COMMUNICATIONS ON
PURE AND APPLIED ANALYSIS
Volume **18**, Number **4**, July **2019**

doi:10.3934/cpaa.2019097

pp. **2163–2195**

WORD COMBINATORICS FOR STOCHASTIC DIFFERENTIAL EQUATIONS: SPLITTING INTEGRATORS

A. Alamo

Departamento de Matemática Aplicada e IMUVA
Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain

J. M. Sanz-Serna*

Departamento de Matemáticas, Universidad Carlos III de Madrid
Avenida de la Universidad 30, E-28911 Leganés (Madrid), Spain

Abstract. We present an analysis based on word combinatorics of splitting integrators for Ito or Stratonovich systems of stochastic differential equations. In particular we present a technique to write down systematically the expansion of the local error; this makes it possible to easily formulate the conditions that guarantee that a given integrator achieves a prescribed strong or weak order. This approach bypasses the need to use the Baker-Campbell-Hausdorff (BCH) formula and shows the existence of an order barrier of two for the attainable weak order. The paper also provides a succinct introduction to the combinatorics of words.

1. **Introduction.** This paper shows how word combinatorics is a useful tool in the analysis of splitting integrators for Ito or Stratonovich systems of stochastic differential equations. In particular we present a technique to write down systematically the expansion of the local error; this makes it possible to easily formulate the conditions that guarantee that a given integrator achieves a prescribed strong or weak order. This approach bypasses the need to use the Baker-Campbell-Hausdorff (BCH) formula and shows the existence of an order barrier of two for the attainable weak order. In the case of Stratonovich systems the technique has already appeared in [1]; the corresponding Ito results appear here for the first time. In addition, while the succinct presentation in [1] focuses on the "recipe" to write down the order conditions, the present paper includes background on the combinatorics of words. In this way we also provide what we hope is a reader-friendly introduction to that area, which has applications outside numerical mathematics in many mathematical tasks, including averaging of periodically or quasiperiodically forced systems of differential equations, reduction of continuous or discrete dynamical systems to normal form, rough path theory, etc. (references are given below).

The importance of splitting integrators [6, 28] has increased continuously in the recent past due to their flexibility to adapt to the structure of the problem being

2000 *Mathematics Subject Classification.* Primary: 65C30, 60H05, 16T05.

Key words and phrases. Stochastic differential equations, splitting integrators, word combinatorics.

J. M. S. has been supported by project MTM2016-77660-P(AEI/FEDER, UE), MINECO (Spain).

* Corresponding author.

solved, be it in the context of multiphysics systems or in the domain of geometric integration (i.e. integration performed under the requirement that the numerical solution has some of the geometric properties possessed by the true solution) [42, 21, 4]. As it is the case with any other one-step integrator, the analysis of a splitting algorithm starts with the study of the local error [9, 22], i.e. the error under the assumption that the computation at time level t_{n+1} starts from information at time t_n that is free of errors. Unfortunately, even in the case where the system being integrated consists of (deterministic) ordinary differential equations, the investigation of the local error may be a daunting task if undertaken in a naive way. Formal series and combinatorial algebra have been very useful tools as we discuss presently; see [44] for a recent survey.

For Runge-Kutta methods, whose history goes back to 1895, the structure of the local error was only understood after Butcher's work in the 1960's [8]; this work made it possible to construct formulas that improve enormously on those known until then. In Butcher's theory, the true and numerical solutions are expanded in series; each term of the series is the product of a power of the step size, a numerical coefficient (elementary weight) and a vector-valued function (elementary differential). There is term in the series associated with each rooted tree. The elementary differentials change with the system being integrated but are common to all Runge-Kutta formulas and to the true solution. The weights change with the integrator but are independent of the system being integrated. B-series [23], formal series indexed by rooted trees, were introduced by Hairer and Wanner as a means to systematize Butcher's approach and to extend it to more general classes of algorithms. B-series are indexed by rooted trees and are combinations of elementary differentials. A key result in the theory of B-series is the rule to compose two B-series to obtain a third. B-series possess many applications in numerical analysis, especially in relation to geometric integration (starting with [11]) and modified equations [10]. (Loosely speaking the modified equation of a numerical integration is the differential equation exactly satisfied by the numerical solution.) Recently B-series have also been used outside numerical mathematics, e.g. to perform high-order averaging of periodic or quasiperiodic systems [12, 13].

For splitting integrations of deterministic systems, the best-known method to investigate the local error [42] uses the BCH formula [43, 21]. This may be considered in indirect approach, in that it does not compare the numerical and true solutions but rather the modified system of the integrator and the true system being solved. The large combinatorial complexity of the BCH formula is certainly a limitation of this technique. An alternative methodology, patterned after Butcher's treatment of the Runge-Kutta case was introduced in [32] (a summary may be seen in [21, section III.3]). A third possibility is the use of *word series* expansions [31, 14, 15, 33, 34, 35, 36, 37]. Word series are patterned after B-series; rather than combining elementary differentials they combine *word basis functions*. They are indexed by words on an alphabet rather than by rooted trees. Their scope is narrower than that of B-series; all problems that may be treated by word series are amenable to analysis via B-series, but the converse is not true. On the other hand, word series, when applicable, are more compact and simpler to use than B-series; in particular the composition rule for word series is much simpler than the corresponding rule for B-series. Word series may be used outside numerical mathematics in tasks such as high-order averaging [14, 15, 34, 36, 37], reduction of dynamical systems to normal form [33], etc. They are very well suited to investigate the local error of splitting algorithms [35] (see also the closely related technique in [5, Section 2.4]).

Turning now our attention to splitting algorithms for stochastic differential equations, the most popular technique is again based in the BCH formula, see e.g. [26, 27]. In [1] we suggested a word-series approach in the case where the equations are interpreted in the sense of Stratonovich. This approach bypasses the use of the BCH formula and it is not difficult to implement in practice. Here we extend the material in [1] in several directions that we now discuss briefly.

This paper contains nine sections. Section 2 recalls the Taylor expansion of the solution of Stratonovich and Ito equations and introduces much of the notation to be used throughout the paper. In Section 3 we present splitting integrators and their local errors. We also discuss briefly the pullback operator associated with a mapping; this is a key notion in what follows, as the local error is investigated here with the help of pullback operators. Section 4 describes the main tool: formal series indexed by words. We employ two kinds of such series: series of differential operators and series of mappings. The central results, i.e. the structure of the strong and weak local error and the strong and weak order conditions, are given in Section 5. In the Stratonovich case the structure has already been presented in [1]; the Ito case is new, as is the detailed discussion of the necessity of the order conditions (Lemma 5.3). As an illustration we show how the structure of the local error made explicit by our methodology may be used advantageously to decide between different splitting algorithms for the Langevin dynamics suggested in the literature. Section 6 deals with the shuffle and quasishuffle products; these play a key role in the combinatorics of words. In our context they are necessary to identify sets of *independent* order conditions, a point not discussed in [1], and to prove the composition rule for word series (Proposition 14). The discussion of the order conditions finishes in Section 7 with the help of the infinitesimal generator. There we show an order barrier of 2 for the weak order attainable by splitting integrators in both the Stratonovich and Ito cases. Of course it is possible to prove the existence of such order barriers only because in the preceding sections we have developed a general theory, capable of investigating the class of all splitting methods; those barriers cannot be established when one works for each individual problem in an ad hoc way. Sections 8 and 9 present some complements; they respectively discuss how the relation between the Ito and Stratonovich interpretations may be understood in terms of word combinatorics and the links between the material in this paper and the theory of Hopf algebras.

We close the introduction with some important points.

- The word "formal" is often used in some disciplines, such as theoretical physics, as somehow synonymous to imprecise or lacking in rigour. In this paper formal series are well defined objects that, after truncation, yield meaningful approximations; they are manipulated rigorously because all the necessary computations involve *finite* sums.
- Our interest is in the *combinatorial* aspects of the theory. Therefore we shall not concern ourselves with the derivation of error bounds or other *analytic* considerations. The interested reader is referred to the appendix of [1] (see also [14]).
- In order not to clatter the exposition, all functions that appear are assumed to be smooth in the whole of the Euclidean space. At some places only a finite number of the terms in some series make sense if the given vector fields have

limited smoothness. In those circumstances one has to replace the series by a finite sum.

2. **Stochastic Taylor expansions.**

We are concerned with Stratonovich,

$$dx = f(x)\,dt + \sum_{i=1}^{n} g_i(x) \circ d\mathcal{B}_i, \tag{1}$$

or Ito,

$$dx = f(x)\,dt + \sum_{i=1}^{n} g_i(x)\,d\mathcal{B}_i, \tag{2}$$

systems of differential equations (see e.g. [30]), where f, g_i, $i = 1, \dots, n$, are smooth vector fields in $\mathbb{R}^d$ and $\mathcal{B}_i$, $i = 1, \dots, n$, are independent scalar Wiener processes. When applying splitting integrators, f is often written a sum $\sum_{j=1}^{m} f_j$; it is then convenient to work hereafter with the formats

$$dx = \sum_{a\in\mathcal{A}_{\rm det}} f_a(x)\,dt + \sum_{A\in\mathcal{A}_{\rm sto}} f_A(x) \circ d\mathcal{B}_A \tag{3}$$

or

$$dx = \sum_{a\in\mathcal{A}_{\rm det}} f_a(x)\,dt + \sum_{A\in\mathcal{A}_{\rm sto}} f_A(x)\,d\mathcal{B}_A. \tag{4}$$

The finite set of indices $\mathcal{A}_{\rm det}$ is called the *deterministic alphabet*; its elements are called *deterministic letters*. The finite set $\mathcal{A}_{\rm sto}$ is the *stochastic alphabet* and its elements are the *stochastic letters*. The set $\mathcal{A} = \mathcal{A}_{\rm det} \cup \mathcal{A}_{\rm sto}$ is called the *alphabet* and is assumed to be nonempty. On the other hand, we include the cases where $\mathcal{A}_{\rm det}$ or $\mathcal{A}_{\rm sto}$ are empty; if $\mathcal{A}_{\rm sto} = \emptyset$ then (3)–(4) is a system of ordinary differential equations. We use lower case a, b, $\dots$ for deterministic letters and upper case A, B, $\dots$ for stochastic letters. The symbols k, ℓ, m, $\dots$ are used to refer to elements of $\mathcal{A}$, i.e. to letters, when it is not necessary to specify if they are deterministic or stochastic.

In this section we recall the expressions of the Taylor expansions of the solutions of (3) or (4) presented in e.g. [25, Chapter 5]. Our treatment is somewhat different, because we deal with the format (3)–(4) rather than with the standard (1)–(2). Specifically, as distinct from [25], we work here with deterministic alphabets $\mathcal{A}_{\rm det}$ that may have several letters and, in the Ito case, introduce introduce a letter $\bar{A}$ for each $A \in \mathcal{A}_{\rm sto}$. In the presentation of the Taylor expansion we shall encounter words, and their differential operators and iterated integrals; these are essential later in the paper.

2.1. **The Stratonovich-Taylor expansion.**

With each letter $\ell \in \mathcal{A}$ we associate a first-order differential operator D_ℓ. By definition, D_ℓ is the Lie operator that maps each smooth function $\chi : \mathbb{R}^d \to \mathbb{R}$ into the function $D_\ell\chi$ that at the point $x \in \mathbb{R}^d$ takes the value

$$D_\ell\chi(x) = \sum_{i=1}^{d} f_\ell^i(x) \frac{\partial}{\partial x^i}\chi(x) = \chi'(x) f_\ell(x) \tag{5}$$

(superscritps denote components of vectors). In (5), the symbol χ' denotes the first (Fréchet) derivative of χ; its value at $x \in \mathbb{R}^d$ is a linear map defined on $\mathbb{R}^d$ and $\chi'(x)f_\ell(x)$ is the image by this linear map of the vector $f_\ell(x) \in \mathbb{R}^d$. Smooth functions $\chi : \mathbb{R}^d \to \mathbb{R}$ will often be referred to as *observables*. Since the Stratonovich calculus follows the rules of ordinary calculus, if $x(t)$ is a solution of (3) and $t_0 \geq 0$, $h \geq 0$,

$$\begin{aligned}\chi(x(t_0+h)) &= \chi(x(t_0)) + \int_{s_1=t_0}^{t_0+h} \chi'(x(s_1))\,dx(s_1)\\ &= \chi(x(t_0)) + \int_{s_1=t_0}^{t_0+h} \sum_{a\in\mathcal{A}_{\rm det}} D_a\chi(x(s_1))\,ds_1\\ &\qquad + \int_{s_1=t_0}^{t_0+h} \sum_{A\in\mathcal{A}_{\rm sto}} D_A\chi(x(s_1))\circ d\mathcal{B}_A(s_1)\\ &= \chi(x(t_0)) + \int_{s_1=t_0}^{t_0+h} \sum_{\ell_1\in\mathcal{A}} D_{\ell_1}\chi(x(s_1))\circ d\mathcal{B}_{\ell_1}(s_1), \qquad (6)\end{aligned}$$

where for deterministic ℓ_1 the notation $\circ d\mathcal{B}_{\ell_1}(s_1)$ means ds_1. In (6), as $h \downarrow 0$, the term $\chi(x(t_0))$ provides the Taylor approximation of order 0 to $\chi(x(t_0+h))$ and the integral gives the corresponding remainder. To obtain additional terms of the Taylor expansion of $\chi(x(t_0+h))$, we first write formula (6) with $D_{\ell_1}\chi(x(s_1))$ in lieu of $\chi(x(t_0+h))$,

$$D_{\ell_1}\chi(x(s_1)) = D_{\ell_1}\chi(x(t_0)) + \int_{s_2=t_0}^{s_1} \sum_{\ell_2\in\mathcal{A}} D_{\ell_2}D_{\ell_1}\chi(x(s_2))\circ d\mathcal{B}_{\ell_2}(s_2),$$

and then substitute in (6) to get

$$\begin{aligned}\chi(x(t)) &= \chi(x(t_0)) + \sum_{\ell_1\in\mathcal{A}} \left(\int_{s_1=t_0}^{t_0+h} \circ d\mathcal{B}_{\ell_1}(s_1)\right) D_{\ell_1}\chi(x(t_0))\\ &\qquad + \sum_{\ell_1,\ell_2\in\mathcal{A}} \int_{s_1=t_0}^{t_0+h} \circ d\mathcal{B}_{\ell_1}(s_1) \int_{s_2=t_0}^{s_1} D_{\ell_2}D_{\ell_1}\chi(x(s_2))\circ d\mathcal{B}_{\ell_2}(s_2).\end{aligned}$$

By iterating this procedure, we find the series

$$\chi(x(t_0)) + \sum_{n=1}^{\infty} \sum_{\ell_1,\ldots,\ell_n\in\mathcal{A}} J_{\ell_n\ldots\ell_1}(t_0+h;t_0) D_{\ell_n}\cdots D_{\ell_1}\chi(x(t_0)), \qquad (7)$$

where $J_{\ell_n\ldots\ell_1}(t_0+h;t_0)$ denotes the *iterated stochastic integral*

$$J_{\ell_n\ldots\ell_1}(t_0+h;t_0) = \int_{s_1=t_0}^{t_0+h} \circ d\mathcal{B}_{\ell_1}(s_1)\cdots \int_{s_n=t_0}^{s_{n-1}} \circ d\mathcal{B}_{\ell_n}(s_n). \qquad (8)$$

Iterated integrals obey the following recursion, $n \geq 2$,

$$J_{\ell_n\ldots\ell_1}(t_0+h;t_0) = \int_{t_0}^{t_0+h} J_{\ell_n\ldots\ell_2}(s;t_0)\circ d\mathcal{B}_{\ell_1}(s). \qquad (9)$$

Remark 1. In the right-hand side of (7) the iterated integrals are constructed from the Brownian processes $\mathcal{B}_A$, $A \in \mathcal{A}_{\rm sto}$, in (3) and do not change if the fields f_ℓ, $\ell \in \mathcal{A}$, (or even their dimension d) change. On the other hand the operators D_ℓ are constructed from the vector fields and do not change with the Brownian processes.

In the deterministic case, iterated integrals were introduced and investigated extensively by Kuo Tsai Chen [16] in the context of his work on topology.

The notation may be simplified by introducing the set $\mathcal{W}$ consisting of all *words* $\ell_n\ell_{n-1}\ldots\ell_1$ constructed with the letters of the alphabet $\mathcal{A}$; $\mathcal{W}$ includes an empty word $\emptyset$ with $n=0$ letters. Elements $\ell \in \mathcal{A}$ are seen as words with a single letter and accordingly $\mathcal{A}$ becomes a subset of $\mathcal{W}$. With each word $w = \ell_n \dots \ell_1$ with $n \geq 1$ letters, we associate the n-th order (linear) *differential operator* $D_w = D_{\ell_n} \cdots D_{\ell_1}$. For the empty word, we define $D_\emptyset$ to be the identity operator Id with $Id\chi = \chi$ for each observable and set $J_\emptyset = 1$. (Then (9) also holds for $n = 1$). With this notation the series in (7) simply reads

$$\sum_{w \in \mathcal{W}} J_w(t_0+h; t_0) D_w \chi(x(t_0)). \tag{10}$$

We note that for a deterministic letter,

$$J_a(t_0+h;t_0) = \int_{t_0}^{t_0+h} ds_1 = h,$$

while in the stochastic case

$$J_A(t_0+h;t_0) = \int_{t_0}^{t_0+h} \circ d\mathcal{B}_A(s_1) = \mathcal{B}_A(t_0+h) - \mathcal{B}_A(t_0)$$

is a Gaussian random variable with standard deviation $h^{1/2}$. For this reason, we attach to each deterministic letter $a \in \mathcal{A}_{\text{det}}$ the *weight* $\|a\| = 1$ and each stochastic letter $A \in \mathcal{A}_{\text{sto}}$ the weight $\|A\| = 1/2$. We then define the weight $\|w\|$ of each word by adding the weights of its letters. The weight of the empty word is 0. The following proposition, whose proof may be seen in [1], lists some properties of the iterated integrals. It shows in particular that, as $h \downarrow 0$, $J_w(t_0+h;t_0)$ may be conceived as having size $\mathcal{O}(h^{\|w\|})$.

Proposition 1. *The iterated Stratonovich integrals $J_w(t_0+h;t_0)$ have the following properties:*

- *The joint distribution of any finite subfamily of the family of random variables $\{h^{-\|w\|} J_w(t_0+h;t_0)\}_{w \in \mathcal{W}}$ is independent of $t_0 \geq 0$ and $h > 0$.*
- $\mathbb{E} \mid J_w(t_0+h;t_0) \mid^p < \infty$*, for each $w \in \mathcal{W}$, $t_0 \geq 0$, $h \geq 0$ and $p \in [0,\infty)$.*
- *For each $w \in \mathcal{W}$ and any finite $p \geq 1$, the (t_0-independent) L^p norm of the random variable $J_w(t_0+h;t_0)$ is $\mathcal{O}(h^{\|w\|})$, as $h \downarrow 0$.*
- $\mathbb{E}\, J_w(t_0+h;t_0) = 0$ *whenever $\|w\|$ is not an integer.*

In view of the proposition we rewrite (10) as:

$$\sum_{\nu \in \mathbb{N}/2} \sum_{w \in \mathcal{W},\, \|w\|=\nu} J_w(t_0+h;t_0) D_w \chi(x(t_0)), \tag{11}$$

where $\mathbb{N}/2 = \{0, 1/2, 1, 3/2, \dots\}$. (For each ν, the inner sum only contains a finite number of terms.) In this way, by discarding the terms with $\nu > \nu_0$ in (11), one obtains the Taylor approximation of order ν_0 for $\chi(x(t))$. Of course the series in (11) in general does not converge to $\chi(x(t_0+h))$; it is a *formal series*, whose truncations provide the required Taylor approximations.

So far it has been assumed that χ is scalar-valued. For a vector-valued χ, the Taylor expansion is also given by (11), with the differential operators D_w defined to act componentwise. The particular choice where $\chi : \mathbb{R}^d \to \mathbb{R}^d$ is taken to be the identity function $x \mapsto x$, yields the expansion of the solution $x(t_0+h)$ itself given by

$$\sum_{\nu \in \mathbb{N}/2} \sum_{w \in \mathcal{W},\, \|w\|=\nu} J_w(t_0+h;t_0) f_w(x(t_0)), \tag{12}$$

where, $f_w(x(t_0))$ denotes the result of applying D_w to the identity function and then evaluating at $x(t_0)$. Note that the functions f_w may be constructed from the fields f_ℓ in (3) with the help of the recursion

$$f_{\ell_n \ldots \ell_1}(x) = f'_{\ell_{n-1}\ldots\ell_1}(x) f_{\ell_n}(x), \qquad n \geq 1, \tag{13}$$

where $f'_{\ell_{n-1}\ldots\ell_1}(x)$ stands for the value at x of the Jacobian matrix of $f_{\ell_{n-1}\ldots\ell_1}$.

2.2. **The Ito-Taylor expansion.** The Taylor expansion of the solution of Ito stochastic differential system was first derived by Platen and Wagner [39]. For (4), formula (6) has to be replaced by

$$\begin{aligned}\chi(x(t_0+h)) &= \chi(x(t_0)) + \int_{s_1=t_0}^{t_0+h} \sum_{a\in\mathcal{A}_{\rm det}} D_a\chi(x(s_1))\,ds_1 \\ &+ \int_{s_1=t_0}^{t_0+h} \sum_{A\in\mathcal{A}_{\rm sto}} D_A\chi(x(s_1))d\mathcal{B}_A(s_1) + \int_{s_1=t_0}^{t_0+h} \sum_{A\in\mathcal{A}_{\rm sto}} D_{\bar{A}}\chi(x(s_1))\,ds_1;\end{aligned} \tag{14}$$

the last term in the right-hand side is the Ito correction, where, for each stochastic letter A, $D_{\bar{A}}$ represents the second-order, linear differential operator

$$D_{\bar{A}}\chi(x) = \frac{1}{2}\sum_{i,j=1}^{d} f_A^i(x) f_A^j(x) \frac{\partial^2}{\partial x^i \partial x^j}\chi(x) = \frac{1}{2}\chi''(x)[f_A(x), f_A(x)]. \tag{15}$$

The symbol χ'' represents the second (Fréchet) derivative of χ; its value $\chi''(x)$ at a point $x \in \mathbb{R}^s$ is a bilinear map defined on $\mathbb{R}^d \times \mathbb{R}^d$ and $\chi''(x)[f_A(x), f_A(x)]$ is the image by this map of the pair of vectors $[f_A(x), f_A(x)]$.

In order to write (14) more compactly, we introduce the *extended alphabet* $\bar{\mathcal{A}} = \bar{\mathcal{A}}_{\rm det} \cup \bar{\mathcal{A}}_{\rm sto}$. The extended set $\bar{\mathcal{A}}_{\rm sto}$ of stochastic letters coincides with the old $\mathcal{A}_{\rm sto}$, i.e. with the set of indices in the second sum in (4); the extended set $\bar{\mathcal{A}}_{\rm det}$ comprises the indices a in the first sum in (4) *and,* in addition, a letter $\bar{A}$ for each $A \in \bar{\mathcal{A}}_{\rm sto} = \mathcal{A}_{\rm sto}$. With these notations, (14) becomes

$$\chi(x(t_0+h)) = \chi(x(t_0)) + \int_{s_1=t_0}^{t_0+h} \sum_{\ell_1\in\bar{\mathcal{A}}} D_{\ell_1}\chi(x(s_1))\,d\mathcal{B}_{\ell_1}(s_1)$$

($d\mathcal{B}_{\ell_1}(s_1) = ds_1$ for $\ell_1 \in \bar{\mathcal{A}}_{\rm det}$); this is the Ito counterpart of the right-most expression in (6). By iterating as in the Stratonovich case, we obtain the series

$$\chi(x(t_0)) + \sum_{n=1}^{\infty} \sum_{\ell_1,\ldots,\ell_n\in\bar{\mathcal{A}}} I_{\ell_n\ldots\ell_1}(t_0+h;t_0) D_{\ell_n}\cdots D_{\ell_1}\chi(x(t_0)) \tag{16}$$

where $I_{\ell_n\ldots\ell_1}(t_0+h;t_0)$ denotes the Ito *iterated stochastic integral*

$$I_{\ell_n\ldots\ell_1}(t_0+h;t_0) = \int_{s_1=t_0}^{t_0+h} d\mathcal{B}_{\ell_1}(s_1)\cdots\int_{s_n=t_0}^{s_{n-1}} d\mathcal{B}_{\ell_n}(s_n).$$

These iterated integrals satisfy the obvious analogue of the recursion (9). Again the iterated integrals do not change if the vector fields are changed and the operators D_ℓ do not change if the Brownian processes are changed.

We now consider the set of words $\overline{\mathcal{W}}$ constructed with the letters of the extended alphabet $\bar{\mathcal{A}}$, and write $D_w = D_{\ell_n}\cdots D_{\ell_1}$ for $w = \ell_n \ldots \ell_1 \in \overline{\mathcal{W}}$, $n > 0$, (recall that

D_ℓ is a second order operator if ℓ is of the form $\bar{A}$, $A \in \mathcal{A}_{\rm sto}$), $D_\emptyset = Id$, $I_\emptyset = 1$. Then (16) has the compact expression

$$\sum_{w \in \overline{\mathcal{W}}} I_w(t_0+h;t_0) D_w \chi(x(t_0)). \tag{17}$$

If letters in $\bar{\mathcal{A}}_{\rm det}$ are again declared to have weight 1 and letters in $\bar{\mathcal{A}}_{\rm sto}$ to have weight 1/2, we have the following result, whose proof is similar to that of Proposition 1:

Proposition 2. *The Ito iterated integrals $I_w(t_0+h;t_0)$ possess the properties of the Stratonovich iterated integrals listed in Proposition 1*

The series (17) is rewritten as

$$\sum_{\nu \in \mathbb{N}/2} \sum_{w \in \overline{\mathcal{W}},\, \|w\|=\nu} I_w(t_0+h;t_0) D_w \chi(x(t_0)), \tag{18}$$

and for the solution itself we have the Taylor series

$$\sum_{\nu \in \mathbb{N}/2,\nu} \sum_{w \in \overline{\mathcal{W}},\, \|w\|=\nu} I_w(t_0+h;t_0) f_w(x(t_0)),$$

where $f_w(x(t_0))$, $w = \ell_n \ldots \ell_1$ denotes the result of successively applying $D_{\ell_1}, \ldots,$ D_{ℓ_n} to the identity function and then evaluating at $x(t_0)$. The f_w satisfy (13) if $\ell_n \in \mathcal{A}_{\rm det} \cup \mathcal{A}_{\rm sto}$ and

$$f_{\ell_n \ldots \ell_1}(x) = \frac{1}{2}\Big(f''_{\ell_{n-1}\ldots\ell_1}(x)\Big)[f_A(x), f_A(x)], \qquad n \geq 1, \tag{19}$$

for $\ell_n = \bar{A}$ with $A \in \mathcal{A}_{\rm sto}$. Since the second derivatives of the identity function vanish we have the following result.

Proposition 3. *If the last (i.e. right-most) letter of $w \in \overline{\mathcal{W}}$ is of the form $\bar{A}$ with $A \in \mathcal{A}_{\rm sto}$, then f_w vanish identically.*

Therefore, after suppressing the f_w that vanish identically, the Taylor expansion may be written:

$$\sum_{\nu \in \mathbb{N}/2,\nu} \sum_{w \in \overline{\mathcal{W}}_0,\, \|w\|=\nu} I_w(t_0+h;t_0) f_w(x(t_0)), \tag{20}$$

where $\overline{\mathcal{W}}_0$ is the subset of $\overline{\mathcal{W}}$ consisting of words whose last letter is not one of the $\bar{A}$, $A \in \mathcal{A}_{\rm sto}$.

3. Analyzing splitting integrators: preliminaries.

3.1. Splitting integrators.

In order to avoid notational complications, let us momentarily consider only the simple instance of (3) given by

$$dx = f_a(x)\, dt + f_A(x) \circ d\mathcal{B}_A. \tag{21}$$

Splitting integrators may be applied to integrate this system if one may solve in closed form the *split systems*

$$dx = f_a(x)\, dt \tag{22}$$

and

$$dx = f_A(x) \circ d\mathcal{B}_A. \tag{23}$$

In the simplest splitting integrator, the Lie-Trotter algorithm, the numerical solution is advanced from its value x_n at a time level t_n to the value x_{n+1} at the next time level t_{n+1} by first integrating (22) in the interval $[t_n, t_{n+1}]$ with initial condition x_n to get a value $\tilde{x}_n$ and then using $\tilde{x}_n$ as initial condition to integrate (23) in the interval $[t_n, t_{n+1}]$ to obtain x_{n+1}. In this way, the simultaneous contributions of f_a and f_A in (21) are replaced by successive contributions. The procedure is best described by introducing, for $0 \le s \le t$, the solution maps $\varphi^{(a)}_{t,s}$, $\varphi^{(A)}_{t,s}$ of (22) and (23); by definition, $\varphi^{(a)}_{t,s}$ (respectively $\varphi^{(A)}_{t,s}$) maps $x \in \mathbb{R}^d$ into the value at time t of the solution of (22) (respectively (23)) with initial value x at time s. Note that, for the autonomous deterministic system (22), $\varphi^{(a)}_{t,s}$ depends on t and s only through the combination (elapsed time) $t - s$, but that is not the case for $\varphi^{(A)}_{t,s}$. In addition $\varphi^{(a)}_{t,s}$ makes sense for $t < s$, but $\varphi^{(A)}_{t,s}$ does not, because stochastic differential equations cannot be evolved backward in time. With this notation in place, one step of the Lie-Trotter algorithm described above is given by $x_{n+1} = \psi_{t_{n+1},t_n}(x_n)$, where

$$\psi_{t_{n+1},t_n} = \varphi^{(A)}_{t_{n+1},t_n} \circ \varphi^{(a)}_{t_{n+1},t_n}. \tag{24}$$

Of course one may also consider the alternative algorithms given by $\varphi^{(a)}_{t_{n+1},t_n} \circ \varphi^{(A)}_{t_{n+1},t_n}$ or the well-known symmetric compositions

$$\varphi^{(a)}_{t_{n+1},t_{n+1/2}} \circ \varphi^{(A)}_{t_{n+1},t_n} \circ \varphi^{(a)}_{t_{n+1/2},t_n}$$

or

$$\varphi^{(A)}_{t_{n+1},t_{n+1/2}} \circ \varphi^{(a)}_{t_{n+1},t_n} \circ \varphi^{(A)}_{t_{n+1/2},t_n}$$

associated with Strang's name ($t_{n+1/2}$ is the midpoint of $[t_n, t_{n+1}]$). More involved splitting algorithms are obtained by composing four or more solution maps of the split systems.

Leaving now the particular instance (21), for a problem of the general form (3) the splitting-integrator mapping $x_{n+1} = \psi_{t_{n+1},t_n}(x_n)$ is a composition of solution operators

$$\varphi^{(i)}_{t_n + d_i(t_{n+1}-t_n),\, t_n + c_i(t_{n+1}-t_n)}, \qquad i = 1, \dots, m. \tag{25}$$

Here c_i and d_i are constants and the superindex i refers to a system of differential equations obtained by taking into account a subset $\mathcal{S}_i$, $i = 1, \dots, m$ of the fields f_ℓ in (3); it has to be supposed that these systems are solvable in closed form. For our purposes here, there is complete freedom when choosing the different $\mathcal{S}_i$; it is possible to have $\mathcal{S}_i = \mathcal{S}_j$ for $i \ne j$ (as in Strang's method where $\mathcal{S}_1 = \mathcal{S}_3$) or to let given vector field f_ℓ appear in $\mathcal{S}_i$ and $\mathcal{S}_j$ with $\mathcal{S}_i \ne \mathcal{S}_j$. It is important to note that it is necessary to assume throughout that

$$c_i < d_i$$

except in the case where $\mathcal{S}_i$ is a deterministic system; stochastic differential equations cannot be evolved backward in time.

The Ito case can be dealt with in the same way; the only difference is that the solution operators of the systems $\mathcal{S}_i$ have to be based on the Ito interpretation.

3.2. **The local error.** An essential part of the analysis of any one-step integrator $x_{n+1} = \psi_{t_{n+1},t_n}(x_n)$ is the study of the corresponding *local error* (or truncation error). By definition, if $\varphi_{t,s}$ denotes the solution operator of the system (3) or (4) being integrated, the local error is the difference

$$\psi_{t_{n+1},t_n}(x_n) - \varphi_{t_{n+1},t_n}(x_n). \tag{26}$$

In what follows we just consider the case $\psi_{t_1,t_0}(x_0)-\varphi_{t_1,t_0}(x_0)$; the case with general n differs from this only in notation. Furthermore we write $t_1 = t_0 + h$, where $h > 0$ is the step-length. Our aim is to understand the behaviour of (26) as $h \downarrow 0$ and this is achieved by Taylor expansion. In the particular case of the Lie-Trotter integrator (24) for the simple system (21), we have therefore to Taylor expand

$$x_1 = \varphi^{(A)}_{t_1,t_0}\Big(\varphi^{(a)}_{t_1,t_0}(x_0)\Big) \tag{27}$$

and compare the result with the expansion of $\varphi_{t_1,t_0}(x_0)$ found in the preceding section. Note that if we write

$$\tilde{x}_0 = \varphi^{(a)}_{t_1,t_0}(x_0), \tag{28}$$

so that

$$x_1 = \varphi^{(A)}_{t_1,t_0}(\tilde{x}_0) \tag{29}$$

the expansions $\sum_i c_i F_i(x_0)$ of (28) at x_0 and $\sum_j d_j G_j(\tilde{x}_0)$ of (29) at $\tilde{x}_0$ are both known; they are particular instances of (12) corresponding to alphabets with the single letter a or A respectively. Then expansion for (27) may be obtained by substituting to get

$$\sum_j d_j G_j\big(\sum_i c_i F_i(x_0)\big),$$

Taylor expanding each $G_j\big(\sum_i c_i F_i(x_0)\big)$ and gathering terms of equal weight. For more complicated splitting integrators there are m mappings being composed and implementing the naive substitution we have described may be a daunting task. We are thus led to the following:

Problem P: Find efficiently the expansion of a composition of mappings $\varphi^{(m)} \circ \cdots \circ \varphi^{(1)}$, *when* $\varphi^{(i)}$, $i = 1, \dots, m$, *have known expansions of the form* (12) *(or* (20) *for the Ito case).*

The solution to this problem presented in the next section is based on expanding pullback operators (see e.g. [37]) rather than mappings.

3.3. **Pullbacks.** Associated with any mapping $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, there is a pullback operator Φ. By definition, Φ maps each observable χ into the observable $\Phi\chi$ whose value at $x \in \mathbb{R}^d$ is $(\Phi\chi)(x) = \chi(y)$ with $y = \varphi(x)$ (φ pushes the point x forward to y, while Φ takes values of the observable "back" from y to x). The pullback operator corresponding to a composition $\varphi^{(2)} \circ \varphi^{(1)}$ is the composition of operators $\Phi^{(1)}\Phi^{(2)}$ (note the reversed order) because

$$\Big(\Phi^{(1)}(\Phi^{(2)}\chi)\Big)(x) = (\Phi^{(2)}\chi)(\varphi^{(1)}(x)) = \chi\Big(\varphi^{(2)}(\varphi^{(1)}(x))\Big).$$

A map and its pullback operator contain the same information: when the operator Φ is known, one may retrieve the underlying map φ by applying Φ to the identity $x \mapsto x$ in $\mathbb{R}^s$. Recovering Φ from φ is similar to what was done *for formal series* rather than for maps to obtain (12) from (11) (or (20) from (18) in the Ito case). Taking this point further, from (11) we may consider that the series

$$\sum_{\nu \in \mathbb{N}/2} \; \sum_{w \in \mathcal{W},\, \|w\| = \nu} J_w(t_0 + h; t_0) D_w, \tag{30}$$

provides the Taylor expansion of the pullback operator of the solution operator φ_{t_0+h,t_0} of (3). For the Ito case (30) is replaced by

$$\sum_{\nu\in\mathbb{N}/2}\ \sum_{w\in\overline{\mathcal{W}},\,\|w\|=\nu} I_w(t_0+h;t_0)D_w. \tag{31}$$

In this way the problem posed above may be reformulated as:

Problem P': Find efficiently the expansion of a composition $\Phi^{(1)}\cdots\Phi^{(m)}$ of pullback operators, when the operators $\Phi^{(i)}$, $i=1,\dots,m$, have known expansions of the form (30) *(or* (31) *for the Ito case).*

The idea of using pullback (differential) operators to analyze local errors is old. Merson [29] used it in 1957 to study Runge-Kutta formulas; however the subsequent treatment in Butcher [8] did away with differential operators and worked only with elementary differentials (mappings). In the stochastic case it is convenient to work both with differential operators and mappings, as it will become clear below.

4. **Series.**

We now solve the problem P' (and by implication P) with the help of some simple algebraic/combinatorial tools.

4.1. **Series for Stratonovich problems.**

4.1.1. *Series.* Words in $\mathcal{W}$ are multiplied by *concatenation*, i.e. if $v=k_1\dots k_m$, $w=\ell_1\dots\ell_n$ are words with m and n letters respectively, their product is the word with $m+n$ letters $vw=k_1\dots k_m\ell_1\dots\ell_n$. In particular $\emptyset\emptyset=\emptyset$, $\emptyset w=w\emptyset=w$. Concatenation is associative but it is not commutative.

The vector space $\mathbb{R}\langle\mathcal{A}\rangle$ consists, by definition, of all linear combinations of words $\sum_{w\in\mathcal{W}}c_w w$ (only a finite number of coefficients $c_w\in\mathbb{R}$ are nonzero). The multiplication of words by concatenation is extended in an obvious way to elements of $\mathbb{R}\langle\mathcal{A}\rangle$,

$$\sum_{v\in\mathcal{W}}c_v v\sum_{w\in\mathcal{W}}d_w w=\sum_{v,w\in\mathcal{W}}c_v d_w uw, \tag{32}$$

and then $\mathbb{R}\langle\mathcal{A}\rangle$ becomes a noncommutative, associative algebra.

In addition, we need to consider the larger noncommutative algebra $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ of formal series. These are formal expressions $\sum_{w\in\mathcal{W}}c_w w$ where it is not any longer assumed that only finitely many coefficients c_w are $\neq 0$. If $S\in\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$, we denote the corresponding coefficients by S_w, i.e. $S=\sum_{w\in W}S_w w$. Formal series are combined linearly in an obvious way and are multiplied as in (32), where we note that the right-hand side is well defined, even if infinitely many c_v and d_w do not vanish, because the number of ways in which a given $u\in\mathcal{W}$ may be written as a concatenation $u=vw$ is finite. More precisely, if we denote by $\mathbb{R}^{\mathcal{W}}$ the set of all sequences of coefficients $\{c_w\}_{w\in\mathcal{W}}$ indexed by words, then the product in (32) is the series $\sum_{u\in\mathcal{W}}e_u u\in\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ with coefficients $\{e_u\}_{u\in\mathcal{W}}$ such that $e_\emptyset=c_\emptyset d_\emptyset$ and , for each nonempty word $u=\ell_1\dots\ell_n$,

$$e_{\ell_1\dots\ell_n}=c_\emptyset d_{\ell_1\dots\ell_n}+\sum_{m=1}^{n-1}c_{\ell_1\dots\ell_m}d_{\ell_{m+1}\dots\ell_n}+c_{\ell_1\dots\ell_n}d_\emptyset. \tag{33}$$

The right-hand side of this formula contains all the ways of writing $u=\ell_1\dots\ell_n$ as a concatenation of two (possibly empty) words. Thus (33) defines a (noncommutative, associative) product in the set $\mathbb{R}^{\mathcal{W}}$ of sequences of coefficients, the so-called *convolution* product, in such a way that the product of series $S\in\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ corresponds to the convolution product of the sequences of coefficients $\{S_w\}_{w\in\mathcal{W}}\in\mathbb{R}^{\mathcal{W}}$.

A general well-known reference to the combinatorics of words is [41].

4.1.2. *Series of differential operators.* Given the vector fields f_ℓ in (3), the concatenation product of words obviously corresponds to the composition of the associated differential operators: $D_{vw} = D_v D_w$ for any $v, w \in \mathcal{W}$ (by definition, $(D_v D_w)\chi = D_v(D_w \chi)$ for each observable χ).

With the series $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ we associate the formal series of differential operators $D_S = \sum_{w\in\mathcal{W}} S_w D_w$. It follows that $S, S' \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$, the product (composition) $D_S D_{S'}$ is the series $D_{SS'}$ associated with SS', whose coefficients, as we know, are given by the convolution product of the coefficients of S and S'. Series of differential operators are a common tool in control theory, see e.g. [19].

With the terminology we have introduced, for each fixed t, t_0 and for each event in the underlying probability space, the expansion in (30) coincides with D_S when the coefficients are $S_w = J_w(t_0 + h; t_0)$, $w \in \mathcal{W}$. Since we have just described how to multiply series D_S, we have solved the problem P' posed in the previous section.

We illustrate the technique by means of a simple example. We integrate the system

$$dx = f_a(x)dt + f_b(x)dt + f_A(x) \circ d\mathcal{B}_A,$$

with the help of the split systems

$$(1)\ dx = f_a(x)dt + f_A(x) \circ d\mathcal{B}_A, \qquad (2)\ dx = f_b(x)dt.$$

We use the Lie-Trotter formula $\varphi^{(2)} \circ \varphi^{(1)}$. According to (30) (when the alphabet is chosen to be $\{a, A\}$), the expansion of $\Phi^{(1)}$ is

$$Id + J_A D_A + J_a D_a + J_{AA} D_{AA} + J_{aA} D_{aA} + J_{Aa} D_{Aa} + J_{AAA} D_{AAA} + \mathcal{O}(2),$$

where $\mathcal{O}(2)$ denotes the terms in the series with weight ≥ 2, and J_A, J_a, ... stand for $J_A(t_0 + h; t_0)$, $J_a(t_0 + h; t_0)$, ... Similarly, the expansion of $\Phi^{(2)}$ is

$$Id + J_b D_b + \mathcal{O}(2).$$

Multiplying out, we obtain the expansion for the product $\Phi^{(1)}\Phi^{(2)}$:

$$\begin{aligned} &Id + J_A D_A + J_a D_a + J_b D_b + J_{AA} D_{AA} \\ &\quad + J_{aA} D_{aA} + J_{Aa} D_{Aa} + J_A J_b D_{Ab} + J_{AAA} D_{AAA} + \mathcal{O}(2). \end{aligned}$$

For the solution of the system being integrated, (30) (when the alphabet is $\{a, b, A\}$) yields

$$\begin{aligned} &Id + J_A D_A + J_a D_a + J_b D_b + J_{AA} D_{AA} \\ &\quad + J_{aA} D_{aA} + J_{bA} D_{bA} + J_{Aa} D_{Aa} + J_{Ab} D_{Ab} + J_{AAA} D_{AAA} + \mathcal{O}(2), \end{aligned}$$

and subtracting we find that the pullback operator associated with the local error has the expansion:

$$(J_A J_b - J_{Ab}) D_{Ab} - J_{bA} D_{bA} + \mathcal{O}(2). \qquad (34)$$

4.1.3. *Word series.* Given the vector fields f_ℓ in (3), with each series $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ we associate the corresponding *word series* $\mathcal{W}_S(x_0)$; this is obtained by applying D_S to the identity map $x \in \mathbb{R}^d \mapsto x$:

$$\mathcal{W}_S(x_0) = \sum_{w\in\mathcal{W}} S_w f_w(x_0).$$

The functions $f_w : \mathbb{R}^d \to \mathbb{R}^d$, $w \in \mathcal{W}$, we already encountered in (12), are called *word basis functions.* Recall that they may be found recursively via (13) from the f_ℓ that appear in the system (3). Word series, introduced and studied in [31, 13, 14, 1, 35, 36, 37], may be seen as equivalent to series of differential operators; the theory of words series is patterned after the theory of B-series [23] familiar to many numerical analysts.

With the terminology above, for fixed t_0 and h and each event in the underlying probability space, the expansion (12) is simply the word series with coefficients given by the iterated integrals J_w. In what follows we shall denote by J the series $J = \sum_w J_w(t_0+h;t_0)w \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$, so that D_J and $\mathcal{W}_J$ are the corresponding series of operators and word series. Formal series of words whose coefficients are iterated integrals are often called *Chen series*; they play a role in several mathematical developments, including the theory of rough paths (see e.g. [2]).

In the example we are discussisng, from (34) we obtain that the local error has the expansion

$$(J_A J_b - J_{Ab}) f_{Ab} - J_{bA} f_{bA} + \mathcal{O}(2), \tag{35}$$

with $f_{Ab} = f'_b f_A$, $f_{bA} = f'_A f_b$.

4.2. **Series for Ito problems.** The preceding material is easily adapted to the Ito system (4). The required changes are few. One considers formal series $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ (words are now based on the extended alphabet) and to each $S = \sum_{w\in\overline{\mathcal{W}}} S_w w$ associates a series of differential operators $D_S = \sum_{w\in\overline{\mathcal{W}}} S_w D_w$. The expansion (31) of the pullback of the solution operator is D_S when the coefficients of the series are chosen to be the Ito iterated integrals. We write this series as D_I and set $I = \sum_{w\in\overline{W}} I_w(t_0+h;t_0)w \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ for the corresponding *Chen series*.

Here is an example. For the Ito system corresponding to the alphabet $\{a,b,A\}$, split as (1) $\{a,A\}$, (2) $\{b\}$, the expansion of of $\Phi^{(1)}$

$$\begin{aligned} Id &+ I_A D_A + I_a D_a + I_{\bar{A}} D_{\bar{A}} + I_{AA} D_{AA} \\ &+ I_{aA} D_{aA} + I_{\bar{A}A} D_{\bar{A}A} + I_{Aa} D_{Aa} + I_{A\bar{A}} D_{A\bar{A}} + I_{AAA} D_{AAA} + \mathcal{O}(2), \end{aligned}$$

the expansion of $\Phi^{(2)}$ is

$$Id + I_b D_b + \mathcal{O}(2),$$

and, multiplying out, the expansion $\Phi^{(1)}\Phi^{(2)}$ is found to be

$$\begin{aligned} &Id + I_A D_A + I_a D_a + I_b D_b + I_{\bar{A}} D_{\bar{A}} + I_{AA} D_{AA} \\ &+ I_{aA} D_{aA} + I_{\bar{A}A} D_{\bar{A}A} + I_{Aa} D_{Aa} + I_A I_b D_{Ab} + I_{A\bar{A}} D_{A\bar{A}} + I_{AAA} D_{AAA} + \mathcal{O}(2). \end{aligned}$$

For the solution of the system being integrated we have

$$\begin{aligned} Id &+ I_A D_A + I_a D_a + I_b D_b + I_{\bar{A}} D_{\bar{A}} + I_{AA} D_{AA} \\ &+ I_{aA} D_{aA} + I_{bA} D_{bA} + I_{\bar{A}A} D_{\bar{A}A} + I_{Aa} D_{Aa} + I_{Ab} D_{Ab} + I_{A\bar{A}} D_{A\bar{A}} \\ &+ I_{AAA} D_{AAA} + \mathcal{O}(2), \end{aligned}$$

and, for the pullback of the truncation error,

$$(I_A I_b - I_{Ab}) D_{Ab} - I_{bA} D_{bA} + \mathcal{O}(2), \tag{36}$$

while for the truncation error itself we have the word series expansion:

$$(I_A I_b - I_{Ab}) f_{Ab} - I_{bA} f_{bA} + \mathcal{O}(2), \tag{37}$$

with $f_{Ab} = f'_b f_A$, $f_{bA} = f'_A f_b$.

5. **The expansion of the local error. Error equations.** In this section we present the Taylor expansion of the local error along with the conditions that have to be imposed to achieve a target strong or weak order of consistency.

5.1. **Expanding the local error.** By applying the technique in the previous section, the Taylor expansion of the mapping $\psi_{t_0+h;t_0}$ that describes a splitting integrator for the Stratonovich system (3) is found as a word series

$$\mathcal{W}_{\widetilde{J}}(x_0) = \sum_{w\in\mathcal{W}} \widetilde{J}_w(t_0+h;t_0) f_w(x_0).$$

Here $\widetilde{J}_w(t_0;t_0+h)$ is either zero or a product of Stratonovich iterated integrals corresponding to words whose concatenation is w (see (35) for an example). Thus, in each product, the iterated integrals being multiplied correspond to words whose weights add up to $\|w\|$. In particular $\widetilde{J}_\emptyset(t_0;t_0+h) = 1$. For the corresponding pullback we have the expansion

$$D_{\widetilde{J}} = \sum_{w\in\mathcal{W}} \widetilde{J}_w(t_0+h;t_0) D_w$$

(see (34)).

Similarly, in the Ito case, $\psi_{t_0+h;t_0}$ has a word series expansion

$$\mathcal{W}_{\widetilde{I}}(x_0) = \sum_{w\in\overline{\mathcal{W}}_0} \widetilde{I}_w(t_0+h;t_0) f_w(x_0)$$

(see (37)) and for the associated pullback the expansion is

$$D_{\widetilde{I}} = \sum_{w\in\overline{\mathcal{W}}} \widetilde{I}_w(t_0+h;t_0) D_w$$

(see (36)).

The proof of the following technical result may be found in [1] for the Stratonovich case; the Ito case is proved similarly.

Proposition 4. *The coefficients $\widetilde{J}_w(t_0+h;t_0)$, $w\in\mathcal{W}$, possess the properties of the exact coefficients $J_w(t_0+h;t_0)$ listed in Proposition 1. The coefficients $\widetilde{I}_w(t_0+h;t_0)$, $w\in\overline{\mathcal{W}}$, possess the properties of the exact coefficients $I_w(t_0+h;t_0)$ listed in Proposition 2.*

By subtracting the expansions of the integrator and the true solution, we immediately obtain the next result. The bound for $\|\delta_w(t_0;h)\|_p$ follows from the third item in Proposition 1 and the corresponding result for $\widetilde{J}_w(t_0+h;t_0)$ in Proposition 4. Note that the halfinteger values of ν drop from (39) in view of the last item in Proposition 1 and the corresponding result for $\widetilde{J}_w(t_0+h;t_0)$.

Theorem 5.1. *For a splitting integrator for the Stratonovich system (3), the local error $\psi_{t_0+h;t_0}(x_0)$ has a word series expansion*

$$W_{\delta(t_0;h)}(x_0) = \sum_{\nu\in\mathbb{N}/2,\nu\neq 0}\ \sum_{w\in\mathcal{W},\|w\|=\nu} \delta_w(t_0;h) f_w(x_0) \tag{38}$$

with coefficients

$$\delta_w(t_0;h) = \widetilde{J}_w(t_0+h;t_0) - J_w(t_0+h;t_0), \qquad w\in\mathcal{W}.$$

For each nonempty $w\in\mathcal{W}$ and any L^p norm $1\le p<\infty$, uniformly in t_0,

$$\|\delta_w(t_0;h)\|_p = \mathcal{O}(h^{\|w\|}), \qquad h\downarrow 0.$$

In addition, for each observable χ, conditional on x_0, the error in expectation

$$\mathbb{E}\big(\psi_{t_0+h;t_0}(x_0)\big) - \mathbb{E}\big(\varphi_{t_0+h;t_0}(x_0)\big)$$

has the expansion

$$\sum_{\nu\in\mathbb{N},\nu\neq 0}\ \sum_{w\in\mathcal{W},\|w\|=\nu} \mathbb{E}\big(\delta_w(t_0;h)\big) D_w\chi(x_0). \tag{39}$$

Of course to obtain good strong approximations, integrators with small error coefficients $\delta_w(t_0;h)$ are to be preferred, all other things being equal, to integrators with large error coefficients. A similar comment applies to weak approximations. We shall illustrate this point in the important case of the Langevin equations

$$\begin{aligned} dq &= M^{-1}p\,dt,\\ dp &= F(q)\,dt - \gamma p\,dt + \sigma M^{1/2}\circ dB(t) \end{aligned}$$

that play a very important role in statistical physics and molecular dynamics. Here M is a diagonal mass matrix with diagonal entries $m_i > 0$, $i = 1,\dots,D$, $\gamma > 0$ is the (constant) friction coefficient, the constant σ governs the fluctuation due to noise, B is a D-dimensional Wiener process, and the force F is conservative, i.e. $F = -\nabla V$ for a suitable scalar-valued potential function V. If we set $x = (p, q)$, $d = 2D$, the Langevin system is a particular case of (3) with

$$f_a(q,p) = (M^{-1}p, 0),\quad f_b(q,p) = (0, F(q)),\quad f_c(q,p) = (0, -\gamma p),$$

and, for $i = 1,\dots,D$,

$$f_{A_i}(q,p) = (0, \sigma\sqrt{m_i}e_i),$$

where e_i is the i-th unit vector in $\mathbb{R}^D$. The deterministic letters a, b and c are respectively associated with inertia, potential forces and friction. If the Langevin system is split into three parts corresponding to $\{f_a\}$, $\{f_b\}$ and $\{f_c,\ f_{A_1},\ \dots,\ f_{A_d}\}$, then each split system may be integrated in closed form (see [1] for details). Leimkuhler and Matthews [26], [27] use the letters A, B and O to refer to these split systems and the acronyms ABOBA and BAOAB for the Strang-like algorithms

$$\varphi^A_{t_0+h;t_0+h/2}\circ\varphi^B_{t_0+h;t_0+h/2}\circ\varphi^O_{t_0+h;t_0}\circ\varphi^B_{t_0+h/2;t_0}\circ\varphi^A_{t_0+h/2;t_0}$$

and

$$\varphi^B_{t_0+h;t_0+h/2}\circ\varphi^A_{t_0+h;t_0+h/2}\circ\varphi^O_{t_0+h;t_0}\circ\varphi^A_{t_0+h/2;t_0}\circ\varphi^B_{t_0+h/2;t_0}$$

respectively. In spite of the similarity between both algorithms, in practice the performance of BAOAB is significantly better than that of ABOBA. It is a very simple matter to compute the expansions (38)–(39) (many word basis vanish due to the structure of the Langevin systems). It turns out that ABOBA yields very large errors for words like A_iba, cba, A_icba, etc. thus explaining the superiority of BAOAB. A full discussion is presented in [1]. Note that the word-series approach makes it possible to study terms of arbitrarily high weight; one is not limited to analysing the leading error terms.

In the Ito case we have the following result:

Theorem 5.2. *For a splitting integrator for the Ito system of differential equations* (4)*, the coefficients*

$$\eta_w(t_0;h) = \widetilde{I}_w(t_0+h;t_0) - I_w(t_0+h;t_0),$$

satisfy, for each nonempty $w\in\overline{\mathcal{W}}$ *and any* L^p *norm* $1\le p<\infty$*, uniformly in* t_0*,*

$$\|\eta_w(t_0;h)\|_p = \mathcal{O}(h^{\|w\|}),\qquad h\downarrow 0.$$

The local error $\psi_{t_0+h;t_0}(x_0)$ *has a word series expansion*

$$W_{\eta(t_0;h)}(x_0) = \sum_{\nu\in\mathbb{N}/2,\nu\neq 0}\ \sum_{w\in\overline{\mathcal{W}}_0,\|w\|=\nu} \eta_w(t_0;h) f_w(x_0) \tag{40}$$

In addition, for each observable χ*, conditional on* x_0*, the error in expectation*

$$\mathbb{E}\big(\psi_{t_0+h;t_0}(x_0)\big) - \mathbb{E}\big(\varphi_{t_0+h;t_0}(x_0)\big)$$

has the expansion

$$\sum_{\nu\in\mathbb{N},\nu\neq 0}\ \sum_{w\in\overline{\mathcal{W}},\|w\|=\nu} \mathbb{E}\big(\eta_w(t_0;h)\big)\, D_w\chi(x_0). \tag{41}$$

Remark 2. For the Langevin system considered above, the Stratonovich and Ito interpretation coincide, due to the additivity of the noise. Since the Ito extended alphabet has more letters than the Stratonovich alphabet, it is more convenient to work with (38)–(39) than with (40)–(41). However it is possible to show the superiority of ABOBA over ABOBA by comparing (40)–(41) for both algorithms. Details will not be given. The relation between Stratonovich and Ito solution is taken up in Section 8 below.

5.2. **Stratonovich order conditions.** If $\mu \in \mathbb{N}/2$, $\mu > 0$, we shall say that the integrator has *strong order* $\geq \mu$ if the series (38) only comprises terms of weight $\geq \mu + 1/2$, i.e. of size $\mathcal{O}(h^{\mu+1/2})$. From Theorem 5.1 it is clear that for $\mu \in \mathbb{N}/2$, $\mu > 0$, the *strong order conditions*,

$$\widetilde{J}_w(t_0+h;t_0) = J_w(t_0+h;t_0), \qquad w\in\mathcal{W},\ \|w\| = 1/2, 1, 3/2, \dots, \mu, \tag{42}$$

are *sufficient* to guarantee strong order $\geq \mu$. Under suitable assumptions on (3), it may be proved that when the order conditions hold the local error actually possesses a $\mathcal{O}(h^{\mu+1/2})$ bound in the L^p norms, $p < \infty$. Here our interest lies in the combinatorial aspects of the theory and will not be concerned with the derivation of such bounds; the interested reader is referred to [1].

Are the strong order conditions (42) *necessary* as well as sufficient to achieve strong order $\geq \mu$? This question may be discussed in two different scenarios:

- *Specific system.* In this case we are only interested in (3) for a fixed, specific choice of dimension d and vector fields f_ℓ in $\mathbb{R}^d$.
- *General system.* Here $\mathcal{A}$ and the coefficients $\widetilde{J}_w$ are fixed and one demands that the series (42) only comprises terms of weight $\geq \mu + 1/2$ for each choice of d and each choice of vector fields f_ℓ, $\ell \in \mathcal{A}$, in (3).

While the general system scenario is not without mathematical interest, in practice it is the specific system case that matters. This point, that would be true for any numerical integrator, is especially so for splitting algorithms: one of the main advantages of the splitting idea is its versatility to be tailored to the specific problem at hand.

In the specific system scenario is possible that for some words w the word basis functions f_w vanish at each x_0. If that is the case, it is not necessary to impose the order conditions $\delta_w = 0$ associated with such words. This is illustrated in [1] in the case of the Langevin dynamics, whose structure implies that many f_w vanish.

In the general system scenario the conditions (42) *are necessary* for strong order $\geq \mu$, in view of the second item of the lemma below that show that the word basis functions are independent.

Lemma 5.3. *Fix the alphabet $\mathcal{A}$ and choose $w \in \mathcal{W}$, $w \neq \emptyset$. There exist a value of the dimension d, vector fields f_ℓ, $\ell \in \mathcal{A}$, in $\mathbb{R}^d$, and a scalar observable χ, which depend on $\mathcal{A}$ and w, such that,*

- $D_w\chi(0) = 1$ *and* $D_u\chi(0) = 0$ *for each nonempty* $u \in \mathcal{W}$, $u \neq w$.
- *The first component* $f_u^1(0)$ *of the vector* $f_u(0) \in \mathbb{R}^d$ *vanishes for each nonempty* $u \in \mathcal{W}$, $u \neq w$, *while* $f_w^1(0) = 1$.

Proof. The first item follows from the second by choosing χ to be the first coordinate mapping $x \mapsto x^1$.

For the second item, the idea of the proof is best understood by means of an example. Suppose that $w = \ell\ell m\ell m$ with $\ell \neq m$. Then set $d = 5$,

$$f_\ell(x) = [0, x^3, 0, x^5, 1]^T, \qquad f_m(x) = [x^2, 0, x^4, 0, 0]^T.$$

(recall that superscripts denote components) and $f_k(x) = 0$ any remaining letters. Thus

$$\sum_{k\in\mathcal{A}} f_k(x) = [x^2, x^3, x^4, x^5, 1]^T.$$

Because second and higher derivatives of the fields vanish, the recurrence (13) shows that for any word $u = k_n \dots k_1$

$$f_u(0) = f'_{k_1} \cdots f'_{k_{n-1}} f_{k_n}(0).$$

Assume that $f_u^1(0) \neq 0$. The Jacobian matrix f'_{k_1} must have a nonzero element in its first row and this implies that $k_1 = m$. Then, by definition of f_m, the first row is $[0, 1, 0, \dots, 0]$, so that the second component of

$$f'_{k_2} \cdots f'_{k_{n-1}} f_{k_n}(0)$$

must be nonzero. This implies that $k_2 = \ell$. By repeating this argument, we conclude that $u = w$ and $f_w^1(0) = 1$.

For a general word w, things are as follows. The dimension d is taken equal to the number of letters in w. The field

$$\sum_{k\in\mathcal{A}} f_k(x) = [x^2, x^3, \dots, x^d, 1]^T$$

is split in such a way that its $d - j + 1$, $j = 1, \dots, d$ component is assigned to the field f_k if k is the letter that occupies the j-th position in w. (In this way there as many nonzero vector fields f_k as distinct letters in w.) ☐

For $\sigma \in \mathbb{N}$, $\sigma > 0$, the *weak order conditions*

$$\mathbb{E}\big(\widetilde{J}_w(t_0 + h; t_0)\big) = \mathbb{E}\big(J_w(t_0 + h; t_0)\big), \qquad w \in \mathcal{W},\ \|w\| = 1, 2, \dots, \sigma, \tag{43}$$

are *sufficient* to ensure that the series in (39) only comprises terms of weight $\geq \sigma+1$, or, as we shall say, the integrator has *weak order* $\geq \sigma$. In a general system scenario the weak order conditions are also *necessary* in view of the first item in the preceding lemma.

5.3. **Ito order conditions.** For the Ito case the strong and week order conditions are

$$\widetilde{I}_w(t_0 + h; t_0) = I_w(t_0 + h; t_0), \qquad w \in \overline{\mathcal{W}}_0,\ \|w\| = 1/2, 1, 3/2, \dots, \mu, \tag{44}$$

and

$$\mathbb{E}\big(\widetilde{I}_w(t_0 + h; t_0)\big) = \mathbb{E}\big(I_w(t_0 + h; t_0)\big), \qquad w \in \overline{\mathcal{W}}, \quad \|w\| = 1, 2, \dots, \sigma, \tag{45}$$

respectively. They guarantee that the series in (40) (respectively (41)) only consists of terms of weight $\geq \mu$ (respectively $\geq \sigma$), or, as we shall say, the integrator has *strong order* $\geq \mu$ (respectively *weak order* $\geq \sigma$).

It is possible to show (but the very long proof will not be reproduced here) that, in the general system scenario and if $\mu = 1/2, 1, 3/2$, the conditions (44) are necessary to achieve strong order $\geq \mu$. Similarly it may be proved that (45) are necessary to have weak order $\geq \sigma$ for general systems if $\sigma = 1, 2, 3$. These particular values of μ and σ are sufficient for establish the order barrier in Theorem 7.2 below. We believe the strong are weak order conditions are necessary for arbitrary μ or σ but a proof is not yet available.

5.4. **Extensions.** In (3) or (4) is assumed that all vector fields f_a and f_A are equally important. In several applications this may not be the case. For instance, consider the system

$$dx = f_a(x)\,dt + \epsilon f_b(x)\,dt + f_A(x) \circ d\mathcal{B}_A, \qquad 0 < \epsilon \ll 1,$$

or its Ito counterpart, where the split systems $\{a, A\}$, $\{b\}$ may be solved in closed form. Thus we are dealing with small perturbation of an integrable system and it makes sense, when expanding the local error, to track not only powers of h but also powers of ϵ, as in done in e.g. [5] in the deterministic scenario. That task is easily accomplished with the tools presented so far. Details will not be given.

6. **The shuffle and quasishuffle products.**

The conditions in (42) are *not independent*; for instance the order condition corresponding to the two-letter word $\ell\ell$ is fulfilled whenever the order condition for ℓ is fulfilled. (Note that the dependence between order conditions and the necessity of the order conditions discussed above are completely different issues.) Similarly there are dependencies within each of the set of conditions (43), (44) and (45). The study of this issue requires the help of the shuffle and quasishuffle products. More generally, these products play a key role when working in many developments involving elements of $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ or $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ [41]. In the deterministic case the shuffle relations between iterated integrated were first noted by Ree [40]. The stochastic scenario was addressed by Gaines [20]. On the other hand there is much literature relating the shuffle and quasishuffle products to stochastic integration, see e.g. [18].

We begin with the Stratonovich/shuffle case. The more complicated Ito/quasishuffle case is presented later.

6.1. **The shuffle product.** To motivate the introduction of the shuffle product, we begin by noting that if $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is any mapping and Φ the associated pullback operator, then, for any pair of scalar-valued observables χ_1, χ_2,

$$\Phi(\chi_1 \cdot \chi_2) = (\Phi\chi_1) \cdot (\Phi\chi_2),$$

where $\cdot$ denotes the standard (pointwise) product of observables, i.e. $(\chi_1 \cdot \chi_2)(x) = \chi_1(x)\chi_2(x)$ for $x \in \mathbb{R}^d$. In other words Φ is *multiplicative*. The series of differential operators D_J and $D_{\widetilde{J}}$ that expand the pullback operators associated with $\varphi_{t_0+h;t_0}$ and $\psi_{t_0+h;t_0}$ are similarly multiplicative. Now, it is easily checked that if $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$, then, in general

$$D_S(\chi_1 \cdot \chi_2) \neq (D_S\chi_1) \cdot (D_S\chi_2)$$

(for instance, if $S = \ell \in \mathcal{A}$, then $D_S(\chi_1 \cdot \chi_2) = (D_S\chi_1) \cdot \chi_2 + \chi_1 \cdot (D_S\chi_2)$). Therefore the coefficients J_w and $J_{\widetilde{w}}$ of the series D_J and $D_{\widetilde{J}}$ must have some special property that tells them apart form "general" coefficients; as we shall see, that property explains the dependence between the order conditions.

In order to identify when a series D_S is multiplicative, we first investigate the action of a differential operator D_w, $w \in \mathcal{W}$, on a product $\chi_1 \cdot \chi_2$. For instance, for $k, \ell, m \in \mathcal{A}$, a trivial computation leads to

$$\begin{aligned} D_{k\ell m}(\chi_1 \cdot \chi_2) \quad = \quad & (D_{k\ell m}\chi_1)\cdot(D_\emptyset\chi_2) + (D_{k\ell}\chi_1)\cdot(D_m\chi_2) \\ & +(D_{km}\chi_1)\cdot(D_\ell\chi_2) + (D_{\ell m}\chi_1)\cdot(D_k\chi_2) \\ & +(D_k\chi_1)\cdot(D_{\ell m}\chi_2) + (D_\ell\chi_1)\cdot(D_{km}\chi_2) \\ & +(D_m\chi_1)\cdot(D_{k\ell}\chi_2) + (D_\emptyset\chi_1)\cdot(D_{k\ell m}\chi_1). \end{aligned}$$

The right-hand side contains eight pairs of words $(k\ell m, \emptyset)$, $(k\ell, m)$, ... What do these pairs have in common? They are precisely the pairs such that when *shuffled* give rise to the word $k\ell m$ in the left-hand side. By definition, the *shuffle product* $u \sqcup\!\sqcup v$ of two words with m and n letters is the sum of the $(m+n)!/(m!n!)$ words that may be formed by interleaving the letters of u with those of v while keeping the letters in the same order as they appear in u and v. For instance $k\ell \sqcup\!\sqcup m = k\ell m + km\ell + \ell km$, $\ell \sqcup\!\sqcup \ell = \ell + \ell = 2\ell$, etc. More formally, the shuffle product of words may defined recursively by the relations [41, Section 1.4]

$$\emptyset \sqcup\!\sqcup \emptyset = \emptyset, \qquad \emptyset \sqcup\!\sqcup \ell = \ell \sqcup\!\sqcup \emptyset = \ell, \quad \ell \in \mathcal{A},$$

and

$$u\ell \sqcup\!\sqcup vm = (u\ell \sqcup\!\sqcup v)m + (u \sqcup\!\sqcup vm)\ell, \qquad u, v \in \mathcal{W}, \quad \ell, m \in \mathcal{A}. \tag{46}$$

The last equality corresponds to the fact that the words arising from shuffling $u\ell$ and vm necessarily end with either the last letter of $u\ell$ or the last letter of vm. Note that for words $u, v \in \mathcal{W}$, the shuffle $u \sqcup\!\sqcup v$ is in general not a word but an element of the space $\mathbb{R}\langle\mathcal{A}\rangle$ of linear combination of words. By linearity, the shuffle product may be trivially extended to a commutative, associative product in $\mathbb{R}\langle\mathcal{A}\rangle$; for instance $(3k+\ell) \sqcup\!\sqcup (\ell - m) = 3k\ell + 3\ell k - 3km - 3mk + 2\ell\ell - \ell m - m\ell$.

At this stage we introduce some additional notation that will be used frequently below. If $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ is a series and $p = \sum_w p_w w \in \mathbb{R}\langle\mathcal{A}\rangle$, we set

$$(S, p) = \sum_w S_w p_w; \tag{47}$$

the sum is well defined because only a finite number of coefficients p_w are $\neq 0$. In the case where p coincides with a word w, (S, w) is just the coefficient S_w; for general p, (S, w) is a linear combination of coefficients S_w. Obviously $(\cdot, \cdot)$ is a real-valued bilinear map. With this notation, we may present the following result (that generalizes the formula for $D_{k\ell m}(\chi_1 \cdot \chi_2)$ above).

Proposition 5. *For any $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ and any pair of observables*

$$D_S(\chi_1 \cdot \chi_2) = \sum_{u,v \in \mathcal{W}} (S, u \sqcup\!\sqcup v)\, D_u\chi_1 \cdot D_v\chi_2.$$

Proof. It is sufficient to prove the case where S coincides with a word. The proof is by induction on the length (number of letters) of the word (not to be confused with its weight). When S is the empty word the result is trivial because, necessarily, in the right-hand side $(S, u \sqcup\!\sqcup v)$ vanishes except if $u = v = \emptyset$ when $(S, u \sqcup\!\sqcup v) = 1$. We assume that the result is true for the word w and prove it for the longer word $w\ell$. Since D_ℓ is a first-order differential operator we may write

$$D_{w\ell}(\chi_1 \cdot \chi_2) = D_w D_\ell(\chi_1 \cdot \chi_2) = D_w(D_\ell\chi_1 \cdot \chi_2 + \chi_1 \cdot D_\ell\chi_2),$$

so that, by the induction hypothesis,

$$D_{w\ell}(\chi_1 \cdot \chi_2) = \sum_{u,v\in\mathcal{W}} (w, u \sqcup\!\sqcup v) \Big(D_{u\ell}\chi_1 \cdot D_v\chi_2 + D_u\chi_1 \cdot D_{v\ell}\chi_2 \Big).$$

Now from the definition of shuffle $(w, u \sqcup\!\sqcup v) = (w\ell, u\ell \sqcup\!\sqcup v) = (w\ell, u \sqcup\!\sqcup v\ell)$ and therefore

$$D_{w\ell}(\chi_1 \cdot \chi_2) = \sum_{u,v\in\mathcal{W}} (w\ell, u\ell \sqcup\!\sqcup v)\, D_{u\ell}\chi_1 \cdot D_v\chi_2 + \sum_{u,v\in\mathcal{W}} (w\ell, u \sqcup\!\sqcup v\ell) D_u\chi_1 \cdot D_{v\ell}\chi_2.$$

The proof concludes by observing that, when $(w\ell, u' \sqcup\!\sqcup v')$ is $\neq 0$, i.e. when $w\ell$ is one of the words arising when shuffling u' and v', the last letter in $w\ell$ must be either the last letter of u' or the last letter of v', so that either $u' = u\ell$ or $v' = v\ell$. □

Since, clearly

$$(D_S\chi_1) \cdot (D_S\chi_2) = \sum_{u,v\in\mathcal{W}} (S, u)(S, v)\, D_u\chi_1 \cdot D_v\chi_2,$$

we may write

$$D_{w\ell}(\chi_1 \cdot \chi_2) - (D_S\chi_1) \cdot (D_S\chi_2) = \sum_{u,v\in\mathcal{W}} \Big((S, u \sqcup\!\sqcup v) - (S, u)(S, v) \Big) D_u\chi_1 \cdot D_v\chi_2. \quad (48)$$

This leads trivially to next result:

Proposition 6. *Consider a series $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$, $S \neq 0$. The series of operators D_S is multiplicative if $S_\emptyset = 1$ and for each pair of words $u, v \in \mathcal{W}$, the so-called shuffle relation*

$$(S, u \sqcup\!\sqcup v) = (S, u)(S, v)$$

holds.

Thus the shuffle relations are equations that link the different coefficients S_w, $w \in \mathcal{W}$. For instance, from the shuffle $\ell \sqcup\!\sqcup \ell = 2\ell\ell$, $\ell \in \mathcal{A}$, we have the shuffle relation $S_\ell^2 = 2S_{\ell\ell}$ and, from the shuffle $k \sqcup\!\sqcup \ell = k\ell + \ell k$, $S_k S_\ell = S_{k\ell} + S_{\ell k}$.

Proposition 6 in tandem with the following result give a new proof of the multiplicativity of D_J that we pointed out above.

Proposition 7. *The Stratonovich iterated integrals $J_w(t_0 + h; t_0)$ satisfy the shuffle relations.*

Proof. For the shuffling of two letters $\ell, m \in \mathcal{A}$, the integration by parts formula

$$\begin{aligned} &\big(\mathcal{B}_\ell(t_0 + h) - \mathcal{B}_\ell(t_0)\big)\big(\mathcal{B}_m(t_0 + h) - \mathcal{B}_m(t_0)\big) \\ &= \int_{t_0}^{t_0+h} \big(\mathcal{B}_\ell(t_0+s) - \mathcal{B}_\ell(t_0)\big) \circ d\mathcal{B}_m(s) + \int_{t_0}^{t_0+h} \big(\mathcal{B}_m(t_0+s) - \mathcal{B}_m(t_0)\big) \circ d\mathcal{B}_\ell(s), \quad (49) \end{aligned}$$

is a statement of the shuffle relation $J_\ell J_m = J_{m\ell} + J_{\ell m}$ (recall that if ℓ or m are not stochastic, then $\mathcal{B}_\ell(t) = t$ or $\mathcal{B}_m(t) = t$ respectively). General shuffles are dealt with by induction based on the recursive definition of the shuffle product in (46) and the recursion (9) for the iterated integrals. □

To present a similar result for the integrator we need a lemma:

Lemma 6.1. *Let $S, T \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$, with $S_\emptyset = T_\emptyset = 1$, satisfy the shuffle relations. Then the product ST has $(ST)_\emptyset = 1$ and satisfies the shuffle relations.*

Proof. Recall that the coefficients of ST are given by the convolution product as in (33), which is based on deconcatenation. The result is a consequence of the following observation: the deconcatenation of the words in a shuffle $u \sqcup\!\sqcup v$ may be found by shuffling the deconcatenations of u and v. An example of this observation follows. Deconcatenation of the shuffle $k\ell \sqcup\!\sqcup m = k\ell m + km\ell + mk\ell$ gives the 12 pairs

$$(k\ell m, \emptyset) + (k\ell, m) + (k, \ell m) + (\emptyset, k\ell m) + (km\ell, \emptyset) + \cdots + (\emptyset, mk\ell).$$

On the other hand by deconcatenating $k\ell$ we obtain $(k\ell, \emptyset) + (k, \ell) + (\emptyset, k\ell)$, and by deconcatenating m obtain $(m, \emptyset) + (\emptyset, m)$. Shuffling now as in

$$\begin{aligned} (k\ell, \emptyset) \sqcup\!\sqcup (m, \emptyset) &= (k\ell \sqcup\!\sqcup m, \emptyset \sqcup\!\sqcup \emptyset) \\ (k\ell, \emptyset) \sqcup\!\sqcup (\emptyset, m) &= (k\ell \sqcup\!\sqcup \emptyset, \emptyset \sqcup\!\sqcup m), \end{aligned}$$

etc. yields the same 12 pairs (the first line of the display gives $(k\ell m + km\ell + mk\ell, \emptyset)$, the second $(k\ell, m)$, etc). To prove the observation in the general case, one may use the recurrence (46).

By using the observation, $(ST, u \sqcup\!\sqcup v)$ may be written as a sum of products

$$\sum_{ij} (S, u_i \sqcup\!\sqcup v_j)(T, u'_i \sqcup\!\sqcup v'_j),$$

($u_i u'_i = u$ and $v_j v'_j = v$) or, since S and T satisy the shuffle relations,

$$\begin{aligned} &\sum_{ij} (S, u_i)(S, v_j)(T, u'_i)(T, v'_j) \\ &\qquad = \sum_i (S, u_i)(T, u'_i) \sum_j (S, v_j)(T, v'_j) = (ST, u)(ST, v). \end{aligned}$$

□

Proposition 8. *For a splitting integrator for the Stratonovich system* (3) *the coefficients* $\widetilde{J}_w(t_0 + h; t_0)$ *satisfy the shuffle relations.*

Proof. The proof is a trivial consequence of the lemma, because $D_{\widetilde{J}}$ is a composition of solution operators D_{J_i} associated with the split systems and therefore, by the preceding proposition, a composition of operators that satisfy the shuffle conditions. □

After the last two propositions, it is easy to see that the Stratonovich strong order conditions are not independent. For instance from the shuffle relations $J_\ell(t_0 + h; t_0)^2 = 2J_{\ell\ell}(t_0 + h; t_0)$ and $\widetilde{J}_\ell(t_0 + h; t_0)^2 = 2\widetilde{J}_{\ell\ell}(t_0 + h; t_0)$, we conclude that the strong order condition $\widetilde{J}_{\ell\ell}(t_0 + h; t_0) = J_{\ell\ell}(t_0 + h; t_0)$ corresponding to the word $\ell\ell$ is fulfilled if the strong order condition $\widetilde{J}_\ell(t_0 + h; t_0) = J_\ell(t_0 + h; t_0)$ holds. Analogously, if $k \neq \ell$ the order condition for $k\ell$ is implied by those of ℓk, k and ℓ, etc. It is possible to obtain *independent* order conditions by keeping only the conditions corresponding to the so-called *Lyndon* words [20] that we describe next. We order the alphabet $\mathcal{A}$ and then order words lexicographically; a Lyndon word is a word that is strictly smaller than all the words obtained by rotating its letters. If the alphabet is $\mathcal{A} = \{a, A\}$ and $a < A$, then aA is a Lyndon word because it precedes the rotated Aa. Similarly aaA is a Lyndon word while aAa and Aaa are not. For this simple alphabet, the Lyndon words with three or fewer letters are a, A, aA, aaA, aAA; their order conditions are independent and imply, via the shuffle relations, the order conditions for aa, AA, aaa, aAa, Aaa, AaA, AAa and AAA.

For reasons of brevity, the independence of the Stratonovich weak order conditions will not be discussed in this paper.

Remark 3. From (48) and Lemma 5.3 the shuffle conditions are *necessary* for D_S to be multiplicative for each choice of d and vector fields f_ℓ, $\ell \in \mathcal{A}$. Hence the last two propositions may be proved in an alternative way: one first notices the multiplicativity of D_J and $D_{\widetilde{J}}$ as expansions of pullbacks associated with the true and numerical solution respectively and then concludes that the shuffle conditions are satisfied because the multiplicativity holds for all choices of vector fields. Recall from Remark 1 that changing the vector fields does not alter the iterated integrals.

6.2. **The quasishuffle product.** As we noted above, the shuffle property of the Stratonovich iterated integrals stems from the formula of integration by parts in (49). For the Ito calculus, formula (49) has to be replaced by

$$\begin{aligned}
&\big(\mathcal{B}_\ell(t_0+h)-\mathcal{B}_\ell(t_0)\big)\big(\mathcal{B}_m(t_0+h)-\mathcal{B}_m(t_0)\big)\\
&=\int_{t_0}^{t_0+h}\big(\mathcal{B}_\ell(t_0+s)-\mathcal{B}_\ell(t_0)\big)d\mathcal{B}_m(s)+\int_{t_0}^{t_0+h}\big(\mathcal{B}_m(t_0+s)-\mathcal{B}_m(t_0)\big)d\mathcal{B}_\ell(s)\\
&\quad+\big[\big(\mathcal{B}_\ell(t_0+h)-\mathcal{B}_\ell(t_0)\big),\big(\mathcal{B}_m(t_0+h)-\mathcal{B}_m(t_0)\big)\big]\,, \qquad (50)
\end{aligned}$$

where the last term represents the quadratic covariation (see e.g. [2, Chapter 5]). If $\ell = m \in \mathcal{A}_{\rm sto}$, then the quadratic covariation in (50) is h; for all other combinations of letters the quadratic covariation vanishes.

The *quasishuffle product* $\bowtie$ to be defined presently is such that for any two letters $\ell, m \in \bar{\mathcal{A}}$, the computation of $\ell \bowtie m$ mimics the integration by parts relation (50). In combinatorial algebra, the definition of a quasishuffle product depends on the choice of a so-called bracket $[\cdot,\cdot]$; different brackets lead to different quasishuffle products as defined by Hoffman [24]. Throughout this paper we only work with one fixed choice of bracket defined as follows. For letters $\ell, m \in \bar{\mathcal{A}}$, $[\ell, m]$ takes the value $\bar{A} \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ if $\ell = m = A \in \mathcal{A}_{\rm sto}$; $[\ell, m] = 0 \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ in all other cases. Then the quasishuffle product of words $u \bowtie v \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ is defined recursively by

$$\emptyset \bowtie \emptyset = \emptyset, \qquad \emptyset \bowtie \ell = \ell \bowtie \emptyset = \ell, \quad \ell \in \bar{\mathcal{A}},$$

and

$$u\ell \bowtie vm = (u\ell \bowtie v)m + (u \bowtie vm)\ell + (u \bowtie v)[x, y], \qquad u, v \in \mathcal{W}, \quad \ell, m \in \bar{\mathcal{A}}.$$

In the particular case $u = v = \emptyset$, the last relation yields $\ell \bowtie m = \ell m + m\ell + [\ell, m]$, a transcription of (50).

The next four results are counterparts of Propositions 5–8. The bilinear form $(\cdot,\cdot)$ in (47), which we defined in $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \times \mathbb{R}\langle\mathcal{A}\rangle$, is now extended to $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \times \mathbb{R}\langle\bar{\mathcal{A}}\rangle$.

Proposition 9. *For any $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ and any pair of observables*

$$D_S(\chi_1 \cdot \chi_2) = \sum_{u,v\in\overline{\mathcal{W}}} (S, u \bowtie v)\, D_u\chi_1 \cdot D_v\chi_2.$$

Proof. One may use the same technique as in Proposition 5. Here the proof is lengthier because it has to contemplate the possibility $\ell = \bar{A}$, $A \in \mathcal{A}_{\rm sto}$ in which case D_ℓ is a second order operator. □

This yields immediately:

Proposition 10. *Consider a series* $S \in \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$, $S \neq 0$. *Then the series of operators* D_S *is multiplicative if* $S_\emptyset = 1$ *and for each pair of words* $u, v \in \mathcal{W}$, *the quasishuffle relation*

$$(S, u \bowtie v) = (S, u)(S, v)$$

holds.

The proofs of the following propositions are similar to those of Propositions 7 and 8 respectively.

Proposition 11. *The the Ito iterated integrals* $I_w(t_0+h; t_0)$ *satisfy the quasishuffle relations.*

Proposition 12. *For a splitting integrator for the Ito system* (4), *the coefficients* $\widetilde{I}_w(t_0 + h; t_0)$ *satisfy the quasishuffle relations.*

The last two propositions show immediately that the Ito strong order conditions are not independent. The dependence between the Ito weak order conditions will be discussed after Proposition 17.

6.3. **Concatenating Chen series.** The shuffle (quasishuffle) relations constrain the values of Stratonovich (Ito) iterated integrals corresponding to different words but based on a common interval $(t_0, t_0 + h)$. Iterated integrals corresponding to adjacent intervals are also interrelated, as we now discuss.

Solution operators of Stratonovich or Ito systems satisfy

$$\varphi_{t_2,t_1} \circ \varphi_{t_1,t_0} = \varphi_{t_2,t_0}, \qquad t_2 \geq t_1 \geq t_0.$$

From here we get the following relations between series of operators

$$D_{J(t_1;t_0)} D_{J(t_2;t_1)} = D_{J(t_2;t_0)}, \quad D_{I(t_1;t_0)} D_{I(t_2;t_1)} = D_{I(t_2;t_0)}, \quad t_2 \geq t_1 \geq t_0;$$

the corresponding relations between elements of $\mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ or $\mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$ (Chen series) are

$$J(t_1; t_0) J(t_2; t_1) = J(t_2; t_0); \quad I(t_1; t_0) I(t_2; t_1) = I(t_2; t_0), \quad t_2 \geq t_1 \geq t_0. \tag{51}$$

The equalities in (51) are, in view of (33), a family of relations between iterated integrals first noted by Chen [16] in the case where there are no stochastic letter. For instance, for words with two letters:

$$J_{\ell m}(t_2; t_0)^2 = J_{\ell m}(t_1; t_0) + J_\ell(t_1; t_0) J_m(t_2; t_1) + J_{\ell m}(t_2; t_1),$$

etc. These relations may alternatively be proved by manipulating the integrals, without going through the series of differential operators as above.

6.4. **Composing word series.** We conclude our study of the shuffle and quasishuffle products by showing that, in some circumstances, the composition $\mathcal{W}_T(\mathcal{W}_S(x))$ of two word series is another word series.

Let us begin with the Stratonovich case. If χ is an observable and $w \in \mathcal{W}$, then $D_w \chi$ is a sum of terms each of which is a derivative $\chi^{(s)}(x)$ acting on combinations of derivatives of the functions f_k, $k \in \mathcal{A}$. A simple example is:

$$D_{\ell m} \chi(x) = \chi''(x)[f_\ell(x), f_m(x)] + \chi'(x) f'_m(x) f_\ell(x).$$

Here, the word ℓm may have weight 1, 3/2 or 2 depending of whether ℓ and m are stochastic of deterministic; the thing to observe is that in each term of the right-hand side of the last equality the f_k $k \in \mathcal{A}$, that appear have a combined weight that matches the weight of ℓm.

If $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ and D_S is the corresponding series of differential operators we may arrange $D_S\chi$ by grouping the terms where the combined weight of the f_k that appear is successively 0, 1/2, 1, 3/2, etc. On the other hand if $\mathcal{W}_S(x)$ is the associated word series and $S_\emptyset = 1$ so that $W_S(x) - x = \mathcal{O}(1/2)$, we may Taylor expand as follows

$$\chi(\mathcal{W}_S(x)) = \chi(x + [W_S(x) - x]) = \chi(x) + \chi'(x)[W_S(x) - x] + \frac{1}{2}\chi''(x)[W_S(x) - x, W_S(x) - x] + \cdots$$

Here the right-hand side may be arranged, as we did in the case of $D_S\chi$, by grouping the terms where the combined weight of the f_k that appear is successively 0, 1/2, 1, 3/2, etc. This arrangement may be carried out because $[W_S(x)-x]^r$ only contributes terms of combined weight $\geq r/2$ and therefore for each weight there is only a finite number of terms to be grouped. It turns out that if S is multiplicative the expansions of $D_S\chi(x)$ and $\chi(W_S(x))$ coincide.

Proposition 13. *Suppose that $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ has $S_\emptyset = 1$ and satisfies the shuffle relations. Then for any observable χ, the expansion of $\chi(W_S(x))$ coincides with $D_S\chi(x)$.*

Proof. If χ is one of the coordinate mappings $x \mapsto x^i$, then the result is true because, by definiton, the i-th component of the word-basis function f_w is obtained by applying D_w to the i-th coordinate mapping. If χ is a product of coordinate mappings, the result holds because D_S acts multiplicatively. By linearity the result is true if χ is a polynomial. Then the result hold for smooth χ because it holds for the Taylor polynomials of any degree of χ around any base point x. □

As a direct consequence we may state:

Proposition 14. *Suppose that $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ has $S_\emptyset = 1$ and satisfies the shuffle relations. Then for any $T \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$, $\mathcal{W}_T(\mathcal{W}_S(x))$ coincides with the words series $W_{ST}(x)$.*

Proof. It is enough to note that, for each word basis function $f_w(x) = D_w id(x)$, according to the preceding proposition, $f_w(W_S(x))$ has the expansion $D_S f_w(x) = D_S D_w id(x)$. □

The Ito case is completely parallel; the only change is that $S \in \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$ has to be demanded to satisfy the quasishuffle relations rather than the shuffle relations.

In fact the computations leading to (34) or (36) are instances of the composition just described.

7. Infinitesimal generators.

It is well known that the infinitesimal generators of (3) or (4) play an important role in the study of these systems, see e.g. [38, Section 2.5]. In this section those generators are described in the language of words. The material has an important implications for the weak order conditions. We begin with Ito systems.

7.1. The Ito generator.

For system (4), we consider the linear combination of *deterministic* letters

$$\mathfrak{G} = \sum_{\ell \in \bar{\mathcal{A}}_{\mathrm{det}}} \ell = \sum_{a \in \mathcal{A}_{\mathrm{det}}} a + \sum_{A \in \mathcal{A}_{\mathrm{sto}}} \bar{A}$$

and define the exponential $\exp(h\mathfrak{G}) \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$, $h \in \mathbb{R}$, as the series

$$\emptyset + h\mathfrak{G} + \frac{h^2}{2}\mathfrak{G}^2 + \cdots,$$

where the powers are based on concatenation, e.g.

$$\mathfrak{G}\mathfrak{G} = \sum_{a,b\in\mathcal{A}_{\mathrm{det}}} ab + \sum_{a\in\mathcal{A}_{\mathrm{det}},B\in\mathcal{A}_{\mathrm{sto}}} a\bar{B} + \sum_{A\in\mathcal{A}_{\mathrm{sto}},b\in\mathcal{A}_{\mathrm{det}}} \bar{A}b + \sum_{A,B\in\mathcal{A}_{\mathrm{sto}}} \bar{A}\bar{B}$$

(note that the right-hand side is simply the sum all the words consisting of two deterministic letters from $\bar{\mathcal{A}}$). The operator $D_{\mathfrak{G}}$ is the *infinitesimal generator* of (4), a linear combination of first and second order differential operators.

Proposition 15. *The expectations of the Ito iterated integrals are given by* $\mathbb{E}I_w(t_0+h; t_0) = 0$ *if* $w \in \overline{\mathcal{W}}$ *has at least one stochastic letter and* $\mathbb{E}I_w(t_0+h;t_0) = h^n/n!$ *if* $w \in \overline{\mathcal{W}}$ *consists of* n *deterministic letters.*

The following relation holds:

$$\mathbb{E}I(t_0+h;t_0) = \exp(h\mathfrak{G}).$$

For any observable and $h > 0$*,*

$$\mathbb{E}\chi(x(t_0+h)) = \exp(hD_{\mathfrak{G}})\chi(x_0),$$

where $x(t)$ *solves* (4) *with* $x(t_0) = x_0$ *and the expectation is conditional on* x_0*.*

Proof. For the first claim we recall that the expectation of Ito integrals vanishes. In addition it is trivially computed that, when all the letters in a word are deterministic, $I_w(t_0+h;t_0) = h^n/n!$, where n represents the number of letters.

By expanding $\exp(h\mathfrak{G}$ as a series, one sees that the second claim is just a reformulation of the first. An alternative proof of this second claim is as follows. As noted before (Proposition 2), the distribution of the random variable $I(t_0+h;t_0)$ is independent of t_0 and therefore we may write $\mathbb{E}I(t_0+h;t_0) = \mathbb{E}I(h)$. The functions $\exp(h\mathfrak{G})$ and $EI(h)$ coincide at $h = 0$, where they take the common value $\emptyset$. By taking expectations in (51), we find the semigroup relation $\mathbb{E}(h_1+h_2) = \mathbb{E}I(h_1)\,\mathbb{E}I(h_2)$ for $h_1, h_2 \geq 0$. Differentiating with respect to h_1 and then setting $h_1 = 0$, $h_2 = h$ yields the linear, constant coefficient differential equation $(d/dt)\mathbb{E}(h) = [(d/dh)\mathbb{E}I(0)]\,\mathbb{E}I(h)$.[1] On the other hand, a straightforward computation leads to $(d/dh)\exp(h\mathfrak{G}) = \mathfrak{G}\exp(h\mathfrak{G})$, and the proof of the second statement concludes by noting that $(d/dh)\mathbb{E}I(0) = \mathfrak{G}$ since $\mathbb{E}I_w(h) = o(h)$ as $h \downarrow 0$ if w has length > 1 and all its letters are deterministic.

For the last claim, just take expectations in (17). ☐

Remark 4. The preceding proposition and the quasishuffle relations among the I_w (Proposition 11) make it possible to compute all the *moments* of the Ito iterated integrals, as first suggested by Gaines [20]. The easiest example is given by the relation $A \bowtie A = 2AA + \bar{A}$ that leads to $I_A^2 = 2I_{AA} + I_{\bar{A}}$; according to the proposition the expectation of the right-hand side equals $0+h$ and therefore $\mathbb{E}I_A^2 = h$, a well known property of the Brownian increment I_A. The values of $\mathbb{E}I_\ell^i$, $\mathbb{E}(I_\ell^i I_{\ell m}^j)$, $\ell, m \in \bar{\mathcal{A}}$, $i, j \in \mathbb{N}$, etc. may be computed similarly after writing the appropriate quasishuffles.

[1]This differential equation in $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ is of course just a system of differential equations for the coefficients $\mathbb{E}I_w(h)$, $w \in \overline{\mathcal{W}}$, that presents no technical difficulty.

Proposition 16. *The expectations* $\mathbb{E}I_w(t_0+h;t_0)$, $w \in \overline{\mathcal{W}}$ *of the Ito iterated integrals satisfy the shuffle relations.*

Proof. This result is an easy consequence of the Proposition 15. With the abbreviation $S = \exp(h\mathfrak{G})$, $(S,u)(S,v)$ and $(S, u \sqcup\!\sqcup v)$ are both 0 if either u or v have a stochastic letter. In other case, if u has m letters and v has n, $(S,u) = h^m/m!$, $(S,v) = h^n/n!$ while $(S, u \sqcup\!\sqcup v)$ is a sum of $(m+n)!/(m!n!)$ coefficients S_w each of them with value $h^{m+n}/((m+n)!$. $\square$

7.2. **Weak order conditions in the Ito case.** We now turn to the series of expectations associated with a splitting integrator specified by the pullback series

$$\widetilde{I}(t_0+h;t_0) = I^{(1)}(t_0+d_1h;t_0+c_1h)\cdots I^{(m)}(t_0+d_mh;t_0+c_mh).$$

In general, the equality

$$\mathbb{E}\widetilde{I}(t_0+h;t_0) = \mathbb{E}I^{(1)}(t_0+d_1h;t_0+c_1h)\cdots \mathbb{E}I^{(m)}(t_0+d_mh;t_0+c_mh) \qquad (52)$$

will *not* hold because the $I^{(i)}(t_0+d_ih;t_0+c_ih)$ are *not independent.* However, as it will shortly become clear, (52) will typically be satisfied. We first present some examples that will help to understand the situation.

Assume that the alphabet $\mathcal{A}$ consists of two letters a and A. Choose a partition of the interval $[0,1]$

$$0 = c_1' < d_1' = c_2' < d_2' = c_3' < \cdots < d_{\nu-1}' = c_\nu' < d_\nu' = 1$$

and let f_A act in the intervals $[t_0+c_i'h, t_0+d_i'h]$, while the deterministic f_a may act on any set of intervals. In this case (52) holds because the Brownian motion $\mathcal{B}_A$ acts on nonoverlapping intervals. This example may be easily extended to the case where there are additional deterministic fields f_b, f_c, ...; in the split systems some of them could be grouped with f_a and some of them grouped with f_A.

As a second example, assume that $\mathcal{A} = \{A, B\}$ and use Strang's splitting with f_A acting first. Here $I^{(1)}$ and $I^{(3)}$ are independent because their intervals do not overlap, while the pairs $I^{(1)}$, $I^{(2)}$ and $I^{(2)}$, $I^{(3)}$ are independent because they use independent Brownian motions. Again this example may be easily generalized by adding additional deterministic and/or stochastic letters.

We have the following general result:

Lemma 7.1. *Assume that* $\mathcal{A}_{\rm sto} \neq \emptyset$ *so that* (4) *does not degenerate into a deterministic differential equations. If a splitting integrator for* (4) *has strong order* > 0 *(i.e.* $\geq 1/2$*, then* (52) *holds.*

Proof. As noted at the end of Section 5, the Ito strong order conditions with $\mu = 1/2$ must be satisfied. Now for each $A \in \mathcal{A}_{\rm sto}$ the strong order condition corresponding to A, shows that $\sum_j I_A(t_0+d_{i_j}h;t_0+c_{i_j}) = I_A(t_0+h;t_0)$, where the sum is extended to all partial systems that include f_A. This implies that, for each fixed A, the corresponding intervals $[c_{i_j}, d_{i_j}]$ cover the interval $[0,1]$ and therefore cannot overlap. $\square$

Schemes that satisfy (52) have the special properties that we study next. To begin with, Lemma 6.1 clearly implies:

Proposition 17. *For splitting integrators for* (4) *that satisfy* (52)*, the expectations coefficients* $\mathbb{E}\widetilde{I}_w(t_0+h;t_0)$ *satisfy the shuffle conditions.*

In turn this result and Proposition 16 show that the weak order conditions are not independent when (52) holds. For instance the weak order condition for $\ell\ell$ is implied by the weak order condition for $\ell \in \overline{W}$, since, as noted repeatedly, $\ell \sqcup\!\sqcup\, \ell = 2\ell\ell$.

In the next proposition we need the *deterministic* system:

$$dx = \sum_{a\in\mathcal{A}_{\rm det}} f_a(x)\,dt + \sum_{A\in\mathcal{A}_{\rm sto}} f_A(x)\,dt, \tag{53}$$

obtained by replacing the differentials $d\mathcal{B}_A$ in the Ito system (4) by dt. It is clear that each splitting algorithm for (4) defines a splitting algorithm for (53) and vice versa.

Proposition 18. *For splitting integrators for* (4) *that satisfy* (52) *and in the general vector field scenario, the following properties are equivalent:*

- *The weak order conditions* (45) *hold for a positive integer σ.*
- *When applied to the deterministic system* (53)*, the integrator has local error $\mathcal{O}(h^{\sigma+1})$.*

Proof. From (52) and Proposition 15

$$\mathbb{E}\widetilde{I}(t_0+h;t_0) = \exp\left(h(d_1-c_1)\mathfrak{G}^{(1)}\right)\cdots\exp\left(h(d_m-c_m)\mathfrak{G}^{(m)}\right),$$

where the $\mathfrak{G}^{(i)}$ are the generators of the partial systems and therefore sums of deterministic letters. Condition (45), requires that, in the series in the last display, the terms corresponding to words with $\leq \sigma$ letters coincide with those of

$$\mathbb{E}I(t_0+h;t_0) = \exp(h\mathfrak{G}).$$

To study the order for (53) we may also use words seeing a deterministic system as the particular case of Ito system where there is no stochastic letter. If we denote by $\bar{A}$ the (deterministic) letter associated with the field f_A, we then have

$$\widetilde{I}(t_0+h;t_0) = \exp\left(h(d_1-c_1)\mathfrak{G}^{(1)}\right)\cdots\exp\left(h(d_m-c_m)\mathfrak{G}^{(m)}\right),$$

and

$$I(t_0+h;t_0) = \exp(h\mathfrak{G}),$$

and order σ requires that the terms involving words with σ or fewer letters in the series in the last two displays coincide. $\square$

The following counterexample shows that, in the last two propositions, hypothesis (52) cannot be dispensed with. For the alphabet $\mathcal{A} = \{a, A\}$, we consider the integrator

$$\varphi^{(A)}_{t_0+h/2,t_0} \circ \varphi^{(a)}_{t_0+h,t_0} \circ \varphi^{(A)}_{t_0+h/2,t_0}.$$

While this is admittedly a contrived example, using the interval $[t_0, t_0+h/2]$ to finish the step (rather than the more natural $[t_0+h/2, t_0+h]$) may have some appeal. On the one hand the distribution of the iterated integrals in $[t_0, t_0+h/2]$ is the same as that in $[t_0+h/2, t_0+h]$) and, on the other hand, working twice with $[t_0, t_0+h/2]$ may make it possible to reuse Brownian increments. For this integrator the hypothesis (52) does not hold. A simple computation, similar to that preceding (36), yields

$$\widetilde{I}(t_0+h;t_0) = 1\emptyset + 2I_A A + I_a a + [I_A^2 + 2I_{AA}]AA + I_{\bar{A}}\bar{A} + \mathcal{O}(3/2)$$

(the iterated integrals in the right-hand side are over $[t_0, t_0 + h/2]$). We note in relation with Lemma 7.1 that here the order condition for A is obviously not satisfied. Taking expectations in the last display,

$$\mathbb{E}\widetilde{I}(t_0 + h; t_0) = 1\emptyset + 0A + ha + \frac{h}{2}AA + h\bar{A} + \mathcal{O}(2).$$

Since $0^2 \neq 2 \times h/2$, for the expectations, the shuffle relation corresponding to $A \sqcup\!\sqcup A = 2AA$ does not hold. On the other hand, from Proposition 15,

$$\mathbb{E}I(t_0 + h; t_0) = 1\emptyset + 0A + ha + 0AA + h\bar{A} + \mathcal{O}(2)$$

so that weak order conditions for $\sigma = 1$ are *not* satisfied. In the deterministic case the algorithm coincides with Strang's splitting with local errors $\mathcal{O}(h^3)$ (i.e. $\sigma = 2$). Thus the weak order does not coincide with the deterministic order.

It turns out that, in the general system scenario, under (52), there is an order barrier: the weak order cannot be better than $\sigma = 2$.

Theorem 7.2. *Assume that* (52) *holds. There is no splitting integrator for* (4) *with weak order* $\sigma \geq 3$.

Proof. By contradiction. As noted at the end of Section 5, the Ito weak conditions with $\sigma = 3$ holds. From Proposition 18 the algorithm is of order ≥ 3 for deterministic problems, which is known to be contradictory with the condition $c_{i_j} < d_{i_j}$ [3]. □

Remark 5. In the deterministic case this order barrier may be overcome by using complex coefficients; a full discussion of the relevant literature may be seen in [4, Section 6.3.3]. To our best knowledge complex coefficients have not yet been tested in the stochastic scenario.

7.3. **The Stratonovich generator.** We briefly outline how the preceding material has to be modified in the Stratonovich case. The expression for the generator is

$$\mathfrak{G} = \sum_{a \in \mathcal{A}_{\rm det}} a + \frac{1}{2} \sum_{A \in \mathcal{A}_{\rm sto}} AA \in \mathbb{R}\langle \mathcal{A} \rangle,$$

and, in analogy with Proposition 15, we have

$$\mathbb{E}J(t_0 + h; t_0) = \exp(h\mathfrak{G}), \tag{54}$$

a formula that may be proved by showing, as in the Ito case, that the left- and right-hand sides satisfy the same initial value problem. As a consequence, one obtains the following formula for the expectation of observables:

$$\mathbb{E}\chi(x(t_0 + h)) = \exp(hD_{\mathfrak{G}})\chi(x_0).$$

Taking the coefficient of the word $w \in \mathcal{W}$ in (54) gives the value of the expectations of the iterated integrals. Clearly $\mathbb{E}J_w(t_0+h; t_0) = 0$ if w is not a concatenation of deterministic letters $a \in \mathcal{A}_{\rm det}$ and pairs AA, $A \in \mathcal{A}_{\rm sto}$ (examples include AAA or $ABAB$ if $A \neq B$). When w is such a concatenation, it is easily shown that

$$\mathbb{E}J_w(t_0 + h; t_0) = \frac{1}{2^{\pi(w)}} \frac{h^{\|w\|}}{\|w\|!}$$

where $\pi(w)$ is the number of pairs that enter in the concatenation (for instance for $AAaBBAA$, $\pi = 3$ and for $AAAA$, $\pi = 2$). Once the expectations $\mathbb{E}J_w(t_0 + h; t_0)$ are known, the shuffle relations in Proposition 6 may be used to compute higher *moments* of the iterated integrals, similarly to what was explained in Remark 4.

As distinct from the $\mathbb{E}I_w$, $w \in \overline{\mathcal{W}}$, studied in Proposition 16, the $\mathbb{E}J_w$, $w \in \mathcal{W}$, do not satisfy the shuffle relations (except of course in the degenerate case where $\mathcal{A}_{\rm sto} = \emptyset$).

For integrators that satisfy the obvious analogue of (52), Proposition 18 also holds in the Stratonovich case and therefore the order barrier in Theorem 7.2 also applies to the Stratonovich interpretation.

8. **Relating the Stratonovich and Ito interpretations.** In this paper, the Stratonovich and Ito theories have been developed in parallel. It is well known that it is actually possible to map one into the other and we now present how to do so by means of words.

8.1. **Relating the Stratonovich and Ito iterated integrals.** Along with the extended alphabet $\bar{\mathcal{A}}$ that we used to carry out the Ito-Taylor expansion, let us now consider a new alphabet $\mathcal{A}^\star$ that consists of all the deterministic letters $a \in \mathcal{A}_{\rm det}$, all the stochastic letters $A \in \mathcal{A}_{\rm sto}$ and, in addition, a deterministic letter $A^\star$ associated with each $A \in \mathcal{A}_{\rm sto}$. After setting $\circ d\mathcal{B}_\ell(s) = ds$ for all deterministic letters, we may define, via (9), Stratonovich iterated integrals J_w for each $w \in \mathcal{W}^\star$, where $\mathcal{W}^\star$ denotes the set of words for the alphabet $\mathcal{A}^\star$. Note that this set of iterated integrals is different from that used to write the Stratonovich-Taylor expansion in (10)–(12) because $\mathcal{W}^\star$ is a larger set than $\mathcal{W}$. With the J_w, $w \in \mathcal{W}^\star$, we construct the Chen series

$$J^\star = \sum_{w \in \mathcal{W}^\star} J_w(t_0 + h; t_0) w.$$

The results in this section require the use of two mappings θ and ρ that we introduce now. We define $\theta : \mathbb{R}\langle\langle \mathcal{A}^\star \rangle\rangle \to \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$ as follows. For letters, we set $\theta(a) = a$ for $a \in \mathcal{A}_{\rm det}$, $\theta(A) = A$ for $A \in \mathcal{A}_{\rm sto}$ and $\theta(A^\star) = \bar{A} - (1/2)AA$ for $A \in \mathcal{A}_{\rm sto}$. For words, we set $\theta(\emptyset) = \emptyset$ and $\theta(\ell_1 \dots \ell_n) = \theta\ell_1 \cdots \theta\ell_n$. We note that, for each $w \in \mathcal{W}^\star$, $\theta(w)$ is a linear combination of words of weight $\|w\|$. Finally, we set $\theta(\sum_w S_w w) = \sum_w S_w \theta w$. Clearly θ is linear and in addition is an algebra morphism, i.e. maps the concatenation $S_1 S_2$ into the concatenation $\theta(S_1)\theta(S_2)$.

We next define a bilinear mapping $\mathbb{R}\langle\langle \mathcal{A}^\star \rangle\rangle \times \mathbb{R}\langle \bar{\mathcal{A}} \rangle \to \mathbb{R}$ as in (47) and define $\rho : \mathbb{R}\langle \bar{\mathcal{A}} \rangle \to \mathbb{R}\langle \mathcal{A}^\star \rangle$ by demanding

$$\big(\theta(S), p\big) = \big(S, \rho(p)\big)$$

for each $S \in \mathbb{R}\langle\langle \mathcal{A}^\star \rangle\rangle$ and each $p \in \mathbb{R}\langle \bar{\mathcal{A}} \rangle$; thus ρ is the linear map obtained from θ by transposition with respect to $(\cdot, \cdot)$. As an example of the computation of ρ, let us find $\rho(AA)$. By definition, $\theta(A^\star) = \bar{A} - (1/2)AA$ and $\theta(AA) = \theta(A)\theta(A) = AA$; for words w other than $A^\star$ and AA, $(\theta(w), AA) = 0$ and therefore $\rho(AA) = AA - (1/2)A^\star$. In general

$$\rho(w) = w + \sum_u \left(-\frac{1}{2}\right)^r u,$$

where the sum is extended to all words that may obtained by replacing pairs of consecutive stochastic letters AA by the corresponding $\bar{A}$ and $r \geq 1$ is the number of pairs replaced. For instance, $\rho(aAAA) = aAAA - (1/2)a\bar{A}A - (1/2)aA\bar{A}$ and $\rho(AAAA) = AAAA - (1/2)\bar{A}AA - (1/2)A\bar{A}A - (1/2)AA\bar{A} + (1/4)\bar{A}\bar{A}$ and $\rho(AB) = AB$ if $A \neq B$.

The maps θ and ρ have been defined so that they encapsulate the relation between Ito and Stratonovich integrals, as shown in the next result, where the first formula expresses each Ito iterated integral as a linear combination of Stratonovich iterated integrals (cf. formula (8) in [20]).

Proposition 19. *For each* $w \in \overline{W}$,

$$I_w(t_0+h;t_0) = \Big(J^\star(t_0+h;t_0), \rho(w)\Big), \tag{55}$$

and, therefore, for each $p \in \mathbb{R}\langle \bar{\mathcal{A}} \rangle$,

$$\Big(I(t_0+h;t_0), p\Big) = \Big(J^\star(t_0+h;t_0), \rho(p)\Big).$$

As a consequence, θ *maps the (Stratonovich) Chen series* $J^\star$ *into the (Ito) Chen series*

$$I = \sum_{w \in \overline{\mathcal{W}}} I_w(t_0+h;t_0) w.$$

Proof. The equality in (55) clearly holds if w is empty or consists of a single stochastic letter. Suppose that it holds for all words with weight $\leq N$, $N \geq 1/2$, and consider a word of of weight $N+1/2$, which we write in the form $wk\ell$. Assume first that $k = \ell = A$ for some $A \in \mathcal{A}_{\rm sto}$. By the recurrence relation between iterated integrals, we have

$$I_{wAA} = \int_{t_0}^{t_0+h} I_{wA}(s)\, d\mathcal{B}_A(s),$$

and then, by the relation between Ito and Stratonovich stochastic integrals (see e.g.[38, Section 3.2], the induction hypothesis and (9), we may write

$$\begin{aligned} I_{wAA} &= \int_{t_0}^{t_0+h} I_{wA}(s) \circ d\mathcal{B}_a(s) - \frac{1}{2}\int_{t_0}^{t_0+h} I_w(s)\, dt \\ &= \int_{t_0}^{t_0+h} \Big(J^\star(s;t_0), \rho(wA)\Big) \circ d\mathcal{B}_A(s) - \frac{1}{2}\int_{t_0}^{t_0+h} \Big(J^\star(s;t_0), \rho(w)\Big)\, ds \\ &= \Big(J^\star, \rho(wA)A\Big) - \frac{1}{2}\Big(J^\star, \rho(w)\bar{A}\Big) \\ &= \Big(J^\star, \rho(wAA)\Big). \end{aligned}$$

For other combinations of k and ℓ one proceeds similarly, starting from

$$I_{wk\ell} = \int_{t_0}^{t_0+h} I_{wk}(s)\, d\mathcal{B}_\ell(s) = \int_{t_0}^{t_0+h} I_{wk}(s) \circ d\mathcal{B}_\ell(s).$$

∎

As a simple instance of (55) we note that, from the relation $\rho(AA) = AA - (1/2)A^\star$ found above, we get $I_{AA} = J_{AA} - (1/2)J_{A^\star}$, i.e. the well-known relation

$$\int_{t_0}^{t_0+h} \mathcal{B}_A(s) d\mathcal{B}_A(s) = \int_{t_0}^{t_0+h} \mathcal{B}_A(s) \circ d\mathcal{B}_A(s) - \frac{1}{2}h = \frac{1}{2}\Big(\mathcal{B}_A(t_0+h)^2 - \mathcal{B}_A(t_0)^2 - h\Big).$$

8.2. **The equivalence Ito–Stratonovich.** Proposition 19 links the Chen series $J^\star$ and I. We investigate next the link between the corresponding series of differential operators. Recall that, associated with each $a \in \mathcal{A}_{\rm det}$ or each $A \in \mathcal{A}_{\rm sto}$, there is a first order (Lie) differential operator (5). On the other hand, letters $\bar{A} \in \bar{\mathcal{A}}$ corresponding to $A \in \mathcal{A}_{\rm sto}$ give rise to second order differential operators (15). We now associate with each letter $A^\star$, $A \in \mathcal{A}_{\text{sto}}$, the first-order differential operator defined by

$$D_{A^\star}\chi(x) = \chi'(x)\Big(-\frac{1}{2}f'_A(x)f_A(x)\Big).$$

Thus $D_{A^\star}$ is the Lie operator of the vector field $-(1/2)f'_A(x)f_A(x)$. With this definition, a simple computation yields

$$D_{A^\star} = D_{\bar{A}} - \frac{1}{2}D_{AA}$$

i.e. $D_{A^\star} = D_{\theta(A^\star)}$. Furthermore, for $a \in \mathcal{A}_{\text{det}}$, $\theta(a) = a$ and, for $A \in \mathcal{A}_{\text{sto}}$, $\theta(A) = A$ and therefore $D_\ell = D_{\theta(\ell)}$ for each $\ell \in \mathcal{A}^\star$. It follows that $D_S = D_{\theta(S)}$ for each series $S \in \mathbb{R}\langle\langle \mathcal{A}^\star\rangle\rangle$. In particular, from the last equality in Proposition 19, we conclude $D_{J^\star} = D_I$, or, in other words, the pullback operator D_I for the Ito equation (4) coincides with the pullback operator $D_{J^\star}$ of the Stratonovich equation

$$dx = \sum_{a\in\mathcal{A}_{\text{det}}} f_a(x)\,dt - \frac{1}{2}\sum_{A\in\mathcal{A}_{\text{sto}}} f'_A(x)f_A(x)\,dt + \sum_{A\in\mathcal{A}_{\text{sto}}} f_A(x)\circ d\mathcal{B}_A. \tag{56}$$

In fact, as is well known, this Stratonovich equation and (4) have the same solutions. This is easily proved: (14) coincides with the result of writing formula (6) for the system (56). Of course, if all the f_A, $A \in \mathcal{A}_{\text{sto}}$ are constant (additive noise), (56) is the same as (3), i.e. (4) and (3) share the same solutions, see e.g. [25, Section 4.9].

9. **Additional algebraic results.** In this section we briefly relate the preceding material to standard results on combinatorial (Hopf) algebras and provide additional algebraic results. *Hopf* algebras are important tools in the study of numerical integrators and in other fields including e.g. renormalization theories; a very readable introduction that requires little algebraic background is presented in [7]. For instance many developments of Butcher's theory of Runge-Kutta methods may be phrased in the language of the Hopf algebra of trees and in fact Butcher anticipated many results on that algebra later rediscovered in different settings. Useful references are [31, 17].

The (associative, commutative) *shuffle* algebra $\mathcal{H}_{\text{sh}}(\mathcal{A})$ of the alphabet $\mathcal{A}$ is defined as follows. As a vector space $\mathcal{H}_{\text{sh}}(\mathcal{A})$ coincides with $\mathbb{R}\langle\mathcal{A}\rangle$. However the product in $\mathcal{H}_{\text{sh}}(\mathcal{A})$ is given by shuffling words rather than by concatenating them. The algebra $\mathcal{H}_{\text{sh}}(\mathcal{A})$ is graded by the weight $\|\cdot\|$. In addition we may consider in $\mathcal{H}_{\text{sh}}(\mathcal{A})$ a *coproduct* by decomposing (deconcatenating) each word $w = \ell_1 \dots \ell_n \in \mathcal{W}$ as

$$\emptyset \otimes \ell_1 \dots \ell_n + \ell_1 \otimes \ell_2 \dots \ell_n + \ell_1 \dots \ell_n \otimes \emptyset.$$

This coproduct is compatible with the shuffle product because, as explained in the proof of Lemma 6.1, the operations of shuffling and deconcatenation commute. Therefore $\mathcal{H}_{\text{sh}}(\mathcal{A})$ is a Hopf algebra graded by the weight.

The dual vector space of $\mathcal{H}_{\text{sh}}(\mathcal{A})$ may be identified with the vector space of formal series $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ via the bilinear form (47). In other words, the linear form on $\mathcal{H}_{\text{sh}}(\mathcal{A})$ that as w ranges in $\mathcal{W}$ associates with w the real number S_w is identified with $\sum S_w w$. With this identification, the concatenation product of series $S \in \mathbb{R}\langle\langle A\rangle\rangle$, or equivalently the product (33) for the coefficents, coincides with the convolution product in the dual of the Hopf algebra. Series S with $S_\emptyset = 1$ that satisfy the shuffle relations are then the linear forms on $\mathcal{H}_{\text{sh}}(\mathcal{A})$ that are multiplication morphisms (i.e. preserve multiplication). The set of those linear forms forms is well known to be a group for the convolution product; this group is called the *shuffle group* and denoted

$\mathcal{G}_{\rm sh}(\mathcal{A})$. Therefore Lemma 6.1 is just the statement that the convolution product of two elements in $\mathcal{G}_{\rm sh}(\mathcal{A})$ lies in $\mathcal{G}_{\rm sh}(\mathcal{A})$.

The *quasishuffle Hopf algebra* $\mathcal{H}_{\rm qsh}(\bar{A})$ is constructed similarly. One endows the vector space $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$ with the quasishuffle product and the deconcatenation coproduct. The series $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ with $S_\emptyset = 1$ that satisfy the quasishuffle relations may then be viewed as forming the *quasishuffle group* $\mathcal{G}_{\rm qsh}$ of linear forms on $\mathcal{H}_{\rm qsh}(\bar{A})$ that are multiplication morphisms.

Theorem 2.5 in [24] shows that the mapping ρ is an isomorphism of $\mathcal{H}_{\rm qsh}(\bar{A})$ onto $\mathcal{H}_{\rm sh}(\mathcal{A}^\star)$. In particular it maps the quasishuffle product into the shuffle product:

$$\rho(u \bowtie v) = \rho(u) \sqcup\!\sqcup\, \rho(v), \qquad \forall u, v \in \overline{W}.$$

This observation and the material in Section 8 make clear that the quasishuffle/Ito results in Propositions 9–12 may be derived from the corresponding shuffle/Stratonovich results by transforming $\sqcup\!\sqcup$ into $\bowtie$ with the help of the inverse isomorphism ρ^{-1}.

REFERENCES

[1] A. Alamo and J. M. Sanz-Serna, A technique for studying strong and weak local errors of splitting stochastic integrators, *SIAM J. Numer. Anal.*, **54** (2106), 3239–3257.

[2] F. Baudoin, *Diffusion Processes and Stochastic Calculus*, European Mathematical Society, Textbooks in Mathematics Vol. **16**, 2014.

[3] S. Blanes and F. Casas, On the necessity of negative coefficients for operator splitting schemes of order higher than two, *Appl. Numer. Maths.*, **54** (2005), 23-37.

[4] S. Blanes and F. Casas, *A Concise Introduction to Geometric Numerical Integration*, CRC Press, Boca Raton, 2016.

[5] 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.

[6] S. Blanes, F. Casas and A. Murua, Splitting and composition methods in the numerical integration of differential equations, *Bol. Soc. Esp. Mat. Apl. SeMA*, **45** (2008), 89–145.

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

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

[9] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, 3^{rd} edition, John Wiley & Sons Ltd., Chichester, 2016.

[10] M. P. Calvo, A. Murua and J. M. Sanz-Serna, Modified equations for ODEs, in *Chaotic Numerics* (eds. P. E. Kloeden and K. J. Palmer), Contemporary Mathematics, Vol. **172**, American Mathematical Society, Providence, (1994), 63–74.

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

[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. T. Chen, Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula, *Ann. of Math.*, **65** (1957), 163–178.

[17] F. Fauvet and F. Menous, Ecalle's arborification-coarborification transforms and Connes-Kreimer Hopf algebra, *Ann. Sci. Ec. Norm. Sup.*, **50** (2017), 39–83.

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

[19] M. Fliess, Fonctionnelles causales non-linéaires et indeterminées noncommutatives, *Bull. Soc. Math. France*, **109** (1981), 3–40.
[20] J. G. Gaines, The algebra of iterated stochastic integrals, *Stochastics*, **49** (1994), 169–179.
[21] E. Hairer, Ch. Lubich and G. Wanner, *Geometric Numerical Integration*, 2^{nd} edition, Springer, Berlin, 2006.
[22] E. Hairer, S. P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I*, 2^{nd} edition, Springer-Verlag, Berlin, 1993.
[23] E. Hairer and G. Wanner, On the Butcher group and general multi-value methods, *Computing*, **13** (1974), 1–15.
[24] M. E. Hoffman, Quasi-shuffle products, *J. Algbr. Comb.*, **11** (2000), 49–68.
[25] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.
[26] B. Leimkuhler and C. Matthews, Rational construction of stochastic numerical methods for molecular sampling, *App. Math. Res. Express*, **2013** (2013), 34–56.
[27] B. Leimkuhler and C. Matthews, Robust and efficient configurational molecular sampling via Langevin Dynamics, *J. Chem. Phys.*, **138** (2013), 174102.
[28] R. I. McLachlan and G. R. W. Quispel, Splitting methods, *Acta Numerica*, **11** (2002), 341–434.
[29] R. H. Merson, An operational method for the study of integration processes, in *Proceedings of the Symposium on Data Processing*, Weapons Researcch Establishement, Salisbury, Australia, (1957), 110.1–110.25.
[30] G. N. Miltstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer, Berlin, 2004.
[31] A. Murua, The Hopf algebra of rooted trees, free Lie algebras and Lie series, *Found. Comput. Math.*, **6** (2006), 387–426.
[32] 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.
[33] A. Murua and J. M. Sanz-Serna, Computing normal forms and formal invariants of dynamical systems by means of word series, *Nonlinear Anal.-Theor.*, **138** (2016), 326–345.
[34] A. Murua and J. M. Sanz-Serna, Vibrational resonance: a study with high-order word-series averaging, *Applied Mathematics and Nonlinear Sciences*, **1** (2016), 239–246.
[35] A. Murua and J. M. Sanz-Serna, Word series for dynamical systems and their numerical integrators, *Found. Comput. Math.*, **17** (2017), 675–712.
[36] A. Murua and J. M. Sanz-Serna, Averaging and computing normal forms with word series algorithms, in *Discrete Mechanics, Geometric Integration and Lie-Butcher Series* (eds. K. Ebrahimi-Fard and M. Barbero Liñan), Springer, Berlin, (2018), 115–137.
[37] A. Murua and J. M. Sanz-Serna, Hopf algebra techniques to handle dynamical systems and numerical integrators, in *Computation and Combinatorics in Dynamics, Stochastics and Control, The Abel Symposium, Rosendal, Norway, August 2016* (eds. E. Celledoni, G. Di Nunno, K. Ebrahimi-Fard and H. Z. Munthe-Kaas), Springer, (2019), to appear.
[38] G. A. Pavliotis, *Stochastic Processes and Applications*, Springer, New York, 2014.
[39] E. Platen and W. Wagner, On a Taylor formula for a class of Ito processes, *Probab. Math. Statist.*, **3** (1982), 37–51.
[40] R. Ree, Lie elements and an algebra associated with shuffles, *Ann. of Math.*, **68** (1958), 210–220.
[41] C. Reutenauer, *Free Lie Algebras*, Clarendon Press, Oxford, 1993.
[42] J. M. Sanz-Serna, Geometric integration, in *State of the Art in Numerical Analysis* (eds. I. S. Duff and G. A. Watson), Clarendon Press, Oxford, (1997), 121–143.
[43] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall, London, 1994.
[44] 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)* (eds. Lei-Guo and Zhiming-Ma), Higher Education Press, Beijing, (2015), 311–331.

Received April 2018; revised November 2018.

E-mail address: alamo.zapatero.alfonso@gmail.com
E-mail address: jmsanzserna@gmail.com