

[23] García-Archilla, B., Sanz-Serna, J. M., Skeel, R. D., Long-time-step methods for oscillatory differential equations, *SIAM J. Sci. Comput.* **20** (1998), pp. 930–963.

[24] Griffiths, D. F., Sanz-Serna, J. M., On the scope of the method of modified equations, *SIAM J. Sci. Statist. Comput.* **7** (1986), 994–1008.

[25] Hairer, E., Backward error analysis of numerical integrators and symplectic methods, *Annals Numer. Math.* **1** (1994), 107–132.

[26] Hairer, E., Lubich, Ch., Wanner, G., *Geometric Numerical Integration, 2nd ed.* Springer, Berlin, 2006.

[27] Hairer E., Murua, A., Sanz-Serna, J. M., The nonexistence of symplectic multiderivative Runge-Kutta methods, *BIT* **34** (1994), 80–87.

[28] Hairer E., Wanner, G., On the Butcher group and general multi-value methods, *Computing* **13** (1974), 1–15.

[29] Jacobson, N., *Lie Algebras.* Dover, New York, 1979.

[30] Kawski, M., Sussmann, H. J., Nonommutative power series and formal Lie algebraic techniques in nonlinear control theory. In *in Operators, Systems, and Linear Algebra*, (ed. by U. Helmke, D. Pratzel-Wolters, E. Zerz). Teubner, Stuttgart, 1997, 111–118.

[31] Lasagni, F. M., Canonical Runge-Kutta methods, *ZAMP* **39** (1988), 952-953.

[32] Lopez-Marcos, M. A., Skeel, R. D., Sanz-Serna, J. M., Cheap enhancement of symplectic integrators. In *Numerical Analysis 1995* (ed. by D. F. Griffiths and G. A. Watson). Pitman Research Notes in Mathematics 344, Longman Scientific and Technical, London, 1996, 107–122.

[33] McLachlan, R., Modin, K., Munthe-Kaas, H., Verdier, O., B-series methods are exactly the local, affine equivariant methods, arXiv.1409v1.

[34] Murua, A., Formal Series and Numerical integrators. Part I: Systems of ODEs and symplectic Integrators, *Appl. Numer. Math.* **29** (1999), 221–251.

[35] Murua,A., The Hopf algebra of rooted trees, free Lie algebras and Lie series, *Found. Comput. Math.* **6** (2006), pp. 387–426.

[36] Murua, A., Sanz-Serna, J. M., Order conditions for numerical integrators obtained by composing simpler integrators, *Phil. Trans. R. Soc. Lond. A* **357** (1999), 1079–1100.

[37] Murua, A., Sanz-Serna, J. M., Word series for dynamical systems and their numerical integrators, arXiv:1502.05528.

[38] Neishtadt, A. I., The separation of motions in systems with rapidly rotating phase, *J. Appl. Math. Mech.* **48** (1984), 133–139.

[39] Reutenauer, C., *Free Lie Algebras.* Clarendon Press, Oxford, 1993.

[40] Sanz-Serna, J. M., Runge-Kutta schemes for Hamiltonian systems, *BIT* **28** (1988), 877–883.

[41] Sanz-Serna, J. M., Geometric integration. In *The State of the Art in Numerical Analysis* (ed. by I. S. Duff and G. A. Watson). Clarendon Press, Oxford, 1997, 121–143.

[42] Sanz-Serna, J. M, Mollified impulse methods for highly oscillatory differential equations, *SIAM J. Numer. Anal.*, **46** (2008), 1040-1059.

[43] Sanz-Serna, J. M., Symplectic Runge-Kutta schemes for adjoint equations, automatic differentiation, optimal control and more, arXiv: 1503.04021.

[44] Sanz-Serna, J. M., Abia, L., Order conditions for canonical Runge-Kutta schemes, *SIAM J. Numer. Anal.* **28** (1991), 1081-1096.

[45] Sanz-Serna, J. M., Calvo, M. P., *Numerical Hamiltonian problems.* Chapman and Hall, London, 1994.

[46] Suris, Y. B., Preservation of symplectic structure in the numerical solution of Hamiltonian systems. In *Numerical Solution of Differential Equations* (ed. by S. S. Filippov). Akad. Nauk. SSSR, Inst. Prikl. Mat., Moscow, 1988, 138–144 (in Russian).

Departamento de Matemáticas, Universidad Carlos III de Madrid, E–28911 Leganés (Madrid), Spain.

E-mail: jmsanzserna@gmail.com

Konputazio Zientziak eta A. A. Saila, Informatika Fakultatea, UPV/EHU, E–20018 Donostia–San Sebastián, Spain.

E-mail: Ander.Murua@ehu.es