Word series for dynamical systems and their numerical integrators

A. Murua* and J.M. Sanz-Serna†

November 30, 2015

Abstract

We study word series and extended word series, classes of formal series for the analysis of some dynamical systems and their discretizations. These series are similar to but more compact than B-series. They may be composed among themselves by means of a simple rule. While word series have appeared before in the literature, extended word series are introduced in this paper. We exemplify the use of extended word series by studying the reduction to normal form and averaging of some perturbed integrable problems. We also provide a detailed analysis of the behaviour of splitting numerical methods for those problems.

Keywords and sentences: Word series, extended word series, B-series, words, Hopf algebras, shuffle algebra, Lie groups, Lie algebras, Hamiltonian problems, integrable problems, normal forms, averaging, splitting algorithms, processing numerical methods, modified systems, resonances.

Mathematics Subject Classification (2010) 34C29, 65L05, 70H05, 16T05

Communicated by Christian Lubich

1 Introduction

In this paper we study word series and extended word series, classes of formal series of functions for the analysis of some dynamical systems and their discretizations. We exemplify the use of extended word series by studying the reduction to normal form of some perturbed integrable problems. We also provide a detailed analysis of the behaviour of splitting numerical methods for those problems. Word series are patterned after B-series [23], a commonly used tool in the study of numerical integrators; while B-series are parametrized by rooted trees, word series are parametrized by words built from the letters of an alphabet. Series of *differential operators* parametrized by words

*Konputazio Zientziak eta A. A. Saila, Informatika Fakultatea, UPV/EHU, E–20018 Donostia–San Sebastián, Spain. Email: Ander.Murua@ehu.es

†(Corresponding author) Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, E–28911 Leganés (Madrid), Spain. Email: jmsanzserna@gmail.com

(including the Chen-Fliess series) are very common in control theory [25] and dynamical systems [18] and have also been used in numerical analysis (see, among others, [27], [28], [16]). Word series are mathematically equivalent to series of differential operators, but being series of *functions*, they are handled in a way very similar to the way B-series are used by numerical analyst. Word series, as defined here, have appeared before in the literature, explicitly [14], [15] or implicitly [13]. Extended word series are introduced in this paper.

B-series, introduced by Hairer and Wanner in 1974 [23], provided the first example of the application of formal series of functions to the theory of numerical integrators (see [36] for a historical survey). B-series, particularly adapted to Runge-Kutta and related methods, give a convenient, systematic way of performing the nontrivial algebraic manipulations needed to write the expansion of the local error in powers of the stepsize. In addition, they facilitate the construction of integrators found by composing simpler integrators; such a construction is required e.g. when investigating the effective order of Runge-Kutta methods [7], [8]. The usefulness of B-series stems from the fact that the composition of two B-series is again a B-series whose coefficients may be written down explicitly and are universal in the sense that they are independent of the particular differential system being integrated. B-series and their extensions grew more important within the notion of geometric integration [33]. In 1994 Calvo and one of the present authors [10] showed how the conditions for a Runge-Kutta scheme to be symplectic may be advantageously derived by examining the corresponding B-series. Hairer's article [21] started the use of B-series to find explicitly modified systems. Since those pioneering contributions the role of B-series and its generalizations [27] in geometric integration has kept growing as it may be seen in the treatise [22].

The word series and extended word series considered here are, when applicable, more convenient than B-series. A reason for this convenience is that they are more compact; in fact the coefficients of a B-series are parametrized by (possibly coloured or decorated) rooted trees and there are many more rooted trees with n vertices than words with n letters. A second advantage of word series and extended word series over B-series is that for words the composition rule (see (11) and (13)) is much simpler than for rooted trees.

An overview of the contributions of this paper is as follows.

Section 2 gives a summary of the rules to manipulate word series. A group $\mathcal{G}$ is introduced that plays the role played by the Butcher group in the theory of B-series. The solution of the differential system being integrated and some numerical methods, including splitting algorithms, may be represented by elements of $\mathcal{G}$. We also identify the Lie algebra $\mathfrak{g}$ associated with $\mathcal{G}$ and the corresponding bracket. This material is very much related to the theory of Hopf algebras [28], [6]; however Section 2 has been written with an audience of computational scientist in mind and a number of more algebraic considerations have been postponed to Section 6.

Extended word series are introduced in Section 3 to cope with perturbed integrable problems; roughly speaking we treat problems that may be seen as arbitrary perturbations of systems that, in suitable variables, may be cast in the form $(d/dt)y = 0$, $(d/dt)\theta = \omega$ (in the language of classical mechanics [2] y/θ would correspond to action/angle variables). We describe the relevant group $\overline{\mathcal{G}}$ and algebra $\overline{\mathfrak{g}}$.

In Section 4 we show how to use extended word series to bring perturbed integrable

problems to *normal form*, i.e. how to change variables to reduce the system being analyzed to a form as simple as possible. As distinct from standard ways of finding normal forms, the extended word series approach does not rely on the vector field being polynomial. Furthermore, the computations required here are *universal* (in the sense of [14]): they are independent of the particular system under consideration.

For highly oscillatory problems the reduction to normal form is very much related to the process of *averaging* out the oscillatory components of the solution and therefore Section 4 extends the material in the series of papers [12], [13], [14], [15]. Furthermore normal forms readily lead to the explicit computation of (formal) invariant quantities of dynamical systems and their discretizations, an issue not covered in this paper and treated in the follow-up article [30].

Section 5 is devoted to the study of general splitting algorithms to simulate perturbed integrable problems. We show how our algebraic approach leads to a convenient expansion of the local error. It is well known that the behaviour of the corresponding global error as h varies is unfortunately extremely complex, as it depends on arithmetic relations between h and the periods present in the solution. Extended word series provide a powerful instrument to analyze that behaviour. In fact, two different approaches are put forward here. In the first, the integrator is processed, i.e. subjected to changes of variables, to remove oscillatory components. The second approach is based on constructing a modified system for the integrator and then bringing the modified system to normal form. Of much interest is the fact that the validity of the modified system holds even if h is not small relative to the periods in the problem (cf. the use by Hairer and Lubich of modulated Fourier expansions [22]). The techniques in this section may be readily applied to the construction and analysis of improved integrators, such as those considered in e.g. [19], [34]; this will be the subject of future work.

Section 6 contains proofs and technical material and there is an Appendix devoted to the practical applicability of splitting integrators.

2 Word series

This section presents word series and provides a summary of the rules for their application. The presentation has computational scientist in mind and focuses on essential features; additional details and proofs are given in Section 6.1, where the approach is more algebraic. Until further notice all functions are assumed to be smooth.

2.1 Definition of word series

We consider the D-dimensional initial-value problem given by

$$x(0) = x_0 \tag{1}$$

and

$$\frac{d}{dt}x = \sum_{a \in A} \lambda_a(t) f_a(x), \tag{2}$$

where t is the (real) independent variable, 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.

It is well known that the solutions of (2) may be expanded formally as follows. Associated with each vector field f_a in (2), there is a first-order linear differential operator E_a: if g is a scalar-valued function, then the function $E_a g$ is defined by

$$E_a g(x) = \sum_{j=1}^{D} f_a^j(x) \frac{\partial}{\partial x^j} g(x) \tag{3}$$

(superscripts denote components of vectors). We shall also let E_a act on vector-valued mappings; it is then understood that the operator is applied componentwise. If $x(t)$ satisfies (2), the chain rule yields

$$\frac{d}{dt} g(x(t)) = \sum_{a \in A} \lambda_a(t) (E_a g)(x(t))$$

or

$$g(x(t)) = g(x(0)) + \sum_{a \in A} \int_0^t dt_1 \, \lambda(t_1) (E_a g)(x(t_1)).$$

The same procedure may be now applied with $(E_a g)(x(t_1))$ in lieu of $g(x(t))$ to rewrite the last equation as

$$\begin{aligned} g(x(t)) \quad = \quad & g(x(0)) + \sum_{a \in A} \int_0^t dt_1 \, \lambda(t_1) (E_a g)(x(0)) \\ & + \sum_{a \in A} \sum_{b \in A} \int_0^t dt_1 \, \lambda(t_1) \int_0^{t_1} dt_2 \, \lambda(t_2) (E_b (E_a g))(x(t_2)). \end{aligned}$$

By continuing this Picard iteration and setting g equal to the identity function $g(x) = x$, we find that the solution of (1)–(2) has the formal expansion

$$x(t) = x_0 + \sum_{n=1}^{\infty} \sum_{a_1, \ldots, a_n \in A} \alpha_{a_1 \cdots a_n}(t) f_{a_1 \cdots a_n}(x_0), \tag{4}$$

where the vector-valued mappings $f_{a_1 \cdots a_n}(x)$ and the scalar-valued functions $\alpha_{a_1 \cdots a_n}$ satisfy the recursions

$$f_{a_1 \cdots a_n}(x) = \partial_x f_{a_2 \cdots a_n}(x) \, f_{a_1}(x), \quad n > 1, \tag{5}$$

($\partial_x f_{a_2 \cdots a_n}(x)$ denotes the value at x of the Jacobian matrix of $f_{a_2 \cdots a_n}$) and

$$\alpha_{a_1}(t) = \int_0^t \lambda_{a_1}(t_1) \, dt_1, \tag{6}$$

$$\alpha_{a_1 \cdots a_n}(t) = \int_0^t \lambda_{a_n}(t_n) \, \alpha_{a_1 \cdots a_{n-1}}(t_n) \, dt_n, \quad n > 1.$$

For future reference we note that

$$\alpha_{a_1\cdots a_n}(t) = \int_0^t dt_n\, \lambda_{a_n}(t_n) \int_0^{t_n} dt_{n-1}\, \lambda_{a_{n-1}}(t_{n-1}) \cdots \int_0^{t_2} dt_1\, \lambda_{a_1}(t_1),$$

or

$$\alpha_{a_1\cdots a_n}(t) = \int\cdots\int_{\mathcal{S}_n(t)} \lambda_{a_1}(t_1)\cdots\lambda_{a_n}(t_n)\, dt_1\cdots dt_n, \tag{7}$$

where the n-fold integral is taken over the simplex

$$\mathcal{S}_n(t) = \{(t_1,\ldots,t_n) \in \mathbb{R}^n : 0 \le t_1 \le \cdots \le t_n \le t\}.$$

Let us present some examples (more may be seen in [28]):

1. In the simplest illustration, the set A has only one element a and the corresponding λ takes the value 1 for each t. Then (2) is the *autonomous* system $(d/dt)x = f_a(x)$. For each n, the inner sum in (4) comprises a single term and from (7) the corresponding coefficient is found to be $\alpha_{a\cdots a}(t) = t^n/n!$. In this case (4) is the standard Taylor expansion of $x(t)$.

2. The *autonomous* system $(d/dt)x = F(x)$ with the right-hand side split as $F(x) = f_a(x) + f_b(x)$ is of the form (2) with $A = \{a, b\}$ and $\lambda_a(t) = \lambda_b(t) = 1$. For each n, the inner sum in (4) comprises 2^n terms and each of them has a coefficient $t^n/n!$. The expansion (4) is the Taylor series for $x(t)$ written in terms of the pieces f_a and f_b rather than in terms of F, a format that is useful in the analysis of splitting numerical integrators. It is of course possible to split F in $m > 2$ parts or even in infinitely many parts; then A has m or infinitely many elements. In all cases the integral (7) has the value $t^n/n!$.

3. Let $\omega > 0$ be a fixed number. If $A = \mathbb{Z}$ (the set of all integers) and, for $k \in \mathbb{Z}$, $\lambda_k(t) = \exp(ik\omega t)$, then (2) is a *non-autonomous* system $2\pi/\omega$-periodic in t with the right-hand side expanded in Fourier series. Here the functions λ_k possess complex values; if the system (2) is real then the mappings f_k take values in $\mathbb{C}^D$ and f_k is the complex conjugate of f_{-k} for each $k \in \mathbb{Z}$. The expansion (4) has been used in [12], [13], [14], [15] to study analytically periodic problems.

4. Also treated in [13], [14], [15] are quasiperiodic problems. If $\omega \in \mathbb{Z}^d$ is a vector of frequencies, these are of the form (2) with $A = \mathbb{Z}^d$ and

$$\lambda_{\mathbf{k}}(t) = \exp(i\mathbf{k}\cdot\omega t), \qquad \mathbf{k} \in \mathbb{Z}^d. \tag{8}$$

   This case will be taken up in the next section.

The notation in (4) may be made slightly more compact by considering A as an *alphabet* and the strings $a_1\cdots a_n$ as *words* with n letters. Then, if $\mathcal{W}_n$ represents the set of all words with n letters, (4) 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

$$x(t) = \sum_{n=0}^{\infty} \sum_{w \in \mathcal{W}_n} \alpha_w(t)\, f_w(x_0) = \sum_{w \in \mathcal{W}} \alpha_w(t)\, f_w(x_0), \tag{9}$$

where $\mathcal{W}$ represents the set of all words.

Subsequent developments will make much use of the set $\mathbb{C}^{\mathcal{W}}$ of all mappings $\delta : \mathcal{W} \to \mathbb{C}$; if $\delta \in \mathbb{C}^{\mathcal{W}}$ and w is a word, then δ_w is a complex number. This set is obviously a vector space for the usual operations between maps: if μ_1, μ_2 are scalars and $\delta_1, \delta_2 \in \mathbb{C}^{\mathcal{W}}$, then $(\mu_1\delta_1 + \mu_2\delta_2) \in \mathbb{C}^{\mathcal{W}}$ is defined by $(\mu_1\delta_1 + \mu_2\delta_2)_w = \mu_1(\delta_1)_w + \mu_2(\delta_2)_w$ for each $w \in \mathcal{W}$.

The expansion (9) motivates the following definition:

Definition 1 *If $\delta \in \mathbb{C}^{\mathcal{W}}$, then its corresponding word series is the formal series*

$$W_\delta(x) = \sum_{w \in \mathcal{W}} \delta_w f_w(x). \tag{10}$$

The scalars δ_w and the functions f_w will be called the coefficients *of the series and* word-basis functions *respectively.*

Clearly the word-basis functions change with the mappings f_a in the system (2) being studied. With this terminology, for each fixed t, the formal series (9) for the solution $x(t)$ of (1)–(2) is a word series whose coefficients $\alpha_w(t)$ are given by (7) and $\alpha_\emptyset(t) = 1$ (these coefficients are independent of the mappings f_a). Also, for each t, the right-hand side of (2) is a word series with coefficients $\beta_a(t) = \lambda_a(t)$ for words with one letter and $\beta_w(t) = 0$ for all other words. As we shall see later, word series W_δ corresponding to other choices of coefficients δ_w are useful in the analysis of dynamical systems and their numerical integrators.

Remark 1 *Word series and moulds.* In Ecalle's terminology, word series are *moulds*, see [17], [18], and the convolution product considered below is the mould product. (Ecalle [17] also considered a composition of moulds —not discussed in this paper— that is analogous to the substitution of B-series [11].)

Remark 2 *Word series and B-series.* Each f_w, $w \neq \emptyset$, is built up from partial derivatives of the f_a, $a \in A$, e.g., if $a, b, c \in A$,

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

The functions $\partial_x f_a(x)\, f_b(x)$, $\partial_{xx} f_a(x)[f_b(x), f_c(x)]$,$\partial_x f_a(x)\, \partial_x f_b(x)\, f_c(x)$ in these expressions are examples of *elementary differentials;* each word basis function $f_w(x)$, with $w \in \mathcal{W}_n$, $n > 0$, is a linear combination with integer coefficients of elementary differentials of order n (i.e. containing n functions $f_a(x)$). There is an elementary differential corresponding to each A-coloured (or A-decorated) rooted tree, i.e. to each

rooted tree where to each vertex it has been assigned an element of A. By expanding each word basis function in terms of elementary differentials, the series (10) becomes a so-called B-series

$$\sum_{\tau} \Delta_\tau \mathcal{F}_\tau(x),$$

where the summation is extended to all A-coloured rooted trees and $\mathcal{F}_\tau(x)$ is the elementary differential corresponding to τ. B-series were introduced by Hairer and Wanner [23] in the simplest case where the alphabet A has only one letter. B-series corresponding to this and larger alphabets are often used in numerical analysis; word series being more compact are better suited to analyze some integrators. For the relation between word series and B-series see [28] and [18] (Ecalle used in this connection the terminology arborifaction-coarborification).

Remark 3 *Word series as power series.* For a system $(d/dt)x = \epsilon \sum_a \lambda_a(t) f_a(x)$, where ϵ is a scalar parameter, (10) becomes the formal power series

$$\sum_{n=0}^{\infty} \epsilon^n \sum_{w \in \mathcal{W}_n} \gamma_w f_w(x).$$

Note that when the alphabet A is infinite the coefficient of ϵ^n is itself an infinite series that has to be understood formally. The format (10) is of course recovered from the power series by setting $\epsilon = 1$; therefore both formats are equivalent. While the papers [13], [15] use the power series format, we prefer to work with (10) as it leads to more compact formulae. Some readers may find it useful to mentally substitute ϵf_a for f_a everywhere in what follows; this may be particularly the case for the perturbed integrable problems considered in Section 3.

2.2 Operations with word series

2.2.1 The convolution product

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

$$(\delta \star \delta')_{a_1 \cdots a_n} = \delta_\emptyset \delta'_{a_1 \cdots a_n} + \sum_{j=1}^{n-1} \delta_{a_1 \cdots a_j} \delta'_{a_{j+1} \cdots a_n} + \delta_{a_1 \cdots a_n} \delta'_\emptyset \tag{11}$$

(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 $\mathbb{1} \in \mathbb{C}^{\mathcal{W}}$ with $\mathbb{1}_\emptyset = 1$ and $\mathbb{1}_w = 0$ for $w \neq \emptyset$).

As we shall see, the operation $\star$ plays an essential role in the manipulation of word series.

2.2.2 The group $\mathcal{G}$

If $w \in \mathcal{W}_m$ and $w' \in \mathcal{W}_n$ are words, $m, n \geq 1$, their *shuffle product* $w \sqcup\!\sqcup w'$ [31] 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. (Examples: for the words ab, cd, the shuffle product is $ab \sqcup\!\sqcup cd = abcd + acbd + cabd + acdb + cadb + cdab$, for the words ab, a, the product is $ab \sqcup\!\sqcup a = aba + aab + aab$.) In addition $\emptyset \sqcup\!\sqcup w = w \sqcup\!\sqcup \emptyset = w$ for each $w \in \mathcal{W}$. The operation $\sqcup\!\sqcup$ is commutative and associative and has 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 $w, w' \in \mathcal{W}$,

$$\gamma_w \gamma_{w'} = \sum_{j=1}^{N} \gamma_{w_j} \qquad \text{if} \qquad w \sqcup\!\sqcup w' = \sum_{j=1}^{N} w_j. \tag{12}$$

The set $\mathcal{G}$ with the operation $\star$ may be regarded in a formal sense (cf. [5]) as a non-commutative Lie group (see Section 6.1.2). For each fixed t, the family of coefficients defined by (7) and $\alpha_\emptyset(t) = 1$ is an element of the group $\mathcal{G}$ (to prove this, consider (12) for $w \in \mathcal{W}_m$ and $w' \in \mathcal{W}_n$ and use (7) to write $\alpha_w \alpha_{w'}$ and each α_{w_j} as integrals over subsets of $\mathbb{R}^{m+n}$, cf. [31, Corollary 3.5]).

When γ belongs to $\mathcal{G}$, the word series $W_\gamma(x)$ has properties that are not shared by general word series. For $\gamma \in \mathcal{G}$, changes of variables $x = C(X)$ commute with the formation of word series as described in [13, Proposition 3.1]. Moreover, 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

$$W_\delta\big(W_\gamma(x)\big) = W_{\gamma\star\delta}(x), \tag{13}$$

i.e. the coefficients of the word series resulting from the substitution are given by the convolution product $\gamma\star\delta$ (this is proved in Section 6.1.3). A similar rule exists of course for B-series, but the recipe there is more complicated than (11) [23], [22, Chapter III].

In the numerical analysis of differential equations, word series with coefficients in $\mathcal{G}$ appear e.g. as expansions in powers of the stepsize of splitting integrators, see Section 5. Then (13) provides the recipe to compose integrators or to compose an integrator and a mapping. For each fixed t, the right-hand side of (2) is an example of a word-series with coefficients in the Lie algebra $\mathfrak{g}$ of the group $\mathcal{G}$ that we study next.

2.2.3 The Lie algebra $\mathfrak{g}$

We denote $\mathfrak{g}$ the set of elements $\beta \in \mathbb{C}^{\mathcal{W}}$ such that $\beta_\emptyset = 0$ and for each pair of nonempty words w, w',

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

It is clear that $\mathfrak{g}$ is a vector subspace of the vector space $\mathbb{C}^{\mathcal{W}}$; furthermore $\mathfrak{g}$ is closed for the skew-symmetric product defined by

$$[\beta, \beta'] = \beta \star \beta' - \beta' \star \beta. \tag{14}$$

This product satisfies the Jacobi identity and therefore endows $\mathfrak{g}$ with a structure of Lie algebra (see Section 6.1.2). In fact $\mathfrak{g}$ is the Lie algebra of the Lie group $\mathcal{G}$: the elements $\beta \in \mathfrak{g}$ coincide with the velocities at $\mathbb{1} \in \mathcal{G}$ of curves in $\mathcal{G}$. In symbols, if $\gamma(t)$, $t \in \mathbb{R}$ is a curve in $\mathcal{G}$ such that $\gamma(0) = \mathbb{1}$, then $\beta \in \mathbb{C}^{\mathcal{W}}$ defined by

$$\beta = \frac{d}{dt}\gamma(t)\Big|_{t=0} \tag{15}$$

(i.e. $\beta_w = (d/dt)\gamma_w(0)$ for for each $w \in \mathcal{W}$) belongs to $\mathfrak{g}$. Moreover any $\beta \in \mathfrak{g}$ arises in this way: the exponential

$$\exp_\star(t\beta) = \mathbb{1} + \sum_{j=1}^{\infty} \frac{t^j}{j!}\beta^{\star j} \tag{16}$$

($\beta^{\star j}$ is the convolution product of j factors all equal to β) defines a curve of elements of $\mathcal{G}$ with velocity β at $t = 0$. The points of this curve actually form a one-parameter subgroup of $\mathcal{G}$ since $\exp_\star(t\beta) \star \exp_\star(t'\beta) = \exp_\star((t+t')\beta)$. The exponent β may be retrieved from the exponential $\gamma = \exp_\star(\beta)$ by means of the logarithm

$$\beta = \log_\star(\gamma) = \sum_{j=1}^{\infty} \frac{(-1)^{(j+1)}}{j}(\gamma - \mathbb{1})^{\star j}. \tag{17}$$

Just as the convolution product with an element of $\mathcal{G}$ corresponds to the operation of substitution of the associated word series (see (13)), the convolution bracket (14) corresponds to the Jacobi bracket (commutator) of the associated word series, for $\beta, \beta' \in \mathfrak{g}$:

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

To prove this, let $\gamma(t)$ be a curve with velocity β as above; then, for any $\delta \in \mathbb{C}^{\mathcal{W}}$,

$$(\partial_x W_\delta(x))W_\beta(x) = \frac{d}{dt}W_\delta\big(W_{\gamma(t)}(x)\big)\Big|_{t=0} = \frac{d}{dt}W_{\gamma(t)\star\delta}(x)\Big|_{t=0} = W_{\beta\star\delta}(x). \tag{18}$$

(We have successively used the chain rule, (13), and the bilinearity of $\star$.)

Since for $\beta \in \mathfrak{g}$, the word series $W_\beta(x)$ belongs to the Lie algebra (for the Jacobi bracket) generated by the mappings (vector fields) f_a, the Dynkin-Specht-Wever formula [24] may be used to rewrite the word series in terms of iterated commutators of these mappings:

$$W_\beta(x) = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{a_1,\ldots,a_n \in A} \beta_{a_1\cdots a_n}[[\cdots[[f_{a_1}, f_{a_2}], f_{a_3}]\cdots], f_{a_n}](x). \tag{19}$$

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

2.2.4 Nonautonomous differential equations in $\mathcal{G}$

Initial value problems

$$\frac{d}{dt}x(t) = W_{\beta(t)}(x(t)), \qquad x(0) = x_0, \tag{20}$$

where for each t, $\beta(t) \in \mathfrak{g}$, are a natural generalization of (1)–(2) (we recall that the right-hand side (2) does not include contributions from basis functions associated with words with more than one letter). These problems may be solved formally by using the ansatz $x(t) = W_{\alpha(t)}(x_0)$ with $\alpha(t) \in \mathcal{G}$ for each t. In view of (13), we may write

$$\frac{d}{dt}W_{\alpha(t)}(x_0) = W_{\beta(t)}(W_{\alpha(t)}(x_0)) = W_{\alpha(t)\star\beta(t)}(x_0), \qquad W_{\alpha(0)}(x_0) = x_0,$$

which leads to the linear, nonautonomous initial value problem

$$\frac{d}{dt}\alpha(t) = \alpha(t) \star \beta(t), \qquad \alpha(0) = \mathbb{1}\,. \tag{21}$$

For the empty word, according to (11), $(d/dt)\alpha_\emptyset = \alpha_\emptyset(t)\beta_\emptyset(t)$; since $\beta(t) \in \mathfrak{g}$ implies $\beta_\emptyset(t) = 0$, we see that $\alpha_\emptyset(t) = 1$. For a word $a \in \mathcal{W}_1$, $(d/dt)\alpha_a(t) = \alpha_\emptyset(t)\beta_a(t) + \alpha_a(t)\beta_\emptyset(t)$, which leads to $\alpha_a(t) = \int_0^t \beta_a(t_1)\,dt_1$. The process may be continued in an obvious way and induction on the number of letters shows that (21) uniquely determines $\alpha_w(t)$ for each $w \in \mathcal{W}$. Furthermore, for each t, the element $\alpha(t) \in \mathbb{C}^{\mathcal{W}}$ defined in this way belongs to $\mathcal{G}$; while this may be established by means of the Magnus expansion (see e.g. [3], [22, Chapter IV]) we provide an elementary proof in Section 6.

Conversely, any curve $\alpha(t)$ of group elements with $\alpha(0) = \mathbb{1}$ solves a problem of the form (21) with

$$\beta(t) = \alpha(t)^{-1} \star \left(\frac{d}{dt}\alpha(t)\right).$$

Since

$$\alpha(t)^{-1} \star \left(\frac{d}{dt}\alpha(t)\right) = \frac{d}{ds}\Big(\alpha(t)^{-1} \star \alpha(t+s)\Big)\Big|_{s=0},$$

for each t, the element $\beta(t)$ defined in this way is a member of $\mathfrak{g}$.

The investigation of normal forms below is based on changing variables. A change of variables $x = W_\kappa(X)$, $\kappa \in \mathcal{G}$, transforms the problem (20) into

$$\frac{d}{dt}X(t) = W_{B(t)}(X(t)), \qquad X(0) = X_0,$$

with $B(t) \star \kappa = \kappa \star \beta(t)$ (or $B(t) = \kappa \star \beta(t) \star \kappa^{-1}$), $X_0 = W_{\kappa^{-1}}(x_0)$ (κ^{-1} is the inverse of κ in the group $\mathcal{G}$); this is a direct consequence of (13) and (18).

2.2.5 The Hamiltonian case

Consider now the particular case where the dimension D of (2) is even and each $f_a(x)$ is a Hamiltonian vector field [35], i.e. $f_a(x) = J^{-1}\nabla H_a(x)$, where J^{-1} is the standard

symplectic matrix. Recall [2] that the Jacobi bracket (commutator) $[J^{-1}\nabla A, J^{-1}\nabla B]$ of two Hamiltonian vectors fields is again a Hamiltonian vector field and that the corresponding Hamiltonian function is the Poisson bracket of the Hamiltonians A and B, defined by $\{A, B\}(x) = \nabla A(x)^T J^{-1} B(x)$. According to (19), for each $\beta \in \mathfrak{g}$, the vector field $W_\beta(x)$ is Hamiltonian

$$W_\beta(x) = J^{-1}\nabla \mathcal{H}_\beta(x)$$

with Hamiltonian function

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

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

$$H_w(x) = \frac{1}{n}\{\{\cdots\{\{H_{a_1}, H_{a_2}\}, H_{a_3}\}\cdots\}, H_{a_n}\}(x). \tag{22}$$

For Hamiltonian systems, changes of variables $x = W_\kappa(X)$, $\kappa \in \mathcal{G}$, are canonically symplectic; after the change of variables the system is again Hamiltonian and the new Hamiltonian function is obtained by changing variables in the old Hamiltonian function [2].

3 Extended word series

In this section we adapt the preceding material to cover perturbed integrable problems.

3.1 Perturbed integrable problems

We now consider systems of the form

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

where $y \in \mathbb{R}^{D-d}$, $0 < d \leq D$, $\omega \in \mathbb{R}^d$ is a vector of frequencies $\omega_j > 0$, $j = 1, \ldots, d$, 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 (see the Appendix). 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 this class as perturbed integrable problems and to f as the perturbation (some readers may prefer to substitute ϵf for f, see Remark 3).

After introducing the functions

$$f_{\mathbf{k}}(y, \theta) = \exp(i\mathbf{k}\cdot\theta)\, \hat{f}_{\mathbf{k}}(y), \qquad y \in \mathbb{R}^{D-d},\ \theta \in \mathbb{R}^d, \tag{23}$$

that satisfy the fundamental identity

$$f_{\mathbf{k}}(y, \theta_1 + \theta_2) = \exp(i\mathbf{k} \cdot \theta_1)\, f_{\mathbf{k}}(y, \theta_2), \tag{24}$$

the system takes the form

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + f(y, \theta) = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + \sum_{\mathbf{k} \in \mathbb{Z}^d} f_{\mathbf{k}}(y, \theta). \tag{25}$$

To find the solution with initial conditions

$$y(0) = y_0, \qquad \theta(0) = \theta_0, \tag{26}$$

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

$$\frac{d}{dt}\begin{bmatrix} y \\ \eta \end{bmatrix} = \sum_{\mathbf{k} \in \mathbb{Z}^d} \exp(i\mathbf{k} \cdot \omega t)\, f_{\mathbf{k}}(y, \eta), \tag{27}$$

a particular instance of the problem considered in Section 2. The alphabet A coincides with $\mathbb{Z}^d$, and, for each 'letter' $\mathbf{k}$, $\lambda_{\mathbf{k}}(t)$ is given by (8). The formula (4) yields

$$\begin{bmatrix} y(t) \\ \eta(t) \end{bmatrix} = \begin{bmatrix} y(0) \\ \eta(0) \end{bmatrix} + \sum_{n=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_n} \alpha_{\mathbf{k}_1 \cdots \mathbf{k}_n}(t)\, f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y(0), \eta(0)), \tag{28}$$

where the coefficients α are still given by (7) (but recall that now the letters a are multiindices $\mathbf{k}$) and the word basis functions are defined by (23) and (5) (the Jacobian in (5) is taken with respect to the D-dimensional variable (y, θ)). We conclude that, in the original variables, the solution flow of (25), has the formal expansion

$$\phi_t(y_0, \theta_0) = \begin{bmatrix} y(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix} + \begin{bmatrix} 0 \\ t\omega \end{bmatrix} + \sum_{n=1}^{\infty} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_n} \alpha_{\mathbf{k}_1 \cdots \mathbf{k}_n}(t)\, f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y_0, \theta_0). \tag{29}$$

Note that the word basis functions are *independent of the frequencies* ω and the coefficients α are *independent of* f. Also from (24) we have the identity:

$$f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y, \theta_1 + \theta_2) = \exp(i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \theta_1)\, f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y, \theta_2), \tag{30}$$

and, in particular

$$f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y, \theta) = \exp(i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \theta)\, f_{\mathbf{k}_1 \cdots \mathbf{k}_n}(y, 0), \tag{31}$$

With the notation of Section 2, we write (29) in the following form (here and later $x = (y, \theta)$):

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

In order to make the formula even more compact, we introduce the vector space $\mathcal{C} = \mathbb{C}^d \oplus \mathbb{C}^{\mathcal{W}}$ and define:

Definition 2 *If $(v,\delta) \in \mathcal{C}$, then its corresponding extended word series is 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 (29) of (25)–(26) 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 in Section 2. Also the right-hand side of (25) is an extended word series $\overline{W}_{(\omega,\beta)}(x)$ with $\beta \in \mathfrak{g} \subset \mathbb{C}^{\mathcal{W}}$ defined by

$$\beta_w = 1 \quad \text{if} \quad w \in \mathcal{W}_1, \qquad \beta_w = 0 \quad \text{if} \quad w \notin \mathcal{W}_1. \tag{32}$$

3.2 Operations with extended word series

3.2.1 The operation ★

The following two linear operators will appear repeatedly. 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

$$(\Xi_v\delta)_w = \exp(i(\mathbf{k}_1 + \cdots + \mathbf{k}_n)\cdot v)\,\delta_w. \tag{33}$$

for $w = \mathbf{k}_1 \ldots \mathbf{k}_n$. The linear operator ξ_v on $\mathbb{C}^{\mathcal{W}}$ is defined as follows: $(\xi_v\delta)_\emptyset = 0$, and for each word $w = \mathbf{k}_1 \cdots \mathbf{k}_n$,

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

Thus Ξ_v and ξ_v are *diagonal* operators with eigenvalues $\exp(i(\mathbf{k}_1 + \cdots + \mathbf{k}_n)\cdot v)$ and $i(\mathbf{k}_1 + \cdots + \mathbf{k}_n)\cdot v$ respectively. Observe that $\Xi_v(\gamma \star \delta) = (\Xi_v\gamma)\star(\Xi_v\delta)$ if $\gamma, \delta \in \mathbb{C}^{\mathcal{W}}$ and that:

$$\frac{d}{dt}\Xi_{tv} = \Xi_{tv}\xi_v = \xi_v\Xi_{tv}.$$

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 $\overline{\mathcal{G}}$. With the help of Ξ_u we define an operation ★ as follows. 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 (13) and (30), it is a simple exercise to check that $\overline{\mathcal{G}}$ acts by substitution on extended word series as follows:

$$\overline{W}_{(v,\delta)}\big(\overline{W}_{(u,\gamma)}(x)\big) = \overline{W}_{(u,\gamma)\bigstar(v,\delta)}(x), \qquad \gamma \in \mathcal{G}. \tag{34}$$

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

[1]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.

3.2.2 The Lie algebra $\overline{\mathfrak{g}}$

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 $\delta \in \mathfrak{g}$. Let us describe the bracket in $\overline{\mathfrak{g}}$. Given $v \in \mathbb{C}^d$ and $\delta \in \mathfrak{g}$, we have the trivial decomposition

$$\overline{W}_{(v,\delta)} = \begin{bmatrix} 0 \\ v \end{bmatrix} + W_\delta(x) = \overline{W}_{(v,0)} + \overline{W}_{(0,\delta)}.$$

By using (31), one can check that the Jacobi bracket of the vector fields $\overline{W}_{(\lambda,0)}$ and $\overline{W}_{(0,\delta)}$ is

$$[\overline{W}_{(v,0)}, \overline{W}_{(0,\delta)}] = \overline{W}_{(0,\xi_v\delta)}.$$

From these relations we conclude that, for arbitrary $(v, \delta), (u, \eta) \in \mathbb{C}^d \oplus \mathfrak{g}$, the Jacobi bracket of the vector fields $\overline{W}_{(v,\delta)}$, $\overline{W}_{(u,\eta)}$ is given by

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

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

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

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

3.2.3 Nonautonomous differential equations in $\overline{\mathcal{G}}$

The initial value problem

$$\frac{d}{dt}x(t) = \overline{W}_{(\omega,\beta(t))}(x(t)), \qquad x(0) = x_0,$$

where $(\omega, \beta(t)) \in \overline{\mathfrak{g}}$ for each t, may be formally solved in a manner that is exactly parallel to treatment given above to (20): $x(t) = \overline{W}_{(t\omega,\alpha(t))}(x_0)$, where $\alpha(0) = \mathbb{1}$ and

$$\frac{d}{dt}(t\omega, \alpha(t)) = (t\omega, \alpha(t))\bigstar(\omega, \beta(t)).$$

Observe that the right-hand side of this equation is, by definition of $\bigstar$, equal to $(t\omega, \alpha(t) \star (\Xi_{t\omega}\beta(t)))$, so that $\alpha(t)$ is the solution of an initial value problem of the form (21) with $\beta(t)$ replaced by $\Xi_{t\omega}\beta(t)$.

A change of variables $x = \overline{W}_{(v,\kappa)}(X)$, $(v, \kappa) \in \overline{\mathcal{G}}$, can be seen to transform the problem into

$$\frac{d}{dt}X(t) = \overline{W}_{(\omega,B(t))}(X(t)), \qquad X(0) = X_0,$$

where now $B(t)$ is determined from

$$B(t) \star \kappa + \xi_\omega\kappa = \kappa \star (\Xi_v\beta(t))$$

and, of course, $x_0 = \overline{W}_{(v,\kappa)}(X_0)$ or $X_0 = \overline{W}_{(v,\kappa)^{-1}}(x_0)$. (Note that $(v, \kappa)^{-1} = (-v, \Xi_{-v}\kappa^{-1})$.)

3.2.4 Perturbed Hamiltonian problems

To end this section, assume in (25), that the dimension D is even with $D/2-d=m\geq 0$ and that the vector of unknowns 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 (23) is a Hamiltonian vector field with Hamiltonian function $H_{\mathbf{k}}(x)$, then the system (25) 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

$$\sum_{j=1}^{d}\omega_j a^j+\sum_{w\in\mathcal{W},\,w\neq\emptyset}\beta_w H_w, \tag{35}$$

with $H_w(x)$ as in (22). Note that the Lie bracket in $\overline{\mathfrak{g}}$ can be used to compute the Poisson bracket of formal Hamiltonian functions of the form (35).

4 Normal forms and averaging

In this section we show how the algebraic machinery introduced above may be applied to build a theory of normal forms [1], [32] for the perturbed integrable problems of the form (25). This theory hinges on the fact that the linear operator $\overline{W}_{(0,\delta)}\mapsto[\overline{W}_{(\omega,0)},\overline{W}_{(0,\delta)}]$ ($\delta\in\mathfrak{g}$) coincides, as we have seen in Section 3.2.2, with the *diagonal* operator $\overline{W}_{(0,\delta)}\mapsto\overline{W}_{(0,\xi_\omega\delta)}$.

Let us consider an autonomous system

$$\frac{d}{dt}x=\frac{d}{dt}\begin{bmatrix}y\\ \theta\end{bmatrix}=\begin{bmatrix}0\\ \omega\end{bmatrix}+W_\beta(x)=\overline{W}_{(\omega,\beta)}(x),\qquad \beta\in\mathfrak{g}. \tag{36}$$

As noted before, this format yields the perturbed problem (25) when β is chosen as in (32). The more general case where β is any element in $\mathfrak{g}$ will be necessary to deal with splitting integrators later. We shall change variables $x=W_\kappa(X)=\overline{W}_{(0,\kappa)}(X)$, $\kappa\in\mathcal{G}$, in order to simplify (36) as much as possible.

Remark 4 There is nothing lost by assuming that $x=W_\kappa(X)$ is a (not extended) word series in the new variables X —or equivalently an extended word series of the special format $x=\overline{W}_{(0,\kappa)}(X)$—. More general changes $x=\overline{W}_{(v,\kappa)}(X)$ do not allow for additional simplications in (36).

From Section 3.2.3, we know that the transformed system is

$$\frac{d}{dt}X = \frac{d}{dt}\begin{bmatrix} Y \\ \Theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix} + W_{\widehat{\beta}}(X) = \overline{W}_{(\omega,\widehat{\beta})}(X), \tag{37}$$

with

$$\xi_\omega \kappa + \widehat{\beta} \star \kappa = \kappa \star \beta. \tag{38}$$

Our aim is to choose $\widehat{\beta} \in \mathfrak{g}$ and $\kappa \in \mathcal{G}$ subject to (38) and such that $\widehat{\beta}$ is as simple as possible; then the system is said to have been brought to normal form. Of course the maximum simplification would be obtained by setting $\widehat{\beta} = 0$, but for this choice of $\widehat{\beta}$ it is not possible to find an appropriate κ; this will be clear in the proof of Theorem 1 and is to be expected from general results on normal forms [1], [32]. More precisely, perturbations that commute with $\overline{W}_{(\omega,0)}$ cannot be eliminated by changing variables. One then has to restrict the attention to $\widehat{\beta} \in \mathfrak{g}$ such that in (37) the unperturbed vector field and the perturbation commute, i.e. $[\overline{W}_{(\omega,0)}, \overline{W}_{(0,\widehat{\beta})}] = 0$. This is equivalent to $\overline{W}_{(0,\xi_\omega \widehat{\beta})} = 0$, or, in terms of the coefficients,

$$i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \omega \, \widehat{\beta}_{\mathbf{k}_1 \ldots \mathbf{k}_n} = 0, \tag{39}$$

for each nonempty word $\mathbf{k}_1 \ldots \mathbf{k}_n$. We have then the following result, which is proved constructively in Section 6.2.

Theorem 1 *There is a change of variables $x = W_\kappa(X)$, $\kappa \in \mathcal{G}$, that reduces the system (36) to the form (37), where $\widehat{\beta} \in \mathfrak{g}$ and $\widehat{\beta}_w = 0$ for all words $w = \mathbf{k}_1 \ldots \mathbf{k}_n$ such that $(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \omega \neq 0$. Furthermore the vector fields $\overline{W}_{(\omega,0)}(X)$ and $\overline{W}_{(0,\widehat{\beta})}(X)$ commute and the solutions of (37) satisfy*

$$X(t) = \phi_t\Big(X(0) + \begin{bmatrix} 0 \\ t\omega \end{bmatrix}\Big) = \phi_t(X(0)) + \begin{bmatrix} 0 \\ t\omega \end{bmatrix},$$

where ϕ_t is the solution flow of the system $(d/dt)X = W_{\widehat{\beta}}(X)$. Equivalently, $X(t) = \overline{W}_{(t\omega,\widehat{\alpha}(t))}(X(0))$, where

$$(t\omega, \widehat{\alpha}(t)) = (t\omega, \exp_\star(t\widehat{\beta})) = \exp_\star(t\widehat{\beta}) \bigstar (t\omega, \, \mathbb{1}\,) = (t\omega, \, \mathbb{1}\,) \bigstar \exp_\star(t\widehat{\beta}).$$

If the system (36) is Hamiltonian, the change of variables is canonical symplectic and the transformed system (37) is Hamiltonian.

When ω is *nonresonant*, i.e. $\mathbf{k} \cdot \omega \neq 0$ for $\mathbf{k} \neq \mathbf{0}$, the theorem implies, in view of (31), that the transformed vector field $\overline{W}_{(\omega,\widehat{\beta})}(X)$ is independent of the angular variables Θ. In other words, (37) is a system where the angles have been *averaged* [1], [2], [32]. In the general situation with a nontrivial resonant module

$$\mathcal{M}_\omega = \{\mathbf{k} \in \mathbb{Z}^d : \mathbf{k} \cdot \omega = 0\},$$

the transformed vector field depends on Θ. However this dependence is only through a number of combinations $\mathbf{l}_1 \cdot \Theta, \ldots, \mathbf{l}_r \cdot \Theta$, $r < d$, where $\mathbf{l}_1, \ldots, \mathbf{l}_r \in \mathbb{Z}^d$ are linearly independent and span the resonant module.

Remark 5 Consider the *highly oscillatory* case where (36) depends on a small parameter δ and $\omega = \mathcal{O}(1/\delta)$, $W_\beta(x) = \mathcal{O}(1)$. The combinations not eliminated by the change of variables have the property that their velocities $(d/dt)\mathbf{l}_i \cdot \Theta$ are $\mathcal{O}(1)$, as distinct from the situation for the original angles with $(d/dt)\theta = \mathcal{O}(1/\delta)$. In this sense, the *fast* angles have been averaged when forming (37).

For convenience, we shall use the expression *oscillatory word* to refer to those words $\mathbf{k}_1 \dots \mathbf{k}_n$ for which $(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \omega \neq 0$. Thus the theorem may be rephrased as saying that the contributions to the vector field corresponding to oscillatory words may be removed from (36) by means of a change of variables. Note that the set $\mathfrak{g}_0$ of all $\widehat{\beta} \in \mathfrak{g}$ such that $\widehat{\beta}_w = 0$ for all oscillatory words is a Lie subalgebra of $\mathfrak{g}$; this follows from the fact that $\widehat{\beta}$ is in $\mathfrak{g}_0$ if and only if $(\omega, 0)$ and $(0, \widehat{\beta})$ commute.

If we now express the commuting vector fields $\overline{W}_{(\omega,0)}(X)$ and $\overline{W}_{(0,\widehat{\beta})}(X)$ in terms of the original variables x by applying the recipe for changing variables given in Subsection 3.2.3, we obtain a decomposition of the right-hand side of (36) as a commuting sum of two terms:

$$\overline{W}_{(\omega,\beta)}(x) = \overline{W}_{(\omega,\kappa^{-1}\star\xi_\omega\kappa)}(x) + W_{\kappa^{-1}\star\widehat{\beta}\star\kappa}(x).$$

The second of these generates a flow

$$x(t) = W_{\kappa^{-1}\star\exp_\star(t\widehat{\beta})\star\kappa}(x(0))$$

where the motion of the fast angles has been averaged. The former generates a quasi-periodic flow

$$x(t) = \overline{W}_{(0,\kappa^{-1})\bigstar(t\omega,\,\mathbb{1}\,)\bigstar(0,\kappa)}(x(0)) = \overline{W}_{(t\omega,\kappa^{-1}\star\Xi_{t\omega}\kappa)}(x(0)).$$

Finally, the flow of (36) is given by

$$x(t) = \overline{W}_{(t\omega,\kappa^{-1}\star\exp_\star(t\widehat{\beta})\star\kappa\star\Xi_{t\omega})}(x(0)).$$

Remark 6 In the nonresonant case, the commuting decomposition of $\overline{W}_{(\omega,\beta)}(x)$ has been obtained in [13, Theorem 5.5] by means of a different (but related) technique.

5 Splitting methods

Splitting algorithms [35], [22] are natural candidates to integrate perturbed integrable problems. In this connection, it is extremely important to emphasize that the practical implementation of splitting methods is *not* necessarily based on the simple format (25). Such a simple format is typically reached after suitable changes of variables and is quite convenient for the analysis. These points are discussed in the Appendix.

Given real coefficients, a_j and b_j, $j = 1, \dots, r$, we study the splitting integrator for (25) defined by

$$\widetilde{\phi}_h = \phi^{(P)}_{b_r h} \circ \phi^{(U)}_{a_r h} \circ \cdots \circ \phi^{(P)}_{b_1 h} \circ \phi^{(U)}_{a_1 h}. \tag{40}$$

Here h is the step-length, $\widetilde{\phi}_h$ the mapping in $\mathbb{R}^D$ that advances the numerical solution over one time step, and $\phi_t^{(U)}$ and $\phi_t^{(P)}$ denote respectively the exact t-flows of the split systems corresponding to the unperturbed dynamics

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ \omega \end{bmatrix}, \tag{41}$$

and the perturbation

$$\frac{d}{dt}\begin{bmatrix} y \\ \theta \end{bmatrix} = f(y,\theta). \tag{42}$$

If we set

$$a = \sum_{j=1}^{r} a_j, \qquad b = \sum_{j=1}^{r} b_j,$$

the integrator is *consistent* if $a = b = 1$.

Since the unperturbed dynamics with frequencies ω_j is reproduced exactly by (40), 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. [19], [34].

5.1 Extended word series expansion of the local 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, using Example 2 in Section 2,

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

where $\tau(t) \in \mathcal{G}$ comprises the Taylor coefficients, i.e. $\tau_w(t) = t^n/n!$ if $w \in \mathcal{W}_n$. The following result makes use of the algebraic formalism to provide explicitly the expansion of the numerical solution.

Theorem 2 *The splitting integrator $\widetilde{\phi}_h$ in (40) possesses the expansion*

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

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

$$\widetilde{\alpha}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = h^n \sum_{1\le j_1\le\cdots\le j_n\le r} \frac{b_{j_1}\cdots b_{j_n}}{\sigma_{j_1\cdots j_n}} \exp(i\,(c_{j_1}\mathbf{k}_1 + \cdots + c_{j_n}\mathbf{k}_n)\cdot\omega h). \tag{43}$$

Here,

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

and,

$$\sigma_{j_1\cdots j_n} = \frac{1}{n!} \quad \text{if} \quad j_1 = \cdots = j_n,$$

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

Proof: From (34), we know that $\widetilde{\phi}_h$ has an expansion in extended word series and that the family of coefficients is given by (pay attention to the ordering)

$$\big(a_1 h\omega,\ \mathbb{1}\big)\bigstar\big(0,\tau(b_1 h)\big)\bigstar\cdots\bigstar\big(a_r h\omega,\ \mathbb{1}\big)\bigstar\big(0,\tau(b_r h)\big);$$

it is enough to compute, according to the definition, the products $\bigstar$ in this expression. □

Remark 7 *The associated quadrature rule.* For words with one letter, the theorem yields:

$$\widetilde{\alpha}_{\mathbf{k}}(h) = h\sum_{1\le j\le r} b_j\ \exp(i\,c_j\mathbf{k}\cdot\omega h).$$

This obviously corresponds to the approximation of the exact coefficient $\alpha_{\mathbf{k}}(h)$ (defined in (6) and (8)) by the (univariate) quadrature rule that on the unit interval has abscissas c_j and weights b_j. This rule will be consistent if $b = 1$, which is implied by the consistency of the integrator.

Remark 8 *Associated cubature rules.* Similarly, for $n > 1$, (43) corresponds to approximating (7) with a cubature rule for the simplex. If the univariate quadrature is consistent, so is the cubature rule for each $n > 1$, because, by using the multinomial expansion,

$$\sum_{1\le j_1\le\cdots\le j_n\le r}\frac{b_{j_1}\cdots b_{j_n}}{\sigma_{j_1\cdots j_n}} = \sum_{n_1+\cdots+n_r=n}\frac{b_1^{n_1}\cdots b_r^{n_r}}{n_1!\cdots n_r!}$$
$$= \frac{1}{n!}\big(\sum_{j=1}^{r} b_j\big)^n = \big(\sum_{j=1}^{r} b_j\big)^n\,\mathrm{Vol}(\mathcal{S}(1)).$$

Discussions perhaps become clearer by introducing scaled coefficients A_w and $\widetilde{A}_w$ such that

$$\alpha_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = h^n A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h),\qquad \widetilde{\alpha}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = h^n \widetilde{A}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h).$$

Note that, by performing the change of variables $t_j = ht'_j$, $j = 1,\ldots,n$, in (7),

$$A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = \int\cdots\int_{\mathcal{S}_n(1)}\exp\big(i(t'_1\mathbf{k}_1+\cdots+t'_n\mathbf{k}_n)\cdot\omega h\big)\,dt'_1\cdots dt'_n, \qquad (44)$$

and that, therefore,

$$|\ A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h)\ | \le \mathrm{Vol}(\mathcal{S}_n(1)) = \frac{1}{n!}.$$

With these preparations, we have proved our next result:

Theorem 3 *The local error of the splitting integrator $\widetilde{\phi}_h$ in (40) possesses the expansion*

$$\widetilde{\phi}_h(x) - \phi_h(x) = \overline{W}_{(h(a-1)\omega,\widetilde{\alpha}(h)-\alpha(h))}(x).$$

i.e.

$$\widetilde{\phi}_h(x_0) - \phi_h(x_0) = \begin{bmatrix} 0 \\ h\big(a-1\big)\omega \end{bmatrix} + \sum_{n=1}^{\infty} h^n \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_n \in \mathbb{Z}^d} \big(\widetilde{A}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) - A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h)\big)\, f_{\mathbf{k}_1\cdots\mathbf{k}_n}(x_0). \tag{45}$$

5.2 Estimates

In order to obtain error estimates it is now necessary to truncate the infinite series in (45) and we shall do so in the next theorem, whose proof is given in Section 6.3. We assume hereafter that:

1. The function $f(x) = f(y,\theta)$ is defined in a set $\Omega = B_R(y_0)\times\mathbb{T}^d$, where $B_R(y_0)$ is the ball $\{y : |y-y_0| < R\} \subset \mathbb{R}^{D-d}$.
2. There exists a finite set of indices $\mathcal{I} \subset \mathbb{Z}^d$ such that for $\mathbf{k} \notin \mathcal{I}$ the Fourier coefficient $\hat{f}_{\mathbf{k}}$ vanishes.
3. There exists an integer $N \geq 2$, such that the Fourier coefficients $\hat{f}_{\mathbf{k}}$ and their partial derivatives of order $\leq N-1$ are continuous and bounded in $B_R(y_0)$.

In the first of these hypotheses the form of the domain Ω is natural since $f(y,\theta)$ is periodic in each of the components of θ. As shown in e.g. [14], the second hypothesis may be relaxed for the conclusions of the theorem to hold; it makes however possible to avoid distracting technicalities. In this connection it should be noted that, for nonlinear problems, even if f has a finite number of Fourier modes the solution $x(t)$ in (29) will include arbitrarily high frequencies (the product of λ's in (7) adds the corresponding wave numbers $\mathbf{k}$).[2]

Theorem 4 *Assume that the system (25) being integrated satisfies the assumptions above. Then there exist positive constants h_0, C,* both independent of ω, *such that:*

1. *For $|h| < h_0$ and arbitrary θ_0, the true solution $\phi_h(x_0)$, $x_0 = (y_0,\theta_0)$, and the numerical solution $\widetilde{\phi}_h(x_0)$ are well defined and lie in Ω.*
2. *The local error at x_0 satisfies*

$$\widetilde{\phi}_h(x_0) - \phi_h(x_0) = \begin{bmatrix} 0 \\ h\big(a-1\big)\omega \end{bmatrix} + \sum_{n=1}^{N-1} h^n \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_n \in \mathcal{I}} \big(\widetilde{A}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) - A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h)\big)\, f_{\mathbf{k}_1\cdots\mathbf{k}_n}(x_0) + \mathcal{R}_h(x_0), \tag{46}$$

where $|\mathcal{R}_h(x_0)| \leq C|h|^N$.

[2]For smooth solutions, terms with high frequency must have small amplitude, a fact that may be exploited in the derivation of error bounds [19], [34]. This point will not be studied here.

The theorem reduces the estimation of the local error to the estimation of the quantities $\widetilde{A}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) - A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h)$. These are errors arising in the quadrature of scalar smooth trigonometric functions and *are completely independent of the function* f.

It is assumed hereafter that the integrator is consistent. We first analyze the local error in the limit $h \to 0$. The condition $a = 1$ implies that the first term in the right-hand side of (46) vanishes. Furthermore, from Remark 7, $\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h) = \mathcal{O}(h)$ as $h \to 0$ and we conclude that $\widetilde{\phi}_h(x) - \phi_h(x) = \mathcal{O}(h^2)$. Note that, for the word with $\mathbf{0} \in \mathbb{Z}^d$ as its only letter, $\widetilde{A}_{\mathbf{0}}(h) - A_{\mathbf{0}}(h) = 0$. Moreover, in view of Remark 8, for each $n \leq N-1$, the n-th term in the sum in (46) is actually $\mathcal{O}(h^{n+1})$ rather than $\mathcal{O}(h^n)$.

If, additionally, the underlying univariate quadrature rule is second-order accurate, i.e. $\sum b_j c_j = 1/2$, then $\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h) = \mathcal{O}(h^2)$, and the integrator will be second order accurate, $\widetilde{\phi}_h(x) - \phi_h(x) = \mathcal{O}(h^3)$, provided that hypothesis 3 holds with $N \geq 3$.

The argument may be taken further to translate accuracy properties of the associated quadrature and cubature rules into accuracy properties of the integrator in the limit $h \to 0$. In this way one recovers the order conditions for splitting methods listed in [4] (cf. [29]). We shall not pursue that path: our interest lies in the size of the local error when h is not small relative to the periods present in the dynamics, a scenario that we discuss next.

It is well known that for a quadrature rule that is exact for polynomials of degree $\leq \sigma$,

$$|\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h)| \leq C|\mathbf{k}\cdot\omega|^{\sigma+1}h^{\sigma+1},$$

for a constant C that only depends on the rule. Therefore for the quadrature errors to be small it is *necessary that* $|h|$ *be small with respect to* $\min(1/|\mathbf{k}\cdot\omega|)$, where the minimum is extended to all $\mathbf{k}$ with $\mathbf{k}\cdot\omega \neq 0$. We reach the unwelcome conclusion that the size of the bound in (46) depends *both* on the size of the perturbation f *and* on ω.

Example 1 Consider the familiar Strang splitting, $r = 2$,

$$a_1 = 1/2, \quad a_2 = 1/2, \quad b_1 = 1, \quad b_2 = 0. \tag{47}$$

The underlying quadrature formula is the second-order accurate midpoint rule. For this integrator, for each $\mathbf{k}$ such that $\mathbf{k}\cdot\omega \neq 0$,

$$\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h) = \exp((1/2)i\mathbf{k}\cdot\omega h) - \frac{\exp(i\mathbf{k}\cdot\omega h) - 1}{i\mathbf{k}\cdot\omega h} \tag{48}$$

(for $\mathbf{k}\cdot\omega = 0$, $\widetilde{A}_{\mathbf{k}}(h) = A_{\mathbf{k}}(h) = 1$). An elementary computation leads to the bound

$$|\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h)| \leq \frac{1}{24}|\mathbf{k}\cdot\omega|^2 h^2,$$

where the constant $1/24$ cannot be improved if the inequality has to hold for arbitrary h. □

Remark 9 The dependence on ω of the local error is not an artifact introduced by our method of analysis. Here is an example. Consider the forced spring (see the Appendix), $(d/dt)p = -\omega^2 q + F$, $dq/dt = p$, where $F \neq 0$ is a time-independent force and $\omega > 0$.

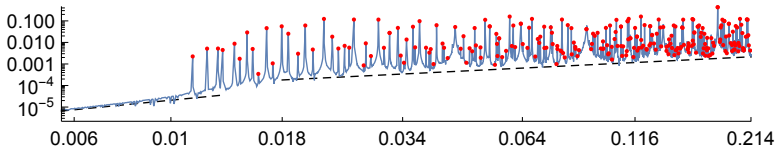

Figure 1: Energy error at time $T = 50$ as a function of h (in doubly logarithmic scale) for Example 2. The discontinuous straight lines correspond to $\mathcal{O}(h^2)$ (left) and $\mathcal{O}(h)$ (right). Small circles have been located at points whose abscissa is a value of h that leads to a first order numerical resonance (Section 5.3). It is apparent that those points give rise to local maxima of the error.

This is the Hamiltonian system with Hamiltonian $H = (1/2)p^2 + (\omega^2/2)q^2 - qF$ or, in action-angle variables

$$H = \omega a - \sqrt{\frac{2a}{\omega}}\sin\theta\, F = \omega a - \frac{1}{2i}\sqrt{\frac{2a}{\omega}}\exp(i\theta)F + \frac{1}{2i}\sqrt{\frac{2a}{\omega}}\exp(-i\theta)F.$$

There are two Fourier modes $k = \pm 1$ in the perturbation.

Choose initial conditions $p_0 = 1$, $q_0 = 0$ (with kinetic energy $1/2$ and no potential energy in the spring). If $h/(1/\omega) = 2\pi$, after one time step, the true solution has $p(h) = 1$ and Strang's method (47) yields an approximation $\widetilde{p}(h) = 1 - hF$; therefore a bound of the form $|\widetilde{p}(h) - p(h)| \leq C|h|^{\sigma+1}$, $|h| < h_0$, with C and h_0 independent of ω cannot exist for $\sigma > 0$. Note that, after m steps, the error in p will be mh!

Example 2 In order to observe the behavior of the Strang's method (47) in problems more involved than the scalar example in the last remark, we have integrated the Hamiltonian problem with $d = 5$ degrees of freedom with Hamiltonian function from [22, Chapter XIII.9]

$$\frac{1}{2}\sum_{j=1}^{5}\left((p^j)^2 + {\omega_j}^2(q^j)^2\right) + U(q),$$

with

$$U(q) = \frac{1}{8}(q^1 q^2)^2 + \left(\frac{1}{20} + q^2 + q^3 + q^4 + \frac{5}{2}q^5\right)^4$$

and $\omega_1 = 1$, $\omega_2 = \omega_3 = 70$, $\omega_4 = 70\sqrt{2}$, $\omega_5 = 2\omega_2$. The system is split by dividing the Hamiltonian into its harmonic (quadratic) part, with linear dynamics, and the perturbation corresponding to U (see the Appendix). We integrated this problem over the (very long) interval $0 \leq t \leq 35000$ with the initial condition (also taken from [22])

$$p_0 = \left(-\frac{1}{5}, \frac{3}{5}, \frac{7}{10}, -\frac{9}{10}, \frac{4}{5}\right), \qquad q_0 = \left(1, \frac{3}{10\omega_2}, \frac{4}{5\omega_2}, -\frac{11\sqrt{2}}{10\omega_4}, \frac{7}{10\omega_2}\right),$$

for which the energy in the harmonic part is 4.225. Figure 1 gives the error in the Hamiltonian at time $t = 50$ as a function of h. Two different regimes are apparent in the figure:

1. For h small, the error in the Hamiltonian is very approximately Ch^2, as it corresponds to the second order of accuracy of Strang's splitting. In this regime the oscillatory nature of the problem is not relevant and the integrator may be analyzed by standard techniques, i.e. expansion of the local error *in powers of* h and transference, using stability, of local error bounds to bounds of the global error.

2. For h large, the error presents a very irregular behavior. This is due to the highly oscillatory character of the solution and, as we shall now describe, may be analyzed via the *word series expansion* of the local error.

5.3 Processing

It is clear that, as distinct from the global error, the quadrature error in (48) varies regularly as h varies. The irregularities in Fig. 1 stem from cancelations, due to the oscillations, of local errors in consecutive time steps. For this reason, sharp error estimates in highly oscillatory problems (see e.g. [19], [34]) do not bound the local error and then sum the bounds; they rather sum first and bound later, so as to take advantage of possible cancelations. We use here an alternative approach that exploits the idea of *processing* that goes back to Butcher [7]. The presentation here follows [26].

If χ_h is a near-identity mapping in $\mathbb{R}^D$ and $\widetilde{\phi}_h$ is an integrator, the mapping

$$\widehat{\phi}_h = \chi_h^{-1} \circ \widetilde{\phi}_h \circ \chi_h \tag{49}$$

defines a *processed* numerical integrator. For $m \geq 1$

$$\widehat{\phi}_h^m = \left(\chi_h^{-1} \circ \widetilde{\phi}_h \circ \chi_h\right)^m = \chi_h^{-1} \circ \widetilde{\phi}_h^m \circ \chi_h; \tag{50}$$

therefore to advance m steps with $\widehat{\phi}_h$ one may preprocess the initial condition to find $\chi_h(x_0)$, advance 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. In practice, the idea of processing is useful if χ_h may be chosen in such a way that $\widehat{\phi}_h$ is more accurate in some sense than the original $\widetilde{\phi}_h$: one then obtains extra accuracy at the (hopefully small) price of having to perform the processing (this gives rise to Butcher's notion of effective order [7], [8]). Here we use the idea of processing as a *technique of analysis*. We shall process the splitting method (40) by means of a mapping χ_h with an expansion in word series: $\chi_h(x) = W_{\kappa(h)}(x)$, $\kappa(h) \in \mathcal{G}$ (see Remark 4). Then the processed integrator will be expressible as an extended word series with coefficients of the form $(h\omega, \widehat{\alpha}(h)) \in \overline{\mathcal{G}}$. By implication, the local error will be of the form (45) with $\widehat{A}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) - A_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h)$ in lieu of $\widetilde{A}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) - A_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h)$. (We have used the obvious notation $h^n \widehat{A}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) = \widehat{\alpha}_{\mathbf{k}_1 \cdots \mathbf{k}_n}$ for the scaled coefficients of the processed method and will similarly set $h^n K_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) = \kappa_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h)$.) Our policy is to determine the processing, i.e. to determine $\kappa(h) \in \mathcal{G}$, in such a way that $\widehat{A}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) - A_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) = 0$, for all *oscillatory words*. We then may hope that the analysis of the processed integrator would be free from the difficulties usually associated with integrators of oscillatory problems. Finally the results on the processed integrator obtained in this way will be translated into results for the original $\widetilde{\phi}_h$.

5.3.1 First-order numerical resonances

The conjugation (49) of the mappings may be translated with the help of (34) into the equation

$$(h\omega, \widehat{\alpha}(h))\bigstar(0, \kappa(h)) = (0, \kappa(h))\bigstar(h\omega, \widetilde{\alpha}(h)) \tag{51}$$

for the coefficients. For words with one letter, (51) implies, according to the definition of $\bigstar$:

$$\exp(i\mathbf{k}\cdot\omega h)\, K_{\mathbf{k}}(h) + \widehat{A}_{\mathbf{k}}(h) = \widetilde{A}_{\mathbf{k}}(h) + K_{\mathbf{k}}(h).$$

There are two cases to be analyzed. We first look at words (including $\mathbf{k} = \mathbf{0}$) that are not oscillatory, i.e. $\mathbf{k}\cdot\omega = 0$. The value $K_{\mathbf{k}}(h)$ drops from (51) and may be regarded as a free parameter. In addition, $\widehat{A}_{\mathbf{k}}(h) = \widetilde{A}_{\mathbf{k}}(h) = 1$ and therefore $\widehat{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h) = 0$. We then consider oscillatory one-letter words $\mathbf{k}$, $\mathbf{k}\cdot\omega \neq 0$, and according to our policy, we try to get $\widehat{A}_{\mathbf{k}}(h) = A_{\mathbf{k}}(h)$. This leads to

$$K_{\mathbf{k}}(h) = \frac{\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h)}{\exp(i\mathbf{k}\cdot\omega h) - 1}, \tag{52}$$

provided that $\exp(i\mathbf{k}\cdot\omega h) \neq 1$. If $\mathbf{k}\cdot\omega \neq 0$ and $\exp(i\mathbf{k}\cdot\omega h) = 1$, we say that a *first-order numerical resonance* occurs. When this happens, $K_{\mathbf{k}}(h)$ drops from (51) and $\widehat{A}_{\mathbf{k}}(h) = \widetilde{A}_{\mathbf{k}}(h)$. As a consequence, in general, $\widehat{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h)$ will not vanish.

If, for given h, there is no first-order numerical resonance, then the expansion of the local error only contains terms corresponding to words with two or more letters. In analogy with Theorem 4 (details will not be given), it is then possible to bound the local error of the processed integrator by Ch^2, with C independent of ω. This in turn will lead to a $C'h$ bound for the global error of the processed integrator and, after taking into account the pre- and postprocessing to a $C''h$ bound for the global error in the method $\widetilde{\phi}_h$ being analyzed. The constant C'' may be chosen to be independent of h, provided that h is bounded away from the resonances; it worsens as h gets closer to a numerical resonance in view of (52). This explains the troughs in Fig. 1.

On the other hand, if for given h there is at least one numerically resonant $\mathbf{k} \in \mathcal{I}$, then, processing is of no help in removing the ω-dependent quadrature error of the original, unprocessed method. This was only to be expected because at a numerical resonance, as shown in Remark 9, the global error in $\widetilde{\phi}_h$ may actually be large (cusps in Fig. 1).

Remark 10 By using the operation $\bigstar$ to compute the expansion of the m-fold compositions $\widetilde{\phi}_h^m$ and ϕ_h^m, we find after some simple algebra that if $\pm\mathbf{l}$ are numerically resonant wavenumbers and $\exp(i\mathbf{k}\cdot\omega h) \neq 1$ for $\mathbf{k} \neq \mathbf{0}, \pm\mathbf{l}$, then the error over m steps has an expansion

$$\begin{aligned}\widetilde{\phi}_h^m(x_0) - \phi_h^m(x_0) = \\ mh\left(\widetilde{A}_{\mathbf{l}}(h) f_{\mathbf{l}}(x_0) + \widetilde{A}_{-\mathbf{l}}(h) f_{-\mathbf{l}}(x_0)\right) + \\ h \sum_{\mathbf{k}\in\mathcal{I}\setminus\{\mathbf{0},\pm\mathbf{l}\}} \frac{\exp(i\mathbf{k}\cdot\omega mh) - 1}{\exp(i\mathbf{k}\cdot\omega h) - 1}\left(\widetilde{A}_{\mathbf{k}}(h) - A_{\mathbf{k}}(h)\right) f_{\mathbf{k}}(x_0) + \cdots\end{aligned}$$

Thus the mh growth as m increases with fixed h we already encountered in Remark 9 holds for general integrators and general differential equations.

5.3.2 Higher-order resonances

Assuming that h does not satisfy any first-order numerical resonance, one may go a step further and look at words with two letters $\mathbf{kl}$; these are oscillatory if $(\mathbf{k}+\mathbf{l})\cdot\omega \neq 0$. Now (51) implies

$$\exp(i(\mathbf{k}+\mathbf{l})\cdot\omega h)\, K_{\mathbf{kl}}(h) + \widehat{A}_{\mathbf{k}}(h)\exp(i\mathbf{l}\cdot\omega h)K_{\mathbf{l}}(h) + \widehat{A}_{\mathbf{kl}}(h) \\ = \widetilde{A}_{\mathbf{kl}}(h) + K_{\mathbf{k}}(h)\widetilde{A}_{\mathbf{l}}(h) + K_{\mathbf{kl}}(h).$$

Whenever $\exp(i(\mathbf{k}+\mathbf{l})\cdot\omega h) = 1$ (second order numerical resonance), the value of $K_{\mathbf{kl}}(h)$ cannot be chosen to ensure that $\widehat{A}_{\mathbf{kl}}(h) = A_{\mathbf{kl}}(h)$. A similar consideration applies for nonoscillatory words with n letters when $\exp(i(\mathbf{k}_1+\cdots+\mathbf{k}_n)\cdot\omega h) = 1$. When no numerical resonance takes place, the equation (51) may be used to find values $\kappa_w(h)$, $w \in \mathcal{W}$ such that, on the one hand, define an element $\kappa(h)$ that belongs to $\mathcal{G}$, (i.e. the shuffle relations hold) and, on the other, ensure that $\widehat{A}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) - A_{\mathbf{k}_1\cdots\mathbf{k}_n}(h) = 0$, for all *oscillatory words*. This will be proved in Remark 13 below by using modified systems.

Remark 11 Since pre- and postprocessing introduce in any case $\mathcal{O}(h)$ errors, the processing technique used here yields $\mathcal{O}(h)$ bounds for the global error of $\widetilde{\phi}_h$ even for values of h where there is no first-order or second-order numerical resonances. Fig. 1 shows that, for this simulation, $\mathcal{O}(h^2)$ global error bounds cannot exist if h is large relative to the periods present in the dynamics.

5.4 Modified equations and modified Hamiltonians

Modified equations [20], [9], [35], [22] provide a useful means to describe the behaviour of numerical integrators.

5.4.1 Modified system using one letter words

We look for a (one-letter word) modified system

$$\frac{d}{dt}\widetilde{x} = \overline{W}_{(\omega,\widetilde{\beta}(h))}(\widetilde{x}), \tag{53}$$

where $\widetilde{\beta} \in \mathfrak{g}$, $\widetilde{\beta}_w = 0$ for $w \in \mathcal{W}_n$, $n > 1$ and the coefficients $\widetilde{\beta}_{\mathbf{k}}(h)$, $\mathbf{k} \in \mathcal{I}$ are chosen in such a way that, for words with one letter, the extended word series expansion of the h-flow $\widetilde{\phi}_h^{[1]}$ of (53) matches the corresponding expansion for the integrator $\widetilde{\phi}_h$. We recall that the system being solved is also of the form (53) with the coefficients β given in (32).

By integrating the system (53) by the procedure outlined in Section 3 and imposing that its flow matches $\widetilde{\phi}_h$ to the desired order, we find the condition

$$\frac{\exp(i\mathbf{k}\cdot\omega h) - 1}{i\mathbf{k}\cdot\omega h}\,\widetilde{\beta}_{\mathbf{k}}(h) = \widetilde{A}_{\mathbf{k}}(h) \tag{54}$$

(it is understood that the fraction takes the value 1 if $\mathbf{k} \cdot \omega = 0$). For $\mathbf{k} = \mathbf{0}$ or for any one letter word that is not oscillatory, this implies $\widetilde{\beta}_{\mathbf{k}}(h) = 1$. For an oscillatory one-letter word $\mathbf{k} \neq \mathbf{0}$, $\mathbf{k} \in \mathcal{I}$, if h is such that $i\mathbf{k} \cdot \omega h = 2\pi j$ for some integer $j \neq 0$ (first order numerical resonance), then the fraction in (54) vanishes and the equation for $\widetilde{\beta}_{\mathbf{k}}(h)$ will in general not be solvable. Thus first-order numerical resonances are obstructions to the construction of the modified system. (In fact the counterexample in Remark 9 proves that modified systems of the form envisaged here do not exist at numerical resonances.) When h is bounded away from resonances, $\widetilde{\phi}_h^{[1]} - \widetilde{\phi}_h = \mathcal{O}(h^2)$ with the implied constant independent of ω (as in Theorem 4).

Example 3 For Strang's method, if $\mathbf{k}$ is oscillatory and there is not a numerical resonance, (54) yields the value

$$\widetilde{\beta}_{\mathbf{k}}(h) = \frac{\mathbf{k} \cdot \omega h}{2\sin(\mathbf{k} \cdot \omega h/2)}. \qquad \square$$

Remark 12 Assume that, for given h, the modified system above has been found. We may then try to find a change of variables $\widetilde{x} = W_{\kappa(h)}(\widetilde{X})$ so that in the new variables the modified vector field matches the field $\overline{W}_{(\omega,\beta)}(X)$ for words with one letter. According to Section 3, we have to impose that $\beta \star \kappa(h) + \xi_\omega \kappa(h)$ and $\kappa(h) \bigstar \widetilde{\beta}(h)$ coincide for words with one letter. This leads to

$$i(\mathbf{k} \cdot \omega)\, \kappa_{\mathbf{k}}(h) + 1 = \widetilde{\beta}_{\mathbf{k}}(h).$$

If $\mathbf{k}$ is not oscillatory, $\kappa_{\mathbf{k}}(h)$ is free because, as noted above, $\widetilde{\beta}_{\mathbf{k}}(h) = 1$. For $\mathbf{k}$ oscillatory, $\kappa_{\mathbf{k}}(h)$ is uniquely determined. By using (54), a little algebra shows that the value of $\kappa_{\mathbf{k}}(h)$ found in this way is the same we obtained in (52). Thus the change of variables $W_{\kappa(h)}$ we used for processing may be seen as determined by the requirement that, in the new variables and for nonoscillatory one-letter words, the modified vector field of the unprocessed integrator reproduces the vector field being integrated.

5.4.2 Other modified systems

More precise modified systems may be constructed by successively adding to the modified vector field contributions from words of 2, 3, ... letters. For the n-th of these modified systems, the modified vector field has $\widetilde{\beta}_w = 0$ for words with more than n letters and we impose that, for words with n or fewer letters, the extended word series expansion of the h-flow $\widetilde{\phi}_h^{[n]}$ matches the corresponding expansion for the integrator $\widetilde{\phi}_h$.

For two-letter words, proceeding as in the case of one-letter word modified systems, we obtain the condition

$$\frac{\exp(i(\mathbf{k}+\mathbf{l}) \cdot \omega h) - 1}{i(\mathbf{k}+\mathbf{l}) \cdot \omega h}\, \widetilde{\beta}_{\mathbf{kl}}(h) + hA_{\mathbf{kl}}(h)\widetilde{\beta}_{\mathbf{k}}(h)\widetilde{\beta}_{\mathbf{l}}(h) = h\widetilde{A}_{\mathbf{kl}}(h).$$

In general, the resulting equation is of the form

$$\frac{\exp(i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \omega h) - 1}{i(\mathbf{k}_1 + \cdots + \mathbf{k}_n) \cdot \omega h}\, \widetilde{\beta}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) - h^{n-1}\widetilde{A}_{\mathbf{k}_1 \cdots \mathbf{k}_n}(h) = \mathcal{O}(h^{n-1}),$$

where the right-hand side depends polynomially on the coefficients $\widetilde{\beta}_w(h)$ of words w with less than n letters. Such an equation can be solved for $\widetilde{\beta}_{\mathbf{k}_1\cdots\mathbf{k}_n}(h)$ provided that there is no numerical resonance, $(\mathbf{k}_1+\cdots+\mathbf{k}_r)\cdot\omega h = 2\pi j$, $r \leq n$, $j \neq 0$. In the limit where the length of the words increases indefinitely one obtains, if there is no numerical resonance of any order, a modified system whose formal h-flow exactly reproduces the expansion of $\widetilde{\phi}_h$.

As explained in Section 2 the modified systems found in this way are Hamiltonian whenever the system being integrated is Hamiltonian. Furthermore the modified Hamiltonian functions are easily expressible in terms of brackets.

Example 4 For the linear forced oscillator in Remark 9, each word basis functions associated with words with two or more letters vanishes. In this case the modified systems above with $n > 1$, coincide with the modified system using only one-letter words and the latter is exact, ie $\widetilde{\phi}_h^{[1]} = \widetilde{\phi}_h$. The (exact) modified Hamiltonian is

$$\frac{1}{2}p^2 + \frac{\omega^2}{2}q^2 - q\,\widetilde{\beta}_1(h)\,F,$$

and therefore in the particular case of Strang's method we have

$$\frac{1}{2}p^2 + \frac{\omega^2}{2}q^2 - q\,\frac{\omega h}{2\sin(\omega h/2)}\,F.$$

For nonresonant h, the effect of using the splitting method is to alter the value of the applied force. Unless $|\omega h| \ll 1$ the misrepresentation of the force introduced by the discretisation will be large. We emphasize that, as distinct from the situation when using conventional modified equations based on series of powers of h, the analysis here does not require h to be small. □

Example 5 For the problem in Example 2 we have measured the variation as t increases of the true energy H of the numerical solution and of the corresponding variations of the energies in the modified one-letter-word Hamiltonian and two-letter-word Hamiltonians. We used $h = 0.7974$ as this, which is more than 12 times larger than the period of the fastest oscillator, avoids first and second order resonances. The results given in Figs. 2–3 clearly bear out how the one-word-letter modified system matches the numerical solution much better than the system being integrated, but not as well as the two-word-letter modified system.

Remark 13 The idea in Remark 12 may be extended. Assume that there is no numerical resonance of any order so that it is possible to find a modified equation whose formal flow exactly reproduces the expansion of the integrator. By using Theorem 1 we may bring the modified system to a normal form where the contribution of all oscillatory words have disappeared. This implies that a processor has been found such that the expansion of the local error of the processed method does not contain contributions of oscillatory words.

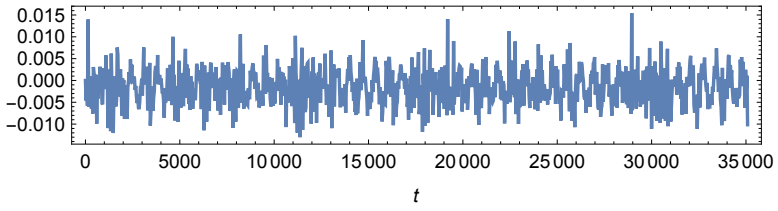

Figure 2: Variation of the true Hamiltonian (energy) evaluated at the numerical solution as a function of t for Example 2 ($h = 0.7974$).

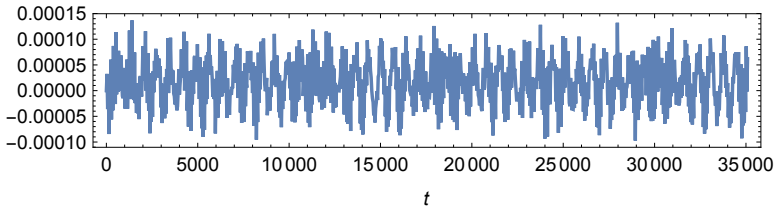

Figure 3: Variation of the one-letter-word Hamiltonian evaluated at the numerical solution as a function of t for Example 2 ($h = 0.7974$). The vertical scale is 100 times larger than in Fig. 2.

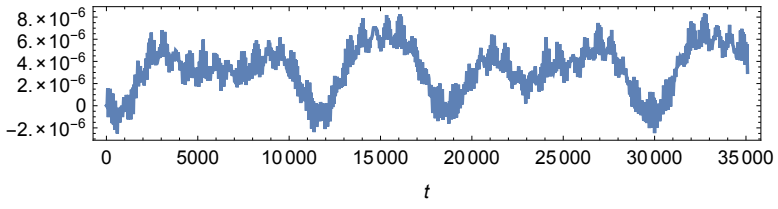

Figure 4: Variation of the two-letter-word Hamiltonian evaluated at the numerical solution as a function of t for Example 2 ($h = 0.7974$).

We close this section with an observation. As noted before, *numerical resonances* ($(\mathbf{k}_1 + \cdots + \mathbf{k}_r) \cdot \omega h = 2\pi j$, $j \neq 0$) obstruct the construction of modified systems; nonoscillatory words ($(\mathbf{k}_1 + \cdots + \mathbf{k}_r) \cdot \omega = 0$) cause no trouble in that connection. On the other hand, the nonoscillatory character of a word *is* an obstruction to its elimination by changing variables. Processing, that as we have just seen is equivalent to finding modified problems *and* then changing variables, is hampered by both numerical resonances and non-oscillatory terms. See in this connection (52), whose denominator vanishes both at numerical resonances *and* for nonoscillatory words.

6 Technical results

This section is devoted to more technical material.

6.1 Algebraic results

6.1.1 Differential operators

In (3) we associated with each vector field f_a in (2) a first-order linear differential operator E_a. With each word $w = a_1 \cdots a_n$, $n > 0$ we now associate the n-th order (linear) differential operator E_w obtained by composition:

$$E_{a_1 \cdots a_n} g(x) = E_{a_1} \cdot E_{a_2} \cdots E_{a_n} g(x);$$

$E_\emptyset$ is defined as the identity operator. Finally, with each $\gamma \in \mathbb{C}^{\mathcal{W}}$ we associate the formal series of linear differential operators

$$D_\gamma = \sum_{w \in \mathcal{W}} \gamma_w E_w.$$

Two non-empty words $w = a_1 \cdots a_m$, $w' = a'_1 \cdots a'_n$ may be *concatenated* [31] to give rise to the word $ww' = a_1 \cdots a_m a'_1 \cdots a'_n$. In addition $\emptyset w = w\emptyset = w$ for each $w \in \mathcal{W}$. Clearly, concatenation of words corresponds to composition of the associated operators: $E_{ww'} = E_w \cdot E_{w'}$.

Each word $a_1 \cdots a_n$ may be *deconcatenated* in $n + 1$ different ways: $\emptyset(a_1 \cdots a_n)$, $(a_1)(a_2 \cdots a_n)$, $\ldots$, $(a_1 \cdots a_n)\emptyset$; these feature in the definition (11) of the convolution product. This observation leads to the following rule for the composition of two series of operators:

$$D_\gamma \cdot D_{\gamma'} = D_{\gamma \star \gamma'}. \tag{55}$$

6.1.2 The shuffle algebra

The product $\sqcup\!\sqcup$ may be extended in a bilinear way from words to linear combinations of words, i.e. if μ_j, $\mu_{j'}$ are scalars:

$$\big(\sum_j \mu_j w_j\big) \sqcup\!\sqcup \big(\sum_{j'} \mu_{j'} w_{j'}\big) = \sum_{j,j'} \mu_j \mu_{j'} w_j \sqcup\!\sqcup w_{j'}.$$

When endowed with this operation, 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 [31], [18], [28]). Note that $\mathrm{sh}(A)$ is graded by the number of letters of the words.

Deconcatenation defines a coproduct and turns $\mathrm{sh}(A)$ into a (commutative, connected, graded) *Hopf algebra* [6]. It is well known that the dual vector space of a Hopf algebra is automatically endowed with a product operation. Here the dual of $\mathrm{sh}(A)$ may be identified in a natural way with $\mathbb{C}^{\mathcal{W}}$ by associating with each linear form ℓ on $\mathrm{sh}(A)$ the family of coefficients $\gamma_w = \ell(w)$, $w \in \mathcal{W}$. After this identification, the product in the dual of $\mathrm{sh}(A)$ coincides with the convolution product $\star$ defined in (11). The sets $\mathcal{G}$ and $\mathfrak{g}$ in Section 2 are then respectively the group of characters and the Lie algebra of infinitesimal characters of the Hopf algebra $\mathrm{sh}(A)$; well known results on Hopf algebras show that $\exp_\star$ in (16) maps $\mathfrak{g}$ onto $\mathcal{G}$ and has an inverse given by $\log_\star$ in (17), see e.g. [28], [18].

6.1.3 The actions of $\mathcal{G}$ and $\mathfrak{g}$ on word series

As shown e.g. in [14], there is a narrow connection between the word basis functions $f_w(x)$ and the operators E_w, $w \in \mathcal{W}$:

$$f_w(x) = E_w x,$$

(in the right-hand side, with an abuse of notation, x denotes the identity function that maps each D-vector into itself). As a consequence we have the following correspondence between word series and series of operators

$$W_\gamma(x) = D_\gamma x. \tag{56}$$

The use of series of *operators* is common in control theory and dynamical systems; word series, being series of *functions*, provide a more convenient way to study numerical integrators.

The operators E_a, $a \in A$, are derivations: $E_a(gh) = (E_a g)h + g(E_a h)$ for each pair of scalar functions g, h. Iteration yields:

$$\begin{aligned} E_{ab}(gh) &= (E_{ab}g)(E_\emptyset h) + (E_b g)(E_a h) + (E_a g)(E_b h) + (E_\emptyset g)(E_{ab} h), \\ E_{abc}(gh) &= (E_{abc}g)(E_\emptyset h) + (E_{bc} g)(E_a h) + (E_{ac} g)(E_b h) + (E_c g)(E_{ab} h) + \\ &\quad (E_{ab}g)(E_c h) + (E_b g)(E_{ac} h) + (E_a g)(E_{bc} h) + (E_\emptyset g)(E_{abc} h), \end{aligned}$$

etc. Note that, in the first of these identities, the pairs of words $(ab, \emptyset)$, (a, b), (b, a), $(\emptyset, ab)$ that feature in the right-hand side are precisely those whose shuffle product gives rise to the word ab that appears in the left-hand side. A similar observation may be made in the second identity. In general, if $w \in \mathcal{W}_m$, $w' \in \mathcal{W}_n$ and $w \sqcup\!\sqcup w' = \sum_j w_j$, then the $w_j \in \mathcal{W}_{m+n}$ are precisely those words for which $(E_w g)(E_{w'} h)$ is one of the 2^{m+n} terms of the expansion of $E_{w_j}(gh)$.[3] This result may be used in combination with the shuffle relations (12) to prove (see e.g. [18], Theorem 2) that, for $\gamma \in \mathcal{G}$,

$$D_\gamma(gh) = D_\gamma(g)\, D_\gamma(h).$$

[3] Algebraically, the action of the operators E_w on products gh defines a coproduct [6]; the shuffle product is obtained from this coproduct by duality [31, Section 1.5].

By considering the coordinate mappings $g(x) = x^j$, $h(x) = x^\ell$ and (56) we conclude that

$$x^i(W_\gamma(x))\, x^i(W_\gamma(x)) = D_\gamma(x^i) D_\gamma(x^\ell) = D_\gamma(x^i x^\ell)$$

and then linearity shows that, for each polynomial mapping P, $P(W_\gamma(x)) = D_\gamma P(x)$. It follows that

$$g(W_\gamma(x)) = D_\gamma g(x), \qquad \gamma \in \mathcal{G}; \tag{57}$$

for any (scalar or vector valued) smooth mapping g. Thus $D_\gamma g(x)$ provides the formal expansion of the composition $g(W_\gamma(x))$ provided that the coefficients γ belong to the group $\mathcal{G}$.

The proof of the formula (13), that defines an action of the group $\mathcal{G}$ on the vector space of all word series, is now easy:

$$W_\delta(W_\gamma(x)) = D_\gamma W_\delta(x) = (D_\gamma \cdot D_\delta)\, x = D_{\gamma \star \delta}\, x = W_{\gamma \star \delta}(x);$$

we have successively used (57), (56), (55) and once more (56).

In (18), the expression $(\partial_x W_\delta(x)) W_\beta$ is the result of applying to the word series $W_\delta(x)$, the first-order differential operator associated with the formal vector field $W_\beta(x)$. The formula then reveals that the action of the algebra $\mathfrak{g}$ on word series corresponds to the operation $\beta \star \delta$.

6.1.4 Linear differential equations

In Section 2 it was proved that the initial value problem (21) has a unique solution with $\alpha(t) \in \mathbb{C}^{\mathcal{W}}$ for each t. We show here that in fact $\alpha(t) \in \mathcal{G}$. We use the following auxiliary result:

Lemma 1 *Assume that $\eta \in \mathbb{C}^{\mathcal{W}}$ is such that, for some positive integer n, and for each $w \in \mathcal{W}_l$, $w' \in \mathcal{W}_m$, $l + m \leq n$, the shuffle relation (12) hold. Then, for $\beta \in \mathfrak{g}$, $w \in \mathcal{W}_l$, $w' \in \mathcal{W}_m$, $l + m \leq n$, with $w \sqcup\!\sqcup w' = \sum_j w_j$:*

$$\sum_j (\eta \star \beta)_{w_j} = \eta_w (\eta \star \beta)_{w'} + (\eta \star \beta)_w \eta_{w'}.$$

Proof: Since $\beta \in \mathfrak{g}$, there exists a curve $\gamma(t)$ in $\mathcal{G}$ such that (15) holds. The hypothesis of the lemma then allows us to write:

$$\sum_j (\eta \star \gamma(t))_{w_j} = (\eta \star \gamma(t))_w (\eta \star \gamma(t))_{w'}.$$

The result is obtained by applying $d/dt|_{t=0}$ to both sides of this equality. □

Now consider the solution $\alpha(t) \in \mathbb{C}^{\mathcal{W}}$ of (21). We shall prove by induction on n that for each t

$$\sum_j \alpha(t)_{w_j} = \alpha(t)_w \alpha(t)_{w'} \tag{58}$$

for $w \in \mathcal{W}_l$, $w' \in \mathcal{W}_m$, $l + m \leq n$, with $w \sqcup\!\sqcup w' = \sum_j w_j$.

This trivially holds for $n = 0$, since $\alpha(t)_\emptyset = 1$ for all t. Assume that (58) is satisfied for some $n \geq 0$, and choose $w \in \mathcal{W}_l$, $w' \in \mathcal{W}_m$, $l + m \leq n + 1$, with $w \sqcup\!\sqcup w' = \sum_j w_j$. From (21) we find

$$\frac{d}{dt}\left(\sum_j \alpha(t)_{w_j} - \alpha(t)_w \alpha(t)_{w'}\right) = \sum_j (\alpha(t) \star \beta(t))_{w_j} - (\alpha(t) \star \beta(t))_w \alpha(t)_{w'} - \alpha(t)_w (\alpha(t) \star \beta(t))_{w'}$$

and the lemma implies that the right-hand side of this equality vanishes. Since (58) holds at $t = 0$, it does so for each value of t.

6.2 Proof of Theorem 1

We simplify the system (36) by performing a sequence of changes of variables with coefficients $\kappa^{([1])}$, $\kappa^{([2])}$, ... in $\mathcal{G}$ so that the change defined by $\kappa^{([n])}$ simplifies the coefficients of the vector field associated with words with n letters and leaves unaltered the coefficients associated with shorter words. The element $\kappa^{([n])}$ is sought in the form $\exp_\star(\lambda^{([n])})$ where $\lambda^{([n])} \in \mathfrak{g}$ and $\lambda_w^{([n])} = 0$ if $w \in \mathcal{W} \setminus \mathcal{W}_n$. If $\beta^{[n-1]}$ and $\beta^{[n]}$ are respectively the vector fields before and after the n-th change of variables, and $w = \mathbf{k}_1 \cdots \mathbf{k}_n$, the equation (38) implies, after taking into account that $\kappa^{([n])}$ vanishes for nonempty words with less than n letters

$$i((\mathbf{k}_1 + \cdots \mathbf{k}_n) \cdot \omega)\kappa_w = \beta_w^{[n-1]} - \beta_w^{[n]}.$$

If $(\mathbf{k}_1 + \cdots \mathbf{k}_n) \cdot \omega \neq 0$ we may choose $\lambda_w^{[n]}$ to enforce $\beta_w^{[n]} = 0$. In other case, we set $\beta_w^{[n]} = \beta_w^{[n-1]}$ and $\lambda_w^{[n]} = 0$. The element $\lambda^{[n]}$ constructed in this way belongs to $\mathfrak{g}$ because the required shuffle relations hold (if shuffling two words leads to resonant words all the coefficients in $\lambda^{[n]}$ vanish; in the nonresonant case the coefficients $\lambda^{[n]}$ are proportional to the corresponding coefficients in $\beta^{[n-1]}$, which satisfy the shuffle relations). In turn $\beta^{([n]} \in \mathfrak{g}$ because

$$\beta^{[n]} = \kappa^{[n]} \star \beta^{[n-1]} \star \left(\kappa^{[n]}\right)^{-1} - (\xi_\omega \kappa^{[n]}) \star \left(\kappa^{[n]}\right)^{-1};$$

both terms of the right-hand side are in $\mathfrak{g}$ (the second is the value at $t = 0$ of

$$(d/dt)\Big(\left(\Xi_{\omega t}\kappa^{[n]}\right) \star \left(\kappa^{[n]}\right)^{-1}\Big)$$

and, as noted above $\Xi_{\omega t}\kappa^{[n]} \in \mathcal{G}$).

6.3 Proof of Theorem 4

For $(y, \theta) \in \Omega$ the function f in (25) is bounded and Lipschitz continuous. Therefore, for $|t|$ small $(y(t), \theta(t)) = \phi_t(y_0, \theta_0)$ is well defined and $|y(t) - y_0| \leq C_1|t|$, where

C_1 is a bound for $|f|$. A simple contradiction argument shows that $|y(h) - y_0| < R$ for $|h| < R/C_1$.

To deal now with the numerical solution, define the intermediate points (stages), $j = 1, \ldots, r$,

$$(y_j, \theta_j) = \phi^{(P)}_{b_j h}\Big(\phi^{(U)}_{a_j h}(y_{j-1}, \theta_{j-1})\Big) = \phi^{(P)}_{b_j h}(y_{j-1}, \theta_{j-1} + a_j h\omega).$$

If $|b_j h| < R/(C_1/r)$, the iteration of the argument used above ensures that $|y_j - y_{j-1}| < R/r$, $j = 1, \ldots, r$ and then the triangle inequality implies that $\widetilde{\phi}_h(y_0, \theta_0) \in \Omega$.

We shall use the notations $\overline{W}^{(N)}_{(t\omega,\alpha(t))}(x)$, $\overline{W}^{(N)}_{(h\omega,\widetilde{\alpha}(t))}(x)$, to refer to the result of suppressing all terms corresponding to words with N or more letters of the extended word series with coefficients $\alpha(t)$, $\widetilde{\alpha}(h)$ respectively (of course the alphabet is now $\mathcal{I}$ rather than $\mathbb{Z}^d$). In addition we set

$$\mathcal{R}^{(T)}_h(x) = \phi_h(x) - \overline{W}^{(N)}_{(h\omega,\alpha(h))}(x), \qquad \mathcal{R}^{(S)}_h(x) = \widetilde{\phi}_h(x) - \overline{W}^{(N)}_{(h\omega,\widetilde{\alpha}(h))}(x)$$

(the superscripts T and S mean 'true' and 'splitting'). Our task is to bound $\mathcal{R}_h(x_0) = \mathcal{R}^{(S)}_h(x_0) - \mathcal{R}^{(T)}_h(x_0)$. For $\mathcal{R}^{(T)}_h(x_0)$, by stopping the iterative procedure (see e.g. [14]) that leads to (28) we find the following representation:

$$\mathcal{R}^{(T)}_h(x_0) \quad = \quad \sum_{\mathbf{k}_1 \cdots \mathbf{k}_N \in \mathcal{W}_N} \int_0^h dt_N \, \exp(i\mathbf{k}_N \cdot \omega t_N) \cdots \int_0^{t_2} dt_1 \, \exp(i\mathbf{k}_1 \cdot \omega t_1) f_{\mathbf{k}_1 \cdots \mathbf{k}_N}(\phi_{t_1}(x_0)).$$

We know that $\phi_{t_1}(x_0) \in \Omega$ for $|h| \le h_0$, and therefore we may guarantee that $|\mathcal{R}^{(T)}_h(x_0)| \le C|h|^N$, with C depending only on $\mathcal{I}$ and bounds for the derivatives of the Fourier coefficients.

For $\mathcal{R}^{(S)}_h(x_0)$ we use a similar device. The key point is that (cf. (28)–(29))

$$\widetilde{\phi}_h(y_0, \theta_0) = (0, h(a_1 + \cdots + a_r)\omega) + (y(h), \eta(h)),$$

where $(y(t), \eta(t))$ is the solution of

$$\frac{d}{dt}\begin{bmatrix} y \\ \eta \end{bmatrix} = \sum_{\mathbf{k} \in \mathbb{Z}^d} \widetilde{\lambda}_{\mathbf{k}}(t) \, f_{\mathbf{k}}(y, \eta),$$

with piece-wise constant functions defined by

$$\widetilde{\lambda}_{\mathbf{k}}(t) = r b_j \exp\big(i\mathbf{k} \cdot \omega(a_1 + \cdots + a_j)h\big), \qquad (j-1)h/r \le t < jh/r, \quad 1 \le j \le r.$$

This differential system associated with $\widetilde{\phi}_h$ is very similar to the system (27) associated with ϕ_h, the difference being that for the former the complex exponentials are frozen at the times $(a_1 + \cdots + a_j)h$. After this observation the residual for $\widetilde{\phi}_h$ is bounded with the technique used for the true ϕ_h.

Acknowledgement. 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 Goverment (Consolidated Research Group IT649-13).

References

[1] V. I. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, 2nd ed., Springer, New York, 1988.

[2] V. I. Arnold, Mathematical Methods of Classical Mechanics, 2nd ed., Springer, New York, 1989.

[3] S. Blanes, F. Casas, J. A. Oteo, and J. Ros, The Magnus expansion and some of its applications, Physics Reports 470 (2009), 151-238.

[4] S. Blanes, F. Casas, A. Farrés, J. Laskar, J. Makazaga, and A. Murua, New families of symplectic splitting methods for numerical integration in dynamical astronomy, Appl. Numer. Math., 68 (2013), 58–72.

[5] G. Bogfjellmo and A. Schmeding, The Lie group structure of the Butcher group, Found. Comput. Math., to appear.

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

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

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

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

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

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

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

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

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

[15] P. Chartier, A. Murua, and J.M. Sanz-Serna, Higher-Order averaging, formal series and numerical integration III: error bounds, Found. Comput. Math., 15(2015), 591-612.

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

[17] J. Ecalle, Les Fonctions Résurgentes, Vols. I, II, III, Publ. Math. Orsay, (1981–1985).

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

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

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

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

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

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

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

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

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

[27] A. Murua, Formal series and numerical integrators, Part I: Systems of ODEs abd symplectic integrators, Appl. Numer. Math., 29 (1999), 221–251.

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

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

[30] A. Murua and J. M. Sanz-Serna, Computing normal forms and formal invariants of dynamical systems by means of word series, Nonlinear Analysis, Theory, Methods and Applications, to appear.

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

[32] J. A. Sanders, F. Verhulst, and J. Murdock, Averaging Methods in Nonlinear Dynamical Systems (2nd. ed.), Springer, New York, 2007.

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

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

[35] J. M. Sanz-Serna and M. P. Calvo, Numerical Hamiltonian Problems, Chapman and Hall, London, 1994.

[36] J. M. Sanz-Serna and A. Murua, Formal series and numerical integrators: some history and some new techniques, in Proceedings of the 8th International Congress on Industrial and Applied Mathematics (ICIAM 2015) (Lei Guo and Zhi-Ming eds.), Higher Edication, Press, Beijing, 2015, pp. 311–331.

Appendix: examples of perturbed integrable problems

Any system

$$\frac{d}{dt}w = Mw + F(w), \tag{59}$$

where M is a skew-symmetric $D \times D$ constant matrix, may be brought by a linear change of variables to the form:

$$\begin{array}{rcll} (d/dt)z^j &=& f^j(z,P,Q), & 1 \le j \le D-2d, \\ (d/dt)P^\ell &=& -\omega_\ell^2 Q^\ell + g^\ell(z,P,Q), & 1 \le \ell \le d, \\ (d/dt)Q^\ell &=& P^\ell + h^\ell(z,P,Q), & 1 \le \ell \le d \end{array} \tag{60}$$

Here d is the number of nonzero eigenvalue pairs $\pm i\omega_\ell$, $\omega_\ell > 0$, of M. In the unperturbed case, $f^i \equiv 0$, $g^\ell \equiv 0$, $h^\ell \equiv 0$, the system consists of d uncoupled harmonic oscillators with frequencies ω_ℓ, together with $D-2d$ trivial equations $(d/dt)z^j = 0$. The introduction of the variables a^ℓ, θ^ℓ such that

$$P^\ell = \sqrt{2\omega a^\ell}\cos\theta^\ell, \qquad Q^\ell = \sqrt{\frac{2a^\ell}{\omega_\ell}}\sin\theta^\ell, \qquad 1 \le \ell \le d, \tag{61}$$

takes now the system to the format (25) with $y = (z^1, \ldots, z^{D-2d}, a)$. The system (59) or (60) is a natural candidate to integration by splitting methods based on separating the linear part (that may be integrated in closed form) from the perturbation. The later may perhaps be treated by means of a numerical integrator with a very fine time step. In favorable instances, the perturbation may be integrated analytically in closed form; this is the situation in the following particular case of (60), commonly found in mechanics (D is even and $z = (p, q)$),

$$\begin{aligned}
(d/dt)p^j &= f^j(q, Q), \qquad 1 \le j \le D/2 - d, \\
(d/dt)q^j &= p^j, \\
(d/dt)P^\ell &= -\omega_\ell^2 Q^\ell + g^\ell(q, Q), \qquad 1 \le \ell \le d, \\
(d/dt)Q^\ell &= P^\ell.
\end{aligned} \tag{62}$$

Under the dynamics of the perturbation, p and P remain constant and q and Q grow linearly with t.

The system (62) is Hamiltonian if the forces f^j, g^ℓ derive from a potential. When that happens, the introduction of the canonical ($dP^j \wedge dQ^j = da^j \wedge d\theta^j$) action/angle variables in (61) preserves the Hamiltonian character of the equations of motion and even the value of the Hamiltonian function.

So far the unperturbed problem has been linear, but nonlinear cases may also be treated. Typically, integrable nonlinear problems may be brought to the form (41) with $\omega = \omega(y)$; a device commonly used e.g. in dynamical astronomy consists in fixing a relevant value y_0 of y, decomposing $\omega(y) = \omega(y_0) + \Delta(y)$ and seeing $(0, \Delta(y))$ as part of the perturbation.

There are many instances of perturbations of nonlinear integrable systems, after the introduction of suitable action/angle variables take the form (25). A well-known example is provided by perturbations of the Keplerian motion of a celestial body.

Remark 14 For (62), as we just noticed, the system corresponding to the perturbation may be solved in closed form in the variables (p, q, P, Q). On the other hand, the analysis in Section 5 operated with a *different* set of variables $x = (y, \theta) = (p, q, a, \theta)$. This causes no difficulty: it is standard practice when using splitting methods that the different split systems are integrated employing different sets of dependent variables. In partial differential equations, parts corresponding to linear, constant-coefficient differential operators are typically integrated in Fourier space and nonlinearities in physical space. Splitting methods are based on true solution flows, which of course *commute with changes of variables*. The situation is very different for, say, Runge-Kutta schemes, where (except for affine changes) changing variables does not commute with the application of the numerical method, and the performance of the integrator very much depends on the choice of dependent variables. For instance, any consistent Runge-Kutta method integrates exactly the unperturbed problem when written as in (41) but incurs in errors when dealing with the unperturbed version of (60) ($f^i \equiv 0$, $g^\ell \equiv 0$, $h^\ell \equiv 0$).