PROCEEDINGS OF THE ROYAL SOCIETY A

rspa.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
Probability, Differential Equations, Computational Mathematics

Keywords:
Lévy processes, efficient integrators, quasi-shuffle algebra

Author for correspondence:
Charles Curry (25/10/18)
e-mail: charles.curry@math.ntnu.no

Algebraic Structures and Stochastic Differential Equations driven by Lévy processes

Charles Curry[1,3], Kurusch Ebrahimi–Fard[2], Simon J.A. Malham[1] and Anke Wiese[1]

[1]Maxwell Institute for Mathematical Sciences and School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK.
[2]Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway.
[3]Current address: Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway.

We construct an efficient integrator for stochastic differential systems driven by Lévy processes. An efficient integrator is a strong approximation that is more accurate than the corresponding stochastic Taylor approximation, to all orders and independent of the governing vector fields. This holds provided the driving processes possess moments of all orders and the vector fields are sufficiently smooth. Moreover the efficient integrator in question is optimal within a broad class of perturbations for half-integer global root mean-square orders of convergence. We obtain these results using the quasi-shuffle algebra of multiple iterated integrals of independent Lévy processes.

1. Introduction

We consider the simulation of Itô stochastic differential systems driven by independent Lévy processes possessing moments of all orders. Our goal is to derive a new strong numerical integration scheme for such systems that is efficient in the following sense. The efficient scheme has a strong error at leading order that is always smaller than the strong error of the corresponding stochastic Taylor approximation. This is true for any such stochastic differential systems of any size, and for all orders of approximation. Moreover, for half-integer orders, our efficient integrator is optimal in the sense that its global root mean-square strong error realizes its smallest possible value in a broad class of perturbations

Royal Society Publishing
Informing the science of the future

compared to the error of the corresponding stochastic Taylor approximation. Lévy processes are a class of stochastic processes which are continuous in probability and possess stationary increments, independent of the past. They are examples of stochastic processes with a well understood structure that incorporate jump discontinuities. Due to the Lévy–Itô decomposition of a Lévy process the systems we consider can be written in the form

$$y_t = y_0 + \sum_{i=0}^{d} \int_0^t V_i(y_s)\mathrm{d}W_s^i + \sum_{i=d+1}^{\ell} \int_0^t V_i(y_{s_-})\mathrm{d}J_s^i,$$

for $y_t \in \mathbb{R}^N$ with $N \in \mathbb{N}$; see for example Applebaum [2]. We assume all the governing autonomous vector fields $V_i\colon \mathbb{R}^N \to \mathbb{R}^N$ are smooth and in general non-commuting. Note that by convention we suppose $W_t^0 \equiv t$. The system is thus driven by the independent Wiener processes W^i for $i = 1, \dots, d$, and the purely discontinuous martingales J^i for $i = d+1, \dots, \ell$ which are expressible in the form

$$J_t^i = \int_0^t \int_{\mathbb{R}} v\bar{Q}^i(\mathrm{d}v, \mathrm{d}s),$$

where $\bar{Q}^i(\mathrm{d}v, \mathrm{d}s) := Q^i(\mathrm{d}v, \mathrm{d}s) - \rho^i(\mathrm{d}v)\mathrm{d}s$. The Q^i are Poisson measures on $\mathbb{R} \times \mathbb{R}_+$ with intensity measures $\rho^i(\mathrm{d}v)\mathrm{d}s$, where the ρ^i are measures on $\mathbb{R}$ with $\rho^i(\{0\}) = 0$ and $\int_{\mathbb{R}}(1 \wedge v^2)\rho^i(\mathrm{d}v) < \infty$. Intuitively, $Q^i(B, (a, b])$ counts the number of jumps of the process J^i taking values in the set B in the time interval $(a, b]$, whilst $\rho^i(B)$ measures the expected number of jumps per unit time the process J^i accrues taking values in the set B. We emphasize that we assume that the J^i possess moments of all orders.

Let us outline our approach to the strong numerical simulation of Lévy-driven stochastic differential systems via steps (1)—(4) below. As we do so we will present the new results we prove in this paper and put them into context. We will also describe the algebraic approach we employ to prove the results and its usefulness.

Step (1). We begin by studying strong integration schemes for stochastic differential systems such as those above. Such schemes are typically derived from the stochastic Taylor expansion for the solution given by

$$y_t = \sum_{w \in A^*} I_w(t)\big[\tilde{V}_w(y_0)\big].$$

A derivation can be found in Platen & Bruti-Liberati [68]; also see Platen [66,67] and Platen & Wagner [69]. It is produced via Picard iteration using the chain rule for the solution process y_t. We show this explicitly at the beginning of § 2. In the expansion the set A^* consists of all the multi-indices (or words) that can be constructed from the set of letters $A := \{0, 1, \dots, d, d+1, \dots, \ell\}$. The terms $I_w(t)\big[\tilde{V}_w(y_0)\big]$ represent iterated integrals of the integrands $\tilde{V}_w(y_0)$ which involve compositions of the vector fields (as first order partial differential and difference operators) governing the Lévy-driven stochastic differential system. See the beginning of § 2 and in particular Theorem 2.1 for a more precise prescription. For sufficiently smooth vector fields, integration schemes of arbitrary strong order of convergence may be constructed by truncating the series expansion to an appropriate number of terms. Platen & Bruti-Liberati [68] show which terms must be retained to obtain a strong integration scheme with a given strong order of convergence. We assume hereafter the governing vector fields are sufficiently smooth for the stochastic Taylor expansion to exist. Our first new result in this paper is Theorem 2.2 in § 2 in which we show that the terms in the stochastic Taylor expansion above can be written in separated form as $I_w(t)\tilde{V}_w(y_0)$. We observe that the time-dependent stochastic information which is encoded in the now integrand-free multiple integrals I_w are separated as scalar multiplicative factors of the structure information which is encoded in the terms $\tilde{V}_w(y_0)$ which involve compositions of the vector fields as first order partial differential operators only. Indeed the separated expansion is essentially achieved by Taylor expanding all the shift operators associated with the purely discontinuous terms in the original stochastic Taylor expansion. Note that the result of this is that the sum is over an expanded set of multi-indices/words $\mathbb{A}^*$ constructed from the alphabet $\mathbb{A}$ which contains A. The additional letters in $\mathbb{A}$ are compensated power brackets associated with the purely discontinuous terms. See § 2 for more precise details as well as Curry, Ebrahimi–Fard, Malham & Wiese [20]. The advantages of this separated form are: (i) the simulation of the stochastic components I_w of the system, for which targeted generic simulation algorithms can be developed, is now separated from the inherent structure components; (ii) it provides the context/basis to develop the more general strong integrators we are about to discuss next and (iii) we can take advantage of well developed algebraic structures which underlie such expansions as we shall see.

As is well known the order of a strong numerical scheme is determined by the set of multiple integrals I_w retained in any approximation. We also emphasize here that in any Monte–Carlo simulation, the bulk of the computational effort goes into simulating the higher order multiple integrals included, in particular those indexed by words of length two or more involving distinct non-zero letters—though see Malham & Wiese [58] for new efficient methods for simulating the Lévy area. Once this set of multiple integrals is fixed (with its associated computational burden) the order of convergence is fixed, i.e. the power of the stepsize in the leading order term of the strong error. More accurate schemes correspond to those that generate smaller coefficients in the leading order term.

Step (2). We then discuss how strong integrators based on truncations of the stochastic Taylor expansion are just one example of a more general class of integrators we shall call map-truncate-invert schemes. General map-truncate-invert schemes were first introduced in Malham & Wiese [57] for Stratonovich drift-diffusion equations and are constructed as follows. We start with the flowmap $\varphi_t\colon y_0 \mapsto y_t$ lying in $\mathrm{Diff}(\mathbb{R}^N)$ which maps the initial data y_0 to the solution y_t at time $t > 0$. Here $\mathrm{Diff}(\mathbb{R}^N)$ denotes the space of diffeomorphisms on $\mathbb{R}^N$. From our discussion above, the separated stochastic Taylor expansion for the flowmap has the form

$$\varphi_t = \sum_{w \in \mathbb{A}^*} I_w(t) \tilde{V}_w.$$

This separated form is key and crucial to all our subsequent developments herein. Now, given any invertible map $f\colon \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$, the corresponding map-truncate-invert scheme is constructed thus: we expand the series $f(\varphi_t)$, truncate according to a chosen grading and then apply f^{-1}. See § 3 for more details, in particular how to construct such integrators in the Lévy-driven context. The case $f = \log$ corresponds to the exponential Lie series integrator of Castell–Gaines for Stratonovich drift-diffusion equations; see Castell & Gaines [12,13]. In practice the $f = \log$ case can be implemented as follows. The series $\psi_t = \log \varphi_t$ lies in the Lie algebra generated by the underlying vector fields with the scalar coefficients given by combinations of multiple Stratonovich integrals. Any truncation $\hat{\psi}_t$ with the multiple integrals replaced by suitable approximate samples is also a vector field. The corresponding solution approximation is $\hat{y}_t = \exp \hat{\psi}_t \circ y_0$. This can be generated by solving the ordinary differential system $u' = \hat{\psi}_t \circ u$ for $u = u(\tau)$ on $\tau \in [0, 1]$ with $u(0) = y_0$. Typically this is achieved using a suitable numerical scheme. The approximation $u(1)$ represents $\hat{y}_t$. Generally, apart from the low order cases stipulated below, the Castell–Gaines method is not asymptotically efficient. This was shown by Malham & Wiese [57] who consequently considered the problem of designing so-called efficient integrators for Stratonovich drift-diffusion equations driven by multiple independent Wiener processes. Here, efficient integrators of a given order have leading order error which is always less than that for the corresponding stochastic Taylor scheme, independent of the governing vector fields. Malham & Wiese [57] then proved that an integration scheme based on the map $f = \mathrm{sinhlog}$ is efficient. These results were obtained in the absence of a drift term and extended to incorporate drift in Ebrahimi–Fard, Lundervold, Malham, Munthe–Kaas & Wiese [21]. It is the generalization of this latter result to Lévy-driven systems that we consider presently.

Step (3). We temporarily revert our discussion to stochastic Taylor approximations. We prove a new result for such schemes which is an important step towards our main result. This concerns the relative accuracy of strong numerical methods for Lévy-driven stochastic differential systems generated by truncating the stochastic Taylor expansion according to word length or mean-square grading. The length of any word $w \in \mathbb{A}$ is the number of letters it contains while its mean-square grading is similar but zero letters are counted twice—consistent with the L^2-norm of I_w. Suppose $R_t^{\mathrm{wl}}(y_0)$ and $R_t^{\mathrm{ms}}(y_0)$ are the remainders generated by truncating the separated stochastic Taylor expansion for y_t by keeping all words of length and mean-square grade less than or equal to n, respectively. Then in § 4 (see Theorem 4.1) we prove that at leading order for all n we have

$$\big\|R_t^{\mathrm{wl}}(y_0)\big\|_{L^2}^2 \leqslant \big\|R_t^{\mathrm{ms}}(y_0)\big\|_{L^2}^2.$$

There is computational effort involved in simulating any such word length stochastic Taylor approximation due to the extra iterated integrals included. However those that are included are indexed by words which include at least one drift (zero) letter, i.e. integrating with respect to $\mathrm{d}t$, or repeated Lévy (non-zero) letters which are computationally less burdensome to compute than those indexed by words containing completely distinct non-zero letters. We assess the trade off of accuracy versus computational effort once we introduce our main result next.

Step (4). Lastly, we present our new antisymmetric sign reverse integrator and prove that it is an efficient integrator for Lévy-driven systems; see Theorem 5.1(a). In addition, we prove that it is optimal at half-integer global root mean-square orders of convergence; see Theorem 5.1(b). It is an example of a map-truncate-invert scheme and the principle ideas underlying its construction and properties are as follows. To begin with we consider the result of applying the first two steps of the map-truncate-invert scheme as described above, i.e. considering a function $f(\varphi_t)$ of the flowmap and then truncating the result according to a given grading up to and including terms of grade n. The pre-remainder is the remainder after truncating $f(\varphi_t)$ in this way. We then show that for word length grading, at leading order, the pre-remainder is equal to the true remainder, i.e. equal to the remainder after applying f^{-1} to the truncated $f(\varphi_t)$. Herein lies the room-for-manouver of which we take advantage. Naturally the map-truncate-invert scheme must emulate the scheme based on the corresponding stochastic Taylor expansion truncated according to the same grading. The question is thus, is the pre-remainder (and thus remainder) associated with the map-truncate-invert scheme at hand less than the remainder associated with the corresponding stochastic Taylor approximation to leading order in the mean-square measure? The answer is "yes" for the antisymmetric sign reverse integrator. This integrator is constructed by taking half the difference of the stochastic Taylor expansion and the stochastic Taylor expansion with all the words indexing the multiple integrals reversed in order and signed according to the length of the words. Hence the pre-remainder and thus remainder are

$$R_t^{\mathrm{ASRI}}(y_0) = \sum_{|w|=n+1} \tfrac{1}{2}\big(I_w - I_{S(w)}\big)V_w(y_0).$$

Here $|w|$ represents the length of the word w and S is the signed reversal map defined as $S\colon a_1a_2\ldots a_n \mapsto (-1)^n\, a_na_{n-1}\ldots a_1$ for any word consisting of letters $a_1, a_2, \ldots, a_n$ from $\mathbb{A}$. We naturally assume $I_{-w} \equiv -I_w$. Our *principal main result* is thus

$$\big\|R_t^{\mathrm{ASRI}}(y_0)\big\|_{L^2}^2 \leqslant \big\|R_t^{\mathrm{wl}}(y_0)\big\|_{L^2}^2,$$

for all y_0, all Lévy-driven systems, to all orders n. Using our result above this means the antisymmetric sign reverse integrator is more accurate than the corresponding stochastic Taylor approximation truncated according to mean-square grading. Furthermore we show that when n is odd, the inequality just above is optimal in the following sense. Suppose we perturb the antisymmetric sign reverse integrator by a general class of perturbations—see § 5 for more details. Then when n is odd we show that the difference between the two mean-square error measures for $R_t^{\mathrm{wl}}(y_0)$ and $R_t^{\mathrm{ASRI}}(y_0)$ above are minimized when the perturbation is zero. One further technical practical aspect comes into play. In the Castell–Gaines case we compute $f^{-1} = \exp$ by solving an ordinary differential system. If all the underlying vector fields are linear f^{-1} may be computed as a matrix-valued compositional inverse function; see Malham & Wiese [57]. However we cannot do this in general for map-truncate-invert schemes. To circumvent this issue we have developed direct-map-truncate-invert schemes with a special case being the direct antisymmetric sign reverse integrator. In direct-map-truncate-invert schemes we utilize the observation above that map-truncate-invert schemes naturally emulate the scheme based on the corresponding stochastic Taylor approximation to the same order, but have a different remainder at leading order as different terms at leading order in the remainder are included in the integrator. Thus in order to simulate a direct-map-truncate-invert scheme we simulate the corresponding stochastic Taylor approximation to the same order as well as the terms that lie in the difference between the stochastic Taylor approximation remainder and map-truncate-invert approximation remainder at leading order. These are additional terms indexed by words of length $n+1$, the terms concerned will not involve iterated integrals involving completely distinct letters (those would belong to the next order integrator). Our discussion above in the last paragraph on the computational burden involved thus applies here. Indeed we round off our results herein with global convergence results in § 6 and numerical simulations demonstrating our conclusions in § 7.

We remark that several challenges are encountered upon generalising from drift-diffusions to Lévy-driven equations. The stochastic Taylor expansions derived by Platen & Bruti–Liberati [68] do not display the separation of geometric and stochastic required for the deployment of map-truncate-invert schemes. Furthermore, due to the discontinuities of the driving paths, it is not possible to use Stratonovich integrals to obtain the usual product rule for iterated integrals. Algebraically, this means that we must use the quasi-shuffle algebra of iterated integrals with respect to Lévy processes (see Curry, Ebrahimi–Fard, Malham & Wiese [20]) rather than the shuffle algebra used in the derivation of map-truncate-invert schemes for drift-diffusions. Further, it is rather intriguing that the efficient integrator in the Lévy-driven (quasi-shuffle) context herein is the antisymmetric sign reverse integrator, i.e. half the difference of the identity and sign-reverse endomorphisms. This coincides with efficient integrator in the drift-diffusion (shuffle) case. In the shuffle case the sign-reverse endomorphism coincides with the antipode. However in the general quasi-shuffle case this is no longer true.

We prove our key results above in the context of combinatorial algebras, in particular the quasi-shuffle Hopf algebra and the convolutional algebra of endomorphisms defined on the quasi-shuffle Hopf algebra. It is this abstraction that not only allows us to prove our main results but also explicitly identify the 'Efficient Integrator'. These algebraic structures underlie the separated stochastic Taylor expansion for the flowmap associated with Lévy-driven stochastic differential systems and all the manipulations we apply to the separated stochastic Taylor expansion when we construct map-truncate-invert schemes and assess their properties. The separated stochastic Taylor expansion is the starting point for our analysis. Key for its derivation is the repeated application of Itô's formula. The stochastic information of our Lévy-driven system is thus given by the Wiener processes and by the purely discontinuous martingales in the Lévy–Itô decomposition of the driving Lévy processes as well as by their quadratic variation processes and power brackets. We encode this stochastic information in our augmented alphabet $\mathbb{A}$: each letter in $\mathbb{A}$ is associated with a Wiener processes or a purely discontinous processes driving the stochastic differential equation or with a compensated power bracket generated from these. For technical reasons it is opportune to use compensated power brackets rather than power brackets. Once we have constructed the alphabet, we can associate words constructed from this alphabet with the corresponding iterated integral. For example, for two Wiener processes W^1 and W^2, the word $\omega = 12$ is associated with the iterated integral $I_{12} = \int_{0\leqslant\tau_1\leqslant\tau_2\leqslant t} \mathrm{d}W^1_{\tau_1}\mathrm{d}W^2_{\tau_2}$. For a Wiener process W^1 and purely discontinuous martingale J^i with $i\in\{d+1,\ldots,\ell\}$, the word $\omega = 1i$ is associated with the iterated integral $I_{1i} = \int_{0\leqslant\tau_1\leqslant\tau_2\leqslant t} \mathrm{d}W^1_{\tau_1}\mathrm{d}J^i_{\tau_2}$. See Curry *et al.* [20] as well as § 2 for more details on the necessity and construction of the augmented alphabet $\mathbb{A}$. Associated with each member of the alphabet $\mathbb{A}$ constructed from the stochastic information is a corresponding partial differential operator. With each word we associate the partial differential operator obtained by compositions of the differential operators associated with the letters constituting the word (see § 2 for more details). Our first main result is that the separated stochastic Taylor expansion for the flowmap above is the sum $\varphi_t = \sum_{w\in\mathbb{A}^*} I_w(t)\tilde{V}_w$ over the real multiplication of the multiple integrals I_w and differential operators $\tilde{V}_w$ which are indexed by all words/multi-indices that can be constructed from the augmented alphabet $\mathbb{A}$ containing the original alphabet A. Now consider functions $f\colon \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$ of the flowmap which have simple power series expansions such as $f(\varphi) = \sum_{k\geqslant 0} c_k \varphi^k$, for some real coefficients c_k. Applying such a function to the flowmap generates linear combinations of products of the multiple integrals and linear combinations of compositions of the differential operators, i.e. operations in the respective algebras generated by the multiple integrals and by the differential operators. The algebra generated by the multiple integrals can be identified with the quasi-shuffle algebra of the words constructed from $\mathbb{A}^*$, again see Curry *et al.* [20] as well as § 2. The algebra generated by the composition of differential operators can be identified with a concatenation algebra of the words. Specifically, if we strip away the I's and $\tilde{V}$'s we can represent the stochastic Taylor series for the flowmap by

$$\sum_{w\in\mathbb{A}^*} w \otimes w.$$

This lies in a product algebra in which the bilinear product of two terms is given by

$$(u \otimes x)(v \otimes y) = (u * v) \otimes (xy),$$

for any words u, v, x, y constructed from $\mathbb{A}^*$. On the left $u * v$ is the quasi-shuffle product of u and v representing the real product $I_u I_v$. On the right xy is the concatentation of x and y representing the composition of the differential operators $\tilde{V}_x$ and $\tilde{V}_y$. Substituting the abstract series representation for the flowmap above into $f(\varphi)$, where f has the power series expansion with real coefficients c_k indicated above, by direct computation and rearrangement we have (see for example Ebrahimi–Fard, *et al.* [21]),

$$f(\varphi) = \sum_{w\in\mathbb{A}^*} F(w) \otimes w,$$

where

$$F(w) = \sum_{k=0}^{|w|} c_k \sum_{\substack{u_1,\ldots,u_k\in\mathbb{A}^*\\ u_1\ldots u_k = w}} u_1 * \cdots * u_k.$$

Here F is a linear endomorphism on words. This last calculation shows us that we can encode functions of the flowmap $f(\varphi)$ by endomorphisms F on the quasi-shuffle algebra of words. A fundamental building block of the endomorphism F is the action

$$w \mapsto \sum_{\substack{u_1,\ldots,u_k\in\mathbb{A}^*\\ u_1\ldots u_k = w}} u_1 * \cdots * u_k,$$

i.e. the sum of the quasi-shuffle product $u_1 * \cdots * u_k$ of any possible separation of w into any subwords $u_1, \ldots, u_k$. This object is the k-th convolution power of the identity endomorphism $\mathrm{id}^{\star k}$ acting on w. We can thus abstract the context one level further, and this abstraction is natural and key. The function f of the flowmap $f(\varphi)$ can be represented by $F(\mathrm{id}) = \sum_{k\geqslant 0} c_k \mathrm{id}^{\star k}$, where the action of $\mathrm{id}^{\star k}$ is $\mathrm{id}^{\star k}(w) = \sum_{u_1\ldots u_k = w} u_1 * \cdots * u_k$. Once the stochastic system under consideration is fixed, i.e. we have prescribed the driving Lévy processes and governing vector fields, the words are fixed. Different integrators are distinguished by their actions on those words and are represented by endomorphisms acting on them. The stochastic Taylor expansion itself corresponds to the case where F is the identity 'id' endomorphism.

We thus utilize two consecutive levels of abstraction herein. The first abstraction is to the product algebra with quasi-shuffles on one side and concatenations on the other. The second abstraction is to the algebra of endomorphisms that act on the Hopf quasi-shuffle algebra. On this algebra we can define an inner product representing the L^2 inner product on stochastic processes. Our abstraction encapsules the information required for our problem under consideration, and we are naturally able to represent map-truncate-invert approximations succinctly and to analyse their properties independent of the underlying system, data and truncation order. This enables us to establish the general results we present.

Lévy processes have many applications. Most notably though they appear in mathematical finance in the construction of models going beyond the Black–Scholes–Merton model to incorporate discontinuities in stock prices. See for example Cont & Tankov [18] and Barndorff–Nielsen, Mikosch & Resnick [4] and the references therein. There are also applications to physical sciences, see for instance Barndorff–Nielsen, Mikosch & Resnick [4]. Work on Euler type methods for simulating Lévy-driven stochastic differential systems including how to incorporate the jump processes can for example be found in Jacod [45], Fournier [26]. See Higham & Kloeden [36,37] for implicit methods and Protter & Talay [71] for weak approximations. The monograph by Platen & Bruti-Liberati [68] provides a comprehensive introduction to numerical simulation schemes for stochastic differential systems driven by Lévy processes and includes financial applications. Barski [5] also develops general high order schemes for Lévy-driven systems. Pathwise integrals of Lévy processes have been constructed in the framework of rough paths by Friz & Shekhar [27].

The problem of minimizing the coefficient in the leading error term in stochastic Taylor integrators was considered by Clark [15] and Newton [64,65] for drift-diffusion equations driven by a single Wiener process; also see Kloeden & Platen [47, Section 13.4]. They derived integration schemes that are asymptotically efficient in the sense that the coefficient of the leading order error is minimal among all schemes. For Stratonovich Wiener-driven stochastic differential equations, Castell & Gaines [12,13] constructed integration schemes using the Chen-Strichartz exponential Lie series expansion (see Strichartz [76]). In the strong order one case for one driving Wiener process, and the strong order one-half case for two or more driving Wiener processes, their schemes are asymptotically efficient in the sense of Clark and Newton.

The quasi-shuffle algebra of iterated stochastic integrals was considered by Gaines [29] for the case of multiple iterated integrals of Wiener processes back in 1994. A few years later Li & Liu [50] considered multiple iterated integrals of Wiener processes and standard Poisson processes. More recently Curry *et al.* [20] proved that the algebra of multiple integrals of a minimal family of semimartingales is isomorphic to the combinatorial quasi-shuffle algebra of words. A set of Lévy processes represents one example. The quasi-shuffle algebra is an extension of the shuffle algebra. Its historical development is of noteworthy interest. Indeed, it was introduced abstractly in a 1979 paper by Newman and Radford [63], where the authors attempt to endow the free coalgebra over an associative algebra with a Hopf algebra structure. A quarter of a century later, and independently from Gaines, in a sequence of papers, Hudson, Parthasarathy and collaborators presented a combinatorial product of iterated quantum stochastic integrals which is equivalent to the Gaines quasi-shuffle product; see Hudson & Parthasarathy [42–44] and Cohen, Eyre & Hudson [17]. This product was coined the sticky-shuffle, see Hudson [41], and was studied from a Hopf algebra viewpoint in Hudson [40]. Independently of these developments, Hoffman [38] comprehensively studied the quasi-shuffle product using a Hopf algebraic framework. The significance of the shuffle algebra was cemented in the work of Eilenberg & Maclane [24], Schützenberger [75] and Chen [14]. See Reutenauer [73] for more details. One advantage of abstracting to the quasi-shuffle algebra is that we can immediately identify the minimum set of iterated integrals that need to be simulated to implement a given accurate strong scheme. This optimizes the total computation time which is dominated by the strong simulation of the iterated integrals. Indeed, Radford [72] proved that the shuffle algebra is generated by Lyndon words. This was extended to the quasi-shuffle algebra by Newman & Radford [63] and independently later by Hoffman [38]. Gaines [29] established this independently for the case of Wiener processes, as did Li & Liu [50] for the case of Wiener processes and standard Poisson processes, while Sussmann [77] had considered a Hall basis. Hence the set of iterated integrals we need to simulate at any given order are identified by Lyndon words.

The study and application of such structures in systems, control and stochastic processes can be traced back to the work of Chen [14], Magnus [55], Kunita [49], Fliess [25], Azencott [3], Sussmann [77], Strichartz [76], Ben Arous [7], Grossman & Larson [30], Reutenauer [73], Castell [11], Gaines [29], Lyons [53], Li & Liu [50], Burrage & Burrage [9], Kawski [46], Baudoin [6] and Lyons & Victoir [54]. Many of these authors focus on the exponential Lie series and its applications. A particular impetus of the study of such stuctures arose in the late nineties when Brouder [8] demonstrated the connection between the Hopf algebra used by Connes & Kreimer [16] for the renormalization problem in perturbative quantum field theory and the Butcher group used to study Runge–Kutta methods by Butcher [10] in the late sixties and early seventies; see Hairer, Lubich & Wanner [35]. Hopf algebraic structures are now a natural lexicon in the study of (to name a few): (i) numerical methods for deterministic systems, such as Runge–Kutta methods and geometric and symplectic integrators—see for example Munthe–Kaas & Wright [62], Lundervold & Munthe–Kaas [52] and McLachlan, Modin, Munthe–Kaas & Verdier [59]; (ii) approximations for stochastic differential equations—see for example Castell & Gaines [12,13], Malham & Wiese [57], Ebrahimi–Fard, *et al.* [21], Ebrahimi–Fard, Malham, Patras & Wiese [22,23]; (iii) rough paths—see for example Hairer & Kelly [34] and Gubinelli & Tindel [32] and (iv) regularity structures for stochastic partial differential equations—see for example Gubinelli [31] and Hairer [33].

In summary in this paper, what is new, representing our *main results*, is that we:

(i) Show the stochastic Taylor series expansion for the flowmap can be written in separated form, i.e. as a series of terms, each of which is decomposable into a multiple Itô integral and a composition of associated differential operators (see § 2; in particular Theorem 2.2);
(ii) Describe the class of map-truncate-invert schemes for Lévy-driven equations and give an algebraic framework for encoding and comparing such schemes. These results generalize those in Malham & Wiese [57] and Ebrahimi–Fard *et al.* [21] (see § 3, Procedure 3.1);
(iii) Prove truncations according to the word length grading give more accurate schemes than approximations of the same order of convergence obtained by truncations according to the mean-square grading (see § 4; in particular Theorem 4.1);
(iv) Show how to compute map-truncate-invert schemes in practice and how to deal with the inversion stage. We call these direct map-truncate-invert schemes (see § 5, Corollary 5.1);
(v) Introduce the antisymmetric sign reverse integrator, a new integration scheme for Lévy-driven equations, represented as half the difference of the identity and sign reverse endomorphisms on the vector space generated by words indexing multiple integrals. This scheme is efficient in the sense that its leading order mean-square error is less than that of the corresponding stochastic Taylor scheme, independent of the governing vector fields (see § 5; in particular Theorem 5.1(a));
(vi) Prove the antisymmetric sign reverse integrator is optimal in the sense outlined above at half-integer global root mean-square orders of convergence (see Theorem 5.1(b)).

We round off this paper by establishing global convergence results from local error estimates in § 6 and providing some explicit antisymmetric sign reverse integrators as well as numerical experiments demonstrating our results in § 7. Further numerical results and details can be found in the electronic supplementary material, including an introduction to the role of quasi-shuffle algebras in the theory of stochastic differential equations.

2. Separated stochastic Taylor expansions

Our setting is a complete, filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, assumed to satisfy the usual hypotheses—see Protter [70, p.3]. For the Itô stochastic differential system driven by Lévy processes presented in the introduction, we assume the initial data $y_0 \in L^2(\Omega, \mathcal{F}_0, P)$.

A stochastic Taylor expansion is an expression for the flowmap as a sum of iterated integrals. We write down the stochastic Taylor expansion for Lévy-driven equations and show how and when it can be written in separated form. Recall the flowmap is defined as the map $\varphi_{s,t} \colon y_s \mapsto y_t$. It acts on sufficiently smooth functions $f \colon \mathbb{R}^N \to \mathbb{R}^N$ as the pullback $\varphi_{s,t}(f(y_s)) := f(y_t)$. We set

$\varphi_t := \varphi_{0,t}$. Itô's formula (see for example Applebaum [2, p. 203] or Protter [70, p. 71]) implies

$$f(y_t) = f(y_0) + \sum_{i=0}^{d} \int_0^t \tilde{V}_i \circ f(y_s)\,\mathrm{d}W_s^i + \sum_{i=d+1}^{\ell} \int_0^t \int_{\mathbb{R}} (\tilde{V}_i \circ f)(y_{s_-}, v)\,\bar{Q}^i(\mathrm{d}v, \mathrm{d}s),$$

where for $i = 1, \dots, \ell$, the $\tilde{V}_i$ are operators defined as follows

$$\tilde{V}_i \circ f := \begin{cases} (V_i \cdot \nabla) f, & \text{if} \quad i = 1, \dots, d, \\ f\big(\,\cdot\, + vV_i(\,\cdot\,)\big) - f(\,\cdot\,), & \text{if} \quad i = d+1, \dots, \ell, \end{cases}$$

and $\tilde{V}_0$ is the operator defined by

$$\tilde{V}_0 \circ f := (V_0 \cdot \nabla) f + \frac{1}{2} \sum_{i=1}^{d} \sum_{j,k=1}^{N} V_i^j V_i^k \partial_{x_j} \partial_{x_k} f + \sum_{i=d+1}^{\ell} \int_{\mathbb{R}} \big[(\tilde{V}_i \circ f)(\cdot, v) - v(V_i \cdot \nabla) f\big]\, \rho^i(\mathrm{d}v).$$

Note for $i = d+1, \dots, \ell$, the $\tilde{V}_i$ introduce an additional dependence on a real parameter v. The stochastic Taylor expansion is derived by expanding the integrands in the Lévy-driven equation using Itô's formula. This procedure is repeated iteratively, where the iterations are encoded as follows. Let A be the alphabet $A := \{0, 1, \dots, \ell\}$. For any such alphabet, we use A^* to denote the free monoid over A—the set of words $w = a_1 \dots a_m$ constructed from letters $a_i \in A$. We write $\mathbb{1}$ for the empty word. For a given word $w = a_1 \dots a_m$ define the operator $\tilde{V}_w := \tilde{V}_{a_1} \circ \cdots \circ \tilde{V}_{a_m}$. Let $\mathfrak{s}(w)$ be the number of letters of w from the subset $\{d+1, \dots, \ell\} \subset A$. For a given integrand $g(t, v) \colon \mathbb{R}_+ \times \mathbb{R}^{\mathfrak{s}(w)} \to \mathbb{R}^N$, we define the iterated integrals $I_w(t)[g]$ inductively as follows. We write $I_{\mathbb{1}}(t)[g] := g(t)$, and

$$I_w(t)[g] := \begin{cases} \displaystyle\int_0^t I_{a_1 \dots a_{m-1}}(s)[g]\,\mathrm{d}W_s^{a_m}, & \text{if} \quad a_m = 0, 1, \dots, d, \\ \displaystyle\int_0^t \int_{\mathbb{R}} I_{a_1 \dots a_{m-1}}(s_-)[g(\cdot, v)]\,\bar{Q}^{a_m}(\mathrm{d}v, \mathrm{d}s), & \text{if} \quad a_m = d+1, \dots, \ell. \end{cases}$$

Iteratively applying the chain rule generates the following (see Platen & Bruti-Liberati [68]).

Theorem 2.1 (Stochastic Taylor expansion)**.** *For a Lévy-driven equation the action of the flowmap φ_t on sufficiently smooth functions $f \colon \mathbb{R}^N \to \mathbb{R}^N$ can be expanded as follows*

$$\varphi_t \circ f = \sum_{w \in A^*} I_w(t)\big[\tilde{V}_w \circ f\big].$$

Remark 2.1 (Stochastic Taylor expansion convergence)**.** For integration schemes derived from the stochastic Taylor expansion to converge, it suffices that the terms $\tilde{V}_w \circ f$ included in the expansion, and those at leading order in the remainder, satisfy global Lipschitz and linear growth conditions (see Platen & Bruti-Liberati [68]). Hereafter we assume these conditions are satisfied.

Remark 2.2 (Platen and Bruti–Liberati form)**.** The stochastic Taylor expansion derived in Platen & Bruti-Liberati [68] is an equivalent though different representation; we show this in the electronic supplementary material.

Given the stochastic Taylor expansion for the flowmap in Theorem 2.1, we now show how we can write it in separated form. One component of the separated form are iterated integrals which are free in the sense of having no integrand. We define these abstractly for an arbitrary given alphabet for the moment, the reason for this will be apparent presently.

Definition 2.1 (Free multiple iterated integrals)**.** *Given a collection of stochastic processes $\{Z_t^{a_i}\}_{a_i \in \mathbb{A}}$, indexed by a given countable alphabet $\mathbb{A}$, free multiple iterated integrals take the form*

$$I_w(t) := \int_{0 < \tau_1 < \cdots < \tau_m < t} \mathrm{d}Z_{\tau_1}^{a_1} \cdots \mathrm{d}Z_{\tau_m}^{a_m},$$

where $w = a_1 \dots a_m$ are words in $\mathbb{A}^*$.

We need to augment $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$ with a set of extended driving processes as follows. A key component in the characterization of the algebra generated by the vector space of free iterated integrals of the driving processes $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$ are the compensated power brackets defined for each $i \in \{d+1, \dots, \ell\}$ and for $p \geq 2$ by

$$J_t^{i^{(p)}} := \int_0^t \int_{\mathbb{R}} v^p \overline{Q}^i(\mathrm{d}v, \mathrm{d}s).$$

Equivalently we have $J_t^{i^{(p)}} = [J^i]^{(p)} - t\int_{\mathbb{R}} v^p\, \rho^i(\mathrm{d}v)$, where $[J_t^i]^{(p)}$ is the pth order nested quadratic covariation bracket of J_t^i, i.e. $[J_t^i]^{(2)} := [J_t^i, J_t^i]$ and $[J_t^i]^{(p)} := [J_t^i, [J_t^i]^{(p-1)}]$ for $p \geqslant 3$. Note that for $p \geqslant 3$ the p-th power bracket $[J_t^i]^{(p)}$ equals the sum of the p-th power of the jumps of J^i to time t. Importantly, the compensated power brackets have the property that if $J^{i^{(p)}}$ is contained in the linear span of $\{t, J_t^i, J_t^{i^{(2)}}, \dots, J_t^{i^{(p-1)}}\}$ for some $p \geq 2$, then $J^{i^{(q)}}$ is also in this linear span for all $q \geq p$. Hence inductively for $p \geq 2$, we augment our family $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$ to include the compensated power brackets $J^{i^{(p)}}$ as long as they are not contained in the linear span of $\{t, J_t^i, J_t^{i^{(2)}}, \dots, J_t^{i^{(p-1)}}\}$. By doing so, we obtain a possibly infinite family of stochastic processes. The iterated integrals of this extended family form the algebra generated by the iterated integrals of our driving processes. See Curry *et al.* [20] for further details. In summary we define our extended alphabet as follows.

Definition 2.2 (Extended alphabet)**.** *We define our alphabet $\mathbb{A}$ to contain the letters $0, 1, \dots, \ell$, associated to the driving processes $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$, and the additional letters $i^{(p)}$ corresponding to any $J^{i^{(p)}}$ contained in the extended family as described above.*

Definition 2.3 (Separated stochastic Taylor expansion)**.** *The flowmap for a Lévy-driven stochastic differential equation possesses a separated stochastic Taylor expansion if it can be written in the form*

$$\varphi_t = \sum_{w \in \mathbb{A}^*} I_w \tilde{V}_w,$$

where $\{I_w\}_{w \in \mathbb{A}^*}$ are the free iterated integrals associated to the extended driving processes and $\tilde{V}_w = \tilde{V}_{a_1} \circ \dots \circ \tilde{V}_{a_m}$ are operators indexed by words that compose associatively.

Remark 2.3 (Separated expansion for jump-diffusion equations)**.** In the case of jump-diffusion equations, i.e. Lévy-driven equations for which all the discontinuous driving processes J^i are standard Poisson processes, the stochastic Taylor expansion is of the separated form. Indeed, as standard Poisson processes have jumps of size one only, the operators $\tilde{V}_i$ with $i = d+1, \dots, \ell$ do not introduce a dependence on an additional parameter. The integrands are thus constant across the range of integration, and hence the expansion is separated.

We can construct a separated stochastic Taylor expansion for the flowmap as follows. By Taylor expansion of the term $f\big(y + vV_i(y)\big)$ appearing in the shift $\tilde{V}_i \circ f$, we have

$$(\tilde{V}_i \circ f)(y, v) = \sum_{m \geq 1} v^m \tilde{V}_{i^{(m)}} \circ f(y),$$

where we write $V_i = \big(V_i^1, \dots, V_i^N\big)^{\mathrm{T}}$ and $f = \big(f^1, \dots, f^N\big)^{\mathrm{T}}$. We define the operators $\tilde{V}_{i^{(m)}}$ by

$$\tilde{V}_{i^{(m)}} \circ f^j := \sum_{k \geq 1} \sum_{i_1 + \dots + i_k = m} \frac{1}{m!} V_i^{i_1} \cdots V_i^{i_k} \frac{\partial^m f^j}{\partial y^{i_1} \cdots \partial y^{i_k}},$$

where the $i_j \in \mathbb{N}$. The product in $V_i^{i_1} \cdots V_i^{i_k}$ is multiplication in $\mathbb{R}$. We then have

$$I_i(t)[(\tilde{V}_i \circ f)(\cdot, v)] = \sum_{m \geq 1} \int_0^t \big(\tilde{V}_{i^{(m)}} \circ f\big)\, \mathrm{d}J_s^{i^{(m)}},$$

for $i = d+1, \ldots, \ell$. Inserting the above into the relation

$$I_w(t)\big[\tilde{V}_w \circ f\big] = I_{a_2 \ldots a_m}(t)\Big[I_{a_1}(\cdot)\big[\tilde{V}_{a_1} \circ (\tilde{V}_{a_2 \ldots a_m} \circ f)\big]\Big]$$

and iterating gives the separated expansion, where the operators $\tilde{V}_a$ are those of the stochastic Taylor expansion for $a \in \{0, 1, \ldots, d\}$, and for $a = i^{(m)}$ are given by the $\tilde{V}_{i^{(m)}}$ defined above. We have thus just established the following result.

Theorem 2.2 (Separated stochastic Taylor expansion: existence). *For a Lévy-driven equation suppose the terms $\tilde{V}_w \circ f$ in the stochastic Taylor expansion for the flowmap are analytic on $\mathbb{R}^N$. Then it can be written in separated form, i.e. as a separated stochastic Taylor expansion.*

Remark 2.4 (Linear vector fields and linear diffeomorphisms). In this special case the separated expansion has an especially simple form; see the electronic supplementary material. Note that the identity map is a special case of a linear diffeomorphism.

Hereafter we assume the existence of a separated Taylor expansion for the flowmap, and all iterated integrals are free iterated integrals.

3. Convolution algebras and map-truncate-invert schemes

We now introduce a class of numerical integration schemes we call map-truncate-invert schemes and show how they can be encoded algebraically. The algebraic structures arise naturally from the products of iterated integrals and the composition of operators appearing in the separated stochastic Taylor expansion. Consider the class of numerical integration schemes obtained from the stochastic Taylor expansion by simulating truncations of the expansion on each subinterval of a uniform discretization of a given time domain $[0, T]$.

Definition 3.1 (Grading function and truncations). *A grading function* $\mathrm{g}\colon \mathbb{A}^* \to \mathbb{N}$ *assigns a positive integer to each non-empty word $w \in \mathbb{A}^*$ and zero to the empty word. A truncation is specified by a grading function and truncation value $n \in \mathbb{N}$. We write $\pi_{\mathrm{g}=n}$, $\pi_{\mathrm{g}\leq n}$ and $\pi_{\mathrm{g}\geq n}$ for the projections of $\mathbb{A}^*$ onto the following subsets: (i) $\pi_{\mathrm{g}=n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : \mathrm{g}(w) = n\}$; (ii) $\pi_{\mathrm{g}\leq n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : \mathrm{g}(w) \leq n\}$ and (iii) $\pi_{\mathrm{g}\geq n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : \mathrm{g}(w) \geq n\}$.*

A numerical integration scheme based on the stochastic Taylor expansion is thus given by successive applications of the approximate flow

$$\sum_{w \in \pi_{\mathrm{g}\leq n}(\mathbb{A}^*)} I_w(t)\tilde{V}_w$$

across the computational subintervals. Now more generally, consider a larger class of numerical schemes that are constructed as follows. For a given invertible map $f\colon \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$, we construct a series expansion for $f(\varphi_t)$ using the stochastic Taylor expansion for φ_t. We truncate the series and simulate the retained iterated Itô integrals across each computational subinterval. An integration scheme is obtained by computing the inverse map f^{-1} of the simulated truncations at each step; see Malham & Wiese [57] and Ebrahimi–Fard *et al.* [21]. We now develop an algebraic framework for studying such map-truncate-invert schemes. The starting point is the quasi-shuffle algebra; see Hudson [41] and Hoffman [38]. This gives an explicit description of the algebra of iterated integrals of the extended driving processes. Let $\mathbb{R}\mathbb{A}$ denote the $\mathbb{R}$-linear span of $\mathbb{A}$, and let $\mathbb{R}\langle\mathbb{A}\rangle$ denote the vector space of polynomials in the non-commuting variables in $\mathbb{A}$.

Definition 3.2 (Quasi-shuffle product). *For a given alphabet $\mathbb{A}$, suppose $[\,\cdot\,,\,\cdot\,]\colon \mathbb{R}\mathbb{A}\otimes\mathbb{R}\mathbb{A}\to\mathbb{R}\mathbb{A}$ is a commutative, associative product on $\mathbb{R}\mathbb{A}$. The quasi-shuffle product on $\mathbb{R}\langle\mathbb{A}\rangle$, which is commutative, is generated inductively as follows: if $\mathbb{1}$ is the empty word then $u*\mathbb{1}=\mathbb{1}*u=u$ and*

$$ua*vb=(u*vb)a+(ua*v)b+(u*v)[a,b],$$

for all words $u,v\in\mathbb{A}^*$ and letters $a,b\in\mathbb{A}$. Here ua denotes the concatenation of u and a.

Remark 3.1 (Word-to-integral isomorphism). The word-to-integral map $\mu\colon w\mapsto I_w$ is an algebra isomorphism. Here the domain is the vector space $\mathbb{R}\langle\mathbb{A}\rangle$ equipped with the quasi-shuffle product. Iterated integrals indexed by polynomials are defined by linearity, i.e. $I_{k_u u+k_v v}=k_uI_u+k_vI_v$, for any constants $k_u,k_v\in\mathbb{R}$ and words $u,v\in\mathbb{A}^*$. This was proved in Curry *et al.* [20], it had already been established by Gaines [29] for drift-diffusions and Li & Liu [50] for jump-diffusions.

Remark 3.2 (Shuffle product). The quasi-shuffle product is a deformation of the shuffle product on $\mathbb{R}\langle\mathbb{A}\rangle$. The deformation is induced by an additional product $[\,\cdot\,,\,\cdot\,]$ defined on $\mathbb{R}\mathbb{A}$. If the product $[\,\cdot\,,\,\cdot\,]$ is identically zero, the quasi-shuffle product is just the shuffle product.

Remark 3.3 (Quadratic covariation bracket). Here we associate the product $[\,\cdot\,,\,\cdot\,]\colon\mathbb{R}\mathbb{A}\otimes\mathbb{R}\mathbb{A}\to\mathbb{R}\mathbb{A}$ underlying the quasi-shuffle with the pullback under the word-to-integral map μ of the quadratic covariation bracket of semimartingales. Explicitly, $[0,a]$ is zero for all a, and $[i^{(p)},j^{(q)}]=\delta_{ij}\big(\lambda(i,p,q)\cdot 0+\big(1_{\{d+1,\ldots,\ell\}}(i)\big)\cdot i^{(p+q)}\big)$, where $\lambda(i,p,q)=1$ if $i\in\{1,\ldots,d\}$, and $\lambda(i,p,q)=\int_{\mathbb{R}}v^{p+q}\rho^i(\mathrm{d}v)$ if $i\in\{d+1,\ldots,\ell\}$. The 0 refers to the letter $0\in\mathbb{A}$, and $1_{\{d+1,\ldots,\ell\}}$ is the indicator function of the set $\{d+1,\ldots,\ell\}$.

Remark 3.4 (Word-to-operator homomorphism). The word-to-operator map $\kappa\colon w\mapsto\tilde V_w$ is an algebra homomorphism. Here the domain is $\mathbb{R}\langle\mathbb{A}\rangle$ equipped with the concatenation product. This follows as $\mathbb{R}\langle\mathbb{A}\rangle$ equipped with the concatenation product is the free associative $\mathbb{R}$-algebra over $\mathbb{A}$, see Reutenauer [73], and each $\tilde V_w$ is given by the associative composition of operators $\tilde V_{a_i}$.

We write $\mathbb{R}\langle\mathbb{A}\rangle_*$ for the quasi-shuffle algebra, otherwise the product on $\mathbb{R}\langle\mathbb{A}\rangle$ is concatenation. The preceding observations combine to give an encoding of integration schemes in the algebra $\mathbb{R}\langle\mathbb{A}\rangle_*\overline{\otimes}\,\mathbb{R}\langle\mathbb{A}\rangle$, where $\overline{\otimes}$ is the completed tensor product of Reutenauer [73, p. 18, 29].

Proposition 3.1 (Algebraic encoding of the flowmap). *For a given Lévy-driven equation, the map $\mu\otimes\kappa$ is a homomorphism from $\mathbb{R}\langle\mathbb{A}\rangle_*\overline{\otimes}\,\mathbb{R}\langle\mathbb{A}\rangle$ to the tensor product of the algebra of iterated integrals of the extended driving processes and the composition algebra generated by the set of operators $\tilde V_w$. Moreover, the flowmap of the equation is the image under $\mu\otimes\kappa$ of the element*

$$\sum_{w\in\mathbb{A}^*}w\otimes w.$$

Truncations of this abstract series representation in $\mathbb{R}\langle\mathbb{A}\rangle_*\overline{\otimes}\,\mathbb{R}\langle\mathbb{A}\rangle$ generate approximations of the flowmap and hence classes of stochastic Taylor numerical integration schemes. Using this context, we now construct an abstract representation for $f(\varphi_t)$ for any given map $f\colon\mathrm{Diff}(\mathbb{R}^N)\to\mathrm{Diff}(\mathbb{R}^N)$ which is expressible as a power series. The key idea is that $f(\varphi_t)$ may be rewritten as the image under $\mu\otimes\kappa$ of a series $\sum F(w)\otimes w$, where $F\in\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$, the space of $\mathbb{R}$-linear maps from the quasi-shuffle algebra $\mathbb{R}\langle\mathbb{A}\rangle_*$ to itself; see Malham & Wiese [57] who established this in the shuffle product context. The explicit form of $F\in\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ is obtained using the convolution algebra associated to the quasi-shuffle product.

Definition 3.3 (Quasi-shuffle convolution product). *For a given quasi-shuffle product $*$ on $\mathbb{R}\langle\mathbb{A}\rangle_*$, the convolution product $\star$ of F and G in $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ is given by*

$$F\star G:=*\circ(F\otimes G)\circ\Delta,$$

where Δ is the deconcatenation coproduct that sends a word w to the sum of all its two-partitions $u \otimes v$ and extends linearly to $\mathbb{R}\langle\mathbb{A}\rangle$, i.e. explicitly for any word $w \in \mathbb{A}^*$ we have $\Delta(w) = \sum_{uv=w} u \otimes v$ and

$$\big(F \star G\big)(w) = \sum_{uv=w} F(u) * G(v).$$

The quasi-shuffle algebra $\mathbb{R}\langle\mathbb{A}\rangle_*$ together with the coproduct Δ forms a bialgebra; see Hoffman [38] and Hudson [40]. In particular, when equipped with the convolution product, the space $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ becomes a unital algebra, where the unit ν is given by the composition of the unit of the quasi-shuffle algebra and the counit of the deconcatenation coalgebra, see Abe [1]. Explicitly, ν is the linear map that sends non-empty words to zero and the empty word to itself. We define the embedding $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*) \to \mathbb{R}\langle\mathbb{A}\rangle_* \overline{\otimes}\, \mathbb{R}\langle\mathbb{A}\rangle$ by

$$F \mapsto \sum_{w \in \mathbb{A}^*} F(w) \otimes w.$$

This is an algebra homomorphism for the quasi-shuffle convolution product; see Reutenauer [73], Curry [19] and Ebrahimi–Fard *et al.* [23]. Given a power series $f(x) = \sum_{k\geq 0} c_k x^k$ with $c_k \in \mathbb{R}$, we define the convolution power series $F^\star(X) := \sum_{k\geq 0} c_k X^{\star k}$, where the $X^{\star k}$ are the kth convolution powers of $X \in \mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$. The following representation of $f(\varphi_t)$ is immediate.

Proposition 3.2 (Convolution power series). *If $f \colon \mathbb{R}\langle\mathbb{A}\rangle_* \overline{\otimes}\, \mathbb{R}\langle\mathbb{A}\rangle \to \mathbb{R}\langle\mathbb{A}\rangle_* \overline{\otimes}\, \mathbb{R}\langle\mathbb{A}\rangle$ has a power series $f(x) = \sum_{k\geq 0} c_k x^k$, we have*

$$f\left(\sum_{w \in \mathbb{A}^*} w \otimes w\right) = \sum_{w \in \mathbb{A}^*} F^\star(\mathrm{id})(w) \otimes w.$$

Further, when the power series f has an inverse $f^{-1}(x) = \sum_{k\geq 0} b_k x^k$ with $b_k \in \mathbb{R}$, the compositional inverse of the convolution power series for F is given by the associated convolution power series F^{-1}.

In other words the pre-image under $\mu \otimes \kappa$ of $f(\varphi_t)$ for any such $f \colon \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$, which we also represent by $f \colon \mathbb{R}\langle\mathbb{A}\rangle_* \overline{\otimes}\, \mathbb{R}\langle\mathbb{A}\rangle \to \mathbb{R}\langle\mathbb{A}\rangle_* \overline{\otimes}\, \mathbb{R}\langle\mathbb{A}\rangle$, is given by the term on the left in the proposition, and thus $f(\varphi_t)$ can be represented by $F^\star(\mathrm{id})$. Combining the algebraic encoding of the flowmap in Proposition 3.1 with Proposition 3.2, we construct the map-truncate-invert scheme associated to a given power series f as follows.

Procedure 3.1 (Map-truncate-invert scheme). Let $f \colon \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$ be an invertible map admitting an expansion as a power series. For a given truncation function $\pi_{\mathrm{g}\leq n}$, the associated map-truncate-invert scheme across a fixed computational interval $[0, t]$ is obtained as follows.

(i) Construct the series $F^\star(\mathrm{id})$;
(ii) Simulate the truncation $\pi_{\mathrm{g}\leq n} \circ F^\star(\mathrm{id})$ given by $\hat{\sigma}_t := \sum_{w \in \pi_{\mathrm{g}\leq n}(\mathbb{A}^*)} I_{F^\star(\mathrm{id})(w)}(t) \tilde{V}_w(\mathrm{id})$.
(iii) Compute the approximation $f^{-1}(\hat{\sigma}_t) \circ y_0$.

Remark 3.5. In $\hat{\sigma}_t$ in (ii) above: The identity map in $F^\star(\mathrm{id})(w)$ is the identity endomorphism in $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$, while the identity map in $\tilde{V}_w(\mathrm{id})$ is the identity diffeomorphism on $\mathbb{R}^N$.

Remark 3.6 (Map-truncate-invert endomorphism). Map-truncate-invert schemes are realizations of the integration scheme associated to the endomorphism $F^{-1} \circ \pi_{\mathrm{g}\leq n} \circ F^\star(\mathrm{id}) \in \mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle)$.

An endomorphism which will prove useful in what follows is the augmented ideal projector.

Definition 3.4 (Augmented ideal projector). *The augmented ideal projector denoted by $\mathfrak{J}$ is given by $\mathfrak{J} := \mathrm{id} - \nu$. In other words, it acts as the identity on non-empty words, but sends the empty word to zero.*

Example 3.1 (Exponential Lie series)**.** Recall from the introduction that the motivating example of a map-truncate-invert scheme was the exponential Lie series integrator, corresponding to $f = \log$. For any endomorphism X that maps the empty word to itself, $\log^\star(X)$ is a power series in $X - \nu$. In particular, we have $\log^\star(\mathrm{id}) = \mathfrak{J} - \mathfrak{J}^{\star 2}/2 + \cdots + (-1)^{k+1}\mathfrak{J}^{\star k}/k + \cdots$.

Remark 3.7. For any word w, say of length k, we can naturally truncate a series in convolutional powers of $\mathfrak{J}$ with real coefficients c_1, c_2, c_3 and so forth, to $c_1\mathfrak{J} + c_2\mathfrak{J}^{\star 2} + \cdots + c_k\mathfrak{J}^{\star k}$. This is because the action of $\mathfrak{J}^{\star(k+1)}$ and subsequent terms in the series is, by convention, zero as a word of length k cannot be partitioned into more than k non-empty parts.

The computation of $f^{-1}(\hat{\sigma}_t)$ from $\hat{\sigma}_t$ is in general non-trivial. For the exponential Lie series integrator, Castell & Gaines [12,13] proposed computing $\exp(\hat{\sigma}_t)(y_0)$ by numerical approximation of the ordinary differential equation $u' = \hat{\sigma}_t(u)$, subject to the initial condition $u(0) = y_0$. An approximate solution may then be recovered from $u(1)$. For the sinhlog integrator in the shuffle product context, Malham & Wiese [57] showed that for linear constant coefficient equations, since $\hat{\sigma}$ is a square matrix, the approximate flow $\exp\sinh^{-1}(\hat{\sigma}_t) = \hat{\sigma}_t + \big(\mathrm{id} + \hat{\sigma}_t^2\big)^{1/2}$ may be computed using a matrix square root. If the vector fields are nonlinear Malham & Wiese [57] suggested expanding the square root to sufficiently high degree terms. In §5 we introduce direct map-truncate-invert schemes which evaluate $F^{-1} \circ \pi_{\mathrm{g}\leq n} \circ F^\star(\mathrm{id})$ directly and circumvent the step involving the computation of f^{-1} in general.

4. Endomorphism inner product, gradings and error analysis

We establish convergence and accuracy of integration schemes at an algebraic level. We first define an inner product on the space of endomorphisms corresponding to the L^2 inner product of the associated approximate flows. We then define the mean-square grading and introduce stochastic Taylor schemes of a specific local and then global order. We subsequently consider word length grading. We show stochastic Taylor integration schemes of a given strong order obtained by truncating according to word length are always more accurate than those obtained by truncating according to the mean-square grade. We measure accuracy by the leading order term in the L^2-norm of the remainder after truncation. To start, to define the endomorphism inner product, we require an algebraic encoding of the expectation of the iterated integrals.

Definition 4.1 (Expectation map)**.** *For any word $w \in \mathbb{A}^*$, the expectation map $E\colon \mathbb{R}\langle\mathbb{A}\rangle \to \mathbb{R}[t]$ is defined by $E\colon w \mapsto t^{|w|}/|w|!$ if $w \in \{0\}^*$ and is zero for all other words. Here $|w|$ denotes the length of the word w, $\{0\}^* \subset \mathbb{A}^*$ is the free monoid over the letter 0 and $\mathbb{R}[t]$ is the polynomial ring over a single indeterminate t commuting with $\mathbb{R}$.*

Remark 4.1 (Expectation of iterated Itô integrals)**.** The expectation map corresponds to the expectation of iterated Itô integrals as follows. First, integrals indexed by words not ending in the letter 0 are martingales and hence have zero expectation. Second, consider integrals indexed by a word with at least one non-zero letter. Fubini's Theorem implies such integrals also have zero expectation; see Protter [70].

We now define an inner product on $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$, following Ebrahimi–Fard *et al.* [21]. Suppose we apply two separate functions to the flowmap which are characterised by the endomorphisms F and G in $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$. The stochastic processes generated by these functions of the flowmap, for given initial data y_0, are $\sum_{w\in\mathbb{A}^*} I_{F(w)}\tilde{V}_w(y_0)$ and $\sum_{w\in\mathbb{A}^*} I_{G(w)}\tilde{V}_w(y_0)$. Our goal is to define an inner product $\langle F, G\rangle$ on $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ which matches the L^2-inner product of these two vector-valued stochastic processes. However we would like all of our results to be independent of the initial data y_0 and the governing vector fields appearing in the $\tilde{V}_w$ terms. We achieve this by replacing the vectors $\tilde{V}_w(y_0)$ with a set of indeterminate vectors indexed by words, $\{\mathbf{V}_w\}_{w\in\mathbb{A}^*}$. We write (u, v) for the inner product of $\mathbf{V}_u$ and $\mathbf{V}_v$, i.e. $(u, v) := \mathbf{V}_u^{\mathrm{T}}\,\mathbf{V}_v$. Let $\mathbf{V}$ denote the infinite square matrix indexed by the words $u, v \in \mathbb{A}^*$ with entries (u, v).

Definition 4.2 (Inner product). *We define the inner product of endomorphisms F and G in* $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ *with respect to* $\mathbf{V}$ *to be*

$$\langle F, G\rangle := \sum_{u,v\in\mathbb{A}^*} E\left(F(u) * G(v)\right)(u,v).$$

All results we subsequently establish will hold independent of $\mathbf{V}$.

Remark 4.2 (Positive definiteness). As the operators $\tilde{V}_i$ typically include second (or higher) order differential operators that send the identity map to zero, distinct endomorphisms may be associated with the same stochastic process. For linear constant coefficient equations for instance, the operators $\tilde{V}_{i^{(m)}}$ are zero operators for all $m > 1$, and the stochastic process associated to any endomorphism is trivial if its image lies in the two-sided ideal in $\mathbb{R}\langle\mathbb{A}\rangle_*$ generated by the letters $i^{(m)}, m > 1$. To obtain positive definiteness of the inner product, we pass to the quotient of $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ under the equivalence relation for which endomorphisms yielding the same stochastic process are equivalent. Positive definiteness on the quotient space follows by similar arguments to those in Ebrahimi–Fard *et al.* [21].

The norm, in this quotient space, of an endomorphism $F \in \mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ is $\|F\| := \langle F, F\rangle^{1/2}$.

The first part of our error analysis uses mean-square grading, see Platen & Bruti-Liberati [68].

Definition 4.3 (Mean-square grading). *For any word $w \in \mathbb{A}^*$, the map $\mathrm{g}^{\mathrm{ms}} \colon w \mapsto 2\zeta(w) + \xi(w)$ is the mean-square grading, where $\zeta(w)$ and $\xi(w)$ are the number of zero and non-zero letters in w, respectively.*

Definition 4.4 (Reduced words). *For a given word w the reduced word $\mathrm{red}(w)$ is defined as the word obtained by deleting any zero letters, and replacing any letters of the form $i^{(m)}$ with i.*

Example 4.1. If $w = 010024^{(3)}30$, we have $\mathrm{red}(w) = 1243$.

The following result will be useful in our subsequent error analysis, see Kloeden & Platen [47].

Lemma 4.1 (Expectation of products). *Let $\mathbb{R}\langle\mathbb{A}\rangle_*$ be the quasi-shuffle algebra based on the set of independent Lévy processes $\{t, W^1, \ldots, W^d, J^{d+1}, \ldots, J^\ell\}$ extended by covariation. For any words $u, v \in \mathbb{A}^*$, if $\mathrm{red}(u) = \mathrm{red}(v)$ then there is a non-zero constant $C = C(u,v)$ such that*

$$E(u * v) = C(u,v)t^{(\mathrm{g}^{\mathrm{ms}}(u)+\mathrm{g}^{\mathrm{ms}}(v))/2},$$

otherwise the expectation of the product $u * v$ is zero.

Proof. Let $\langle p, w\rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$ denote the coefficient of a given word w in the polynomial p. We then have

$$E(u * v) = \sum_{w\in\{0\}^*} \langle u * v, w\rangle_{\mathbb{R}\langle\mathbb{A}\rangle} \frac{1}{|w|!} t^{|w|}.$$

Consider the generating relation of the quasi-shuffle product in Definition 3.2, which states that $ua * vb = (u * vb)a + (ua * v)b + (u * v)[a, b]$. We see that if a summand in the polynomial $ua * vb$ is to be a multiple of $w \in \{0\}^*$, we require either $a = 0$, $b = 0$ or $\langle [a, b], 0\rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$ non-zero. As the quadratic covariation of t with any Lévy process vanishes, all zero letters contribute to the expectation only through the first two terms in the quasi-shuffle generating relation above. In contrast, non-zero letters a and b contribute only through the third term. We see that each zero letter appears exactly once in any $w \in \{0\}^*$ for which $\langle u * v, w\rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$ is non-zero, and any pair of non-zero letters a, b contribute one letter together. In particular, we have $E(u * v) = C(u,v)t^{(\mathrm{g}^{\mathrm{ms}}(u)+\mathrm{g}^{\mathrm{ms}}(v))/2}$, where $C(u,v)$ may equal zero.

By the independence of the driving processes, we have $[i, j]$, $[i^{(p)}, j]$ and $[i^{(p)}, j^{(q)}]$ are all zero for all i, j, p, q, $i \neq j$. Moreover, we have $[J^{i^{(p)}}, J^{i^{(q)}}]_t = J_t^{i^{(p+q)}} + t\int_{\mathbb{R}} v^{p+q}\rho^i(\mathrm{d}v)$. In particular, we have that $\langle [i^{(p)}, i^{(q)}], 0\rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$ is non-zero for all $i > d$ and all $p, q \geq 1$ for which $i^{(p)}$ and $i^{(q)}$ exist. Moreover, as $[W^i, W^i]_t = t$, we have $[i, i] = 0$ for $i = 1 \ldots, d$.

The expression $u*v$ is a linear combination of words, each of which arises from a choice of one of the three terms at each stage in the inductive quasi-shuffle generating relation. Indeed the kth letter in a given word thus obtained is the letter a, b or $[a,b]$ chosen at the kth application of the inductive definition. For this word to consist of only zeros, we must choose the first or the second term as long as either a or b is zero. When a and b are both non-zero, we must choose the third term. From the preceding paragraph, this will be a sum featuring a multiple of the zero letter provided both letters are either equal to i, where $i=1,\ldots,d$, or are $i^{(p)}$ and $i^{(q)}$ respectively for some $i>d$ and $p,q\geq 1$. Continuing this procedure, we see that we will obtain a word comprising only zeros if and only if the reduced words $\mathrm{red}(u)$ and $\mathrm{red}(v)$ are equal. □

We wish to compare the errors of different integration schemes across a given time step at the algebraic level, using the inner product defined above. As the flowmap corresponds to the identity in $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$, we define the remainder endomorphism as follows.

Definition 4.5 (Remainder endomorphism). *For any endomorphism $H\in\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ encoding an approximation of the flowmap, we define the associated remainder endomorphism R to be*

$$R:=\mathrm{id}-H.$$

Example 4.2. A simple example is $H=\pi_{\mathrm{g}\leq n}$ corresponding to a stochastic Taylor expansion truncated according to a given grading 'g'. In this case $R=\mathrm{id}-\pi_{\mathrm{g}\leq n}=\pi_{\mathrm{g}\geq n+1}$. More generally we might have $H=F^{-1}\circ\pi_{\mathrm{g}\leq n}\circ F(\mathrm{id})$ so $R=\mathrm{id}-F^{-1}\circ\pi_{\mathrm{g}\leq n}\circ F(\mathrm{id})$.

For the rest of this section we focus on numerical schemes constructed by truncating the stochastic Taylor expansion, first, according to mean-square grading and, second, according to word length grading. We note the flowmap $\varphi_t\in L^2$, see Lemma 6.1. Let $\hat{\varphi}_t^{\mathrm{ms}}$ denote the truncated stochastic Taylor expansion, truncated according to mean-square grading. We denote the corresponding remainder by $R_t^{\mathrm{ms}}:=\varphi_t-\hat{\varphi}_t^{\mathrm{ms}}$ and note $R_t^{\mathrm{ms}}\in L^2$, see Platen & Bruti–Liberati [68]. Locally the numerical scheme is of mean-square order n if

$$R_t^{\mathrm{ms}}=\sum_{w\in\pi_{\mathrm{g}\geq n+1}(\mathbb{A}^*)}I_w(t)\tilde{V}_w.$$

In particular this implies, for sufficiently small t, that $\|R_t^{\mathrm{ms}}(y_0)\|_{L^2}^2=(1+|y_0|^2)\cdot\mathcal{O}(t^{n+1})$ for any initial data $y_0\in\mathbb{R}^N$. We naturally apply the numerical scheme $\hat{\varphi}_t^{\mathrm{ms}}$ successively over a suitably fine discretization of the global time interval of integration to obtain a suitably accurate numerical approximation. This means we need to determine the global order of convergence for such a scheme. Milstein's Theorem [60,61] provides a mechanism for inferring global convergence estimates from local ones in the case of drift-diffusion equations. This can be extended to Lévy driven equations and is provided in Theorem 6.1 in §6. We call it the Generalized Milstein Theorem. To apply this theorem we require local expectation estimates for $R_t^{\mathrm{ms}}(y_0)$. We find that

$$E\big(R_t^{\mathrm{ms}}(y_0)\big)=\sum_{k\geq\lfloor n/2\rfloor+1}\frac{t^k}{k!}\tilde{V}_{0^k}(y_0).$$

Here we used that $E\big(I_w(t)\big)$ is only non-zero for words $w\in\{0\}^*$. The notation 0^k denotes such a word w of length k, and $\mathrm{g}^{\mathrm{ms}}(0^k)=2k$. Using the linear growth estimates we observe for some constant $K>0$ we have

$$\Big|E\big(R_t^{\mathrm{ms}}(y_0)\big)\Big|\leq K\,(1+|y_0|^2)^{1/2}\left(\sum_{k\geq\lfloor n/2\rfloor+1}\frac{t^k}{k!}\right).$$

For any finite t the sum is convergent. In particular for small t the upper bound is $\mathcal{O}(t^{\lfloor n/2\rfloor+1})$. Recall from above that $\|R_t^{\mathrm{ms}}(y_0)\|_{L^2}^2=(1+|y_0|^2)\cdot\mathcal{O}(t^{n+1})$. We now refer to the Generalized Milstein Theorem 6.1 in §6. Matching parameters we see that $p_1=\lfloor n/2\rfloor+1$ and $p_2=(n+1)/2$. The theorem states that the approximation $\hat{\varphi}_t^{\mathrm{ms}}$ with remainder R_t^{ms} above will converge globally at rate $p_2 - 1/2$ if $p_1 \geq p_2 + 1/2$. While this is true for when n is even, it does *not* hold when n is odd. This is simply due to the fact that pure deterministic terms in the stochastic Taylor series remainder have a whole integer less root mean-square global order of convergence compared to their local order of convergence. To rectify this we can simply modify our scheme $\hat{\varphi}_t^{\mathrm{ms}}$ to

$$\hat{\varphi}_t^{\mathrm{ms}} = \sum_{\mathrm{g}^{\mathrm{ms}}(w)\leq n} I_w(t)\tilde{V}_w + I_{0^{n^*}}(t)\tilde{V}_{0^{n^*}},$$

where $n^* := \lfloor (n+1)/2 \rfloor$ if n is odd and zero if it is even. This means that the leading order deterministic term in R_t^{ms} is of the same order as previously when n is even, but is of order $\lfloor (n+1)/2 \rfloor + 1 = \lfloor (n+3)/2 \rfloor$ when n is odd. By inspection we observe that $p_1 \geq p_2 + 1/2$ is now satisfied. The modified scheme $\hat{\varphi}_t^{\mathrm{ms}}$ above has mean-square global order of convergence n and the terms included exactly match those specified in Platen & Bruti–Liberati [68, p. 290].

We now consider word length grading which we employ for our main result in the next section.

Definition 4.6 (Word length grading). *For a given word $w \in \mathbb{A}^*$, the word length grading is denoted $|w|$ or g^{wl}, and defined to be the number of letters in w.*

Remark 4.3 (Computational effort). The bulk of the computation effort, when implementing accurate strong numerical schemes derived from truncations of series representations of the flowmap, is associated with the simulation of the iterated integrals I_w. See Lord, Malham & Wiese [51] and Malham & Wiese [56] for more details in the drift-diffusion case. In particular the iterated integrals involving the most distinct non-deterministic letters require the most effort. Hence the additional computational cost required to simulate all the iterated integrals $\{I_w : |w| \leq n\}$ as opposed to the subset $\{I(w) \colon \mathrm{g}^{\mathrm{ms}}(w) \leq n\}$, is minimal.

One benefit of truncating to word length as opposed to mean-square grading is the following.

Theorem 4.1 (Mean-square versus word length graded truncations). *Let $R_t^{\mathrm{wl}}(y_0)$ and $R_t^{\mathrm{ms}}(y_0)$ denote the remainders generated by truncating the separated stochastic Taylor expansion for y_t respectively using the mean-square and word length gradings, both truncated at the same given grade n. Then at leading order for all n we have*

$$\big\|R_t^{\mathrm{wl}}(y_0)\big\|_{L^2}^2 \leqslant \big\|R_t^{\mathrm{ms}}(y_0)\big\|_{L^2}^2.$$

Proof. First, recall the definitions for the mean-square and word length gradings. We observe for any word $w \in \mathbb{A}^*$ we have $\mathrm{g}^{\mathrm{ms}}(w) = 2\zeta(w) + \xi(w)$ and $\mathrm{g}^{\mathrm{wl}}(w) = \zeta(w) + \xi(w)$. Hence we have $\mathrm{g}^{\mathrm{wl}} \leq \mathrm{g}^{\mathrm{ms}}$. This implies that for any alphabet $\mathbb{A}$ constructed from at least one deterministic and one non-deterministic letter, we have $\pi_{\mathrm{g}^{\mathrm{ms}}\leq n}(\mathbb{A}^*) \subset \pi_{\mathrm{g}^{\mathrm{wl}}\leq n}(\mathbb{A}^*)$. Thus correspondingly for their complements $\pi_{\mathrm{g}^{\mathrm{wl}}\geq n+1}(\mathbb{A}^*) \subset \pi_{\mathrm{g}^{\mathrm{ms}}\geq n+1}(\mathbb{A}^*)$. Second, let R^{ms} and R^{wl} denote the remainder *endomorphisms* associated with truncating the stochastic Taylor expansion by mean-square and word length gradings, respectively. Hence for the inner products $\langle R^{\mathrm{ms}}, R^{\mathrm{ms}} \rangle$ and $\langle R^{\mathrm{wl}}, R^{\mathrm{wl}} \rangle$ we sum over words in $\pi_{\mathrm{g}^{\mathrm{ms}}\geq n+1}(\mathbb{A}^*)$ and $\pi_{\mathrm{g}^{\mathrm{wl}}\geq n+1}(\mathbb{A}^*)$, respectively. The difference remainder endomorphism $\hat{R} := R^{\mathrm{ms}} - R^{\mathrm{wl}}$ is at leading order non-zero on words in $\pi_{\mathrm{g}^{\mathrm{ms}}=n+1}(\mathbb{A}^*) \cap \pi_{\mathrm{g}^{\mathrm{wl}}\leq n}(\mathbb{A}^*)$. For any word w in this set we have $2\zeta(w) + \xi(w) = n+1$ and $\zeta(w) + \xi(w) \leq n$, which imply $\zeta(w) > 0$ and thus $\xi(w) < n+1$. Now consider the inner product $\langle R^{\mathrm{wl}}, \hat{R} \rangle$. Any word u on which R^{wl} is non-trivial at leading order has length $\zeta(u) + \xi(u) \geq n+1$. Any word v on which $\hat{R}$ is non-trivial at leading order, we have just shown that $\xi(v) < n+1$. Hence we must have $\xi(u) \neq \xi(v)$ and so $\mathrm{red}(u) \neq \mathrm{red}(v)$. We deduce that $\langle R^{\mathrm{wl}}, \hat{R} \rangle$ is zero. The result then follows from the relation $\|R^{\mathrm{ms}}\|^2 = \|R^{\mathrm{wl}}\|^2 + 2\langle \hat{R}, R^{\mathrm{wl}} \rangle + \|\hat{R}\|^2$. This result holds whether we include the 0^{n^*} in $\hat{\varphi}_t^{\mathrm{ms}}$ or not. □

5. Antisymmetric sign reverse integrator

We now present the error analysis of map-truncate-invert schemes, concluding with the description of the antisymmetric sign reverse integrator and the main theorem concerning its efficiency. We also present the direct map-truncate-invert schemes alluded to at the end of §3. We begin by introducing the pre-remainder endomorphism associated to any map-truncate-invert scheme, that is the remainder terms associated to the truncation of the series $F^{\star}(\mathrm{id})$. We relate the pre-remainder and remainder of such schemes and use this relation to compute the inverse step of a direct map-truncate-invert scheme. We follow with the introduction of the antisymmetric sign reverse integrator and its explicit characterization. We then prove our main theorem on the efficiency of the antisymmetric sign reverse integrator.

Definition 5.1 (Pre-remainder endomorphism)**.** *Let $F\colon \mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)\to\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ be an invertible map and suppose $\pi_{\mathrm{g}\le n}$ is a projection corresponding to a truncation. The pre-remainder endomorphism Q associated to $\pi_{\mathrm{g}\le n}\circ F(\mathrm{id})$ is defined by*

$$Q := F(\mathrm{id}) - \pi_{\mathrm{g}\le n}\circ F(\mathrm{id}).$$

This definition allows for more general maps F than power series in the convolution algebra, which we require presently. The relationship between the pre-remainder and remainder is critical to the error analysis of map-truncate-invert schemes.

Hereafter we assume a grading g which is preserved by the quasi-shuffle product in question. By this we mean that for all words $u,v\in\mathbb{A}^*$, the polynomial $u*v$ is homogeneous of degree $\mathrm{g}(u)+\mathrm{g}(v)$, i.e. it is a sum of words of grade $\mathrm{g}(u)+\mathrm{g}(v)$; see Ebrahimi–Fard *et al.* [21]. In this case we observe that if π represents one of the projectors $\pi_{\mathrm{g}\le n}$, $\pi_{\mathrm{g}=n}$ or $\pi_{\mathrm{g}\ge n}$ according to such a grading g, then $\pi\circ\mathfrak{J}^{\star k}=\mathfrak{J}^{\star k}\circ\pi$ for any $k\in\mathbb{N}$.

Remark 5.1 (Gradings preserved by quasi-shuffles)**.** The power bracket grading introduced in Curry *et al.* [20, Section 4], defined on uncompensated brackets, is grading preserving for any quasi-shuffle. This assigns the value 2 to the letter 0, the value 1 to letters $1,\ldots,\ell$, and the sum of the gradings for the quadratic covariation bracket between any two words. Mean-square grading is preserved for drift-diffusions, however it is not preserved in general for any quasi-shuffle containing discontinuous terms. Importantly, word length grading g^{wl} is preserved when the quasi-shuffle product is in fact the shuffle product. *Importantly*, this latter case will underlie our eventual application below.

Lemma 5.1 (Remainder and pre-remainder relation)**.** *Suppose $f(1+x)=\sum_{k\ge1}c_kx^k$ is a power series with inverse $f^{-1}(x)=1+\sum_{k\ge1}b_kx^k$, where $b_1\equiv 1/c_1$. Let $\pi_{\mathrm{g}\le n}$ be the projection according to the grading preserved by the quasi-shuffle product. Let R and Q be the remainder and pre-remainder, respectively, associated to the map-truncate-invert endomorphism $F^{-1}\circ\pi_{\mathrm{g}\le n}\circ F^{\star}(\mathrm{id})$. Then we have*

$$R=\frac{1}{c_1}Q+\text{h.o.t.}.$$

Remark 5.2 (Higher order terms)**.** The notation 'h.o.t.' in Lemma 5.1 is used to denote higher order terms in the following sense. Suppose our goal is to apply the endomorphism $G^{\star}(\mathrm{id})=C_1\mathfrak{J}+C_2\mathfrak{J}^{\star2}+C_3\mathfrak{J}^{\star3}+\cdots$ to all words w with $\mathrm{g}(w)\le n$. If we write $G^{\star}(\mathrm{id})=C_1\mathfrak{J}+\cdots+C_n\mathfrak{J}^{\star n}+\mathcal{O}(H)$, this implies the endomorphism H annihilates all words w with $\mathrm{g}(w)\le n$. In this sense the term(s) in $\mathcal{O}(H)$ represent higher order terms. For example, suppose for a particular word w we have $\mathrm{g}(w)=n$. The term $\mathfrak{J}^{\star(n+1)}$ splits w into a sum of all its possible non-empty $(n+1)$-partitions quasi-shuffled together of the form $w_1*w_2*\cdots*w_{n+1}$. Since the grading is preserved by the quasi-shuffle product, we have $\mathrm{g}(w_1)+\cdots+\mathrm{g}(w_{n+1})=n$. However since the partitions w_i for all $i=1,\ldots,n+1$ are all non-empty so that $\mathrm{g}(w_i)\ge1$, we have a contradiction. By convention we suppose $\mathfrak{J}^{\star(n+1)}$ annihilates such a word w and so $\mathfrak{J}^{\star(n+1)}$ represents a higher order term in the sense we have outlined.

Proof. We set $P := \pi_{\mathrm{g}\leq n} \circ F^{\star}(\mathrm{id})$. Then we see the pre-remainder $Q = \pi_{\mathrm{g}\geq n+1} \circ F^{\star}(\mathrm{id})$ and thus also $P + Q = F^{\star}(\mathrm{id})$. We observe that since $R = F^{-1}(P+Q) - F^{-1}(P)$ we find

$$R = \sum_{k\geq 1} b_k \left((P+Q)^{\star k} - P^{\star k}\right) = b_1\, Q + b_2\,(P \star Q + Q \star P) + \text{h.o.t.}.$$

Then we have $P \star Q = \big(\pi_{\mathrm{g}\leq n} \circ F^{\star}(\mathrm{id})\big) \star \big(\pi_{\mathrm{g}\geq n+1} \circ F^{\star}(\mathrm{id})\big)$. Since $F^{\star}(\mathrm{id}) = c_1\, \mathfrak{J} + \text{h.o.t.}$ we get $P \star Q = \big(c_1\, \pi_{\mathrm{g}\leq n} \circ \mathfrak{J}\big) \star \big(\pi_{\mathrm{g}\geq n+1} \circ (c_1\, \mathfrak{J} + \cdots + c_{n+1}\, \mathfrak{J}^{\star(n+1)})\big) + \text{h.o.t.}$. Hence we deduce that $P \star Q$ annihilates any words of grade less than $n+2$. Thus the term $b_2 P \star Q$, as well as similarly $b_2\, Q \star P$, represent higher order terms. Using that $b_1 \equiv 1/c_1$ gives the result. □

A similar calculation was used to establish efficiency of the sinhlog integrator in Ebrahimi–Fard *et al.* [21]. We use it to compute the inverse step of a map-truncate-invert scheme as follows.

Corollary 5.1 (Direct map-truncate-inverse schemes)**.** *Suppose $f(1+x) = \sum_{k\geq 1} c_k x^k$ is an invertible power series and let $\pi_{\mathrm{g}\leq n}$ be a truncation according to the grading preserved by the quasi-shuffle product. Then $F^{-1} \circ \pi_{\mathrm{g}\leq n} \circ F^{\star}(\mathrm{id})$, up to higher order terms, is given by*

$$\pi_{\mathrm{g}\leq n} + \pi_{\mathrm{g}=n+1} \circ \Big(\mathfrak{J} - \frac{1}{c_1} F^{\star}(\mathrm{id})\Big).$$

Proof. Starting with the definition of the remainder endomorphism R, and then subsequently using our results above, we observe

$$\begin{aligned} F^{-1} \circ \pi_{\mathrm{g}\leq n} \circ F^{\star}(\mathrm{id}) &= \mathrm{id} - R \\ &= \pi_{\mathrm{g}\leq n} + \pi_{\mathrm{g}\geq n+1} - \frac{1}{c_1} Q + \text{h.o.t.} \\ &= \pi_{\mathrm{g}\leq n} + \pi_{\mathrm{g}\geq n+1} \circ \Big(\mathrm{id} - \frac{1}{c_1}\, F^{\star}(\mathrm{id})\Big) + \text{h.o.t.} \\ &= \pi_{\mathrm{g}\leq n} + \pi_{\mathrm{g}=n+1} \circ \Big(\mathfrak{J} - \frac{1}{c_1}\, F^{\star}(\mathrm{id})\Big) + \text{h.o.t.}. \end{aligned}$$

In the last step we used that $\mathfrak{J}$ is the identity on non-empty words. □

Remark 5.3 (Direct map-truncate-invert in practice)**.** Note that using the power series representation for $F^{\star}(\mathrm{id})$ the direct map-truncate-inverse scheme in Corollary 5.1 is given by

$$\pi_{\mathrm{g}\leq n} - \frac{1}{c_1} \pi_{\mathrm{g}=n+1} \circ \left(c_2 \mathfrak{J}^{\star 2} + c_3 \mathfrak{J}^{\star 3} + \cdots\right) = \pi_{\mathrm{g}\leq n} - \frac{1}{c_1}\left(c_2 \mathfrak{J}^{\star 2} + \cdots + c_{n+1} \mathfrak{J}^{\star(n+1)}\right) \circ \pi_{\mathrm{g}=n+1}.$$

We observe in this formula the action of direct map-truncate-invert schemes. We simulate the corresponding truncated stochastic Taylor scheme represented by $\pi_{\mathrm{g}\leq n}$ and add the corresponding additional terms shown. Note the additional terms $\mathfrak{J}^{\star 2}$, $\mathfrak{J}^{\star 3}$ and so forth only involve products over lower order multiple Itô integrals that have already been simulated in $\pi_{\mathrm{g}\leq n}$. For instance, for a word w with $\mathrm{g}(w) = n+1$, the term $\mathfrak{J}^{\star 2}(w)$ is the sum of all products of Itô integrals of the form $I_u I_v$ with non-empty words u and v such that $uv = w$ and using grade preservation, $\mathrm{g}(u) + \mathrm{g}(v) = n+1$. In particular $\mathrm{g}(u) \leq n$ and $\mathrm{g}(v) \leq n$, so I_u and I_v are already included in $\pi_{\mathrm{g}\leq n}$. The question arises as to whether a given direct map-truncate-inverse scheme converges. For the integrator we construct presently, this will be an automatic consequence of the convergence of the corresponding stochastic Taylor integrator according to word length grading.

We turn our attention to establishing our main result concerning an efficient integrator for equations driven by Lévy processes. For convenience we introduce the following bracketing notation for the bracket $[\,\cdot\,,\,\cdot\,]$ in the quasi-shuffle product. For letters a we set $[a] := a$, while for words $w = a_1 \ldots a_k$ we set $[w] := [a_1, [a_2, \ldots, [a_{k-1}, a_k] \ldots]]$. The commutativity of the bracket means the order of the letters a_i is irrelevant and then its associativity means that all k-fold brackets are equivalent to the canonical form of left-to-right bracketing shown.

Definition 5.2 (Reversal, sign reversal and quasi-shuffle antipode endomorphisms $|S|$, S and $\hat{S}$)**.** *We define three endomorphisms on $\mathbb{R}\langle\mathbb{A}\rangle$ as follows. If $a_i \in \mathbb{A}$ for $i = 1, \ldots, n$ are letters, then we define the: (i) Reversal map: $|S|\colon a_1 \ldots a_n \mapsto a_n \ldots a_1$; (ii) Sign reversal map: $S\colon a_1 \ldots a_n \mapsto (-1)^n a_n \ldots a_1$; and (iii) Quasi-shuffle antipode: $\hat{S}\colon a_1 \ldots a_n \mapsto (-1)^n \sum [u_1]\,[u_2] \ldots [u_k]$, where for the quasi-shuffle antipode the sum is over all possible factorizations of $a_n \ldots a_1 = u_1 u_2 \ldots u_k$ into non-empty subwords u_i for all $k = 1, \ldots, n$.*

Remark 5.4 (Antipode and quasi-shuffle Hopf algebra)**.** The quasi-shuffle antipode is the quasi-shuffle convolution reciprocal map of the identity, i.e. $\mathrm{id} \star \hat{S} = \hat{S} \star \mathrm{id} = \nu$. Indeed, with this in hand, the quasi-shuffle algebra $\mathbb{R}\langle\mathbb{A}\rangle_*$, with deconcatenation Δ as coproduct, is also a Hopf algebra. For more details, see Hudson [40] and Hoffman [38].

Remark 5.5. For the shuffle algebra where the bracket $[\,\cdot\,,\,\cdot\,]$ is trivial, we have $\hat{S} = S$.

Malham & Wiese [57] and Ebrahimi–Fard *et al.* [21] considered drift-diffusion equations interpreted in the Stratonovich sense—which corresponds to the shuffle product case. The scheme of main interest therein was the map-truncate-invert integration scheme generated from the map $f = \mathrm{sinhlog}$. The shuffle convolution power series associated to the series $f(1+x) = \mathrm{sinhlog}(1+x) = x + \frac{1}{2}\sum_{k\geq 2}(-1)^{k-1}x^k$ is expressible in the form $F^{⧢}(X) = \frac{1}{2}(X - X^{⧢(-1)})$. Here $⧢$ represents the convolution product corresponding to the shuffle product. In the shuffle case, $\mathrm{id} ⧢ S = S ⧢ \mathrm{id} = \nu$ and we therefore have the identity $\mathrm{sinhlog}^{⧢}(\mathrm{id}) = \frac{1}{2}(\mathrm{id} - S)$, and similarly $\mathrm{coshlog}^{⧢}(\mathrm{id}) = \frac{1}{2}(\mathrm{id} + S)$. The efficiency results of Malham & Wiese [57] and Ebrahimi–Fard *et al.* [21] for the sinhlog integrator rely on the identity $\langle\!\langle \pi_{\mathrm{g}=n} \circ \mathrm{sinhlog}^{⧢}(\mathrm{id}), \pi_{\mathrm{g}=n} \circ \mathrm{coshlog}^{⧢}(\mathrm{id})\rangle\!\rangle = 0$. Here $\langle\!\langle\,\cdot\,,\,\cdot\,\rangle\!\rangle$ is an inner product on $\mathrm{End}(\mathbb{R}\langle\mathbb{A}\rangle_{⧢})$ similar to $\langle\,\cdot\,,\,\cdot\,\rangle$ but where the underlying expectation map differs due to the use of Stratonovich integrals. The above result utilizes the shuffle algebra structure of multiple Stratonovich integrals with respect to continuous semimartingale integrators. To obtain an appropriate extension of the sinhlog integrator to Lévy-driven equations, we broaden our definition of map-truncate-invert schemes. In particular, we introduce the antisymmetric sign reverse integrator as follows.

Definition 5.3 (Antisymmetric sign reverse integrator)**.** *The antisymmetric sign reverse integrator is the direct map-truncate-invert scheme associated with $\frac{1}{2}(\mathrm{id} - S)$, truncating according to word length.*

Remark 5.6. We cannot in general give an expression for $\frac{1}{2}(\mathrm{id} - S)$ as a convolution power series in $\mathfrak{J}$ in any quasi-shuffle convolution algebra with non-trivial bracket. To see this, compare the action of S and a map $\alpha\mathfrak{J} + \beta\mathfrak{J}^{\star 2}$ on a general word of length two. We see that $S(ab) = ba$ and the identity $(\alpha\mathfrak{J} + \beta\mathfrak{J}^{\star 2})(ab) = (\alpha + \beta)ab + \beta ba + \beta[a, b]$. Comparing these two expressions, we see that $-\alpha = \beta = 1$, and that $[a, b] = 0$ for all a, b. This implies that $[\,\cdot\,,\,\cdot\,]$ must be trivial.

The following representation generates a practical implementation of the antisymmetric sign reverse integrator. In particular, it includes the inverse step.

Lemma 5.2 (Direct representation of the antisymmetric sign reverse integrator)**.** *The direct antisymmetric sign reverse integrator is given by*

$$\sum_{|w|\leq n} I_w(t)\tilde{V}_w + \sum_{|w|=n+1} I_{\mathrm{coshlog}^{⧢}(w)}(t)\tilde{V}_w,$$

or equivalently

$$\sum_{|w|\leq n} I_w(t)\tilde{V}_w + \sum_{|w|=n+1} \Big(I_{\mathrm{coshlog}^{\star}(\mathrm{id})(w)}(t) + \tfrac{1}{2} I_{(S-\hat{S})(w)}(t)\Big)\tilde{V}_w.$$

Proof. The map $\frac{1}{2}(\mathrm{id} - S)$ is identified with $\mathrm{sinhlog}^{⧢}(\mathrm{id})$. The shuffle product preserves the word length grading so $\pi_{\mathrm{g}=n+1} \circ S = S \circ \pi_{\mathrm{g}=n+1}$ and thus $\pi_{\mathrm{g}=n+1} \circ \mathrm{sinhlog}^{⧢}(\mathrm{id}) = \mathrm{sinhlog}^{⧢}(\mathrm{id}) \circ$

$\pi_{g=n+1}$. Hence applying Corollary 5.1 with $F = \text{sinhlog}^{\sqcup\!\sqcup}$ and using that $c_1 = 1$ in this case and the identity $\text{sinhlog}^{\sqcup\!\sqcup}(\text{id}) + \text{coshlog}^{\sqcup\!\sqcup}(\text{id}) = \text{id}$, we have

$$\begin{aligned}
&\big(\text{sinhlog}^{\sqcup\!\sqcup}\big)^{-1} \circ \pi_{g\le n} \circ \text{sinhlog}^{\sqcup\!\sqcup}(\text{id}) \\
&\quad = \pi_{g\le n} + \big(\text{id} - \text{sinhlog}^{\sqcup\!\sqcup}(\text{id})\big) \circ \pi_{g=n+1} + \text{h.o.t.} \\
&\quad = \pi_{g\le n} + \big(\text{coshlog}^{\sqcup\!\sqcup}(\text{id})\big) \circ \pi_{g=n+1} + \text{h.o.t.} \\
&\quad = \pi_{g\le n} + \big(\text{coshlog}^{\star}(\text{id})\big) \circ \pi_{g=n+1} + \big(\text{coshlog}^{\sqcup\!\sqcup}(\text{id}) - \text{coshlog}^{\star}(\text{id})\big) \circ \pi_{g=n+1} + \text{h.o.t.} \\
&\quad = \pi_{g\le n} + \big(\text{coshlog}^{\star}(\text{id})\big) \circ \pi_{g=n+1} + \tfrac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1} + \text{h.o.t.},
\end{aligned}$$

using that $\text{coshlog}^{\sqcup\!\sqcup}(\text{id}) = \frac{1}{2}(\text{id} + S)$ and $\text{coshlog}^{\star}(\text{id}) = \frac{1}{2}(\text{id} + \hat{S})$. Ignoring higher order terms, using the convolution algebra embedding and applying $\mu \otimes \kappa$ then gives the desired result. □

Remark 5.7 (Antisymmetric sign reverse integrator in practice)**.** The terms $\big(\text{coshlog}^{\star}(\text{id})\big) \circ \pi_{g=n+1}$ and $\frac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1}$ correspond to polynomials of words of lower word length n or less. For the former term this is straightforward following analogous arguments to those in Remark 5.3—noting that the projection operator $\pi_{g=n+1}$ is applied before $\text{coshlog}^{\star}(\text{id})$. For the latter term, $(S - \hat{S}) \circ (a_1 \dots a_{n+1})$ consists of a sum of terms, each of which contains at least one bracket. For example one term would be $(-1)^{n+1}[a_{n+1}, a_n]a_{n-1} \dots a_1$ which is a word of length n, and so forth. Hence the additional terms $\big(\text{coshlog}^{\star}(\text{id})\big) \circ \pi_{g=n+1}$ and $\frac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1}$ required for implementing the direct antisymmetric sign reverse integrator consist of lower order Itô integrals we have already simulated to construct the stochastic Taylor integrator $\pi_{g\le n}$.

Lemma 5.3. *Let $\pi_{g=n}$ be the projection according to the word length grading. Then for any $\mathbf{V}$, any words u and v and endomorphisms X and Y with $\pi = \pi_{g=n}$, we have:*

(i) $E \circ |S| \equiv E$;
(ii) $|S|(u * v) \equiv \big(|S|(u)\big) * \big(|S|(v)\big)$;
(iii) $\langle |S| \circ X, Y\rangle = \langle X, |S| \circ Y\rangle$;
(iv) $\langle X, Y\rangle = \langle |S| \circ X, |S| \circ Y\rangle$;
(v) $\|\pi \circ S\|^2 = \|\pi \circ \text{id}\|^2$;
(vi) $\big\langle \frac{1}{2}(\text{id} - S)\,,\, \frac{1}{2}(\text{id} + S)\big\rangle = 0$.

Remark 5.8. An immediate consequence of item (iii) in Lemma 5.3 is that $|S|$ is self-adjoint with respect to the inner product. Further if we combine this result with item (i) then for any endomorphism Z we have $\langle \text{id}, Z\rangle = \langle |S|, Z\rangle \Leftrightarrow |S| \circ Z = Z$.

Proof. We prove each result item by item. Result (i) follows from the fact that $E(w) = 0$ unless $w \in \{0\}^*$, so we have $E(w) = E\big(|S|(w)\big)$. From Hoffman & Ihara [39, Prop. 4.2] we know that the reversal map $|S|$ is an automorphism for any quasi-shuffle algebra, which establishes result (ii). Result (iii) follows by direct computation as follows. For any endomorphisms X and Y, results (ii) and then (i) imply $E\left((|S| \circ X)(u) * Y(v)\right) = E\left(|S|\big(X(u) * (|S| \circ Y)(v)\big)\right) = E\left(X(u) * (|S| \circ Y)(v)\right)$. Using the definition of the inner product establishes the result. Result (ii) implies that $E\big(|S|(u) * |S|(v)\big) = E\big(|S|(u * v)\big) = E(u * v)$, giving result (iv). Finally using this, for any words $u, v \in \mathbb{A}^*$ we have $E\big((\pi_{g=n} \circ S)(u) * (\pi_{g=n} \circ S)(v)\big) = (-1)^{2n} E\big(|S|(\pi_{g=n} \circ u) * |S|(\pi_{g=n} \circ v)\big) = E\big((\pi_{g=n} \circ u) * (\pi_{g=n} \circ v)\big)$, giving result (v). Result (vi) follows from (v) using the bilinearity of the inner product. □

Finally, our main result is as follows. We consider the following perturbed version of the antisymmetric sign reverse integrator, $\frac{1}{2}(\text{id} - S) + \epsilon Z$, where Z is any endomorphism satisfying $|S| \circ Z = Z$ on words of length $n + 1$ and ϵ is a real-valued parameter.

Theorem 5.1 (Main result: Efficiency of the antisymmetric sign reverse integrator)**.** *Consider the flowmap of a stochastic differential system driven by independent Lévy processes with moments of all orders.*

Assume the flowmap possesses a separated stochastic Taylor expansion, and the vector fields are sufficiently smooth to ensure convergence of the truncated stochastic Taylor integration schemes to all orders. Then:

(a) The antisymmetric sign reverse integrator at a given truncation level n is efficient in the sense that its local leading order mean-square errors are always smaller than those of the truncated stochastic Taylor scheme of the same order according to word length grading, independent of the driving vector fields and of the initial conditions, i.e. at leading order we have

$$\|R_t^{\mathrm{ASRI}}(y_0)\|_{L^2}^2 \leqslant \|R_t^{\mathrm{wl}}(y_0)\|_{L^2}^2;$$

(b) When n is odd the antisymmetric sign reverse approximation is optimal in the following sense. If we perturb the antisymmetric sign reverse integrator by the perturbations described above, then the difference between the two quantities shown in the inequality above is minimized when the perturbation is zero.

Proof. It suffices to show the leading order remainder endomorphism $\pi_{\mathrm{g}=n+1} \circ \mathrm{id}$ associated with the truncated stochastic Taylor expansion has greater norm than the remainder endomorphism $\pi_{\mathrm{g}=n+1} \circ \frac{1}{2}(\mathrm{id} - S)$ associated with the antisymmetric sign reverse integrator, independent of $\mathbf{V}$. For convenience we set $Q_\epsilon := \frac{1}{2}(\mathrm{id} - S) + \epsilon\, Z$. Using the identity $\mathrm{id} = Q_\epsilon + \frac{1}{2}(\mathrm{id} + S) - \epsilon\, Z$, and the results of Lemma 5.3, setting $\pi = \pi_{\mathrm{g}=n+1}$ we have

$$\begin{aligned}
\|\pi \circ \mathrm{id}\|^2 &= \Big\langle Q_\epsilon + \tfrac{1}{2}(\mathrm{id}+S) - \epsilon\, Z\,,\, Q_\epsilon + \tfrac{1}{2}(\mathrm{id}+S) - \epsilon\, Z \Big\rangle \\
&= \big\|Q_\epsilon\big\|^2 + 2\big\langle Q_\epsilon, \tfrac{1}{2}(\mathrm{id}+S)\big\rangle - 2\epsilon\langle Q_\epsilon\,,\, Z\rangle + \big\|\tfrac{1}{2}(\mathrm{id}+S)\big\|^2 \\
&\quad - 2\epsilon\big\langle \tfrac{1}{2}(\mathrm{id}+S)\,,\, Z\big\rangle + \epsilon^2\|Z\|^2 \\
&= \big\|Q_\epsilon\big\|^2 + \big\|\tfrac{1}{2}(\mathrm{id}+S)\big\|^2 + 2\big\langle \tfrac{1}{2}(\mathrm{id}-S)\,,\, \tfrac{1}{2}(\mathrm{id}+S)\big\rangle + 2\epsilon\big\langle Z\,,\, \tfrac{1}{2}(\mathrm{id}+S)\big\rangle \\
&\quad - 2\epsilon\big\langle \tfrac{1}{2}(\mathrm{id}-S)\,,\, Z\big\rangle - 2\epsilon^2\|Z\|^2 - 2\epsilon\big\langle \tfrac{1}{2}(\mathrm{id}+S)\,,\, Z\big\rangle + \epsilon^2\|Z\|^2 \\
&= \big\|\pi \circ Q_\epsilon\big\|^2 + \big\|\pi \circ \tfrac{1}{2}(\mathrm{id}+S)\big\|^2 - \epsilon\big\langle \pi \circ \tfrac{1}{2}(\mathrm{id}-S)\,,\, \pi \circ Z\big\rangle - \epsilon^2\|\pi \circ Z\|^2.
\end{aligned}$$

Note that in all the intermediate calculations above all the endomorphisms shown, Q_ϵ, id, S and Z and so forth should be preceded by '$\pi\circ$' which we left out for clarity. First we observe that when $\epsilon = 0$ we conclude that $\big\|\pi \circ \frac{1}{2}(\mathrm{id} - S)\big\|^2 \le \|\pi \circ \mathrm{id}\|^2$, giving the first result. Second we observe that when n is odd with $\pi = \pi_{\mathrm{g}=n+1}$, then using that $\langle \mathrm{id}, Z\rangle = \langle |S|, Z\rangle \Leftrightarrow |S| \circ Z = Z$ from Remark 5.8 above, the linear term in ϵ on the right above is zero. This establishes the second result of the theorem. □

Remark 5.9 (Characterization of the perturbations)**.** A wide range of endomorhpisms Z satisfy the condition $|S| \circ Z = Z$ on words of length $n+1$. For example the endomorphism $\mathfrak{J}^{\diamond(n+1)}$, where '$\diamond$' is the convolutional product associated with any quasi-shuffle product such as the one above or simply the shuffle, satisfies the stated condition. Another example is $\frac{1}{2}(\mathrm{id} + |S|)$.

Remark 5.10. Using Theorem 4.1 on mean-square versus word length graded truncations, the result above implies that the antisymmetric sign reverse integrator is more accurate than the corresponding stochastic Taylor approximation truncated according to mean-square grading.

Remark 5.11 (Global strong mean-square order of convergence)**.** Using Theorem 6.1 from § 6 below, we conclude that the antisymmetric sign reverse integrator in Theorem 5.1 converges with global strong mean-square order of convergence n.

The argument underlying our optimality claim for the antisymmetric sign reverse integrator (or the direct antisymmetric sign reverse integrator) when n is odd is as follows. Consider the case when $n = 1$ corresponding to global order $1/2$ convergence. In this case the antisymmetric sign reverse integrator coincides with the Castell–Gaines integrator which we already know is an efficient integrator. The antisymmetric sign reverse integrator in this instance is effectively $(\nu + \mathfrak{J})(w) \equiv \mathrm{id}(w)$ on words of length $|w| = 1$ but has a remainder consisting of words w of length $|w| = 2$ generated by the form $(\nu + \mathfrak{J} - \frac{1}{2}\mathfrak{J}^{\sqcup\!\sqcup 2})(w)$. It is natural to perturb this integrator by adding the term ϵZ with $Z=\mathfrak{J}^{\sqcup\sqcup 2}$ which satisfies the condition $|S|\circ Z=Z$ on words of length 2—our premise is not to perturb the term $\mathfrak{J}$ which would affect the role of the single letters, i.e. the role of the individual processes, at this level. Then Theorem 5.1 shows that $\epsilon=0$ gives the optimal efficient integrator, consistent with the fact that it is an efficient integrator. Next we consider the case $n=2$. In this case the antisymmetric sign reverse integrator is $(\nu+\mathfrak{J}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 2})(w)$ on words of length $|w|\leqslant 2$ with a remainder consisting of words w of length $|w|=3$ generated by the form $(\nu+\mathfrak{J}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 2}+\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 3})(w)$. Since it is natural to ask/require our integrator to retain its efficiency properties if we truncate it to generate the corresponding integrator at the next order down, we do not perturb $(\nu+\mathfrak{J}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 2})$ and consider perturbing the integrator by adding the perturbation ϵZ with $Z=\mathfrak{J}^{\sqcup\sqcup 3}$ which satisfies the required property for words of length 3. In this instance Theorem 5.1 tells that the optimally efficient integrator may be realized for a non-zero value of ϵ, however for $\epsilon=0$ it *is* an efficient integrator independent of the underlying stochastic system, so we stick with that form. Continuing to the case $n=3$, the antisymmetric sign reverse integrator is $(\nu+\mathfrak{J}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 2}+\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 3})(w)$ on words of length $|w|\leqslant 3$ with remainder consisting of words of length 4 generated by $(\nu+\mathfrak{J}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 2}+\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 3}-\frac{1}{2}\mathfrak{J}^{\sqcup\sqcup 4})(w)$. Following the sequence of arguments we established for the previous case implies we consider perturbing the antisymmetric sign reverse integrator by $\epsilon\mathfrak{J}^{\sqcup\sqcup 4}$. Theorem 5.1 implies that $\epsilon=0$ gives the optimal efficient integrator, and so forth.

Remark 5.12 (Modified direct antisymmetric sign reverse integrator)**.** The optimality result concerns odd values of n. In the case of even n, there is a small modification of the direct antisymmetric sign reverse integrator which performs well in practical tests. Consider the coefficient of $\tilde{V}_{a^{n+1}}$ for some letter a, i.e. the operators of leading order corresponding to words made up of only one letter. This coefficient is $\frac{1}{2}I_{(1+(-1)^{n+1})a^{n+1}}$, which is equal to $I_{a^{n+1}}$ in the case of odd n, but is zero for even n. For odd n, the term is exactly that appearing in the stochastic Taylor expansion of higher order, whilst for even n, this term vanishes completely. It is also possible in the case of even n to generate the terms $I_{a^{n+1}}$ from iterated integrals of lower word order. In the modified direct antisymmetric sign reverse integrator scheme for even n, we add these additional terms, i.e. the integrator is given by

$$\sum_{|w|\leq n} I_w(t)\tilde{V}_w+\sum_{\substack{|w|=n+1\\ w\neq a^{n+1}}} I_{\text{coshlog}^{\sqcup\sqcup}(w)}(t)\tilde{V}_w+\sum_{w=a^{n+1}} I_{a^{n+1}}(t)\tilde{V}_w.$$

Adding the additional terms produces a better approximation as, by Lemma 4.1, each $w=a^{n+1}$ is orthogonal to all other words, so reproducing the coefficient of every such I_w in the exact remainder improves the accuracy of the scheme. As we do not require the simulation of additional iterated integrals, we guarantee higher accuracy for minimal extra computational effort.

6. Global convergence

The error estimates derived in §4 are local estimates. We require a means of obtaining global convergence from local error estimates. In the case of drift-diffusion equations, such a mechanism was established by Milstein [60,61]. In this section we present a natural generalization of Milstein's theorem to equations driven by Lévy processes. The proof of Milstein's theorem requires the following bounds and continuity results for the exact flow. The proofs are standard and do not rely on the analysis of the previous sections; they are not reproduced here, for details see Situ [74, p. 76–78], Fujiwara & Kunita [28, p. 84–86], Applebaum [2, p. 332–336].

Lemma 6.1 (Continuity and bounds for flows of Lévy-driven equations)**.** *Let $\varphi_{s,t}$ be the flow of a Lévy-driven equation. For any $\mathcal{F}_s$-measurable $x,y\in L^2$, and $s<t$ in $[0,T]$, there exists a constant $K>0$ such that: (i) $\|\varphi_t(y)\|_{L^2}^2\leq K(1+\|y\|_{L^2}^2)$; (ii) $\|\varphi_{s,t}(y)-y\|_{L^2}^2\leq K(t-s)(1+\|y\|_{L^2}^2)$; (iii) $\|\varphi_t(x)-\varphi_t(y)\|_{L^2}^2\leq e^{Kt}\|x-y\|_{L^2}^2$ and (iv) $\|\varphi_{s,t}(x)-\varphi_{s,t}(y)-(x-y)\|_{L^2}^2\leq K(t-s)\|x-y\|_{L^2}^2$.*

We now present the generalization of Milstein's theorem to Lévy-driven equations. Essentially, it states that any Markovian integration scheme for which the local mean-square error converges with order p, converges globally in the mean-square sense with order $p-1/2$, provided the expectation of the local error is at least of order $p+1/2$.

Theorem 6.1 (Generalized Milstein Theorem)**.** *Let $\hat{\varphi}_{s,t}$ be an approximate flow defined for values $s \leq t$ such that $s,t \in \{h, 2h, \ldots, T\}$, where h is the step size of the scheme. Suppose that the expectation of the local error of the approximation is of order p_1, and that the local mean-square error is of order p_2, i.e. for any $t < T$ there exists a constant K such that*

$$\left|E\big(\varphi_{t,t+h}(y) - \hat{\varphi}_{t,t+h}(y)\big)\right| \leq K(1+|y|^2)^{1/2}h^{p_1},$$

$$\|\varphi_{t,t+h}(y) - \hat{\varphi}_{t,t+h}(y)\|_{L^2} \leq K(1+|y|^2)^{1/2}h^{p_2},$$

where $p_2 \geq 1/2$ and $p_1 \geq p_2 + 1/2$. Suppose further that $\hat{\varphi}_{s,t}$ is independent of $\mathcal{F}_s$ for all $s < t$. Then the approximation converges globally to the exact flow with strong order $p_2 - 1/2$, i.e. there exists a constant K such that for all $t \in \{h, 2h, \ldots, T\}$ we have $\|\varphi_t(y_0) - \hat{\varphi}_t(y_0)\|_{L^2} \leq K(1+\|y_0\|^2_{L^2})^{1/2}h^{p_2-1/2}$.

The proof given in Milstein [60,61] for equations driven by Wiener processes relies only on the local accuracy of the approximation and the continuity and growth bounds for the exact flow given in the preceding lemma. By Lemma 6.1, these properties of the exact flow hold for Lévy-driven equations, and hence the proof of Milstein immediately generalizes to give the above theorem. We remark that the assumption that $\hat{\varphi}_{s,t}$ is independent of $\mathcal{F}_s$ is required, as the local error estimates must hold across each computational interval $[s,t]$ where the expectation is taken with respect to $\mathcal{F}_s$. All the schemes we consider fulfil this property.

7. Example integrators and numerical simulations

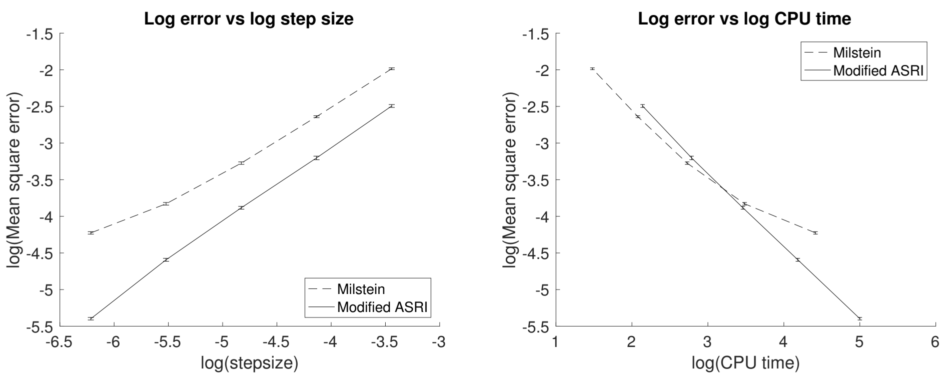

Figure 1: The order 1 modified antisymmetric sign reverse integrator scheme is more accurate than the Milstein scheme for any step size, as shown in the left panel. Further it is more efficient than the Milstein scheme for sufficiently small error tolerances, as can be seen in the right panel.

We present some example antisymmetric sign reverse integrators (see Lemma 5.2) as well as some numerical experiments which demonstrate the results proved above.

Example 7.1. Consider the order one antisymmetric sign reverse approximation. Since $\text{coshlog}^\star(\text{id}) = \nu + \frac{1}{2}\mathfrak{J}^{\star 2} - \frac{1}{2}\mathfrak{J}^{\star 3} + \cdots$ and $(S - \hat{S})(a_1a_2a_3) = \frac{1}{2}\big([a_3a_2]a_1 + a_3[a_2a_1] + [a_3a_2a_1]\big)$

the direct antisymmetric sign reverse approximation has the form

$$\sum_{|w|\le 2} I_w \tilde{V}_w + \sum_{w=a_1a_2a_3} \tfrac{1}{2}\big(I_{a_1}I_{a_2a_3} + I_{a_1a_2}I_{a_3} - I_{a_1}I_{a_2}I_{a_3} + I_{[a_3a_2]a_1} + I_{a_3[a_2a_1]} + I_{[a_3a_2a_1]}\big)\tilde{V}_w.$$

The terms in the second sum with $a1 = a2 = a3$ are zero and need not be evaluated. The modified antisymmetric sign reverse integrator includes the following additional terms,

$$\sum_{w=aaa} \frac{1}{3}\big(I_{aa}I_a - I_{[a,a]}I_a + I_{[a,a,a]}\big)\tilde{V}_w.$$

Example 7.2. We now examine the case of order one half integrators. Here $\mathrm{coshlog}^{\star}(\mathrm{id}) = \nu + \frac{1}{2}\mathfrak{J}^{\star 2}$ and $(S - \hat{S})(a_1a_2) = -\frac{1}{2}[a_2a_1]$, hence the approximation takes the form

$$\sum_{|w|\le 1} I_w \tilde{V}_w + \sum_{w=a_1a_2} \tfrac{1}{2}\big(I_{a_1}I_{a_2} + I_{a_2}I_{a_1} - I_{[a_2a_1]}\big)\tilde{V}_w$$

Consider the equation $\mathrm{d}y_t = A_0 y_t \mathrm{d}t + A_1 y_t \mathrm{d}W_t^1 + V_2(y_t)\mathrm{d}W_t^2 + A_3 y_t \mathrm{d}\tilde{N}_t$, where W_t^1, W_t^2 are Wiener processes and $\tilde{N}_t$ is a standard Poisson process with intensity λ. Here V_2 is the nonlinear vector field $V_2(x_1, x_2, x_3, x_4) = \big(\sin(x_1), \cos(x_2), x_4, -\sin(x_3)\big)^T$, whilst the constant coefficient linear vector fields are defined by the $\mathbb{R}^{4\times 4}$-valued matrices A_0, A_1 and A_2. Their explicit form is given in the electronic supplementary material. We compare the global mean square error $E(\sup_{0\le t\le T} |y_t - \hat{y}_t|^2)^{1/2}$, estimated by sampling 1000 paths, for two approximations of $\hat{y}_t$: the Milstein scheme and the order 1 modified antisymmetric sign reverse integrator. The intensity of $\tilde{N}_t$ was taken to be $\lambda = 50$ and the initial condition was $y_0 = (1, 0.8, 0.6, 0.4)^T$. The code employed is available in the electronic supplementary material together with similar plots for the drift-diffusion case, linear vector fields and so forth. We observe in Figure 1 (left panel) that the antisymmetric sign reverse integrator is more accurate than the Milstein scheme for any given step size, in accordance with the theory. The situation is more nuanced when the mean square error is plotted against the computational time as in Figure 1 (right panel). Generically the Milstein scheme outperforms the modified antisymmetric sign reverse integrator for large step sizes, when the computational effort of the schemes is dominated by the evaluation of the vector fields. For smaller stepsizes, the computational cost of simulating iterated integrals dominates, upon which the modified antisymmetric sign reverse integrator outperforms the Milstein scheme.

Ethics

No research on humans or animals was conducted.

Data accessibility

The datasets supporting this article have been uploaded as part of the supplementary material.

Competing interests

We have no competing interests.

Authors' contributions

The research presented herein was a joint and equal effort by all authors.

Acknowledgement

C.C. would like to thank Michael Tretyakov and Seva Shneer for useful comments. We are very grateful to all the referees whose comments and suggestions significantly helped to improve the presentation of the manuscript; in particular for pointing us towards reference [63] on the quasi-shuffle product.

Funding statement

The authors did not utilize any funding sources.

References

1. Abe, E. 1980 *Hopf algebras*, Cambridge University Press.
2. Applebaum, D. 2004 *Lévy Processes and Stochastic Calculus*, Cambridge University Press.
3. Azencott, R. 1982 Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman, *Seminar on Probability XVI, Lecture Notes in Math.*, **921**, Springer, 237–285.
4. Barndorff–Nielsen, O., Mikosch, T., Resnick, S., ed. 2001 *Lévy processes: theory and applications*, Springer, New York.
5. Barski, M. 2015 Approximations for solutions of Lévy-type stochastic differential equations, arXiv:1512.06572v1
6. Baudoin, F. 2004 *An introduction to the geometry of stochastic flows*, Imperial College Press.
7. Ben Arous, G. 1989 Flots et series de Taylor stochastiques, *Probab. Theory Relat. Fields*, **81**, 29–77.
8. Brouder, C., 2000, Runge–Kutta methods and renormalization, *Eur. Phys. J. C* **12**, 521–534.
9. Burrage, K., Burrage, P. M. 1999 High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula, *Phys. D* **133**, 34–48.
10. Butcher, J.C., 1972 An algebraic theory of integration methods, *Math. Comp. 26*, 79–106.
11. Castell, F. 1993 Asymptotic expansion of stochastic flows, *Probab. Theory Related Fields*, **96**, 225–239.
12. Castell, F., Gaines, J., 1995 An efficient approximation method for stochastic differential equations by means of the exponential Lie series, *Math. Comp. Simulation* **38**, 13–19. (doi:10.1016/0378-4754(93)E0062-A)
13. Castell, F., Gaines, J., 1996 The ordinary differential equation approach to asymptotically efficient schemes for solutions of stochastic differential equations, *Ann. Inst. H. Poincaré Probab. Statist.* **32**, 231–250.
14. Chen, K.-T. 1968 Algebraic paths., *Journal of Algebra*, **10**, 8–36.
15. Clark, J.M.C. 1982 An efficient approximation for a class of stochastic differential equations, in *Advances in Filtering and Optimal Stochastic Control, Cocoyoc, Mexico, 1982, Lecture Notes in Control and Information Sciences*, **42**, Berlin etc: Springer. (doi:10.1007/BFb0004526)
16. Connes, A., Kreimer, D. 1998 Hopf algebras, renormalization and noncommutative geometry, *Commun. Math. Phys.* **199**, 203–242
17. Cohen, P.B., Eyre, T. W. M., Hudson, R. L. 1995 Higher order Itô product formula and generators of evolutions and flows, *International Journal of Theoretical Physics* **34**, 1–6.
18. Cont, R., Tankov, P. 2004 *Financial Modelling with Jump Processes*, Chapman & Hall/CRC.
19. Curry, C., 2014 *Algebraic structures in stochastic differential equations*, PhD thesis, Heriot–Watt University.
20. Curry, C., Ebrahimi–Fard, K., Malham, S.J.A., Wiese, A. 2014 Lévy processes and quasi-shuffle algebras, *Stochastics* **86**(4), 632–642. (doi:10.1080/17442508.2013.865131)
21. Ebrahimi–Fard, K., Lundervold, A., Malham, S.J.A., Munthe–Kaas, H., Wiese, A. 2012 Algebraic structure of stochastic expansions and efficient simulation, *Proc. R. Soc. A* **468**, 2361–2382. (doi:10.1098/rspa.2012.0024)
22. Ebrahimi–Fard, L., Malham, S.J.A., Patras, F., Wiese, A., 2015 Flows and stochastic Taylor series in Itô calculus, *J. Phys. A.* **48**, 495202 (17pp). (doi:10.1088/1751-8113/48/49/495202)
23. Ebrahimi–Fard, L., Malham, S.J.A., Patras, F., Wiese, A., 2015 The exponential Lie series for continuous semimartingales, *Proc. R. Soc. A* **471**. (doi:10.1098/rspa.2015.0429)
24. Eilenberg, S, Mac Lane, S. 1953. On the groups $H(\pi, n)$. *Annals of Mathematics* **58**(1), 55–106.
25. Fleiss, M. 1981 Functionelles causales non linéaires et indéterminées non-commutatives, *Bulletin de la Société Mathématique de France* **109**, 3–40
26. Fournier, N. 2012, Simulation and approximation of Lévy-driven stochastic differential equations, *ESAIM: Probability and Statistics* **15**, 233–248.
27. Friz, P., Shekhar, A. 2012 General Rough Integration, Lévy Rough Paths and a Lévy-Kintchine type formula, *arXiv:1212.5888*.

28. Fujiwara, T., Kunita, H. 1985 Stochastic differential equations of jump type and Lévy processes in diffeomorphisms group, *J. Math. Kyoto Univ.* **25**(1), 71–106.
29. Gaines, J. 1994 The algebra of iterated stochastic integrals, *Stochast. and Stochast. Rep.* **49**, 169–179. (doi:10.1080/17442509408833918)
30. Grossman, R., Larson, R.G. 1992 The realization of input-output maps using bialgebras, *Forum Math.* **4**, 109–121.
31. Gubinelli, M. 2010 Ramification of rough paths, *J. Differential Equations* **248**, 693–721.
32. Gubinelli, M., Tindel, S. 2010 Rough evolution equations, *Ann. Probab.* **38**(1), 1–75.
33. Hairer, M. 2014 A theory of regularity structures *Inventiones mathematicae* **198**(2), 269–504.
34. Hairer, M., Kelly, D. 2015 Geometric versus non-geometric rough paths, *Annales de l'I.H.P. Probabilités et Statistiques*, **51**(1), 207–251.
35. Hairer, E., Lubich, C., Wanner, G. 2002 *Geometric Numerical Integration*, Springer Series in Computational Mathematics.
36. Higham, D.J., Kloeden, P.E. 2005 Numerical methods for nonlinear stochastic differential equations with jumps, *Numer. Math.* **101**, 101–119. (doi 10.1007/s00211-005-0611-8)
37. Higham, D.J., Kloeden, P.E. 2006 Convergence and stability of Implicit methods for jump-diffusion systems, *International Journal of Numerical Analysis and Modelling* **3**(2), 125–140.
38. Hoffman, M.E. 2000 Quasi-shuffle products, *Journal of Algebraic Combinatorics* **11**, 49–68. (doi:10.1023/A:1008791603281)
39. Hoffman, M.E., Ihara, K. 2012 Quasi-shuffle products revisited, Preprint. Max-Planck-Institut für Mathematik Bonn.
40. Hudson, R. L. 2009 Hopf-algebraic aspects of iterated stochastic integrals, *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **12** 479–496.
41. Hudson, R. L. 2012 Sticky shuffle product Hopf algebras and their stochastic representations, in "New trends in stochastic analysis and related topics", *Interdiscip. Math. Sci.* **12**, 165–181, World Sci. Publ., Hackensack, NJ.
42. Hudson, R. L., Parthasarathy, K. R. 1984 Quantum Itô's formula and stochastic evolutions, *Communications in Mathematical Physics* **93**, 301–323.
43. Hudson, R. L., Parthasarathy, K. R. 1993 The Casimir chaos map for $U(N)$, *Tatra Mountains Mathematical Proceedings* **3**, 1–9.
44. Hudson, R. L., Parthasarathy, K. R. 1994 Casimir chaos in a Boson Fock space, *Journal of Functional Analysis* **119**, 319–339.
45. Jacod, J. 2004 The Euler scheme for Lévy driven stochastic differential equations: limit theorems *Ann. Probab.* **32**(3), 1830–1872.
46. Kawski, M. 2001 The combinatorics of nonlinear controllability and noncommuting flows, *Lectures given at the Summer School on Mathematical Control Theory*, Trieste, September 3–28.
47. Kloeden, P.E., Platen, E. 1999 *Numerical Solution of Stochastic Differential Equations* (3rd Printing), Berlin etc: Springer.
48. Kloeden, P.E., Platen, E., Wright, I. 1992 The approximation of multiple stochastic integrals, *Stochastic Anal. Appl.* **10**(4), 431–441. (doi:10.1080/07362999208809281)
49. Kunita, H. 1980 On the representation of solutions of stochastic differential equations, *Lecture Notes in Math.* **784**, Springer–Verlag, 282–304.
50. Li, C.W., Liu, X.Q. 1997 Algebraic structure of multiple stochastic integrals with respect to Brownian motions and Poisson processes, *Stochastics and Stoch. Reports* **61**, 107–120.
51. Lord, G., Malham, S.J.A. & Wiese, A. 2008 Efficient strong integrators for linear stochastic systems, *SIAM J. Numer. Anal.* **46**(6), 2892–2919.
52. Lundervold, A. and Munthe–Kaas, H. 2011 Hopf algebras of formal diffeomorphisms and numerical integration on manifolds, *Contemporary Mathematics* **539**, 295–324
53. Lyons, T. 1998 Differential equations driven by rough signals, *Rev. Mat. Iberoamericana* **14**(2), 215–310.
54. Lyons, T., Victoir, N. 2004 Cubature on Wiener space, *Proc. R. Soc. Lond. A* **460**, 169–198.
55. Magnus, W. 1954 On the exponential solution of differential equations for a linear operator, *Comm. Pure Appl. Math.* **7**, 649–673.
56. Malham, S.J.A., Wiese, A. 2008 Stochastic Lie group integrators, *SIAM J. Sci. Comput.* **30**, 597–617. (doi:10.1137/060666743)
57. Malham, S.J.A., Wiese, A. 2009 Stochastic expansions and Hopf algebras, *Proc. R. Soc. A* **465**, 3729–3749. (doi:10.1098/rspa.2009.0203)
58. Malham, S.J.A., Wiese, A. 2014 Efficient almost-exact Levy area sampling, *Statistics and Probability Letters* **88**, 50–55.

59. McLachlan, R.I., Modin, K., Munthe–Kaas, H., Verdier, O. 2016 What are Butcher series, really? arXiv:1512.00906v2
60. Milstein, G.N. 1987 A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations, *Theor. Prob. Appl.* **32**(4), 738–741.
61. Milstein, G.N. 1995 *Numerical integration of stochastic differential equations*, Mathematics and Its Applications **313**, Kluwer Academic Publishers.
62. Munthe–Kaas, H.Z., Wright, W.M. 2007 On the Hopf algebraic structure of Lie group integrators, *Foundations of Computational Mathematics* **8**(2), 227–257
63. Newman, K., Radford, D. E. 1979 The cofree irreducible Hopf algebra on an algebra, *American Journal of Mathematics* **101**(5), 1025–1045.
64. Newton, N.J. 1986 An asymptotically efficient difference formula for solving stochastic differential equations, *Stochastics*, **19**, 175–206 (doi:10.1080/17442508608833423)
65. Newton, N.J. 1991 Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations, *SIAM J. Appl. Math.* **51**, 542–567. (doi:10.1137/0151028)
66. Platen, E. 1980 Approximation of Itô integral equations, In *Stochastic differential systems,* Volume 25 of *Lecture notes in Control and Inform. Sci.*, 172–176. Springer. (doi:10.1007/BFb0004008)
67. Platen, E. 1982 A generalized Taylor formula for solutions of stochastic differential equations, *SANKHYA A* **44**(2), 163–172.
68. Platen, E., Bruti–Liberati, N. 2010 *Numerical Solution of Stochastic Differential Equations with Jumps in Finance.* Berlin etc: Springer.
69. Platen, E., Wagner, W. 1982 On a Taylor formula for a class of Itô processes, *Probab. Math. Statist.* **3**(1), 37–51.
70. Protter, P. 2003 *Stochastic Integration and Differential Equations (2nd ed.)* Berlin, etc: Springer.
71. Protter, P., Talay, D. 1997 The Euler scheme for Lévy driven stochastic differential equations, *Ann. Probab.* **25**, 393-âĂŞ423.
72. Radford, D.E. 1979 A natural ring basis for the shuffle algebra and an application to group schemes, *Journal of Algebra* **58**, 432–454. (doi:10.1016/0021-8693(79)90171-6)
73. Reutenauer, C. 1993 *Free Lie Algebras*, London Mathematical Society Monographs, Clarendon Press, Oxford.
74. Situ, R. 2005 *Theory of Stochastic Differential Equations with Jumps and Applications: Mathematical and Analytical Techniques with Applications to Engineering*, Springer.
75. Schützenberger, M.-P. 1958. Sur une propriété combinatoire des algèbres de Lie libres pouvant être utilisée dans un problème de mathématiques appliquées., *Séminaire P. Dubreil. Algèbre et théorie des nombres* **12**(1), 1–23.
76. Strichartz, R.S. 1987 The Campbell–Baker–Hausdorff–Dynkin formula and solutions of differential equations, *J. Funct. Anal.* **72**, 320–345. (doi:10.1016/0022-1236(87)90091-7)
77. Sussmann, H.J. 1986 A product expansion for the Chen series, *Theory and Applications of Nonlinear Control Systems*, C.I. Byrnes and A. Lindquist (editors), Elsevier Science Publishers B.V. (North Holland).