ALGEBRAIC STRUCTURES IN STOCHASTIC DIFFERENTIAL EQUATIONS

by

Charles Curry

Submitted for the degree of
Doctor of Philosophy

Department of Mathematics
School of Mathematical and Computer Sciences
Heriot-Watt University

October 2014

Abstract

We define a new numerical integration scheme for stochastic differential equations driven by Lévy processes with uniformly lower mean square remainder than that of the scheme of the same strong order of convergence obtained by truncating the stochastic Taylor series. In doing so we generalize recent results concerning stochastic differential equations driven by Wiener processes. The aforementioned works studied integration schemes obtained by applying an invertible mapping to the stochastic Taylor series, truncating the resulting series and applying the inverse of the original mapping. The shuffle Hopf algebra and its associated convolution algebra play important roles in the their analysis, arising from the combinatorial structure of iterated Stratonovich integrals. It was recently shown that the algebra generated by iterated Itô integrals of independent Lévy processes is isomorphic to a quasi-shuffle algebra. We utilise this to consider map-truncate-invert schemes for Lévy processes. To facilitate this, we derive a new form of stochastic Taylor expansion from those of Wagner & Platen, enabling us to extend existing algebraic encodings of integration schemes. We then derive an alternative method of computing map-truncate-invert schemes using a single step, resolving difficulties encountered at the inversion step in previous methods.

Acknowledgements

First and foremost I would like to thank my supervisors Simon Malham and Anke Wiese for all their help throughout this project. I also wish to thank Kurusch Ebrahimi–Fard for inviting me to Madrid on a number of occasions and many resulting fruitful discussions. I would also like to thank Elena Celledoni, Hans Munthe-Kaas and Brynjulf Owren for a highly enjoyable and interesting stay in Trondheim. I also wish to thank Dominique Manchon for many interesting and helpful discussions in both Madrid and Trondheim, and Frédéric Patras for valuable discussions in Edinburgh and Madrid.

Contents

List of Figures

Chapter 1

Introduction

The main result of this thesis is the description of a new integration scheme for stochastic differential equations of the form

$$dY_t = V_0(Y_t)dt + \sum_{i=1}^{d} V_i(Y_t)dZ_t^i, \tag{1.1}$$

where $V_0, V_1, \ldots V_i$ are sufficiently smooth vector fields on $\mathbb{R}^N$ and $(Z_t^1, \ldots, Z_t^d)$ is a vector of independent Lévy processes possessing moments of all orders. The resulting integration scheme is best applied in the following situations. The first comprises the situation where the equation is a jump diffusion, that is where the Lévy processes are either Wiener processes or standard Poisson processes. The second main application is to linear equations, that is those for which the driving vector fields $V_i(Y_t) = A_i Y_t$, where A_i are constant $N \times N$ matrices. Lévy processes are a class of stochastic processes, comprising processes continuous in probability and possessing stationary increments, independent of the past. Particularly, they are examples of stochastic processes with a well understood structure that nonetheless may incorporate jump discontinuities. Stochastic differential equations of the above form have many applications, notably in mathematical finance for the construction of models going beyond the celebrated the Black-Scholes-Merton model

$$dY_t = aY_t dt + bY_t dW_t, \tag{1.2}$$

where W_t are Wiener processes, in incorporating the discontinuities in stock prices that are a financial reality. Mandelbrot (1963) used Lévy processes to model cotton prices, see McCulloch (1996) for a discussion. Merton (1976) proposed an alternative model utilizing Lévy processes. More recently models driven by Lévy processes have been studied by, among others, Bates (1996), Kou (2002), Madan, Carr & Chang (1998), Rydberg (1997), Boyarchenko & Levendorskii (2002, 2010). There are also applications to physical sciences, see for instance Bardou et al. (2002).

Explicit solutions for stochastic differential equations are rare, especially those with discontinuities and multiple driving noise processes. We accordingly focus on numerical integration schemes. The derivation of our integration scheme utilizes methods of algebraic combinatorics. Particularly, Hopf algebras play an important part in our analysis. Hopf algebras, modules possessing simultaneously unital associative algebra and counital coassciative coalgebra structures obeying certain compatibility relations, first arose in the works of Hopf on algebraic topology. They have since found applications in areas such as geometric integration schemes (see Grossman & Larson 1989, Murua 2006, Munthe-Kaas & Wright 2007), efficient stochastic integrators, (Malham & Wiese 2009, Ebrahimi–Fard et al. 2012), control theory (Kawski 2001, Gray & Duffaut Espinosa 2011, Gray, Duffaut Espinosa & Ebrahimi–Fard 2014), stochastic partial differential equations (Hairer, 2014) and perturbative quantum field theory (Connes & Kreimer 2000, Manchon 2008).

For certain classes of stochastic differential equation, Platen (1980, 1982) and Wagner & Platen (1982) have shown that the solution may be expressed as an infinite series of iterated integrals. For instance, the solution of the autonomous Stratonovich equation

$$dY_t = V_0(Y_t)dt + \sum_{i=1}^{d} V_i(Y_t) \circ dW_t^t, \tag{1.3}$$

where the $\circ$ indicates that the integrals are interpreted in the Stratonovich sense, may be written in the form

$$Y_t = \sum_w J_w(t) V_w \circ \mathrm{id}(Y_0). \tag{1.4}$$

The sum is over all words $w = a_1 \ldots a_n \in \mathbb{A}^*$, where $\mathbb{A}^*$ is the free monoid over the alphabet $\mathbb{A} = \{0, 1, \ldots, d\}$, the letter 0 corresponds to the deterministic process t and $\{1, \ldots, d\}$ to the d independent Wiener processes driving the system. The $V_w = V_{a_1} \circ \ldots \circ V_{a_n}$ are partial differential operators arising from the composition of vector fields, where V_i acts on the space of smooth functions $f : \mathbb{R}^n \to \mathbb{R}^n$ as the Lie derivative $f \mapsto \nabla_{V_i}(f)$. The J_w are multiple stochastic integrals

$$J_w(t) = \int_{0<\tau_{n-1}<\ldots<\tau_1<t} \circ dW^{a_1}_{\tau_n} \ldots \circ dW^{a_n}_{\tau_1}. \tag{1.5}$$

This is derived by expanding the vector fields $V(Y_s)$ iteratively in terms of their evaluation at the initial data using Itô's formula, we have

$$V(Y_t) = V(Y_0) + \int_0^t \nabla_{V_0}(V(Y_s))ds + \sum_i \int_0^t \nabla_{V_i}(V(Y_s))dW^i_s, \tag{1.6}$$

which we may then substitute into the defining equation. It is at this level that the algebraic structures we study enter, as iterated path integrals have a rich algebraic pedigree. For a given path of bounded variation $\gamma = (\gamma_1(t), \ldots \gamma_m(t))$, $t \in [a, b]$ in $\mathbb{R}^N$, K.–T. Chen (1957) considered the iterated path integrals indexed by words in $\{1, \ldots, m\}^*$, defined inductively by

$$I^{\gamma}_{a_1 \ldots a_n} = \int_a^b I^{\gamma|_{[a,t]}}_{a_1 \ldots a_{n-1}} d\gamma_{a_n}(t). \tag{1.7}$$

In doing so, he extended the celebrated Baker-Campbell-Hausdorff formula that $\log(e^x e^y)$ is a Lie element when x and y are non-commuting indeterminates (see Reutenauer 1993, Chapter 3.1). He showed that the formal power series

$$\theta(\gamma) := \log\left(1 + \sum_{p=1}^{\infty} \sum I^{\gamma}_{a_1 \ldots a_n} x_{a_1} \ldots x_{a_n}\right) \tag{1.8}$$

in non-commuting indeterminates $x_1, \ldots x_m$ is a Lie element; that is it consists of a series of polynomials arising from linear combinations of commutators of elements. Strichartz (1987) gave the explicit combinatorial expression for the above series,

$$\theta(\gamma) = \sum_{n=1}^{\infty} \sum_{\sigma \in \mathcal{G}_n} \frac{(-1)^{d(\sigma)}}{n^2 \binom{n-1}{d(\sigma)}} I^{\gamma}_{\sigma^{-1}(a_1 \ldots a_n)} [x_{a_1}, [x_{a_2}, \ldots [x_{a_{n-1}}, x_{a_n}] \ldots], \tag{1.9}$$

where $\mathcal{G}_n$ is the symmetric group on n elements, and $d(\sigma)$ is the number of descents of the permutation σ, defined to be the cardinality of the set $\{i : \sigma(i+1) < \sigma(i)\}$. Note that we have here adopted the standard version of the Chen-Strichartz formula using the left-to-right Lie bracketing $[x_{a_1}, \ldots, [x_{a_{n-1}}, x_{a_n}] \ldots]$, Strichartz originally used the right-to-left bracketing $[\ldots [x_{a_1}, x_{a_2}], \ldots, x_{a_n}]$; see Castell (1993). Ree (1958) showed that Chen's results were best understood in the following algebraic aspect. For any path $\gamma = (\gamma_1, \ldots, \gamma_m)$, the mapping from the space of polynomials in the noncommuting indeterminates $a_1 \ldots, a_m$ to iterated path integrals

$$\varphi : x_{a_1} \ldots x_{a_n} \mapsto I^{\gamma}_{a_1 \ldots a_n} \tag{1.10}$$

is a homomorphism for a certain product defined on the space of non-commuting polynomials called the shuffle product; the name derives from earlier similar products considered by Eilenberg and Mac Lane (1953) in the context of algebraic topology. This ultimately expresses an extension of the integration-by-parts rule for integration operators

$$I(x)I(y) = I(I(x)y) + I(xI(y)). \tag{1.11}$$

The shuffle algebra was subsequently studied extensively by Schützenberger (1958), Chen (1968) and others. Chen's iterated integrals and their algebraic properties have since appeared in a variety of contexts; the Chen-Fliess approach to control theory resolves inputs and outputs as series of iterated integrals (see Fliess 1981, 1983, Kawski 2001, Gray & Duffaut Espinosa 2008) and Lyons' theory of rough paths (Lyons 1998, Lyons et al. 2007) considers differential equations driven by an extended notion of a continuous path, effectively incorporating the higher iterated integrals. There iterated integrals are interpreted as functionals in the tensor algebra

$T(V)$; differential equations may be solved using approximations to the extended paths in a certain topology.

Gaines (1994) and Castell & Gaines (1995, 1996) applied the results of Chen, Ree and Strichartz to autonomous stochastic differential equations driven by Wiener processes: the sample paths are almost surely not of bounded variation, but where the integrals are expressed in Stratonovich form, Wagner & Platen's iterated integrals obey the same algebraic relations as Chen's iterated integrals; the formula of Chen-Strichartz still applies. Castell & Gaines demonstrated the power of the numerical scheme resulting from using a finite truncation of the Chen-Strichartz series; the solution Y may be computed from this truncated series using a standard numerical scheme for ordinary differential equations.

Castell & Gaines (1996) integration scheme was examined further by Lord, Malham & Wiese (2008) and Malham & Wiese (2008); they showed that it did not always improve upon schemes derived from simple truncation of the stochastic Taylor series, as proposed by Wagner & Platen. Malham & Wiese (2009) took the perspective that further integration schemes with desirable properties could be studied using the properties of the shuffle algebra. We define the flowmap for a given stochastic differential equation to be the map $\varphi(t,\omega) : Y_0 \mapsto Y_t$ sending arbitrary initial data to the solution at time t for a given ω, see eg Azencott (1982), Ben Arous (1989). We can go further and consider the flowmap as acting on smooth functions through the pullback $\varphi_*(F)(Y_0) := F(\varphi)(Y_0)$, we can then generalize (1.4) as follows, see Baudoin (2004):

$$\varphi_*(F) = \sum_w J_w V_w \circ F. \tag{1.12}$$

A numerical integration scheme is defined by the truncation of the above series. Malham & Wiese (2009) considered a class of integration schemes indexed by invertible maps $f : \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$; for such a mapping they construct an integration scheme by truncating the series $f(\varphi_t)$ and then applying f^{-1}. The integration scheme of Castell & Gaines corresponds to taking the mapping $f = \log$.

The series $f(\varphi_t)$ is then given by the Chen-Strichartz formula and the inversion may be computed readily by solving an ordinary differential equation using a standard deterministic numerical method, such as an appropriate Runge-Kutta scheme. Malham & Wiese (2009) showed that the integration scheme corresponding to the series $f = \text{sinhlog}$ possesses desirable asymptotic efficiency results; such integrators always possess a smaller coefficient of leading order remainder than schemes comprising truncations of the stochastic Taylor series. These results were obtained in the absence of a drift term, and extended to incorporate drift in Ebrahimi–Fard et al. (2012). We generalize these results to stochastic differential equations driven by Lévy processes. To achieve this, we will have to account for the fact that the integration-by-parts relation (1.11) does not hold for stochastic integrals with respect to discontinuous semimartingale integrators. We will therefore need to look beyond the shuffle algebra. Such considerations are in a sense natural, consider the finite difference operator:

$$\delta_h : f \mapsto \frac{f(x+h) - f(x)}{h}. \tag{1.13}$$

It does not obey the Leibniz rule, but instead a deformed Leibniz relation

$$\delta_h(fg) = \delta_h(f)g + f\delta_h(g) + h(fg). \tag{1.14}$$

Furthermore, if $f(x) \to 0$ sufficiently quickly as $x \to \infty$, the operator δ_h possesses an inverse, the summation operator $I_h : g \mapsto -\sum_{n=0}^{\infty} hg(x+nh)$. This operator obeys a deformed version of the integration-by-parts rule:

$$I(f)I(g) = I(I(f)g) + I(fI(g)) + hI(fg). \tag{1.15}$$

This identity, the Rota-Baxter relation with weight h (see, eg Ebrahimi–Fard & Patras, 2013) is similar to the stochastic integration-by-parts formula. Cartier (1972) gave a presentation of the free commutative Rota-Baxter algebra incorporating a product which is a deformed version of the shuffle product. Hoffman (1999) later formalized this as the quasi-shuffle product and showed it possessed many of the same properties of the shuffle algebra shown by Schützenberger and Chen; he showed it possesses a Hopf algebra structure isomorphic to the shuffle Hopf algebra. Quasi-shuffle algebras have been subsequently studied from an operad theory perspective (see Loday 2007) and found applications in the study of multiple zeta values (Zudilin, 2003) and non-geometric rough paths (Hairer & Kelly 2014).

Recently, Curry et al. (2014) demonstrated an isomorphism between the algebra generated by iterated integrals of independent Lévy processes possessing moments of all orders and a certain quasi-shuffle algebra. We utilise this result to extend formalisms of Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012) to numerical schemes for stochastic differential equations driven by Lévy processes. In accomplishing this, we show how the algebraic structures allow Malham & Wiese's map-truncate-invert schemes to be computed in a single step. We then define a new integration scheme using this single-step method which may be seen as the correct generalization of the sinhlog integrator to discontinuous stochastic differential equations. We will also clarify an issue concerning methods of truncating stochastic Taylor series.

The thesis is structured as follows. We begin with a review of stochastic differential equations. The reader familiar with theory of stochastic calculus and stochastic differential equations may omit this chapter. We will discuss the motivations of stochastic differential equations driven by Wiener processes. We then define stochastic integrals with respect to semimartingales. We proceed to review the Itô calculus for semimartingales, and discuss stochastic differential equations with respect to semimartingales. We then introduce Lévy processes and review their properties, giving the Lévy decomposition of a Lévy processes as the sum of a scalar multiple of time, a Wiener process and an integral with respect to a Poisson random measure. To facilitate this, we conclude with a discussion of integration with respect to random measures.

We have already highlighted the importance of shuffle and quasi-shuffle Hopf algebras to our work. We therefore follow with a chapter discussing shuffle and quasi-shuffle algebras. We begin with definitions and illustrative examples. We then define bialgebras and Hopf algebras and discuss the Hopf algebraic structures associated to shuffle and quasi-shuffle algebras, in particular giving Hoffman's isomorphisms of shuffle and quasi-shuffle Hopf algebras. We proceed to discuss the Lie theoretic aspects of the shuffle algebra, culminating in Radford's theorem that the Lyndon words form a basis for the shuffle algebra, and Hoffman's demonstration that they likewise form a basis of the quasi-shuffle algebra. We close the chapter with a presentation of the results published in Curry et al. (2014) that the algebra generated by iterated integrals of Lévy processes with moments of all orders is isomorphic to a certain quasi-shuffle algebra.

In the following chapter, 'Numerical methods for stochastic differential equations', we will examine the map-truncate-invert schemes of Malham & Wiese (2009), and show how they may be generalized to stochastic differential equations driven by Lévy processes. We begin by examining Wagner & Platen's stochastic Taylor expansion, an analogue of the deterministic Taylor expansion from which we may derive integration schemes of arbitrary order under the assumption that the coefficient functions of the equation are sufficiently smooth (see Wagner & Platen 1982, Platen 1982). For instance, it is sufficient that the coefficient functions are uniformly bounded and have uniformly bounded derivatives of all orders. The stochastic Taylor expansion is derived in the context of equations driven by Wiener processes and Poisson random measures; we will show how these may be adapted to Lévy-driven equations using the Lévy decomposition given above. Furthermore, following work of Malham & Wiese (2009), we will discuss how a large class of integration schemes may be induced from the resulting solution series by applying an invertible mapping, truncating the series and then inverting. We require further results to utilize the resulting expansions. Particularly, the algebraic study of integration schemes requires the separation of stochastic and geometric information in the stochastic Taylor expansion. This property is present naturally in Wagner-Platen expansions of stochastic systems with continuous driving processes (see Baudoin 2004), but not generally for discontinuous equations. The derivation from the Wagner-Platen expansion of solution series with the desired separation presented here is new. Having accomplished this, we show how we can encode algebraically the class of map-truncate-invert schemes introduced by Malham & Wiese (2009). Here the convolution algebra associated with the shuffle and quasi-shuffle bialgebras takes centre stage. This work is new, although it is a natural generalization of the encoding given in Ebrahimi–Fard et al. (2012) for integration schemes in the context of continuous systems. We conclude with a new examination of different ways of truncating the derived expansions, giving an important optimality result.

Having finished our presentation of the role of algebra in deriving and comparing integration schemes, we proceed to consider the implications of this perspective. We begin in chapter 5, entitled ‘Map-truncate-invert schemes’, by summarizing the results obtained through these methods by Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012). We conclude with a counterexample showing that the sinhlog integrator they derived for continuous systems does not necessarily retain its desirable properties when immediately generalized to discontinuous, Lévy-driven systems. In the following chapter, we give a detailed explicit derivation of the relative errors of a general power-series integrator and the integrators of Wagner & Platen. This is presented for drift-diffusion equations with autonomous driving vector fields satisfying the smoothness hypotheses necessary for convergence of integration schemes deriving from truncated stochastic Taylor series. In doing so, we show the necessity of expanding the space of maps considered in map-truncate-invert schemes beyond those given by power series. This material is new, as is all the material of the following chapter, ‘One-step schemes and sign-reverse integrator’, where we give a new method for implementing map-truncate-invert schemes in a single step. This removes the difficulties inherent in the inversion steps of such schemes, and as such is an important refinement of the sinhlog integration scheme. Furthermore, it permits us to look beyond the power-series integrators previously considered. This allows us to give a new integration scheme for autonomous Lévy-driven equations with coefficient functions satisfying the smoothness hypotheses for the convergence of the stochastic Taylor series. The resulting scheme may be seen as a generalization of Malham & Wiese's sinhlog integrator for continuous systems, and show it possesses the same desirable efficiency properties.

The final full chapter, 'Practical implementation', includes a discussion of practical issues inherent in the implementation of integration schemes for Lévy-driven integrators. The opening section is a review; much of the material may be found in Cont & Tankov (2004), note however the important recent results of Fournier (2009) and their application. The chapter concludes with some numerical experiments. In the conclusion, we discuss the applicability the integration schemes derived here and possible generalizations.

The material presented in Section 3.2 has been published in Curry et al. (2014). The material presented in Chapter 4, sections 3–7 and Chapters 6–8 is new. In summary, the new results presented are as follows. From Chapter 4:

- The existence of expansions for Lévy-driven equations derived from Wagner & Platen's stochastic Taylor expansions displaying separation of stochastic and geometric information.
- A comparison of truncation methods using the word order grading and the mean-square grading of Wagner & Platen.
- The construction of a class of numerical integration schemes for Lévy-driven equations we call map-truncate-invert schemes, and and algebraic framework for encoding and comparing such schemes. These results are a natural generalization of results presented in Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012).

From Chapter 6:

- An explicit derivation comparing errors for truncated stochastic Taylor schemes and more general power series schemes for strong order 0.5 and strong order 1.0 integrators of Itô drift-diffusion equations.

From Chapter 7:

- A method of resolving the map-truncate-invert schemes, first introduced in Malham & Wiese (2009) and generalized here, to a single step scheme.
- The derivation of a new single-step integration scheme for stochastic differential equations driven by Lévy processes that is efficient in the sense that it minimizes the coefficient of leading order in the mean-square remainder compared to the truncated stochastic Taylor scheme, for all Lévy-driven equations with sufficiently smooth coefficients.

The results of Chapters 4 and 7 will be presented in the forthcoming paper Curry et al. (2014b), currently under preparation.

Chapter 2

Review of Stochastic Integration and Differential Equations

Here we review the theory of stochastic calculus and stochastic differential equations. We first consider the prototypical examples arising from the Wiener process and discuss the motivations behind these examples. We then give a more technical account of stochastic integration with respect to semimartingales. We discuss the implications of this theory in the following section, entitled 'Stochastic Calculus'. We then give an account of stochastic differential equations driven by semimartingales. There follows an introduction to Lévy processes, the stochastic processes at the heart of this thesis. Motivated by the Lévy decomposition, we will then study integration with respect to random measures. We conclude with a brief account of numerical methods for stochastic differential equations. For a more detailed treatment of white noise and stochastic processes in general, see Doob (1953) or Parzen (1962). For a full account of stochastic integration and stochastic differential equations, see Ikeda & Watanabe (1989) and Protter (2003). For a standalone account of Lévy processes, see Bertoin (1996) or Sato (2013). For a comprehensive account of stochastic integration and stochastic differential equations with respect to Lévy processes, see Appelbaum (2009). Detailed accounts of numerical methods for stochastic differential equations driven by Wiener processes may be found in Milstein (1995), Kloeden & Platen (1999) and Milstein & Tretyakov (2004). For a full treatment of numerical methods for equations with jumps, see Bruti–Liberati &

Platen (2010).

2.1 White Noise and Wiener Processes

The construction of mathematical models often requires a rationalization of random forces. Consider the motion of a small particle on a water surface, a speck of dust or perhaps a grain of pollen. According to Langevin, the motion may be described by an equation of the form

$$\frac{dX_t}{dt} = -aX_t + b\xi_t, \tag{2.1}$$

with the term $-aX_t$ arising from friction, and ξ_t modelling the 'noisy' influence of molecular force (see Langevin, 1908). The latter is to be understood in a probabilistic sense. Indeed, we require the notion of a stochastic process, defined to be a parametrized set of random variables $\{X_t\}_{t\in[0,T]}$. The importance of the central limit theorem suggests we consider those stochastic processes for which all of the joint distribution functions are Gaussian; we call such a process a Gaussian process. In many applications, the practical homogeneity of space or time suggests considering noise processes obeying certain stationarity properties (see Neyman & Scott, 1959). Precisely, we consider processes for which the means $E(X_t)$ are constant and the covariances obey $\mathrm{Cov}(X_s, X_t) = c(t-s)$, where c is a real function. Note the identity $\mathrm{Var}(X_t) = c(0)$, as $\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t)$. Processes with such properties are said to be stationary in the wide sense (see Doob, 1953). In equations like Langevin's arising from physical considerations, the noise ξ_t is often modelled as a Gaussian process, stationary in the wide sense (see Papoulis, 1991). In studying wide sense stationary processes, it is often helpful to consider the spectral density function, defined as the Fourier transform of the autocorrelation of the process:

$$S(\nu) = \int_{-\infty}^{\infty} E\{X_{t+s}X_t\}e^{-i\nu s}ds. \tag{2.2}$$

Ornstein & Uhlenbeck (1930) suggested from physical considerations that the noise process ξ_t in Langevin's equation should be modelled as a wide-sense stationary process with covariance function $c(s) = S_0\delta(s)$. Such processes would have constant spectral density, and hence by analogy with optics are called white noise processes (see Parzen, 1962). Whilst desirable from a physical perspective, the delta function in the defining law requires some further interpretation. For this purpose, we define a standard Wiener process W_t to be a Gaussian process with mean zero at all times, and independent increments $\Delta_{s,t} = W_t - W_s$ obeying $\mathrm{Var}(\Delta_{t,s}) = |t-s|$. These are also known as Brownian motions. Doob (1942) showed that Langenvin's equation, driven by Gaussian white noise, could be written rigorously in the form

$$dX_t = -aX_t dt + bdW_t, \tag{2.3}$$

a so-called stochastic differential equation. In reality, this is to be interpreted as an integral equation of the form

$$X_t = X_0 - \int_0^t aX_s ds + \int_0^t bdW_t. \tag{2.4}$$

To illustrate Doob's perspective, consider the process defined by

$$X^h(t) = \frac{1}{h}\Delta_{t,t+h} = \frac{W_{t+h} - W_t}{h}, \tag{2.5}$$

it may be shown to be zero mean Gaussian process, stationary in the wide sense with spectral density $S_h(\nu) = \left(\frac{\sin(\pi\nu h)}{\pi\nu h}\right)^2$, see Doob (1953, p.525). We see that $S_h(\nu) \to 1$ as $h \to 0$, and hence the Gaussian white noise may be considered as a derivative of the Wiener process, in a suitably weak sense. However, the sample paths $t \mapsto W(t,\omega)$ are almost surely nowhere differentiable (see Hida, 1980). In applications, the noisy forcing terms often depend on the state of the system. The simplest nontrivial dependency is often termed 'multiplicative noise' and corresponds to an equation of the form

$$dX_t = aX_t dt + bX_t dW_t. \tag{2.6}$$

This is sometimes called a Geometric Brownian motion. The celebrated Black-Scholes-Merton model assumes that stock prices follow such an equation (see Black & Scholes 1973 and Merton 1973). Again, this is to be interpreted as an integral equation, but we require considerable care in defining terms of the form $\int X_s dW_s$. The intention is to consider this integral as a random variable, defined pathwise as an integral. We may define the Riemann-Stieltjes integral $\int_a^b f dg$ as the limit of sums

$$S_n = \sum_{t_k, t_{k+1} \in \pi_n} f(t_k)(g(t_{k+1}) - g(t_k)), \tag{2.7}$$

where $\pi_n = \{a = t_0 < t_1 < \ldots < t_l = b\}$ is a sequence of partitions with mesh size tending to zero. It is however a standard result of calculus (see Protter 2003, p. 43) that a necessary and sufficient condition for the above sums to converge for all continuous functions f is that g be of finite variation on bounded intervals; in other words for all finite intervals I

$$\sup_{\pi} \sum_{t_k, t_{k+1} \in \pi} |g(t_{k+1}) - g(t_k)| < \infty, \tag{2.8}$$

where the supremum is over all partitions π of I. The sample paths of Wiener process almost surely are of unbounded variation on any interval (Protter 2003, p. 19), so we may not define integrals $\int X_s dW_s$ as pathwise Riemann-Stieltjes integrals. Nonetheless, Itô (1944) was able to show that we may define a 'stochastic integral' as a limit of sums

$$X_0 W_0 + \sum_{\sigma_n} X_{T_i}(W^{T_{i+1}} - W^{T_i}), \tag{2.9}$$

where $\sigma_n = \{T_0 \leq T_1 \leq \ldots \leq T_l\}$ is a sequence of random partitions with mesh size tending to zero, and $W^{T_i} := W_{T_i \wedge t}$ denotes the process stopped at the random time T_i. The convergence is uniform on compacts in probability, defined as follows.

Definition 2.1.1 *A sequence of processes X_n converges to X uniformly on compacts in probability if, for any given $t > 0$ and $\epsilon > 0$,*

$$P\left(\sup_{0 \leq s \leq t} |X_n(s) - X(s)| \geq \epsilon\right) \to 0. \tag{2.10}$$

2.2 Stochastic Integration and Semimartingales

More generally, our theoretical setting is a complete probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $(\mathcal{F}_t)_{0\leq t\leq\infty}$, an increasing family of σ-algebras that intuitively represents the time evolution of the information known concerning the system. For example, in stochastic differential systems driven by noise process, such as Langevin's equation, all the stochastic information is contained in the state of the noise process(es) up to the given time, we therefore use the filtration $\mathcal{F}_t = \sigma(W_s, s < t)$. It is assumed that the filtration satisfies the 'usual hypotheses': that $\mathcal{F}_0$ contains all P-null sets of $\mathcal{F}$ and the filtration is right continuous, meaning $\mathcal{F}_t = \cap_{u>t}\mathcal{F}_u$ for all $t < \infty$. We will restrict ourselves to considering adapted processes.

Definition 2.2.1 *A stochastic process X is said to be adapted to a given filtration $(\mathcal{F}_t)$ if the random variables X_t are $\mathcal{F}_t$-measurable for all t.*

Roughly speaking this means the state of the process at a given time does not depend on the future state of the system. We further require the notion of a stopping time.

Definition 2.2.2 *A random time $T(\omega)$ on a filtered probability space is a stopping time if, for every $t \geq 0$, the event $\{T \leq t\}$ is $\mathcal{F}_t$-measurable.*

Definition 2.2.3 *For a given stopping time T, we define the stopping time σ-algebra $\mathcal{F}_T$ to be the sub-σ-algebra of $\mathcal{F}$ given by*

$$\{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t,\ \forall t \geq 0\} \tag{2.11}$$

Intuitively, the stopping time σ-algebra $\mathcal{F}_T$ is the smallest σ-algebra containing all right-continuous adapted processes with left limits, sampled at T, see Protter (2003, Theorem I.6). The stochastic integral may then be constructed for adapted integrands and integrators as follows: firstly, on the space of simple predictable processes X, that is adapted stochastic processes with a representation of the form

$$X_t = X_0 1_0(t) + \sum_{i=1}^{n} \chi_i 1_{(T_i, T_{i+1})} \tag{2.12}$$

where (T_i) is a finite, non-decreasing sequence of stopping times and (χ_i) a set of almost surely finite, $\mathcal{F}_{T_i}$-measurable random variables; we take

$$\int X dY = X_0 Y_0 + \sum_{i=1}^{n} \chi_i (Y_{T_{i+1}} - Y_{T_i}). \tag{2.13}$$

The topology of uniform convergence on compacts is metrizable (see Protter 2003 or Ikeda & Watanabe 1989); moreover the space of left-continuous, adapted processes with right limits is complete with respect to this metric, and contains the simple predictable processes as a dense subset. We may then extend the above integral map to all left-continuous, adapted integrands with right limits by continuity. The question arises, for which integrators Y_t is the integral map $X \mapsto \int X dY$ continuous with respect to uniform convergence on compacts in probability? Finite variation processes are included, as the stochastic integral then corresponds to the pathwise Riemann-Stieltjes integral (see Protter 2003, Theorem II.17). Moreover Itô showed that Wiener processes also possess this continuity property (see Protter 2003, Theorem II.8). A deeper result is that continuity of the integral map is a 'local' property: a process X possesses a property locally if there exists a sequence of stopping times (T_n) increasing to ∞ almost surely, for which all the stopped processes X^{T_n} possess the property (see Protter 2003, Theorem II.6). Resolution of the question of continuity of the integral map requires Doob's theory of martingales.

Definition 2.2.4 *A process M_t is a martingale with respect to the filtration $\mathcal{F}_t$ if each M_t is an L^1 random variable such that*

$$E\{M_t | \mathcal{F}_s\} = M_s \text{ (a.s)}, \tag{2.14}$$

for all $s \leq t$. Processes obeying the above properties but with the respective inequalities $\leq, \geq$ replacing the equality in the above relation are called supermartingales and *submartingales respectively.*

Due to the aforementioned local property of the continuity of the integral map, we also require the notion of a local martingale.

Definition 2.2.5 *A process M_t is a local martingale if there exists a sequence of increasing stopping times T_n obeying $\lim_{n\to\infty} = \infty$ a.s., such that $X^{T_n}1_{\{T_n>0\}}$ is a martingale for each n.*

Doob highlighted the importance of martingales to Itô's treatment of stochastic integration; a Wiener process is a martingale with respect to the filtration it generates, and it is from this property that 'Itô's isometry'

$$E\left\{\left(\int_0^t X_s dW_s\right)^2\right\} = E\left\{\int_0^t X_s^2 ds\right\}, \tag{2.15}$$

arises (see Ikeda & Watanabe 1989, p.49). The importance of this identity in Itô's first papers on stochastic integration suggested an integral might be defined for more general martingales; this was achieved in successively more general terms by Courrège (1962), Kunita & Watanabe (1967) and Doleans-Dadé & Meyer (1970). A full treatment of these results requires the beautiful but technical theory of Doob and Meyer relating (sub- and super-) martingales to (sub- and super-) harmonic functions, see Doob (1980) for a comprehensive study. The conclusion, reached independently by Bitcheler and Dellacherie in the 1970s (see Protter 2003, Theorem III.47) is that the integral map $X \mapsto \int XdY$ is continuous if and only if Y is a semimartingale, defined in the following sense.

Definition 2.2.6 *An adapted, right-continuous process with left limits Y is a semimartingale if it may be written as a sum $Y_t = Y_0 + M_t + A_t$, where M is a local martingale and A a finite variation process, obeying $M_0 = A_0 = 0$.*

We conclude our treatment by noting that the stochastic integral with respect to a semimartingale admits a wider class of integrands than left-continuous processes. Specifically, the allowable integrands consist of predictable processes.

Definition 2.2.7 *The predictable σ-algebra is defined to be the smallest σ-algebra on $\mathbb{R}_+ \times \Omega$ for which all left-continuous adapted processes with right limits are measurable.*

Definition 2.2.8 *A stochastic process is said to be predictable if it is measurable with respect to the predictable σ-algebra.*

See, eg Protter (2003, Chapter IV) for a discussion of stochastic integration with respect to predictable integrands.

2.3 Stochastic Calculus

It is important to note that the definition of the stochastic integral $\int X dY$ adopted above possesses properties different from those of deterministic integrals. Particularly, the analogue of the deterministic integration-by-parts rule $fg = \int f dg + \int g df$ does not hold. Indeed, for given semimartingales X, Y, Meyer's quadratic covariation bracket, defined by

$$[X, Y] = XY - \int X_- dY - \int Y_- dX \tag{2.16}$$

is in general non-zero. Note the use of the left-hand limits $X_{t-} := \lim_{s \uparrow t} X_s$ in the above, this is as we require the integrands to be predictable. The bracket is evidently bilinear and symmetric and thus obeys a polarization identity

$$[X, Y] = \frac{1}{2}([X + Y] - [X] - [Y]), \tag{2.17}$$

where $[X] := [X, X]$ is called the quadratic variation. The quadratic variation may be shown to be an increasing process; the quadratic covariation of two semimartingales may be then written as a difference of increasing processes. This property characterizes finite variation processes, hence $[X, Y]$ is a finite variation process and therefore semimartingale for any semimartingales X, Y (see Protter 2003, Chapter II.6). The quadratic covariation bracket therefore endows the vector space of semimartingales with a commutative algebra structure. We may represent the quadratic covariation as a limit of Riemann sums of the form

$$[X,Y] = X_0Y_0 + \lim_{n\to\infty}\sum_i (X^{T^n_{i+1}} - X^{T^n_i})(Y^{T^n_{i+1}} - Y^{T^n_i}), \tag{2.18}$$

see Protter (2003, p.68). This has further consequences for the transformation properties of the stochastic integral. Particularly, let $\sigma_n = \{0 = T^n_0 \leq \ldots \leq T^n_{k_n}\}$ be a sequence of random partitions tending to the identity, meaning $\lim_n \sup_k T^n_k = \infty$ a.s., and $\sup_k |T^n_{k+1} - T^n_k| \to 0$ a.s., see Protter (2003, p.64). For a continuous semimartingale X_t, we may then derive a change-of-variables rule by writing $f(X_t) - f(X_0)$ as a limit of telescoping sums,

$$f(X_t) - f(X_0) = \sum_{i=0}^{k_n} f(X_{T^n_{i+1}}) - f(X_{T^n_i}). \tag{2.19}$$

Assuming f is sufficiently smooth, we may expand each summand using Taylor's formula, we obtain

$$f(X_t) - f(X_0) = \sum_i f^{'}(X_{T^n_i})(X_{T^n_{i+1}} - X_{T^n_i}) + \frac{1}{2}\sum_i f^{''}(X_{T^n_i})(X_{T^n_{i+1}} - X_{T^n_i})^2 + \ldots, \tag{2.20}$$

see Protter (2003). In conventional calculus, we would expect the second term to tend to zero. Instead, examining the Riemann-sum properties of the quadratic variation, it may be seen to converge uniformly on compacts in probability to $\frac{1}{2}\int_0^t f^{''}(X_s)d[X]_s$. In particular, the following change-of-variables rule, called Itô's formula applies:

$$f(X_t) - f(X_0) = \int_0^t f^{'}(X_s)dX_s + \frac{1}{2}\int_0^t f^{''}(X_s)d[X]_s. \tag{2.21}$$

The change-of-variables is further complicated by the presence of jumps. For twice differentiable functions f, the most general form of Itô's formula is as follows. For a vector of semimartingales, $(X_1, \ldots, X_d)$, we have

$$f(X_t) - f(X_0) = \sum_i \int_{0+}^t \frac{\partial f}{\partial x_i}(X_{s-})dY^i_s + \frac{1}{2}\sum_{i,j}\int_{0+}^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{s-})d[X^i, X^j]^c_s \tag{2.22}$$

$$+\sum_{s\leq t}\left\{f(Y_s)-f(Y_{s-})-\sum_i \frac{\partial f}{\partial x_i}(Y_{s-})\Delta Y_s^i\right\}, \tag{2.23}$$

see Protter (2003, Chapter II.7). In modelling applications where a stochastic differential equation arises as an idealized limit of differential equations obeying the usual laws of calculus, these transformation properties are undesirable. For instance, in Langevin-type equations where the noise results from an external force, such as electromagenetic wave propagation in a turbulent atmosphere, the noise will not be truly white, but modelled by a stationary process with an autocorrelation function supported on an interval an order of magnitude smaller than the time scale of the system. In this case, the resulting equation is well-defined and transforms according to the usual rules of calculus. Accordingly, the behaviour of the system is not accurately reproduced by the white noise idealization when the equation is interpreted in the Doob sense using Itô's integral, see van Kampen (1981). In addition, stochastic differential equations using the above definition of the integral are not in general invariant under coordinate changes, for physical applications this presents difficulties. These considerations suggest the desirability of developing a stochastic integral which obeys the usual laws of calculus. This was accomplished independently for continuous integrators in unpublished work of D.L. Fisk, 1963 (see Jarrow & Protter 2004 for a discussion) and by Stratonovich (1966), and further extended by Meyer (1977). Following Meyer, we define the Stratonovich integral by

$$\int_0^t Y_s := dX_s = \int_0^t Y_s dX_s + \frac{1}{2}[Y,X]_t^c, \tag{2.24}$$

where $[Y,X]^c$ denotes the continuous part of the quadratic covariation bracket, and the $\circ$ indicates that the integral is interpreted in the Stratonovich sense. We require that the covariation process $[Y,X]$ is defined, as is the case for instance for semimartingale integrators Y. For a continuous semimartingale X, we see that the usual integration-by-parts rule holds, and in addition we obtain the usual change-of-variables rule

$$f(X_t)-f(X_0)=\int_0^t f'(X_s)\circ dX_s. \tag{2.25}$$

Again assuming the continuity of X_t, the Stratonovich integral may then be written as a limit of Riemann sums of the form

$$\int_0^t Y_{s-} \circ dX_s = \lim_{n\to\infty} \sum_i \frac{1}{2}(Y_{T_i^n} + Y_{T_{i+1}^n})(X^{T_{i+1}^n} - X^{T_i^n}). \tag{2.26}$$

This was Stratonovich's original approach, see Stratonovich (1966) or Protter (2003). For discontinuous integrators there are methods of representing the Stratonovich integral as a limit of Riemann sums but they are somewhat complicated and we will not reproduce them here, see Protter (2003, pp. 291–295). The Stratonovich integral is sometimes defined for Wiener processes as a limit of time averaged Riemann sums of the form

$$\int f(W_s) \circ dW_s = \lim \sum_i f(W_{\frac{1}{2}(t_i+t_{i+1})})(W_{t_{i+1}} - W_{t_i}), \tag{2.27}$$

this representation does not hold for general semimartingale integrators, even under the assumption of continuity, see Yor (1977). We must also note that the Stratonovich integral does not obey the usual laws of calculus where the integrators have jumps, we obtain change-of-variables formulae such as

$$f(X_t) - f(X_0) = \int_0^t f^{'}(X_s) \circ dX_s + \sum_{s\leq t} \left(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\right), \tag{2.28}$$

see Protter (2003).

2.4 Stochastic Differential Equations

Langevin's equation is then a very specific case of the following stochastic differential equation for an R^N-valued process Y, driven by a set of real-valued semimartingales $Z = (Z^1, \ldots, Z^d)$ and maps $V_i : T \times R^N \to R^N$

$$dY_t = V_0(t, Y_t)dt + \sum_{i=1}^{d} V_i(t, Y_t)dZ_t^i. \tag{2.29}$$

As we have discussed, this is to be interpreted as an integral equation:

$$Y_t = Y_0 + \int_0^t V_0(s, Y_s)ds + \sum_{i=1}^{d} \int_0^t V_i(s, Y_s)dZ_s^i \tag{2.30}$$

We must pause to reflect on the meaning of this equation. Suppose we are free to prescribe a particular evolution of the driving processes, $\{Z_t^1, \dots Z_t^d, t \geq 0\}$; a pairing (Y, Z) satisfying the above equation is called a weak solution. If, however, we require the solution Y as a functional of a general process Z, this is termed a strong solution. In this thesis we will focus exclusively on strong solutions. It remains to consider the existence and uniqueness of strong solutions to the above equations. As in the case of ordinary differential equations, the relevant conditions on the driving fields V_i include Lipschitz conditions:

$$||V_i(t, x) - V_i(t, y)|| \leq \alpha_i ||x - y|| \tag{2.31}$$

for some constants α_i and all t, x, y, and linear growth conditions:

$$||V_i(t, x)|| \leq \beta_i(1 + ||x||), \tag{2.32}$$

where β_i are constants. In some cases the latter conditions may be deducible from the former. For a full treatment of existence and uniqueness results, see Protter (2003, Chapter 5.3), and for more specific results pertaining to equations driven by Lévy processes see Appelbaum (2004, chapter 6.2). For results specific to equations driven by Wiener processes, see (Milstein 1995 or Kloeden & Platen, 1998). We will always assume the driving fields V_i are sufficiently regular to ensure the existence and uniqueness of strong solutions.

2.5 Lévy processes

Whilst the Wiener process is an important model of driving noise, it is not suitable for all purposes. For instance, sample paths of Wiener processes and solutions to equations driven by Wiener noise almost surely possess a continuous modification, see Protter (2003, p.17). In many circumstances this is a desirable property, but we may wish to incorporate jumps into our model, for example in modelling financial markets, where sudden movements of prices are a reality. The Wiener process is also scale invariant, in the sense that it possess similar statistical properties across all time scales; again, this may not reflect the properties of the noise we wish to model. See for instance the introductory chapter of Cont & Tankov (2004) for a detailed discussion of the shortcomings of modelling financial markets using Wiener processes, and the desirability of incorporating jump discontinuities in the driving noise. Analytically, we know that we can define a stochastic integral for semimartingales, and that stochastic differential equations driven by semimartingales admit solutions if certain criteria are satisfied. Semimartingales are a very broad class of process and we will restrict ourselves to a subclass of driving semimartingales called Lévy processes. This includes a wide variety of driving noise, including Wiener processes as a special case but permitting also discontinuous processes, and is analytically more tractable than the general semimartingale.

Definition 2.5.1 *An adapted process $(X_t)_{t\in[0,\infty)}$ is defined to be a Lévy process if it satisfies the following conditions: (assuming $s < t \in [0,\infty)$)*

1. $X_0 = 0$ *(a.s.)*
2. *Continuous in probability: we have* $\lim_{h\to 0} P(|X_{t+h} - X_t| \geq \epsilon) = 0$ *for all* $\epsilon > 0, t < \infty$.
3. *Increments independent of past: for all $t > s$, $X_t - X_s$ is independent of $\mathcal{F}_s$, where $(\mathcal{F}_t)_{0\leq t\leq\infty}$ is the filtration generated by X.*
4. *Stationary increments: for all $t > s$, we have $X_t - X_s \sim X_{t-s}$.*

Lévy processes may be considered the continuous time generalizations of random walks, defined to be discrete time stochastic processes consisting of sums of independent identically distributed random variables. The justification of this observation is as follows: consider the discrete time process $S_n(h)$ resulting from sampling a given Lévy process X_t at times $t = nh$. We then have $S_n(h) = \sum_{k=0}^{n-1} \Delta_k$, where

$\Delta_k = X_{(k+1)h} - X_{kh}$ are independent, identically distributed random variables by the definition of Lévy processes. We observe that for any value of t, the distribution of X_t may be represented as a sum of n independent identically distributed random variables for all positive integers n (see Bertoin 1996, p.11). Probability distributions possessing this property are called infinitely divisible. Indeed, infinitely divisible distributions are in one-to-one correspondence with Lévy processes, as for any infinitely divisible distribution F we may define a Lévy process X such that F is the law of X_1 (see Protter 2003, Chapter I.4). Recall that the law of a $\mathbb{R}^N$-valued random variable X is characterized by its characteristic function $\Phi(z) = E\{\exp(iz.X)\}$, where $z.X$ denotes the Euclidean inner product. The stationary, independent increments possessed by a Lévy process ensure its characteristic process $\Phi_t(z) = E\{\exp(iz.X_t)\}$ is multiplicative in the sense that $\Phi_{t+s}(z) = \Phi_t(z)\Phi_s(z)$. The stochastic continuity then shows that $t \mapsto \Phi_t(z)$ must be an exponential function: for any $\mathbb{R}^N$-valued Lévy process X_t we have

$$E\{\exp(iz.X_t)\} = \exp(t\phi(z)), \tag{2.33}$$

where $\phi(z) : \mathbb{R}^N \to \mathbb{R}$ is a continuous function (see Bertoin, 1996). Observe that the characteristic function of an infinitely divisible distribution is then of the form $e^{\phi(z)}$. In fact, we can also characterize the form of ϕ, but we first require a definition:

Definition 2.5.2 (Lévy measure) *Let X be an $\mathbb{R}^N$-valued Lévy process. Define a measure on $\mathbb{R}^N$ by*

$$\nu(A) = E(|\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}|) \tag{2.34}$$

where A is a Borel-measurable set and ΔX_t is the jump process defined by $\Delta X_t = X_t - \lim_{s\uparrow t} X_s$.

Intuitively, $\nu(A)$ measures the expected number per unit time of jumps the process X_t accrues taking values in the set A. Suppose further that (X_t) is real-valued. We obtain the following:

$$E\{\exp(iz.X_t)\} = \exp(t\phi(z)), \tag{2.35}$$

where we have, for some $\gamma, \sigma > 0$

$$\phi(z) = i\gamma z - \frac{1}{2}\sigma z^2 + \int_{-\infty}^{\infty} (e^{izx} - 1 - izx1_{|x|\leq 1})\nu(dx), \tag{2.36}$$

where the measure ν satisfies $\int_{\mathbb{R}}(1 \wedge |x|^2)dx < \infty$. The above identity is the Lévy-Khintchine formula (see Protter, 2003). A similar expression may be derived for $\mathbb{R}^N$-valued processes, see Jacod & Shiryaev (1987) or Bertoin (1996) for details. Related is Lévy's decomposition of a Lévy process: suppose that (X_t) is a real-valued Lévy process. We may then write

$$X_t = \alpha t + \sigma W_t + \int_0^t \int_{\mathbb{R}} x(Q(ds, dx) - ds\nu(dx)), \tag{2.37}$$

where α, σ are real numbers, W_t is a standard Wiener process (see Appelbaum 2004, Theorem 2.4.16). The integral is with respect to the compensated random measure $Q(ds, dx) - ds\nu(dx)$, where ν is the Lévy measure and Q is defined such that $Q([a, b], A)$ counts the number of jumps of the process X_t taking values in the set A in the time $[a, b]$. In the following section we will explore the meaning of this.

2.6 Integration and Point Processes

The treatment in this section largely follows Ikeda & Watanabe (1989). Let $(\mathcal{E}, \mathcal{B}_{\mathcal{E}})$ be a measurable space, where $\mathcal{B}_{\mathcal{E}}$ is the Borel σ-algebra on $\mathcal{E}$. We call a function $p : D_p \to \mathcal{E}$ such that $D_p \subset (0, \infty)$ a point function with marks in $\mathcal{E}$. The intention is to encode the occurence at certain times of an event, or 'mark'. For our purposes, this will always be a real number, intended to represent the magnitude of a jump in the value of some process. To any point function p we associate a counting measure $Q_p(dt, dx)$ on $(0, \infty) \times \mathcal{E}$ with the product Borel σ-algebra $\mathcal{B}((0, \infty)) \times \mathcal{B}_{\mathcal{E}}$, defined such that

$$Q_p((0, t] \times U) = |\{s \in D_p \cap (0, t] : p(s) \in U\}|. \tag{2.38}$$

In words, $Q_p((0, t] \times U)$ counts the number of times $s \leq t$ for which there is a corresponding mark $p(s)$ taking values in U. Let $\pi_{\mathcal{E}}$ be the space of point functions with marks in $\mathcal{E}$, and define $\mathcal{B}(\pi_{\mathcal{E}})$ to be the smallest σ-algebra for which all mappings $p \mapsto Q_p((0,t] \times U)$, $t > 0, U \in \mathcal{B}_{\mathcal{E}}$ are measurable. We then define a point process on $\mathcal{E}$ to be a $(\pi_{\mathcal{E}}, \mathcal{B}(\pi_{\mathcal{E}}))$-valued random variable. We say a point process $p(\omega)$ is Poisson if the associated random counting measure $Q_p(\omega)$ is a Poisson random measure, meaning for all $B \in \mathcal{B}((0,\infty)) \times \mathcal{B}_{\mathcal{E}}$ we have

$$P(Q_p(B) = n) = \frac{\lambda(B)^n e^{-\lambda(B)}}{n!}, \tag{2.39}$$

where λ is a non-random measure on $(0,\infty) \times \mathcal{E}$ called the intensity. It is furthermore required that $Q_p(B_1), \ldots, Q_p(B_n)$ be mutually independent for any disjoint, measurable $B_1, \ldots B_n$. A point process is said to be stationary if the random variables $p(\omega)$ and $\theta_t p(\omega)$ are equal in law for all t, where θ_t is the shift operator acting on point functions by $\theta_t p(s) = p(s+t)$. A Poisson point process is stationary if and only if its intensity measure may be written as

$$\lambda(dt, dx) = dt\nu(dx), \tag{2.40}$$

for some measure ν defined on $\mathcal{E}$ (See Ikeda & Watanabe 1989, p. 43). Suppose once more we have a filtration structure $(\mathcal{F}_t)$ on the underlying probability space. If each $Q_p(t,U)$ is $\mathcal{F}_t$-measurable, then we say the point process $p(\omega)$ is adapted to the filtration. Where it exists we then define, for a point process p with associated counting measure Q_p, a stochastic integral

$$\int_0^t \int_{\mathcal{E}} f(s,x,\omega) Q_p(ds,dx) = \sum_{s \leq t} f(s, p(s), \omega). \tag{2.41}$$

To discuss the properties further, we require the notion of the compensator of process. Rao's theorem (see Rao 1969 or Protter 2003, p. 119) implies that for any process X of locally integrable variation there is a unique predictable, finite variation process $\hat{X}$ such that $X - \hat{X}$ is a local martingale. As the process $t \mapsto Q_p(t,U)$ is of locally integrable variation, it possesses a compensator, which we define to be $\hat{Q}_p(t,U)$. Under mild conditions on $\mathcal{E}$, we may define an associated measure $\hat{Q}_p$

giving rise to the compensator. For a stationary Poisson point process with counting measure Q_p and intensity $\lambda(dt, dx) = dt\nu(dx)$, we have $\hat{Q}_p(t, U) = t\nu(dx)$, see Ikeda & Watanabe (1989, p. 60). We may then define a compensated measure $\tilde{Q}_p$ by $\tilde{Q}_p = Q_p - \hat{Q}_p$. The key point is that we may be able to define an integral with respect to the compensated measure when the integral with respect to the original measure does not exist. For instance, suppose X_t is a real-valued Lévy process with zero Wiener component in its Lévy decomposition and jump measure Q. If X_t is of infinite variation on a given interval $[0, t]$, the sum of jumps $\sum_{s\leq t} \Delta X_s$ is undefined. The integral $\int_0^t \int_{\mathbb{R}} xQ(dx, ds)$ is therefore similarly undefined. Nonetheless, the Lévy decomposition $X_t = \alpha t + \int_0^t \int_{\mathbb{R}} x(Q(dx, ds) - ds\nu(dx))$ will still hold. We begin by defining

$$\int_0^t \int_{\mathcal{E}} f(s, x, \omega)\tilde{Q}_p(ds, dx) \tag{2.42}$$

$$= \int_0^t \int_{\mathcal{E}} f(s, x, \omega)Q_p(ds, dx) - \int_0^t \int_{\mathcal{E}} f(s, x, \omega)\hat{Q}_p(ds, dx), \tag{2.43}$$

for an appropriate class of functions f for which the integrals on the right exist, see Ikeda & Watanabe (1989). The process $t \mapsto \int_0^t \int_{\mathcal{E}} f(s, x, \omega)\tilde{Q}_p(ds, dx)$ is then a martingale. We endow the space of square-integrable martingales with a metric topologizing uniform convergence on compacts in probability. The integral $\int_0^t \int_{\mathcal{E}} f(s, x, \omega)\tilde{Q}_p(ds, dx)$ may then be extended to a wider class of functions by using a sequence of truncated versions of f for which the uncompensated integral is defined and appealing to convergence in the metric space (see Ikeda & Watanabe 2009). Important is the following result, recalling Itô's isometry. For a process X, we define the conditional quadratic variation $\langle X \rangle$ to be the compensator of the quadratic variation $[X]$, where this exists. The following identity holds:

$$\langle \int_0^t \int_{\mathcal{E}} f(s, x, \omega)\tilde{Q}_p(ds, dx) \rangle = \int_0^t \int_{\mathcal{E}} f^2(s, x, \omega)\hat{Q}_p(ds, dx). \tag{2.44}$$

2.7 Numerical Methods

Explicit strong solutions for stochastic differential equations are rare, especially those involving multiple or discontinuous driving processes. We shall therefore restrict our focus to numerical simulation of solutions. Indeed, we consider approximations which are functionals of the driving processes, or more generally their multiple iterated integrals; in practice these processes will be simulated using a pseudorandom number generator and the approximate solution computed from the simulation by a recursive algorithm. The approximate solutions we generate over a time interval $[0, T]$ are time discrete processes defined at a discretization grid of stopping times $0 = \tau_0 < \tau_1 < \ldots < \tau_N = T$. We shall restrict ourselves throughout to equidistant, deterministic grids $\tau_n = nh$, where h is a given stepsize. We wish to minimize the root mean square error of an approximation $\hat{Y}_t$ to the solution Y_t

$$\epsilon \coloneqq \sup_{t<T} ||\hat{Y}_t - Y_t||_{L^2}. \tag{2.45}$$

More precisely, consider a family of approximations $\hat{Y}(h)$ parametrized by their stepsize h. The basic requirement is that the scheme should converge, that is to say $\epsilon(h) \to 0$ as $h \to 0$. We say that the scheme $\hat{Y}(h)$ converges with strong order γ at time T if there exist constants C and h_0, such that for all $h < h_0$ we have

$$\epsilon(h) \coloneqq \sup_{0 \le t \le T} ||\hat{Y}_t(h) - Y_t||_{L^2} \le C h^{\gamma}. \tag{2.46}$$

The strong order of convergence of a numerical scheme based on truncation of the stochastic Taylor series, or modifications thereof, is limited by the set of multiple iterated integrals included. By Radford's theorem and its extension to quasi-shuffle algebras (see Chapter 3.1.4 of this thesis), the iterated integrals have a basis comprising those indexed by Lyndon words. Our methods therefore cannot improve upon the strong order of convergence (see Ebrahimi–Fard et al. 2012). Indeed, Clark & Cameron (1980) showed that any numerical approximation of a drift-diffusion equation that is measurable with respect to the σ-algebra generated by a set of Wiener increments of fixed stepsize possesses a maximum generic order of convergence of one half. Order one convergence may be obtained when there is only one driving Wiener process, or certain commutativity conditions are satisfied by the driving vector fields, see Clark & Cameron (1980). Clark (1982) and Newton (1986, 1991) accordingly derived numerical integration schemes for one-dimensional drift-diffusion equations, depending on the driving Wiener process only through its values at discretization points, that minimize the mean square error among the class of such schemes in the following sense.

Definition 2.7.1 *A numerical scheme $\hat{Y}^0$ for an equation of the form stochastic differential equation driven by a one-dimensional Wiener process W_t is said to be asymptotically efficient if for all such schemes $\hat{Y}$ with N steps, and any given non-zero vector y,*

$$\liminf_N \frac{E\{(y^T(Y_T - \hat{Y}_N))^2 | W_h, \dots, W_{Nh}\}}{E\{(y^T(Y_T - \hat{Y}^0_N))^2 | W_h, \dots, W_{Nh}\}} \geq 1. \tag{2.47}$$

This characterizes the requirement that the leading order coefficient of the error ϵh, expressed as a power series in h, is minimal among all such schemes. Castell & Gaines (1996) showed that their integration scheme, based on the truncated exponential Lie series, is asymptotically efficient in the above sense. The requirement that the approximation be measurable with respect to the σ-algebra generated by the increments of the driving processes is too strong for our purposes. Indeed, it is only satisfied for schemes based on truncated Taylor expansions if the Lyndon basis for the set of iterated integrals included contains no words of length greater than one. Particularly, the integration schemes of Milstein (1974) and Wagner & Platen (1982) of higher order of convergence are excluded. Lord, Malham & Wiese (2008) and Malham & Wiese (2008) examined whether higher order Castell & Gaines schemes could be shown to satisfy the following weaker error optimality result.

Definition 2.7.2 *A numerical scheme for a stochastic differential equation driven by Lévy processes is said to be efficient if the local mean square error accrued at each timestep is less than the mean square error of the integration scheme of the same order of convergence arising from truncation of the stochastic Taylor series, for all* *driving vector fields in a prescribed class.*

A justification for minimising the local flow remainder to obtain optimality at the global level for stochastic integration scheme may be found in Malham & Wiese (2009). Malham & Wiese (2008) showed that the order $\frac{3}{2}$ Castell & Gaines scheme is efficient for all drift-diffusion equations driven by a single Wiener process, with smooth, autonomous vector fields. Lord, Malham & Wiese (2008) provided a counterexample showing that, in the absence of commutativity conditions, the order one Castell & Gaines scheme is not efficient for equations with multiple driving Wiener processes. Malham & Wiese (2009) derived an integration scheme for such equations and proved an efficiency result for their scheme. The main result of this thesis will be the derivation of an integration scheme obeying an efficiency result for a certain class of equations driven by Lévy processes.

Chapter 3

Integration Algebras

We now consider the properties of shuffle and quasi-shuffle algebras, recalling that they arise in the study of multiple iterated integrals. We begin with definitions and illustrative examples. We then define bialgebras and Hopf algebras and discuss the Hopf algebraic structures associated to shuffle and quasi-shuffle algebras, in particular giving Hoffman's isomorphisms of shuffle and quasi-shuffle Hopf algebras. We proceed to discuss the Lie theoretic aspects of the shuffle algebra, culminating in Radford's theorem that the Lyndon words form a basis for the shuffle algebra, and Hoffman's demonstration that they likewise form a basis of the quasi-shuffle algebra. For a fuller treatment of the above topics in relation to the shuffle algebra, see Reutenauer (1993). For further insights into the quasi-shuffle algebra, see the original papers Hoffman (2000) and Hoffman & Ihara (2012). The quasi-shuffle algebra is treated from a more abstract perspective in Loday (2007). There are several good textbooks treating Hopf algebras in general, see Sweedler (1969) and Abe (1980). For a more general discussion of the combinatorics of the free monoid and related structures, see Lothaire (1983). We close the chapter with a presentation of the results published in Curry et al. (2014) that the algebra generated by iterated integrals of Lévy processes with moments of all orders is isomorphic to a certain quasi-shuffle algebra.

3.1 Shuffle and Quasi-shuffle Algebras

We begin by giving the definitions of the shuffle and quasi-shuffle algebras. We follow by discussing how the duality between these products and the standard concatenation product $x \times y = xy$ of non-commuting indeterminates gives rise to rich algebraic structures known as Hopf algebras. The isomorphism between the shuffle and quasi-shuffle algebras is an isomorphisms of Hopf algebras; we give this isomorphism explicit, along with an isomorphism of the graded dual Hopf algebras. In the following section, we show that both shuffle and quasi-shuffle algebras possess a basis in terms of Lyndon words, this follows from a consideration of the Lie-theoretic properties of the shuffle algebra and we will give an illustrative sketch of the reasoning behind this.

3.1.1 Definitions

Let $\mathbb{A}^*$ be the free monoid on the set or 'alphabet' $\mathbb{A}$, comprising all words $w = a_1 \dots a_n$ with letters $a_i \in \mathbb{A}$. The product is concatentation of words: $a_1 \dots a_n \times a'_1 \dots a'_m = a_1 \dots a_n a'_1 \dots a'_m$. The identity for the concatenation product is the empty word, denoted by $\mathbb{1}$. For a given field K, we may extend this product K-linearly to the space $K\langle\mathbb{A}\rangle$ of noncommuting polynomials in elements of $\mathbb{A}$, with coefficients in K. The resulting algebra is the free associative K-algebra on $\mathbb{A}$.

Definition 3.1.1 *The shuffle product on $K\langle\mathbb{A}\rangle$ is defined inductively on words as follows.*

$$w \sqcup\!\sqcup \mathbb{1} = \mathbb{1} \sqcup\!\sqcup w = w, \tag{3.1}$$

where $w \in \mathbb{A}^*$, and for letters $a, b \in \mathbb{A}$ and words $u, v \in \mathbb{A}^*$,

$$au \sqcup\!\sqcup bv = a(u \sqcup\!\sqcup bv) + b(au \sqcup\!\sqcup v). \tag{3.2}$$

The K-module $K\langle\mathbb{A}\rangle$ then acquires a commutative algebra structure with the shuffle product, we will write $K\langle\mathbb{A}\rangle_{\sqcup\!\sqcup}$ for this algebra. An alternative characterization of the shuffle product is as follows. Let $w = a_1 \dots a_n$ be a word of length n. To each subset

$I = \{i_1 < \ldots < i_k\} \subset [n] := \{1, \ldots, n\}$ we associate a subword $w|I := a_{i_1} \ldots a_{i_k}$. We then define, for words $u_1, u_2 \in \mathbb{A}^*$ of lengths m and n respectively,

$$u_1 \sqcup\!\sqcup\, u_2 = \sum_{\substack{I_1 \cup I_2 = [m+n] \\ |I_1| = m,\, |I_2| = n}} w(I_1, I_2), \tag{3.3}$$

where $w(I_1, I_2)$ is the word of length $m + n$ uniquely determined by $w|I_j = u_j$, for $j = 1, 2$. Indeed, for any $k \leq m + n$, either $k \in I_1$ or $k \in I_2$. Let j be the index of the subset to which k belongs, and suppose that k is the lth element in I_j. The kth letter of the constructed word $w(I_1, I_2)$ is then the lth letter of u_j. This is sufficient to determine the word, as the above argument holds for all $k \leq m+n$, and hence determines each letter of $w(I_1, I_2)$.This characterization of the shuffle product corresponds to the intuitive idea of summing over all words made up of the totality of letters of u_1 and u_2, preserving the order of the letters of each word u_i in the larger word individually.

Example 3.1.2 *Let $\mathbb{A} = \{a, b, c\}$. We then have*

$$\begin{aligned} ab \sqcup\!\sqcup\, cb &= a(b \sqcup\!\sqcup\, cb) + c(ab \sqcup\!\sqcup\, b) \\ &= a(b(\mathbb{1} \sqcup\!\sqcup\, cb) + c(b \sqcup\!\sqcup\, b)) + c(a(b \sqcup\!\sqcup\, b) + b(ab \sqcup\!\sqcup\, \mathbb{1})) \\ &= a(bcb + c(b(\mathbb{1} \sqcup\!\sqcup\, b) + b(b \sqcup\!\sqcup\, \mathbb{1}))) + c(ab(b \sqcup\!\sqcup\, \mathbb{1}) + ab(\mathbb{1} \sqcup\!\sqcup\, b) + bab) \\ &= abcb + 2acbb + 2cabb + cbab. \end{aligned} \tag{3.4}$$

Alternatively, the decompositions of $\{1, 2, 3, 4\}$ as an intersection of two subsets containing two elements are

$$(1, 2) \cup (3, 4),\ (1, 3) \cup (2, 4),\ (1, 4) \cup (2, 3),\ (2, 3) \cup (1, 4),\ (2, 4) \cup (1, 3),\ (3, 4) \cup (1, 2).$$

For each of the above decompositions $\{1, 2, 3, 4\} = I_1 \cup I_2$, the words $w(I_1, I_2)$ are given respectively by

$$abcb,\ acbb,\ acbb,\ cabb,\ cabb,\ cbab.$$

Summing over the above, we obtain again $ab \sqcup\!\sqcup\, cb = abcb + 2acbb + 2cabb + cbab$.

Now let $K\mathbb{A}$ be the free K-module over $\mathbb{A}$, and suppose that $K\mathbb{A}$ possesses of itself a commutative K-algebra structure, with product denoted M.

Definition 3.1.3 *The quasi-shuffle product on $K\langle\mathbb{A}\rangle$ induced by the commutative product M on $K\mathbb{A}$ is a deformation of the shuffle product, defined inductively as follows.*

$$w * \mathbb{1} = \mathbb{1} * w = w, \tag{3.5}$$

for all $w \in \mathbb{A}^*$, and for letters $a, b \in \mathbb{A}$ and words $u, v \in \mathbb{A}^*$,

$$au * bv = a(u * bv) + b(au * v) + M(a,b)(u * v). \tag{3.6}$$

We may adapt the characterization of the shuffle product using partitions of $[n]$ to give an alternative, equivalent definition of the quasi-shuffle product if we remove the requirement that the partitions are non-intersecting. Precisely, we have for words $u = u_1 \ldots u_m$, $v = v_1 \ldots v_n$,

$$u * v = \sum_{l=m\vee n}^{m+n} \sum_{\substack{I \cup J = [l] \\ |I|=m,\, |J|=n}} \tilde{w}(I,J), \tag{3.7}$$

where $m \vee n$ denotes the maximum of m and n, and $\tilde{w}(I,J)$ is determined as follows.

$$\tilde{w}(I,J)|\{k\} = \begin{cases} u_{i_k} & k \in I \notin J \\ v_{j_k} & k \in J, \notin I \\ M(u_{i_k}, v_{j_k}) & k \in I, \in J. \end{cases} \tag{3.8}$$

Note that the quasi-shuffle definition recovers the shuffle product in the case where M is the trivial product $M(a,b) = 0$ for all a, b.

Example 3.1.4 *Let $\mathbb{A} = \{a_0, \ldots, a_n, \ldots,\} \cong \mathbb{Z}_+$, the group of non-negative integers under multiplication, and endow $\mathbb{R}\mathbb{A}$ with the commutative algebra structure induced by the group operation, ie. $a_i.a_j = a_{i+j}$. The induced quasi-shuffle algebra on $\mathbb{A}$ is sometimes known as the stuffle algebra. It arises, for instance in the study* *of multiple zeta values. Given a word $w = a_{i_1} \dots a_{i_l}$, we define a map $\zeta : U \to \mathbb{R}$ by*

$$\zeta(w) = \sum_{n_1 > n_2 > \dots n_l \geq 1} \frac{1}{n_1^{i_1} n_2^{i_2} \dots n_l^{i_l}}, \tag{3.9}$$

where U is the subset of $\mathbb{A}^*$ comprising words of non-zero length where the first letter $a_1 \geq 2$. The images $\zeta(a_{i_1} \dots a_{i_l})$, sometimes written $\zeta(i_1, \dots, i_l)$ are called multiple zeta values. The map ζ is a homomorphism for the above quasi-shuffle product (see Zudilin, 2003). For instance, the relation

$$a_2 * a_2 = 2a_2a_2 + a_4 \tag{3.10}$$

holds, from which we derive the following non-trivial identity concerning multiple zeta values

$$\zeta(2)^2 = 2\zeta(2, 2) + \zeta(4). \tag{3.11}$$

3.1.2 Shuffle Hopf Algebras

We have defined two products on words: the concatenation $\mathrm{conc}(u, v) = uv$ and the shuffle $\mathrm{sh}(u, v) = u \sqcup\!\sqcup v$. These products are compatible in a sense through duality. We will ultimately show that this compatibility may be understood in the framework of Hopf algebras. First, we must discuss the meaning of duality in this context.

Definition 3.1.5 *For a given polynomial $P \in K\langle\mathbb{A}\rangle$, write (P, w) for the coefficient of the word w in P. This bracketing extends as follows to give an inner product on $K\langle\mathbb{A}\rangle$.*

$$(P, Q) \coloneqq \sum_{w \in \mathbb{A}^*} (P, w)(Q, w). \tag{3.12}$$

Definition 3.1.6 *For a field K and alphabet $\mathbb{A}$, we then define $K\langle\langle\mathbb{A}\rangle\rangle$ to be the space of all formal series*

$$S = \sum_{w \in \mathbb{A}^*} (S, w)w. \tag{3.13}$$

We then define a product on $K\langle\langle\mathbb{A}\rangle\rangle$ by

$$(ST, w) = \sum_{w = uv} (S, u)(T, v), \tag{3.14}$$

this extends the concatenation product on $K\langle\mathbb{A}\rangle$*.*

The inner product on $K\langle\mathbb{A}\rangle$ then extends to a pairing $K\langle\langle\mathbb{A}\rangle\rangle \times K\langle\mathbb{A}\rangle \to K$ defined by

$$(S, P) = \sum_{w \in \mathbb{A}^*} (S, w)(P, w), \tag{3.15}$$

as the sum on the right is finite. Indeed, this pairing identifies $K\langle\langle\mathbb{A}\rangle\rangle$ with the dual space of $K\langle\mathbb{A}\rangle$. Note then that the elements of $\mathbb{A}^*$ form an orthonormal basis for this inner product. We then further identify $K\langle\mathbb{A}\rangle \otimes K\langle\mathbb{A}\rangle$ with a subspace of its dual using the unique inner product for which $\mathbb{A}^* \otimes \mathbb{A}^*$ is an orthonormal basis.

The space $K\langle\mathbb{A}\rangle$ has a natural graded algebra structure $K\langle\mathbb{A}\rangle = \bigoplus_n K\mathbb{A}_n$, where $K\mathbb{A}_n$ consist of the homogenous polynomials of degree n. A graded vector space $V = \bigoplus_n V_n$ admits a graded dual space defined by $V^o = \bigoplus V_n^*$, where V_n^* are the dual spaces of the graded components. The graded dual space is in general a subspace of the dual space. Considering the pairing defined above, we see that each of the graded components $K\mathbb{A}_n$ of $K\langle\mathbb{A}\rangle$ are self-dual; it follows that $K\langle\mathbb{A}\rangle$ is the graded dual of itself. To further explore the compatibility of shuffle and concatenate, we require the definition of a coalgebra, dual to the definition of an algebra. Let A be an unitary associative K-algebra with product $\mu : A \otimes A \to A$ and unit $\eta : K \to A$. The associative and unitary properties may be characterized by the commutativity of the following diagrams.

$$\begin{tikzcd}
A \otimes A \otimes A \arrow[r, "\mu \otimes 1"] \arrow[d, "1 \otimes \mu"'] & A \otimes A \arrow[d, "\mu"] \\
A \otimes A \arrow[r, "\mu"] & A
\end{tikzcd}
\qquad
\begin{tikzcd}
K \otimes A \arrow[r, "\eta \otimes 1"] \arrow[dr, "\cong"'] & A \otimes A \arrow[d, "\mu"] & A \otimes K \arrow[l, "1 \otimes \eta"'] \arrow[dl, "\cong"] \\
& A &
\end{tikzcd}$$

Following a standard procedure of category theory, we define the dual notion of a coalgebra by inverting the arrows in the diagrams above. Specifically, a K-module C equipped with coproduct $\Delta : C \to C \otimes C$ and counit $\epsilon : A \to K$ is said to be a counital coassociative K-coalgebra if the following diagrams commute.

$$\begin{tikzcd}
C \otimes C \otimes C & & C \otimes C \arrow[ll, "\Delta \otimes 1"'] \\
C \otimes C \arrow[u, "1 \otimes \Delta"] & & C \arrow[ll, "\Delta"'] \arrow[u, "\Delta"']
\end{tikzcd}
\qquad
\begin{tikzcd}
K \otimes C & C \otimes C \arrow[l, "\epsilon \otimes 1"'] \arrow[r, "1 \otimes \epsilon"] & C \otimes K \\
& C \arrow[ul, "\cong"] \arrow[u, "\Delta"] \arrow[ur, "\cong"'] &
\end{tikzcd}$$

In what follows, we will show both the concatenation and shuffle algebras admit dual coalgebras. Moreover, for each of these algebra structures, we observe a compatibility relation between itself and the coalgebra dual to the other algebra.

Definition 3.1.7 *Define $\delta : K\langle \mathbb{A} \rangle \to K\langle \mathbb{A} \rangle \otimes K\langle \mathbb{A} \rangle$ to be the unique concatenation homomorphism determined by $\delta : a \mapsto a \otimes \mathbb{1} + \mathbb{1} \otimes a$ for all $a \in \mathbb{A}$.*

We now compute the action of δ on a general word:

$$\begin{aligned} \delta(a_1 \dots a_n) &= \delta(a_1) \dots \delta(a_n) && (3.16) \\ &= (a_1 \otimes \mathbb{1} + \mathbb{1} \otimes a_1) \dots (a_n \otimes \mathbb{1} + \mathbb{1} \otimes a_n). && (3.17) \end{aligned}$$

Now note that the above product is a sum over tensor products of words $u \otimes v$, where each a_i appears either in u or v, and u or v each comprise a subword of $a_1 \dots a_n$. In other words, we have

$$\delta(w = a_1 \dots a_n) = \sum_{\substack{I_1 \cup I_2 = [n] \\ I_1 \cap I_2 = \emptyset}} w|I_1 \otimes w|I_2. \tag{3.18}$$

Recalling the definition of the inner product on $K\langle \mathbb{A} \rangle$, we then interpret our previous characterization of δ as follows:

$$\delta(w) = \sum_{u, v \in \mathbb{A}^*} (w, u \sqcup\!\sqcup\, v) u \otimes v. \tag{3.19}$$

Any linear map $\mu : V \to W$ induces a transpose $\mu^* : W^* \to V^*$ defined by $\mu^*(f) = f \circ \mu$. It follows that the shuffle product $\mathrm{sh} : K\langle \mathbb{A} \rangle \otimes K\langle \mathbb{A} \rangle \to K\langle \mathbb{A} \rangle$ induces a transpose

$$\mathrm{sh}^* : K\langle \mathbb{A} \rangle^* \to (K\langle \mathbb{A} \rangle \otimes K\langle \mathbb{A} \rangle)^*. \tag{3.20}$$

Furthermore, the restriction of sh^* to $K\langle\mathbb{A}\rangle \subset K\langle\mathbb{A}\rangle^*$ under the inner product above is mapped $K\langle\mathbb{A}\rangle \otimes K\langle\mathbb{A}\rangle$, the identification of the latter with a subset of the dual space performed as indicated above. The resulting restriction $\Delta : K\langle\mathbb{A}\rangle \mapsto K\langle\mathbb{A}\rangle \otimes K\langle\mathbb{A}\rangle$ obeys the duality relation

$$(\mathrm{sh}(u \otimes v), w) = (u \otimes v, \Delta(w)). \tag{3.21}$$

Comparing this relation to the characterization (3.19) of the map δ, we see that the two maps δ and Δ are the same. We may hence regard δ as a coproduct dual to the shuffle product and will henceforth refer to it as the de-shuffle coproduct.

Definition 3.1.8 *The de-concatenation coproduct δ' on $K\langle\mathbb{A}\rangle$ is defined by*

$$\delta'(w) = \sum_{u,v \in \mathbb{A}^*} (w, uv) u \otimes v. \tag{3.22}$$

It may be shown (see Chen 1968, Theorem 1.8 or Reutenauer 1993, Proposition 1.9) that δ' is the unique shuffle homomorphism sending $a \mapsto a \otimes \mathbb{1} + \mathbb{1} \otimes a$ for letters $a \in \mathbb{A}$.

Definition 3.1.9 (Bialgebra) *A K-module possessing simultaneoulsy an algebra structure with product μ and unit η, and a coalgebra structure with coproduct δ and counit η is defined to be a bialgebra if either of the following equivalent conditions hold:*

1. *μ and η are coalgebra morphisms,*

2. *Δ and ϵ are algebra morphisms.*

Define unit and counit maps on $K\langle\mathbb{A}\rangle$ by

$$\eta : \left\{ \begin{array}{ccc} K & \to & K\langle\mathbb{A}\rangle \\ k & \mapsto & k.\mathbb{1} \end{array} \right. , \qquad \epsilon : \left\{ \begin{array}{ccc} K\langle\mathbb{A}\rangle & \to & K \\ P & \mapsto & (P, \mathbb{1}) \end{array} \right. . \tag{3.23}$$

The above duality results relating the concatenation and shuffle products and their dual coproducts δ, δ' ultimately imply the following, see Reutenauer (1993, Proposition 1.9).

Theorem 3.1.10 *The space of non-commuting polynomials $K\langle\mathbb{A}\rangle$ possesses two bialgebra structures: the first with concatenation as product, δ as coproduct; for the second we have shuffle as product and δ' as coproduct. In both cases, η is the unit and ϵ the co-unit.*

Definition 3.1.11 *Given a k-algebra (A, μ) and k-coalgebra (C, Δ), we may define a convolution algebra on the space of k-module morphisms $C \to A$ by*

$$f \star g \coloneqq \mu \circ (f \otimes g) \circ \Delta. \tag{3.24}$$

For any product μ and coproduct Δ on the space $K\langle\mathbb{A}\rangle$, the space $\mathrm{End}(K\langle\mathbb{A}\rangle)$ of K-module endomorphisms thus inherits a convolution algebra structure.

Define the product algebra $\mathcal{A} \coloneqq K\langle\mathbb{A}\rangle\bar{\otimes}K\langle\mathbb{A}\rangle_{\sqcup\!\sqcup}$, where the product is concatenation on the left and shuffle on the right, and $\bar{\otimes}$ denotes the completion with respect to the inner product previously defined on $K\langle\mathbb{A}\rangle\bar{\otimes}K\langle\mathbb{A}\rangle$. The following result plays an important role in the algebraic encoding of integration schemes.

Lemma 3.1.12 *We may canonically embed the space $\mathrm{End}(K\langle\mathbb{A}\rangle) \to \mathcal{A}$ under the mapping $\psi : f \mapsto \sum_w w \otimes f(w)$. Furthermore this embedding is a homomorphism when we endow the space $\mathrm{End}(K\langle\mathbb{A}\rangle)$ with the convolution product induced by* sh *and δ'.*

Proof: This result is given in Reutenauer (1993, p. 29) for the other convolution structure with concatenation product and de-shuffle coproduct. The proof given here shows that both results are special cases of a more general theorem valid for a general convolution algebra. For a given product μ on $K\langle\mathbb{A}\rangle$, if its dual coproduct μ^* may be restricted to a mapping $K\langle\mathbb{A}\rangle \to K\langle\mathbb{A}\rangle \otimes K\langle\mathbb{A}\rangle$, we obtain the duality relation

$$(\mu(p \otimes q), r) = (p \otimes q, \mu^*(r)). \tag{3.25}$$

Let μ be such a product, and ν a coproduct defined on $K\langle\mathbb{A}\rangle$. It follows that, for given $f, g \in \mathrm{End}(K\langle\mathbb{A}\rangle)$ and associated $\sum u \otimes f(u)$, $\sum v \otimes g(v) \in K\langle\mathbb{A}\rangle_\mu\bar{\otimes}K\langle\mathbb{A}\rangle_\nu$, the product of the above series is given by

$$\sum_{u,v} \mu(u \otimes v) \otimes \nu(f \otimes g)(u \otimes v) = \sum_{u,v,w} (\mu(u \otimes v), w) w \otimes \nu(f \otimes g)(u \otimes v) \tag{3.26}$$

$$= \sum_{u,v,w} w \otimes (u \otimes v, \mu^*(w))\nu(f \otimes g)(u \otimes v) \tag{3.27}$$

$$= \sum_{w} w \otimes v \circ (f \otimes g) \circ \mu^*(w). \tag{3.28}$$

The result as stated follows a fortiori.

We conclude the section by stating the main theorem concerning the Hopf algebraic properties of the shuffle algebra and related structures.

Definition 3.1.13 (Hopf algebra) *A bialgebra is defined to be a Hopf algebra if the identity mapping of its convolution algebra admits a convolution inverse.*

The following result is due to Chen (1968), see also Reutenauer (1993).

Theorem 3.1.14 *There exist two bialgebra structures on $K\langle\mathbb{A}\rangle$, one with concatenation as product and δ as coproduct, the other with shuffle as product and δ' as coproduct. In both cases the unit is $\eta : k \mapsto k.\mathbb{1}$ and the counit $\epsilon : P \mapsto (P, \mathbb{1})$. The mapping $S : a_1 \dots a_n \mapsto (-1)^n a_n \dots a_1$ is the inverse of the identity mapping* $\mathrm{id} = \eta \circ \epsilon$ *in both associated convolution algebras. In particularly, both bialgebra structures are Hopf algebras with antipode S.*

3.1.3 Quasi-shuffle Hopf Algebras

Let $K\langle\mathbb{A}\rangle_*$ be the quasi-shuffle algebra induced by the commutative binary operation M, defined on the K-linear span of $\mathbb{A}$. It may be shown that this too possesses a Hopf algebraic structure when equipped with the de-concatenation coproduct δ'. The proof is by construction of a Hopf algebra isomorphism to the shuffle Hopf algebra with deconcatenation coproduct. First, we require a description of a class of endomorphisms of noncommuting polynomials introduced by Hoffman (2000). For a word $w = a_1 \dots a_n$, we write

$$[w] = M(a_1, M(a_2, \dots, M(a_{n-1}, a_n) \dots)). \tag{3.29}$$

Definition 3.1.15 *To a given formal power series with zero constant term $f = c_1 t + c_2 t^2 + \dots$ we associate a map $\tilde{f} : w \mapsto c_{|w|}[w]$. We then define a mapping Ψ from the space of formal series to endomorphisms of noncommuting polynomials by*

$$\Psi(f)(w) = \sum_{k=1}^{|w|} \tilde{f}^{\odot k}(w), \tag{3.30}$$

where $\odot$ is the convolution product of Hoffman & Ihara (2012, Section 4) defined by $(f \odot g)(w) = \sum_{w=uv} f(u)g(v)$, that is the convolution arising from the concatenation product and de-concatenation coproduct.

This construction is equivalent to that introduced in Hoffman (2000) and hence Ψ is an homomorphism for the composition structures of the space of formal power series and $\mathrm{End}(K\langle A\rangle)$ by Hoffman & Ihara (2012, Theorem 3.1).

Definition 3.1.16 *The Hoffman exponential and logarithm maps are defined to be the endomorphisms*

$$\exp_H = \Psi(e^t - 1) \quad \text{and} \quad \log_H = \Psi(\log(1+t)). \tag{3.31}$$

Example 3.1.17 *As an illustration, we compute a Hoffman mapping in the stuffle algebra defined in Example 3.1.4. For any word $w = a_{i_1} \dots a_{i_l}$, we have $[w] = a_k$, where $k = \sum_j i_j$. We compute*

$$\log_H(a_1 a_2 a_5) = \tilde{f}(a_1 a_2 a_5) + \tilde{f}(a_1)\tilde{f}(a_2 a_5) + \tilde{f}(a_1 a_2)\tilde{f}(a_5) + \tilde{f}(a_1)\tilde{f}(a_2)\tilde{f}(a_5). \tag{3.32}$$

From the series $\log(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \dots$ we obtain the coefficients $c(1) = 1, c(2) = -\frac{1}{2}, c(3) = \frac{1}{3}$. It follows that

$$\log_H(a_1 a_2 a_5) = \frac{1}{3} a_8 - \frac{1}{2}(a_1 a_7 + a_3 a_5) + a_1 a_2 a_5. \tag{3.33}$$

We are in a position to state the following, obtained in the setting of graded Hopf algebras in Hoffman (2000), and in the non-graded case in Hoffman & Ihara (2012).

Theorem 3.1.18 *Let $K\langle\mathbb{A}\rangle_*$ be the quasi-shuffle algebra induced by the commutative binary operation M, defined on the K-linear span of $\mathbb{A}$. When equipped with the deconcatenation product δ, unit $\eta : k \mapsto k.\mathbb{1}$ and counit $\epsilon : P \mapsto (P, \mathbb{1})$, $K\langle\mathbb{A}\rangle_*$ inherits a Hopf algebra structure. Furthermore, the Hoffman map $\log_H$ is a Hopf algebra isomorphism from this Hopf algebra to the shuffle Hopf algebra with deconcatenation coproduct. Its inverse is given by $\exp_H$.*

The K-module $K\langle\mathbb{A}\rangle$ possesses two Hopf algebraic structures relating to the shuffle product: one with shuffle as product and de-concatenate as coproduct, the other with concatenate as product and de-shuffle as coproduct. These are related by duality: precisely, when $\mathbb{A}$ is a finite set, the latter Hopf algebra is the graded dual of the first, when $K\langle\mathbb{A}\rangle$ is equipped with the natural grading structure arising from the degree of homogenous polynomials, see Hoffman (2000). If the quasi-shuffle algebra $K\langle\mathbb{A}\rangle_*$ possesses a graded algebra structure, the Hoffman isomorphisms are isomorphisms of graded Hopf algebras. The graded dual of $K\langle\mathbb{A}\rangle_*$ therefore exists and is a Hopf algebra isomorphic to the concatenation Hopf algebra with shuffle coproduct. We now give this structure explicitly.

Definition 3.1.19 *For a quasi-shuffle algebra $K\langle\mathbb{A}\rangle$ over a finite alphabet $\mathbb{A}$, define the de-quasi-shuffle coproduct to be the unique concatenation homomorphism acting on letters as*

$$\delta'' : a \mapsto a \otimes \mathbb{1} + \mathbb{1} \otimes a + \sum_{\substack{u,v\in\mathbb{A} \\ M(u,v)=a}} u \otimes v. \tag{3.34}$$

The restriction to finite alphabets ensures that the sum appearing in the above definition is finite. We obtain the following, see Hoffman (2000).

Theorem 3.1.20 *Suppose $K\langle\mathbb{A}\rangle = \bigoplus K\langle\mathbb{A}\rangle_n$ possesses a grading structure, that $\mathbb{A}$ is a finite alphabet and that $K\mathbb{A}$ possesses a commutative algebra structure with grade-preserving product M. Then $K\langle\mathbb{A}\rangle$ possesses a Hopf algebra structure with* *concatenation as product, de-quasi-shuffle as coproduct, unit $\eta : k \mapsto k.\mathbb{1}$ and counit $\epsilon : P \mapsto (P, \mathbb{1})$. Moreover, there is a graded Hopf algebra isomorphism from this space to the concatenation Hopf algebra with de-shuffle coproduct, graded by word length.*

This is a direct consequence of the existence of the Hoffman maps as graded Hopf algebra isomorphisms and the relationship of the shuffle and concatenate Hopf algebras as graded duals of each other. The resulting dual isomorphisms are given by

$$\exp_H^*(a) = \sum_{n\geq 1} \frac{1}{n!} \sum_{M(a_1, M(\ldots, M(a_n)\ldots) = a} a_1 \ldots a_n, \tag{3.35}$$

$$\log_H^*(a) = \sum_{n\geq 1} \frac{(-1)^{n-1}}{n} \sum_{M(a_1, M(\ldots, M(a_n)\ldots) = a} a_1 \ldots a_n. \tag{3.36}$$

Note that, as $\exp_H : K\langle\mathbb{A}\rangle_{ш} \to K\langle\mathbb{A}\rangle_*$, we have $\exp_H^* : (K\langle\mathbb{A}\rangle_*)^* \to (K\langle\mathbb{A}\rangle_{ш})^*$, and similarly $\log_H^* : (K\langle\mathbb{A}\rangle_{ш})^* \to (K\langle\mathbb{A}\rangle_*)^*$.

We conclude our discussions of the Hopf algebraic properties of quasi-shuffle algebras by considering the form of the antipode, that is the inverse of the identity in the associated convolution algebra. We require this as the convolution algebra takes an important role in our analysis. There are two convolution algebras to consider; we here discuss that arising from the quasi-shuffle product and deconcatenation coproduct. We begin by defining two maps: the reversing endomorphism $R : a_1 \ldots a_n \mapsto a_n a_{n-1} \ldots a_1$, and the sign map $T = \Psi(-t) : w \mapsto (-1)^{|w|} w$. The antipode given previously for the shuffle Hopf algebra and its graded dual then decompose as $S = TR$. It was noted in Hoffman & Ihara (2012) that the reversing map R is an algebra automorphism for both the shuffle and quasi-shuffle algebras. We can say more:

Lemma 3.1.21 *Let $K\langle\mathbb{A}\rangle_*$ be a quasi-shuffle algebra. The reversing map R commutes with all maps $\Psi(f)$, where f is a power series.*

This follows from the commutativity of $[,]$. Note particularly that R then commutes with both $\exp_H$ and $\log_H$. Now define the map $\Sigma = \Psi\left(\frac{t}{1-t}\right)$, noting that $\frac{t}{1-t} =$

$t+t^2+t^3+\ldots$. The homomorphism property of Ψ implies that $\Sigma = \exp_H \circ T \circ \log_H \circ T$ (see Hoffman & Ihara 2012, Corollary 3.2). Let $\hat{S}$ denote the antipode for the convolution algebra with quasi-shuffle product and de-concatenate coproduct. Note the identity $\hat{S} = \exp_H \circ S \circ \log_H = \exp_H T \log_H R$. As R and $\log_H$ commute, we obtain the following result of Hoffman (2000), noting that $T^2 = \mathrm{id}$.

Lemma 3.1.22 *Let $K\langle\mathbb{A}\rangle_*$ be a quasi-shuffle algebra. The Hopf algebraic properties guarantee the existence of an antipode for the convolution algebra with quasi-shuffle product and de-concatenate coproduct. The antipode $\hat{S}$ then takes the form*

$$\hat{S} = \Sigma T R. \tag{3.37}$$

When the convolution algebra with concatenate as product and de-quasi-shuffle as coproduct may be defined and possesses an antipode, it does not take the form given above. We will not use this antipode, its form may be found in Hoffman (1999, p. 9). Note the contrast with the shuffle case, where the antipodes for the two convolution algebra structures considered are the same.

3.1.4 Lie Theoretic Aspects and Bases

We now examine the relation of the shuffle algebra to the free Lie algebra. In doing so, we will show that the shuffle algebra and quasi-shuffle algebra both admit a basis of Lyndon words.

Definition 3.1.23 *To any polynomials $P, Q \in K\langle A\rangle$, we may define a corresponding Lie bracket $[P, Q] = PQ - QP$, where PQ is the concatenation of P and Q. We say a given polynomial is a Lie polynomial if it is an element of $\mathcal{L}(A)$, the submodule of $K\langle A\rangle$ generated from A by the bracket $[,]$.*

It has been shown (see Lothaire, 1981) that $\mathcal{L}(A)$ is the free Lie algebra on the set A. The following results are foundational in the theory of free Lie algebras (see Lothaire, 1981 or Reutenauer, 1993).

Lemma 3.1.24 (Friedrichs' criterion) *A polynomial $P \in K\langle\mathbb{A}\rangle$ is a Lie polynomial if and only if it is a primitive element for the de-shuffle coproduct δ, that is*

$$\delta(P) = P \otimes \mathbb{1} + \mathbb{1} \otimes P. \tag{3.38}$$

In defining the Lyndon words, we will require two important notions from the theory of free monoids. A word $w \in \mathbb{A}^*$ is said to be primitive if it is not a concatenation power of another word. Two words $x, y \in \mathbb{A}^*$ are called conjugate if there exist $u, v \in \mathbb{A}^*$ such that $x = uv$ and $y = vu$. We now define the lexicographical order, a means of extending an order on a set $\mathbb{A}$ to the free monoid $\mathbb{A}^*$.

Definition 3.1.25 *Given an existing total order $<$ on an alphabet $\mathbb{A}$, the lexicographical order on $\mathbb{A}^*$ extends this to give an order of $\mathbb{A}^*$ such that given $u, v \in \mathbb{A}^*$,*

1. *$u < v$ if $v \in u\mathbb{A}^+$,*
2. *$u < v$ if (for some $a < b \in \mathbb{A}$; $l, p, q \in \mathbb{A}^*$)* $\left\{ \begin{array}{c} u = lap \\ v = lbq \end{array} \right\}$,

where $\mathbb{A}^+$ is the subset of $\mathbb{A}^*$ comprising non-empty words, and $u\mathbb{A}^+$ is the left-coset by u of $\mathbb{A}^+$, that is the space of all words of length two or greater with first letter u.

We are now in a position to give the following definition:

Definition 3.1.26 (Lyndon words) *Given an ordered alphabet $\mathbb{A}$, a word $w \in \mathbb{A}^*$ is a Lyndon word if it is primitive and minimal with respect to its conjugacy class, where the minimality is with respect to the lexicographical order on $\mathbb{A}^*$.*

The Lyndon words are a particular case of Hall words (see Reutenauer 1993, Chapter 4), and thus form a basis of the free Lie algebra. Furthermore, any word w has a non-increasing factorization into Lyndon words $w = l_1 \dots l_m$ (see Reutenauer 1993). For any Lyndon word l we may define an associated polynomial P_l. Explicitly, the Lyndon words of length greater than one are characterized by the possession of a factorization $l = uv$, where u, v are both Lyndon words, and $u < v$. If we take v to be the maximal Lyndon word for this construction, this factorization is called the standard factorization, and coincides with the notion of standard factorization of Hall words (see Reutenauer, Chapter 5.1).

Definition 3.1.27 *To a given Lyndon word l with standard factorization $l = uv$, we then define inductively the associated Hall polynomials*

$$P_l = [P_u, P_v]. \tag{3.39}$$

For a general word with non-increasing Lyndon factorization $w = l_1 \dots l_m$, we define $P_w = P_{l_1} \dots P_{l_m}$.

The following is a corollary of a more general result concerning Hall polynomials (see Reutenauer 1993, Theorem 4.9):

Lemma 3.1.28 *The polynomials P_w, indexed by $w \in \mathbb{A}^*$, defined above form a basis for the free associative K-algebra $K\langle\mathbb{A}\rangle$.*

To any basis $(P_w)_{w\in\mathbb{A}^*}$ of the algebra $K\langle\mathbb{A}\rangle$ is associated a dual basis $(S_w)_{w\in\mathbb{A}^*}$ of the space of formal series $K\langle\langle\mathbb{A}\rangle\rangle$, canonically isomorphic to the dual space $K\langle\mathbb{A}\rangle^*$, this is defined such that, for any word $u \in \mathbb{A}^*$, we have

$$u = \sum_{w\in\mathbb{A}^*} (S_w, u) P_w. \tag{3.40}$$

The following result is a corollary of a theorem of Schützenberger (see Schützenberger 1958, Reutenauer 1993, Theorem 5.3) pertaining to the dual basis associated to Hall polynomials, first noted in the context of Lyndon words by Melancon and Reutenauer (1989).

Lemma 3.1.29 *Let $(S_w)_{w\in\mathbb{A}^*}$ be the basis of $K\langle\langle\mathbb{A}\rangle\rangle$ dual to the basis $(P_w)_{w\in\mathbb{A}^*}$ of Hall polynomials associated to the Lyndon words. The following results provide an inductive method of computing the series S_w.*

1. *$S_1 = \mathbb{1}$.*

2. *Let l by a Lyndon word with first letter a, ie $l = aw$, where $w \in A^*$. Then $S_l = aS_w$.*

3. *Let $w = l_1^{i_1} \dots l_k^{i_k}$ be a decreasing product of powers of distinct Lyndon words.*

We then have

$$S_w = \frac{1}{i_1! \ldots i_k!} S_{l_1}^{\sqcup\!\sqcup i_1} \sqcup\!\sqcup \ldots \sqcup\!\sqcup S_{l_k}^{\sqcup\!\sqcup l_k}. \tag{3.41}$$

We do not give a complete proof, the second part of the above lemma particularly is rather technical. Instead, we show how the third part arises naturally from Friedrichs' criterion, hence the focus on Lie theoretic elements. Write $w = w_1 \ldots w_i$, where the w_j are non-increasing, not necessarily distinct Lyndon words. By duality, we have

$$(S_{w_1} \sqcup\!\sqcup \ldots \sqcup\!\sqcup S_{w_i}, P_u) = (S_{w_1} \otimes \ldots \otimes S_{w_i}, \delta_i(P_u)), \tag{3.42}$$

where δ_i is the i-fold de-shuffle, the unique concatenation homomorphism $K\langle\langle \mathbb{A} \rangle\rangle \to K\langle \mathbb{A} \rangle^{\bar{\otimes} i}$ sending a letter $a \in A \mapsto a \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} + \mathbb{1} \otimes a \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} + \mathbb{1} \otimes \ldots \otimes \mathbb{1} \otimes a$ (see Reutenauer, p. 25). A mild extension of Friedrichs' criterion shows that Lie polynomials are primitive elements for the i-fold de-shuffle, i.e.

$$\delta_i(P) = (P \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1})(\mathbb{1} \otimes P \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}) \ldots (\mathbb{1} \otimes \ldots \otimes \mathbb{1} \otimes P). \tag{3.43}$$

Writing $u = u_1 \ldots u_n$ as a non-increasing product of Lyndon words, we have $P_u = P_{u_1} \ldots P_{u_n}$, where each P_{u_j} is a Lie polynomial. It follows that

$$\delta_i(P_u) = \prod_{k=1}^{n} (P_{u_k} \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}) \ldots (\mathbb{1} \otimes \ldots \otimes \mathbb{1} \otimes P_{u_k}). \tag{3.44}$$

As we have $(S_w, P_u) = \delta_{w,u}$, the required result follows shortly. The distinguishing property of Lyndon words amongst the various Hall sets is the following triangularity property (see Reutenauer, Theorem 5.1):

Lemma 3.1.30 *Let w be factored as a non-increasing product of Lyndon words, with (P_w) the associated basis of Hall polynomials, and (S_w) the dual basis. We have*

1. $P_w = w+$ *greater words,*
2. $S_w = w+$ *smaller words.*

Theorem 3.1.31 (Radford) *For any given total order on the alphabet $\mathbb{A}$, the set of associated Lyndon words form a basis for the shuffle algebra $K\langle\mathbb{A}\rangle_{\sqcup\!\sqcup}$.*

Combining Lemmas 3.1.29 and 3.1.30, we have for words with decreasing Lyndon factorization $w = l_1^{i_1} \dots l_k^{i_k}$,

$$\frac{1}{i_1! \dots i_k!} l_1^{\sqcup\!\sqcup\, i_1} \sqcup\!\sqcup \dots \sqcup\!\sqcup\, l_k^{\sqcup\!\sqcup\, i_k} = w + \sum_{u<w} \alpha_u u. \tag{3.45}$$

The result then follows immediately.

Corollary 3.1.32 (Hoffman) *For any given total order on the alphabet $\mathbb{A}$, the set of associated Lyndon words form a basis for the quasi-shuffle algebra $K\langle\mathbb{A}\rangle_*$ induced by a given commutative product M.*

This was already noted by Li & Liu (1997) for a specific case of quasi-shuffle product; it follows by an inductive rewriting of the basis $\{\exp_H(l)\}$, where l are Lyndon words. The key point is that the quasi-shuffle product has a 'triangularity property' of its own,

$$u * v = u \sqcup\!\sqcup\, v + \sum \text{words of smaller length.} \tag{3.46}$$

3.2 Quasi-shuffle algebra of iterated Lévy integrals

The shuffle algebra arises naturally from the consideration of iterated path integrals (see Ree, 1958), this was shown to extend to the case of iterated Stratonovich integrals of Brownian paths in Gaines (1994). In Curry et al (2014), it was shown that the iterated integrals of independent Lévy processes $\{Z^0(t) = t, Z^1, \dots Z^d\}$ possessing moments of all orders,

$$I_{a_1 \dots a_n} = \int_{0<\tau_{n-1}<\dots<\tau_1<t} dZ_{\tau_n}^{a_1} \dots dZ_{\tau_1}^{a_n} \tag{3.47}$$

generate a quasi-shuffle algebra where the underlying commutative algebra structure corresponds to the quadratic covariation algebra. More precisely, there exists a quasi-shuffle algebra $K\langle\tilde{\mathbb{A}}\rangle_*$ generated by an alphabet $\tilde{\mathbb{A}}$ containing the set $\mathbb{A} = \{0, 1, \ldots, d\}$ and commutative product $[,]$ defined on the free K-module $K\tilde{\mathbb{A}}$; furthermore, there exists a word-to-integral map $w \mapsto I_w$ which is an algebra isomorphism from the quasi-shuffle algebra $K\langle\tilde{\mathbb{A}}\rangle_*$ to the algebra of iterated integrals generated by the Lévy processes $\{Z^0, \ldots, Z^d\}$. The Itô integration-by-parts rule

$$\int_0^t dX_s \int_0^t dY_s = \int_0^t \int_0^\tau dX_\tau dY_s + \int_0^t \int_0^\tau dY_\tau dX_s + \int_0^t d[X,Y]_s \tag{3.48}$$

implies that the commutative product $[,]$ on $K\tilde{\mathbb{A}}$ must correspond to the quadratic covariation bracket. The Lévy processes $\{Z^0, \ldots, Z^n\}$ are not in general closed under the quadratic covariation bracket, we must therefore augment the alphabet A and extend the word-to-integral map to obtain a quasi-shuffle isomorphism. All the results in this section are from Curry et al. (2014), although in the concluding comments we will present the additional observation that the compensated power bracket processes may be taken to replace the power bracket processes in the quasi-shuffle algebra. Following Nualart & Schoutens (2000) and Jamshidian (2005), we accordingly consider the iterated quadratic covariation brackets of a semimartingale.

Definition 3.2.1 (Power bracket) *To any semimartingale X, we associate the power bracket processes $[X]^{(n)}$ defined inductively by $[X]^{(1)} = X$ and $[X]^{(n)} = [X^{(n-1)}, X]$.*

Important and related are the Teugels martingales defined by Nualart & Schoutens (2000).

Definition 3.2.2 (Teugels martingale) *For a given Lévy process Z possessing moments of all orders and Lévy decomposition*

$$Z_t = \alpha t + \sigma W_t + \int_0^t \int_{\mathbb{R}} x(Q(ds,dx) - ds\nu(dx)), \tag{3.49}$$

we define the associated Teugels martingales by

$$Y^{(n)} = \int_0^t \int_{\mathbb{R}} x^n(Q(ds,dx) - s\nu(dx)). \tag{3.50}$$

Note that the Lévy decomposition incorporates an integral with respect to the compensated random measure $Q(ds,dx) - ds\nu(dx)$, see Section 2.5 for the random measures appearing in the Lévy decomposition and Section 2.6 for integration with respect to compensated random measures. The definition of the Teugels martingales given above differs slightly from that of Nualart & Schoutens (2000), they are equivalent by the following lemma. Also note that Nualart & Schoutens assume the stronger condition that the Lévy process possess an exponential moment. This is to ensure the density of polynomials of the Lévy increments in the space of square-integrable random variables, measurable with respect to the completion of the filtration generated by the Lévy processes. This result is required only for the results of Nualart & Schoutens (2000) concerning martingale representation. We accordingly require only the weaker assumption that the Lévy process possess moments of all orders. The relation between the power bracket processes and Teugels martingales of a Lévy process as defined above is given in Curry et al. (2014).

Lemma 3.2.3 *For a given real-valued Lévy process Z_t possessing moments of all orders, the power bracket processes $[Z]^{(n)}$ with $n \geq 2$ are given by*

$$[Z]^{(n)}_t = 1_{\{n=2\}}\sigma^2 t + \int_0^t \int_{\mathbb{R}} x^n Q(ds,dx). \tag{3.51}$$

In particular, the power brackets and Teugels martingales associated to Z are related by

$$[Z]^{(n)}_t = (1_{\{n=2\}}\sigma^2 + \lambda_n)t + Y^{(n)}_t, \tag{3.52}$$

where $\lambda_n \coloneqq \int_{\mathbb{R}} x^n \nu(dx)$.

Proof: For general semimartingales X, Y with continuous part X^c, Y^c, we have

$$[X,Y]_t = \langle X^c, Y^c \rangle_t + \sum_{s \leq t} (\Delta X_s)(\Delta Y_s), \tag{3.53}$$

this is Theorem 1.4.52 of Jacod & Shiryaev (1987). Using the Lévy decomposition, we obtain

$$[Z]^{(2)}_t = \langle \sigma W, \sigma W \rangle_t + \sum_{s \leq t} (\Delta Z_s)^2. \tag{3.54}$$

It is standard result that $\langle W, W\rangle_t = t$, see eg Ikeda & Watanabe (1989, Theorem II.6.1). Furthermore, for all measurable sets Λ and functions f such that the integral is defined, we have

$$\int_{\Lambda} f(x)Q(dx, ds) = \sum_{s\leq t} f(\Delta Z_s)1_{\Lambda}(\Delta Z_s), \tag{3.55}$$

see Protter (2003). As by hypothesis X possesses moments of all orders, the above integral is defined for all $f(x) = x^m$ where $m \geq 2$ is an integer; the first part of the result then follows for the case $n = 2$. The first part of the result follows by induction using the relation

$$[[Z]^{(n-1)}, Z] = \sum_{s\leq t}(\Delta Z_s)^{n-1}(\Delta Z_s), \tag{3.56}$$

again derived from Theorem 1.4.52 of Jacod & Shiryaev. The second part follows from the relation

$$\int_0^t \int_{\mathbb{R}} x^n(Q(ds, dx) - ds\nu(dx)) = \int_0^t \int_{\mathbb{R}} x^n Q(ds, dx) - t\int_{\mathbb{R}} x^n \nu(dx), \tag{3.57}$$

which holds by definition when the integrals on the right hand side above exist (see the discussion of integration with respect to point processes in Section 2.7 of this thesis, or Chapter II. 3 of Ikeda & Watanabe, 1989). This is guaranteed by the hypothesis that Z possesses moments of all orders.

The idea is to extend the alphabet $\tilde{\mathbb{A}}$ by incorporating new letters corresponding to each power bracket process that is not a linear combination of processes with corresponding letters already in the alphabet. To obtain an algebra isomorphism to a quasi-shuffle algebra, we will require that such processes are not linear combinations of multiple iterated integrals with respect to processes corresponding to letters in the alphabet. The following analysis, taken from Curry et al. (2014) shows that such relations do not exist. In doing so, we will require the notion of strong orthogonality of locally square integrable martingales. Recall that, for a locally square integrable martingale M, we may define the angle bracket $\langle M\rangle$ to be the unique predictable increasing process such that $M^2 - \langle M \rangle$ is a local martingale; this extends by polarization: $\langle M, N \rangle := \frac{1}{4}(\langle M+N \rangle - \langle M - N \rangle)$, see Davis (2005).

Definition 3.2.4 *Two locally square integrable martingales M, N are said to be strongly orthogonal if $\langle M, N \rangle = 0$.*

Note that an equivalent characterization is that M and N are strongly orthogonal if and only if their product $M_t N_t$ is a local martingale, see Davis (2005). The angle brackets of the Teugels martingales of a Lévy processes are given by

$$\langle Y^{(i)}, Y^{(j)} \rangle = (\lambda_{i+j} + 1_{\{n=2\}}\sigma^2)t, \tag{3.58}$$

see Davis (2005, Equation 1.7), and hence induce an inner product $\ll, \gg$ on the linear space generated by all Teugels martingales $Y^{(i)}$ of Z obeying

$$\ll Y^{(i)}, Y^{(j)} \gg = (\lambda_{i+j} + 1_{\{n=2\}}\sigma^2). \tag{3.59}$$

If we define an inner product $\ll, \gg$ on the space $\mathbb{R}\langle x \rangle$ of real polynomials over the single indeterminate x by

$$\ll P, Q \gg = \sigma^2 P(0)Q(0) + \int_{\mathbb{R}} P(x)Q(x)x^2\nu(dx), \tag{3.60}$$

then $\ll x^{i-1}, x^{j-1} \gg = \lambda_{i+j} + 1_{\{i=j=1\}}\sigma^2$, and the mapping $x^{i-1} \mapsto Y^{(i)}$ is an isometry of the two inner product spaces, see Nualart & Schoutens (2000). The following result is due to Nualart & Schoutens (2000).

Lemma 3.2.5 *Let $\{Y^{(i)}\}$ be the Teugels martingales of a Lévy process Z. There exist pairwise strongly orthogonal square-integrable martingales $\{H^{(i)}\}$ and constants $c_{n,i}$ with $c_{n,n} = 1$ such that*

$$Y^{(n)} = c_{n,1}H^{(1)} + c_{n,2}H^{(2)} + \ldots + c_{n,n}H^{(n)}. \tag{3.61}$$

The proof of Nualart & Schoutens (2000) is constructive. Indeed, any orthogonalization of the set $\{1, x, x^2, \ldots\}$ with respect to the inner product $\ll, \gg$ induces an orthogonalization of the set $\{Y^{(1)}, Y^{(2)}, \ldots\}$ with respect to the inner product $\ll, \gg$. As orthogonality in the inner product space of Teugels martingales implies strong orthogonality, we obtain the result, see Nualart & Schoutens (2000). For the remainder of this section, we let $H^{(i)}$ be the orthogonalization of the Teugels martingales of the above form, defined inductively by

$$H^{(1)} := Y^{(1)}, \tag{3.62}$$

$$H^{(n)} := Y^{(n)} - \sum_{k=1}^{n-1} \int \frac{d\langle Y^{(n)}, H^{(k)}\rangle}{d\langle H^{(k)}\rangle} dH^{(k)}, \tag{3.63}$$

this is the orthogonalization utilized in Jamshidian (2005) and is equivalent to defining H^i as the orthogonal projection onto the orthogonal complement of the direct sum of the stable spaces $\mathcal{S}(Y^j)$ with $j < n-1$, see Protter (2003, Chapter IV.3) and Jamshidian (2005) for the definition of stable spaces and further details. We will not require this interpretation. The following result is from Curry et al. (2014, Remark 1).

Lemma 3.2.6 *Suppose that $H^{(1)}, \ldots, H^{(n)} \neq 0$. The angle brackets $\langle H^{(k)}\rangle_t$ are scalar multiples of t and for any square-integrable, predictable process φ_s*

$$|| \int_0^t \varphi_s dH_s^{(n)} ||_{L^2} = 0 \tag{3.64}$$

if and only if $\varphi_s = 0$ for all s. Moreover, if $\{\varphi^i\}, i = 1, \ldots, n$ are left-continuous processes obeying

$$\sum_{i=1}^{n} \int \varphi_s^i dH_s^{(i)} = 0, \tag{3.65}$$

then $\varphi_s^i = 0$ for all i, s.

Proof: Each $H^{(n)}$ is in the linear span of $\{Y^{(1)}, \ldots, Y^{(n)}\}$, hence the characterization of the angle brackets of $Y^{(i)}$ gives the first result. The second statement follows from the extended form of Itô's isometry

$$|| \int_0^t \varphi_s dH_s^{(n)} ||_{L^2}^2 = E\left(\int_0^t (\varphi_s)^2 d\langle H^{(n)}\rangle \right), \tag{3.66}$$

see Ikeda & Watanabe (1989, Proposition II.2.4). Indeed, writing $\langle H^{(n)} \rangle = c_n t$ by the first statement, we obtain

$$|| \int_0^t \varphi_s dH_s^{(n)} ||_{L^2}^2 = c_n \int_0^t ||\varphi_s||_{L^2}^2 ds. \tag{3.67}$$

Note that c_n is necessarily positive as $\langle H^{(n)} \rangle$ is an increasing process by definition. The result follows from the positive definitiness of the L^2 norm. Furthermore, we have

$$|| \sum_{i=1}^n \int_0^t \varphi_s dH_s^{(n)} ||_{L^2}^2 = \sum_{i,j} E\left\{ \left(\int_0^t \varphi_s^i dH_s^{(i)} \right) \left(\int_0^t \varphi_s^j dH_s^{(j)} \right) \right\}. \tag{3.68}$$

Stochastic integrals of strongly orthogonal martingales are strongly orthogonal, see Protter (2003, Chapter IV.3) and Nualart & Schoutens (2000, p. 116). The pairwise strong orthogonality of the $H^{(i)}$ then extends to give pairwise strong orthogonality of the integrals $\int \varphi_s^i dH_s^{(i)}$, and the above identity reduces to

$$|| \sum_{i=1}^n \int_0^t \varphi_s dH_s^{(n)} ||_{L^2}^2 = \sum_{i=1}^n || \int_0^t \varphi_s^i dH_s^{(i)} ||_{L^2}^2. \tag{3.69}$$

Under the hypothesis of the third statement, we have

$$|| \int_0^t \varphi_s^i dH_s^{(i)} ||_{L^2}^2 = 0 \tag{3.70}$$

for all $i = 1 \ldots n$. The result then follows from the second statement.

The following result from Curry et al. (2014, Lemma 3.1) is a direct consequence of the definition of the orthogonalization $\{H^{(i)}\}$ and the relationship between power brackets and Teugels martingales of a Lévy process.

Lemma 3.2.7 *Let $k \geq 1$. The following properties are equivalent:*

1. $H^{(k)} = 0$.

2. $H^{(n)} = 0$ *for all* $n \geq k$.

3. $[Z]^{(k)}$ *is in the linear span of* $\{t, [Z]^{(1)}, \ldots, [Z]^{(k-1)}\}$.

4. $[Z]^{(n)}$ *is in the linear span of* $\{t, [Z]^{(1)}, \ldots, [Z]^{(k-1)}\}$ *for all* $n \geq k$.

5. $[Y]^{(k)}$ *is in the linear span of* $\{t, [Y]^{(1)}, \ldots, [Y]^{(k-1)}\}$.

6. $[Y]^{(n)}$ *is in the linear span of* $\{t, [Y]^{(1)}, \ldots, [Y]^{(k-1)}\}$ *for all* $n \geq k$.

Proof: We begin by showing that the first result is equivalent to the fifth. This follows from the definition of the $H^{(i)}$ as we have

$$Y^{(n)} = c_{n,1}H^{(1)} + c_{n,2}H^{(2)} + \ldots + c_{n,n}H^{(n)} \tag{3.71}$$

for some constants $c_{n,i}$. The third result is also equivalent to the fifth by Lemma 3.2.3. By the same arguments we obtain the equivalence of the second, fourth and sixth statements. The second statement is clearly stronger than the first. To obtain the full result it is then sufficient to show that the third statement implies the fourth. Writing

$$[Z]^{(k+1)} = [Z^{(k)}, Z] = \left[\sum_{i=1}^{k-1} c_i Z^{(i)}, Z\right] = \sum_{i=1}^{k-1} c_i Z^{(i+1)}, \tag{3.72}$$

we see that $[Z]^{(k+1)}$ is in the span of $\{t, [Z]^{(1)}, \ldots, [Z]^{(k)}\}$ and hence also in the span of $\{t, [Z]^{(1)}, \ldots, [Z]^{(k-1)}\}$. The result follows by induction.

The next important result from Curry et al. (2014, Theorem 3.2) is derived from the above lemmas, and is critical in ensuring the absence of relations between power bracket processes and linear combinations of multiple stochastic integrals with respect to power bracket processes.

Lemma 3.2.8 *Let Z be a Lévy process possessing moments of all orders, and $n \geq 1$. Let $\{\varphi^i\}, i = 0, \ldots, n-1$ be left-continuous processes such that*

$$[Z]_t^{(n)} = \sum_{i=1}^{n-1} \int_0^t \varphi_s^i d[Z]_s^{(k)} + \int_0^t \varphi_s^0 ds. \tag{3.73}$$

Then the φ^i are constant for all $i = 0, \ldots n-1$.

Proof: By Lemma 3.2.7, we may assume that $H^{(1)}, \ldots, H^{(n-1)} \neq 0$ without loss of generality. If a semimartingale admits a decomposition as the sum of a predictable finite variation process and a local martingale, it does so uniquely (see Protter, 2003). Note that the relation between power brackets and Teugels martingales is such a decomposition. Rewritten in terms of Teugels martingales, the above equation then yields separate equations for the predictable, finite variation part and local martingale part; we have

$$\lambda_n = \varphi_t^1 \alpha + \varphi_t^0 + \sum_{k=2}^{n-1} \varphi_t^k (\alpha_k + \sigma^2 1_{\{k=2\}}), \tag{3.74}$$

and

$$Y_t^{(n)} = \sum_{i=1}^{n-1} \int_0^t \varphi_s^i dY_s^{(i)}, \tag{3.75}$$

The latter equation may be rewritten in the form

$$H^{(n)} - \sum_{j=1}^{n-1} \int_0^t \sum_{i=1}^{n-1} \varphi_s^i c_{ij} - c_{nj} dH_s^{(j)} = 0. \tag{3.76}$$

It follows from Lemma 3.2.6 that such an identity holds only if $H^{(n)} = 0$ and $\sum_{i=1}^{n-1} \varphi_s^i c_{ij} - c_{nj} = 0$, hence the φ^i are constant for all $i = 0, \ldots, n-1$.

We are now in a position to construct the quasi-shuffle algebra $\mathbb{R}\langle\tilde{\mathbb{A}}\rangle_*$. We augment the alphabet $\mathbb{A} := \{0, 1, \ldots, d\}$ as follows. For each Z^i, suppose there exists a least integer $k(i)$ such that the power bracket process $[Z^i]^{(k)}$ is in the linear span of $\{t, [Z^i]^{(1)}, \ldots, [Z^i]^{(k(i)-1)}\}$. Then by Lemma 3.2.7, all $[Z^i]^{(n)}$ are in the linear span of $\{t, [Z^i]^{(1)}, \ldots, [Z^i]^{(k(i)-1)}\}$. Accordingly, we augment the alphabet $\mathbb{A}$ to include the letters $\{i^{(2)}, \ldots, i^{(k(i))}\}$. If there is no such integer k, we must include the countable set $\{i^{(2)}, \ldots, i^{(n)}, \ldots\}$. Repeating this procedure, we obtain a possibly infinite extended alphabet $\tilde{\mathbb{A}}$.

Definition 3.2.9 *Given a set of independent Lévy processes $\{Z^0(t) = t, Z^1, \ldots Z^d\}$ possessing moments of all orders, the Lévy alphabet extension $\tilde{\mathbb{A}}$ is defined to be the alphabet constructed from $\mathbb{A} := \{0, 1, \ldots, d\}$ by the process above.*

The main result of Curry et al. (2014) is the following.

Theorem 3.2.10 *Let $\{Z^0(t) = t, Z^1, \ldots Z^d\}$ be independent Lévy processes possessing moments of all orders, and let $\tilde{\mathbb{A}}$ be the associated Lévy alphabet extension of*

$\{0, 1, \ldots, d\}$. The word-to-integral map μ sending $i^{(n)} \mapsto I_{i^{(n)}} := [Z^i]^{(n)}$, and acting on words by

$$\mu : w \mapsto I_w := \int_{0<\tau_{n-1}<\ldots<\tau_1<t} dI_{a_1}(\tau_n) \ldots dI_{a_n}(\tau_1) \tag{3.77}$$

is an algebra isomorphism from the quasi-shuffle algebra $\mathbb{R}\langle\tilde{\mathbb{A}}\rangle_*$ to the algebra of iterated integrals of Lévy processes and their power brackets. The commutative algebra structure on $\mathbb{R}\tilde{\mathbb{A}}$ underlying the quasi-shuffle algebra $\mathbb{R}\langle\tilde{\mathbb{A}}\rangle_*$ is defined to be the pullback under the word-to-integral map of the quadratic covariation bracket,

$$[a, b] := \mu^{-1}([\mu(a), \mu(b)]). \tag{3.78}$$

Proof: If this pullback may be defined, the isomorphism property follows from the Itô integration-by-parts formula. It remains to show the injectivity of μ. For this, we require that no $[Z^i]^{(n)}$, $n = 1, 2, \ldots$ may be written as a linear combination of multiple iterated integrals (with multiplicity greater than one). Now, the independence of Z^i and Z^j for $i \neq j$ ensures that such a relation could only exist using multiple iterated integrals with respect to $Z^{(i)}$ and the power brackets $[Z^i]^{(k)}$. This is not possible by Lemma 3.2.8.

Note that, by the proof of Lemma 3.2.8, if we modify the above word-to-integral map so that letters in the extended alphabet are sent to the associated Teugels martingale rather than power bracket process, ie. $i^{(n)} \mapsto (Y^i)^{(n)}$, the injectivity remains. Indeed, we have the following.

Corollary 3.2.11 *Let $\{Z^0(t) = t, Z^1, \ldots Z^d\}$ be independent Lévy processes possessing moments of all orders, and let $\tilde{\mathbb{A}}$ be the associated Lévy alphabet extension of $\{0, 1, \ldots, d\}$. The word-to-integral map $\tilde{\mu}$ sending $i^{(n)} \mapsto I_{i^{(n)}} := (Y^i)^{(n)}$, and acting on words by*

$$\tilde{\mu} : w \mapsto I_w := \int_{0<\tau_{n-1}<\ldots<\tau_1<t} dI_{a_1}(\tau_n) \ldots dI_{a_n}(\tau_1) \tag{3.79}$$

is an algebra isomorphism from the quasi-shuffle algebra $\mathbb{R}\langle\tilde{\mathbb{A}}\rangle_*$ to the algebra of *iterated integrals of Lévy proceses and their power brackets. The commutative algebra structure on $\mathbb{R}\tilde{\mathbb{A}}$ underlying the quasi-shuffle algebra $\mathbb{R}\langle\tilde{\mathbb{A}}\rangle_*$ is defined to be the pullback under the word-to-integral map of the quadratic covariation bracket,*

$$[a, b] \coloneqq \tilde{\mu}^{-1}([\tilde{\mu}(a), \tilde{\mu}(b)]). \tag{3.80}$$

Proof: Recall that for each Z^i, we defined $k(i)$ to be the least integer such that the power bracket process $[Z^i]^{(k)}$ is in the linear span of $\{t, [Z^i]^{(1)}, \ldots, [Z^i]^{(k-1)}\}$. By Lemma 3.2.7, all of the Teugels martingales $(Y^{(i)})^{(n)}$ are in the linear span of $\{t, (Y^i)^{(1)}, \ldots, (Y^i)^{(k(i)-1)}\}$. We note further that Lemma 3.2.8 concerning the non-existence of certain relations amongst power bracket processes was obtained from the analogous result for relations amongst Teugels martingales, using the uniqeness of decomposition of special semimartingales. It follows that the mapping $\tilde{\mu}$ is injective, and the result follows as per that of theorem above.

We have now demonstrated the existence of two quasi-shuffle algebra structures isomorphic to the algebra generated by the multiple iterated integrals of a finite set of independent Lévy processes with moments of all orders. We are now in a position to use the properties of quasi-shuffle algebras reviewed in the first part of this thesis in the forthcoming derivation and analysis of numerical integration schemes.

Chapter 4

Numerical methods

In this chapter we generalize the map-truncate-invert integration schemes of Malham & Wiese (2009) to stochastic differential equations driven by Lévy processes. We will then show how we may encode these schemes algebraically and compare different schemes within this class at an algebraic level. These methods are natural generalizations of those expounded in Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012), and feature in the forthcoming paper Curry et al. (2014b).

We recall that the starting point for map-truncate-invert schemes is the stochastic Taylor expansion for a Stratonovich drift-diffusion equation, giving rise to the expression of the pulled-back flowmap in the form

$$\varphi_*(f) = \sum_w J_w V_w \circ f, \tag{4.1}$$

where J_w are the multiple iterated Stratonovich integrals and V_w the compositions of vector fields. Here the geometric information encoded in the composition of vector fields is deterministic, and the randomness enters only through the scalar random variables $J_w(\omega)$. There is a complete separation of geometric and stochastic information, see Baudoin (2004) for a discussion of this perspective.

To study map-truncate-invert schemes for stochastic differential equations driven by Lévy processes, we require a more general form of stochastic Taylor expansion.

This has been derived by Platen (1982) for stochastic differential equations driven by independent Wiener processes and Poisson point processes. We begin by giving this expansion, and in the following section discuss the conditions under which truncations of the expansion yield integration schemes of a given order. We follow with a discussion of the application of these results to equations driven by Lévy processes. We note that the resulting expansions do not possess the aforementioned separation of geometric and stochastic information except in certain special cases. We discuss these cases separately, before detailing a new stochastic Taylor expansion derived from the Wagner-Platen expansion that exhibits the desired separation. We accordingly derive a number of expansions for the pulled-back flowmap of similar form to that given above in the continuous case.

We then proceed with the construction of an algebraic framework for the comparison of map-truncate-invert schemes. We begin by showing that the expansions of the flowmap we have derived may be understood in the context of certain formal product algebras; this is accomplished by giving a formal description of the expanded flowmap in the various forms given previously and constructing homomorphisms to product algebras $\mathbb{R}\langle A\rangle \otimes \mathbb{R}\langle A\rangle$. Here $\mathbb{R}\langle A\rangle$ is the space of non-commuting polynomials constructed from a given set A, and the product on the left will be the free associative product, corresponding to the convolution of words, whilst that on the right will be a quasi-shuffle product. Using the previous results, we then discuss the class of map-truncate-invert schemes, introduced using the embedding of convolution algebra structures in product algebras discussed in section 3.1. We justify these schemes with convergence results. We then construct a framework for comparing such schemes as follows. For a given integration scheme, we define an associated algebraic remainder endomorphism and an inner product structure on the convolution algebra; this inner product is designed to encode the L^2 inner product of the associated approximate flows. Our goal in the following chapters will be to obtain an integration scheme minimizing the norm of the associated remainder. To this end, we present an important theorem ensuring that convergent map-truncate-invert schemes obeying certain properties have an associated remainder endomorphism that is smaller in the L^2 sense than that of the truncated stochastic Taylor expansion, for all Lévy-driven equations with sufficiently smooth coefficients and integrators with moments of all orders. We also comment on the accumulation of local errors in the global error of an integration scheme, following Malham & Wiese (2009).

We conclude by discussing the appropriate manner for truncating stochastic Taylor series appearing in integration schemes, according to the various goals of minimizing computational cost, maximizing the order of convergence and minimizing the leading order remainder.

4.1 Stochastic expansions

Recall that the stochastic Taylor expansion for drift-diffusion equations given in the introduction to this thesis (see also Kloeden & Platen, 1999) is derived by expanding the driving vector fields using Itô's formula,

$$V(Y_t) = V(Y_0) + \int_0^t \nabla_{V_0}(V(Y_s))ds + \sum_i \int_0^t \nabla_{V_i}(V(Y_s))dW_s^i. \tag{4.2}$$

The most general forms of Itô's formula, in particular all those applying to discontinuous semimartingales feature terms containing non-integral sums of the form $\sum_{s\le t} f(Y_s) - f(Y_{s-}) - \partial_x f(Y_{s-})\Delta Y_s^i$, for instance for a twice differentiable scalar function f we have

$$\begin{aligned} f(Y_t) - f(Y_0) &= \sum_i \int_{0+}^t \frac{\partial f}{\partial x_i}(Y_{s-})dY_s^i + \frac{1}{2}\sum_{i,j}\int_{0+}^t \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_{s-})d[Y^i, Y^j]_s^c \\ &+ \sum_{s\le t}\left\{ f(Y_s) - f(Y_{s-}) - \sum_i \frac{\partial f}{\partial x_i}(Y_{s-})\Delta Y_s^i \right\}, \end{aligned} \tag{4.3}$$

see Protter (2003). Attempts to generalize the construction of stochastic Taylor expansions to discontinuous systems must reflect this. Platen (1982) showed that the natural setting to accommodate this were the jump stochastic differential equations first considered in Itô (1951), of the form:

$$\begin{aligned} Y_T &= Y_{t_0} + \int_{t_0}^{T} V_0(s, Y_{s-})ds + \sum_{i=1}^{d} \int_{t_0}^{T} V_i(s, Y_{s-})dW_s^i \\ &+ \int_{t_0}^{T} \int_{\mathbb{R}} V_{-1}(s, Y_{s-}, v)[Q^i(ds, dv) - \nu^i(dv)ds], \end{aligned} \tag{4.4}$$

where Q^i are stationary Poisson random measures on the space $\mathbb{R}_+ \times \mathbb{R}$, with compensators $m^i(ds, dv) = ds \times \nu^i(dv)$. Here $V_0, \ldots, V_{-i}, \ldots, V_i$ are vectors of real-valued, Borel-measurable functions. When the stochastic integrals with respect to the uncompensated measure exist, the above equation is equivalent to

$$\begin{aligned} Y_T &= Y_{t_0} + \int_{t_0}^{T} \tilde{V}_0(s, Y_{s-})ds + \sum_{i} \int_{t_0}^{T} V_i(s, Y_{s-})dW_s^i \\ &+ \sum_{i} \int_{t_0}^{T} \int_{\mathbb{R}} V_{-i}(s, Y_{s-}, v)Q^i(ds, dv), \end{aligned} \tag{4.5}$$

where $\tilde{V}_0$ and V_0 are related by

$$V_0(t, x) = \tilde{V}_0(t, x) + \sum_{i} \int_{\mathbb{R}} V_{-i}(t, x, v)\nu^i(dv). \tag{4.6}$$

The following result is standard, see eg. Ikeda & Watanabe (1989, p.245).

Theorem 4.1.1 *Suppose each V_i, $i \in \{0, \ldots, d\}$ obey global Lipschitz and linear growth conditions. Suppose moreover that, for all $t \in [0, T]$ and $x, y \in \mathbb{R}^N$, each V_{-j}, $j \in \{1, \ldots, n\}$ obeys the following jump versions of the Lipschitz and linear growth conditions.*

$$\int_{\mathbb{R}} |V_{-j}(t, x, v) - V_{-j}(t, y, v)|^2 \nu^j(dv) \leq C_1 |x - y|^2, \tag{4.7}$$

$$\int_{\mathbb{R}} |V_{-j}(t, x, v)|^2 \nu^j(dv) \leq C_2(1 + |x|^2). \tag{4.8}$$

Then the jump equation (4.4) admits a unique strong solution.

We must adapt our approach to iterated integrals to accommodate the mark dependence of the system. For a given word $w \in \{-n, \ldots, -1, 0, 1, \ldots, d\}^*$, let $s(w)$

be the number of negative letters, that is those that correspond to integrals with respect to a jump measure.

Definition 4.1.2 *The iterated integrals are defined for a given integrand*

$$g(t, \omega, v_1, \ldots, v_{s(w)}) : \mathbb{R} \times \Omega \times \mathbb{R}^{s(w)} \to \mathbb{R} \tag{4.9}$$

inductively by

$$I_{-i}[g(\cdot)](t_0;t) = \int_{t_0}^{t} \int_{\mathbb{R}} g(s, \omega, v)[Q^i(ds, dv) - \nu^i(dv)ds], \tag{4.10}$$

$$I_a[g(\cdot)](t_0;t) = \int_{t_0}^{t} g(s, \omega) dW_s^a, \tag{4.11}$$

where $a \in \{0, \ldots, n\}$*, and we employ the convention* $W_t^0 = t$*, and*

$$I_{a_1 \ldots a_n}[g(\cdot)](t_0;t) = I_{a_n}[I_{a_1 \ldots a_{n-1}}[g(\cdot)](t_0;\cdot)](t_0;t). \tag{4.12}$$

We may omit the t_0 dependency on occasion, in which case it should be assumed that $t_0 = 0$. We require for each iterated integral I_w a space of allowable integrals.

Definition 4.1.3 *For a given set of n stationary Poisson random measures* Q^i *with compensators* $ds \times \nu^i$*, we define spaces* $\mathcal{H}_w$ *of adapted stochastic processes, indexed by words* $w \in \{-n, \ldots, -1, 0, 1, \ldots, d\}^*$*, as follows. Let*

$$\mathcal{H}_1 = \{g : \sup E(g(t, \omega)) < \infty\}, \tag{4.13}$$

$$\mathcal{H}_0 = \{g : E\left(\int_0^T |g(s, \omega)| ds\right) < \infty\}. \tag{4.14}$$

For any positive letter $i \in \{1, \ldots, d\}$*, define*

$$\mathcal{H}_i = \{g : E\left(\int_0^T (g(s, \omega))^2 ds\right) < \infty\}. \tag{4.15}$$

For any negative letter $-j : \{-1, \ldots, -n\}$, define

$$\mathcal{H}_{-j} = \{g : E\left(\int_0^T \int_{\mathbb{R}} |g(s,\omega,v)|^2 \nu^j(dv)ds\right) < \infty\}. \tag{4.16}$$

For any given word $w = a_1 \ldots a_m$ of length greater than one, we define inductively

$$\mathcal{H}_w = \{g : I_{a_1 \ldots a_{m-1}}[g(\cdot)] \in \mathcal{H}_{a_m}\}. \tag{4.17}$$

Recall that the stochastic Taylor expansion was derived from iterative expansions of the driving vector fields in terms of their initial data. In this case, we must utilize a more general form of Itô's formula (see eg Protter, p.81), writing $V_i(x) = \sum_j V_i^j(x)\partial_{x_j}$ we obtain the following expansion of the vector fields

$$\begin{aligned} V(Y_t) &= V(Y_0) + \sum_i \int_0^t \nabla_{V_i}(V(Y_{s-}))dW_s^i \qquad (4.18) \\ &+ \int_0^t \left(\frac{\partial V}{\partial t}(Y_s) + \nabla_{V_0}(V(Y_s)) + \frac{1}{2}\sum_i V_i^j(Y_s)V_i^k(Y_s)\frac{\partial^2 V}{\partial x_j \partial x_k}(Y_s)\right. \\ &+ \left.\sum_i \int_{\mathbb{R}} \left(V(Y_s + V_{-i}(s, Y_s, v)) - V(Y_s) - \nabla_{V_{-i}}(V(Y_s))\right)\nu^i(dv)\right) ds \\ &+ \sum_i \int_0^t \int_{\mathbb{R}} (V(Y_{s-} + V_{-i}(s-, Y_{s-}, v)) - V(Y_{s-})[Q^i(ds,dv) - \nu^i(dv)ds]. \end{aligned}$$

We are thus led to define the composition of operators $\tilde{V}_w = \tilde{V}_{a_1} \circ \ldots \circ \tilde{V}_{a_n}$, where the $\tilde{V}_i$ act on functions as the usual Lie derivative $f \mapsto \nabla_{V_i}(f)$, but $\tilde{V}_{-i}(v)$ is the shift operator

$$\tilde{V}_{-i}(v) : f(x) \mapsto f(x + V_{-i}(x,v)) - f(x) \tag{4.19}$$

and $\tilde{V}_0$ is the second order differential operator

$$\tilde{V}_0 : f(x,t) \mapsto \frac{\partial f(x,t)}{\partial t} + \nabla_{V_0}(f(x,t)) + \frac{1}{2}\sum_i V_i^j(x)V_i^k(x)\frac{\partial^2 f}{\partial x_j \partial x_k}(x,t) \tag{4.20}$$

$$+\sum_i \int_{\mathbb{R}} \left(\tilde{V}_{-i}(v) \circ f(x,t) - \nabla_{V_{-i}(v)}(f(x,t))\right)\nu(dv). \tag{4.21}$$

We have introduced the tildes to stress that the action for non-positive letters differs from the usual action of vector fields on functions. Note that each application of $\tilde{V}_{-i}$ introduces a dependency on $v \in \mathbb{R}$. In practice, this will always be integrated out, but it is for this reason that we may not immediately separate the vector fields and integrals as per the continuous case. Before giving the full Wagner-Platen expansions, we must pause to consider more closely the possible truncated forms and remainders. The recursive derivation of an expansion of the form

$$\varphi = \sum_{w} I_w(V_w) \tag{4.22}$$

requires that if we include a word w we must also include $w-$. The negative sign following the word indicates the removal of the final letter, so $(a_1 \dots a_n)- = a_1 \dots a_{n-1}$. A subset $\mathcal{A} \subset \mathbb{A}^*$ of the free monoid is said to be hierarchical (see Kloeden & Platen, p. 180) if it is non-empty, uniformly bounded in length, meaning $\sup_w |w| < \infty$, and for each word $w \in \mathcal{A}$ we have in addition $w- \in \mathcal{A}$. Given a hierarchical set $\mathcal{A}$, we define its associated remainder set $\mathrm{Rem}(\mathcal{A}) \subset \mathbb{A}^*$ to be the set of words $w \notin \mathcal{A}$ such that $w- \in \mathcal{A}$.

Definition 4.1.4 (Wagner-Platen expansion) *Let Y_t be the unique strong solution process of an equation of the form* (4.4)*, with driving vector fields*

$$V_{-i}, \dots, V_{-1}, V_0, V_1, \dots V_d. \tag{4.23}$$

For a given hierarchical set $\mathcal{A} \subset \{-n, \dots, -1, 0, 1, \dots, d\}^*$, stopping times ρ, τ obeying $0 \leq \rho \leq \tau$ and function $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, the associated Wagner-Platen expansion is given by

$$f(\tau, Y_\tau) = \sum_{w \in \mathcal{A}} I_w[\tilde{V}_w \circ f(\cdot, Y_\cdot)](\rho, \tau) + \sum_{w \in \mathrm{Rem}(\mathcal{A})} I_w[\tilde{V}_w \circ f(\cdot, Y_\cdot)](\rho, \tau), \tag{4.24}$$

assuming the function f and the driving vector fields are sufficiently smooth that all the integrands $V_w \circ f \in \mathcal{H}_w$.

4.2 Integration schemes

Let $t_n := nh$ be a uniform discretization of the interval $[0, T]$. A given hierarchical set $\mathcal{A}$ then induces the following numerical integration scheme,

$$\hat{Y}_{t_{n+1}} = \sum_{w \in \mathcal{A}} I_w[\tilde{V}_w \circ \mathrm{id}(\hat{Y}_n)](t_n, t_{n+1}), \tag{4.25}$$

assuming the Wagner-Platen expansion exists across each interval $[t_n, t_{n+1}]$. We now define a grading function $g : A^* \mapsto \mathbb{R}_+$ such that the Wagner-Platen expansions corresponding to hierarchical sets $g(w) \leq m$ give integration schemes of strong order m.

Definition 4.2.1 (Mean-square grading) *Given a word $w \in \{-n, \ldots, d\}^*$, let $\xi(w)$ be the number of occurrences of the letter 0, and $\zeta(w)$ the number of non-zero letters. We then define*

$$g(w) = \begin{cases} \xi(w) + \frac{1}{2}\zeta(w), & w \notin \{0\}^* \\ \xi(w) - \frac{1}{2}, & w \in \{0\}^*. \end{cases} \tag{4.26}$$

Next, we require conditions on the smoothness of the vector fields sufficient to ensure the convergence of the integration schemes to be considered. For a given word $w \in \{-n, \ldots, d\}^*$, let $j(w)$ be the subword of w comprising only the negative letters. For instance, we have $j(-1, 2, -1, 3, 4, -2) = -1, -1, -2$. We write $j(w)_i$ for the ith letter of the word $j(w)$. Note that $|j(w)| = s(w)$, where we recall that $s(w)$ counts the number of negative letters in the word w.

Hypothesis 4.2.2 *For a given set of iterated integrals (I_w) and hierarchical set $g(w) \leq m$, we say that the vector fields (V_w) obey the smoothness hypotheses if the following conditions hold.*

1. *The compositions $\hat{V}_w \circ \mathrm{id}$ exist and obey $\hat{V}_w \circ \mathrm{id} \in \mathcal{H}_w$ for all $w : g(-w) \leq m$, and are continuously differentiable with respect to space and twice continuously differentiable with respect to time for all $w : g(w) \leq m$.*

2. *For all $w : g(w) \leq m$, $t \in [0, T]$, $x, y \in \mathbb{R}^d$ and $u \in \mathbb{R}^{s(w)}$, the compositions*

$\hat{V}_w \circ \mathrm{id}$ *obey the jump Lipschitz condition*

$$|\hat{V}_w \circ \mathrm{id}(t,x,u) - \hat{V}_w \circ \mathrm{id}(t,y,u)| \leq C_2(u)|x-y|, \tag{4.27}$$

where $C_2(u)$ *is* $\nu^{j(w)_1} \times \ldots \times \nu^{j(w)_{s(w)}}$*-integrable.*

3. *For all* $w : g(w) \leq m$*,* $t \in [0,T]$*,* $x, y \in \mathbb{R}^d$ *and* $u \in \mathbb{R}^{s(w)}$*, the compositions* $\hat{V}_w \circ \mathrm{id}$ *obey the linear growth condition*

$$|\hat{V}_w \circ \mathrm{id}(t,x,u)|^2 \leq C_3(u)(1+|x|^2), \tag{4.28}$$

where $C_3(u)$ *is* $\nu^{j(w)_1} \times \ldots \times \nu^{j(w)_{s(w)}}$*-integrable.*

Note that the smoothness hypotheses are satisfied for a given hierarchical set $g(w) \leq m$ if the vector fields V_w are $2(m+2)$-times continuously differentiable functions that are uniformly bounded, with uniformly bounded derivatives, see Bruti-Liberati & Platen (2010). They are also clearly satisfied where the vector fields are constant and linear, of the form $V_i(x) = A^i.x$, for constant matrices A^i. The fundamental result on strong convergence of the above integration scheme is the following, see Bruti-Liberati & Platen (2005).

Theorem 4.2.3 *Let* $\hat{Y}_t$ *be integration scheme corresponding to the Wagner-Platen expansion with timestep* h *and hierarchical set* $g(w) \leq m$*. Assume the initial condition* Y_0 *is an* L^2*-random variable, and that*

$$||Y_0 - \hat{Y}_0||_{L^2} \leq C_1 h^m. \tag{4.29}$$

Suppose further that the vector fields V_w *satisfy the smoothness hypotheses for the set* $g(w) \leq m$*. Then*

$$E(\sup_{t\in[0,T]} |Y_t - \hat{Y}_t|^2) \leq K h^{2m}, \tag{4.30}$$

for some finite positive constant K*, independent of* h*, and the integration scheme* $\hat{Y}_t$ *converges with strong order* h*.*

4.3 Lévy flows and Taylor expansions

The main purpose of this thesis is to use algebraic methods to study integration schemes for stochastic differential equations of the form

$$Y_t = Y_0 + \int_0^t V_0(Y_s)ds + \sum_{i=1}^{d} \int_0^t V_i(Y_{s-})dZ_s^i \tag{4.31}$$

where $(Z^1, \ldots, Z^d)$ are independent Lévy processes, and the $V_i : \mathbb{R}^N \to \mathbb{R}^N$ are sufficiently smooth, non-commuting, autonomous vector fields. Note that, whilst the Wagner-Platen expansions given in the previous section allow for time-dependent coefficients, we will exclusively treat the autonomous case for the remainder of this thesis. Such a system may be naturally rephrased in the setting of the previous section by recalling the Lévy decomposition $Z_t^i = \alpha^i t + \sigma^i W_t + J^i t$, where α^i, σ^i are constants, W^i are independent Wiener processes and J^i are purely discontinuous martingales of the form

$$J_t^i = \int_0^t \int_{\mathbb{R}} x(Q^i(ds, dv) - ds\nu^i(dv)), \tag{4.32}$$

where the Q^i are Poisson measures with intensity measure $dt \times \nu^i(dv)$. We may therefore without loss of generality consider the equation (4.31) to be of the form

$$Y_t = Y_0 + \int_0^t V_0(Y_s)ds + \sum_{i=1}^{d} \int_0^t V_i(Y_s)dW_s^i + \sum_{i=-1}^{-n} \int_0^t V_i(Y_{s-})dJ_s^{-i}, \tag{4.33}$$

where the J^i are of the form (4.32). This is in line with the equations considered in the previous section, where we have multiple driving measures and vector fields $V_{-i}(x; v) = vV_{-i}(x)$. In the introductory chapter, we remarked that algebraic structures enter due to the existence of expansions of the flowmap of the above equation in the form

$$\varphi(t) = \sum_w V_w(t)I_w(t). \tag{4.34}$$

The product structure of the V_w and I_w is known, as the V_w compose associatively, and the I_w generate a quasi-shuffle algebra. The pullback action of the flowmap on diffeomorphisms of $\mathbb{R}^N$ may therefore be encoded symbolically in a formal product algebra structure. We must note that the Wagner-Platen expansions as expounded in the previous section are not of this form as they do not generally exhibit the requisite separation of geometric and stochastic information; the integrands $V_w(t)$ may not be taken outside the integrals $I_w(t)$ due to the mark dependencies. Indeed, consider the iterated integral $I_{-1,1}[V_{-1,1}](t)$ arising in the Wagner-Platen expansion corresponding to $f = \mathrm{id}$, it is given by

$$\int_0^t \left(\int_0^s \int_{\mathbb{R}} (V_1(Y_0 + vV_{-1}(Y_0)) - V_1(Y_0))(Q^1(d\tau, dv) - d\tau\nu^1(dv)) \right) dW_s^1. \quad (4.35)$$

However, the quasi-shuffle algebra of iterated Lévy integrals only includes those integrals of the form

$$\int_0^t \int_{\mathbb{R}} f(v)(Q^i(ds, dv) - ds\nu^i(dv)) \quad (4.36)$$

for which the function $f(v)$ is a polynomial. Indeed $f(v) = v$ corresponds to J^1, and $f(v) = v^n$ the Teugels martingale $(Y^i)^{[n]}$. As the integral enclosed in brackets in (4.35) is not necessarily of this form, the Wagner-Platen expansion is not generally of the form (4.34), with the integrals I_w being elements of the quasi-shuffle algebra of iterated Lévy integrals. The problem is that the space of iterated integrals featuring in the Wagner-Platen expansions is strongly influenced by the geometric data of the system, and relationships between the iterated Wagner-Platen integrals are in general hard to come by.

Definition 4.3.1 *The flowmap for a stochastic differential equation of the form (4.33) possesses a separated stochastic Taylor expansion if it may be written in the form*

$$\varphi = \sum_{w \in \tilde{\mathbb{A}}} \tilde{V}_w I_w, \quad (4.37)$$

where $\tilde{\mathbb{A}}$ is the extended Lévy alphabet, and I_w are the associated multiple iterated integrals.

We will now consider various methods of obtaining the desired separation of geometric and stochastic data. Firstly, we will note several forms of equation for which the integrands in the Wagner-Platen expansion exhibit either trivial or strictly polynomial mark dependency. In these cases, the above concerns do not apply. They are enumerated as follows:

1. Drift-diffusion equations, where there is no discontinuous term and hence no mark dependency.

2. Jump-diffusion equations, where the discontinuous driving processes are standard Poisson processes, hence possess jumps of identical magnitude.

3. Linear equations, where the driving fields are linear, and hence induce linear multiplicative mark dependencies upon shifting.

More generally, we show that for sufficiently smooth vector fields, shifts by vector fields of the form $V_{-i}(x;v) = vV_{-i}(x)$ may be expanded via Taylor's theorem. The resulting expression is polynomial in the mark dependency, so will be sufficient for our purposes.

4.3.1 Drift-diffusion equations

Suppose we have a Lévy-driven equation for which all the driving processes are continuous. The Lévy decomposition then indicates that we are considering a drift-diffusion equation, where the driving processes are either independent Wiener processes or deterministic processes that are scalar multiples of time. This is the situation considered in the papers Malham & Wiese (2009), Ebrahimi–Fard et al (2012), but here the Wiener integrals are interpreted in the Itô, rather than Stratonovich sense. We will proceed to show that the Wagner-Platen expansion naturally induces a representation of the flowmap of such equations in a concatenate-quasi-shuffle product algebra. The absence of driving jump measures means we need not concern ourselves with shift operators; the vector fields may therefore be taken outside the iterated integrals. The actions of V_i and $\tilde{V}_i$ coincide for non-zero letters i, it follows that $\tilde{V}_w$ differs from the standard composition V_w only in that each $\tilde{V}_0$ acts as the second-order differential operator

$$\tilde{V}_0 : f(x) \mapsto \nabla_{V_0}(f(x)) + \frac{1}{2}\sum_{i>0} V_i^j(x)V_i^k(x)\frac{\partial^2 f(x)}{\partial x_j \partial x_k}. \tag{4.38}$$

Furthermore, from Theorem 3.2.10 we know that the iterated integrals generate a quasi-shuffle algebra induced by the quadratic covariation algebra. More explicitly, the product is seen to be that given in Gaines (1994): $I_w I_w = I_{w*w'}$, where the latter product is the quasi-shuffle on $\mathbb{A}^* = \{0, 1, \ldots, d\}^*$ given by the inductive formula

$$ua * vb = (u * vb)a + (ua * v)b + \mathbb{I}_{a=b\neq 0}(u * v)0, \tag{4.39}$$

for $u, v \in \mathbb{A}^*$, $a, b \in \mathbb{A}$. It follows that the if the vector fields satisfy the smoothness hypotheses to all orders, the Wagner-Platen expansion induces a separated stochastic Taylor expansion for the pulled back flowmap, where the $\tilde{V}_w$ are differential operators arising from the composition of operators $\tilde{V}_{a_n} \circ \ldots \circ \tilde{V}_{a_1}$ described above.

4.3.2 Jump-diffusions

Suppose that the process J^{-1} is a standard Poisson process. Then the integrands in the Wagner-Platen expansion exhibit no mark-dependence, furthermore we have

$$\int_{\mathbb{R}} \nu(dv) = \lambda, \tag{4.40}$$

where $\nu(dv)$ is the Lévy measure and λ the intensity of the Poisson process. More generally, suppose our driving random measures have finite support. The associated integral then decomposes as a sum of independent Poisson processes. Accordingly, we consider equations of the form

$$Y_t = Y_0 + \int_0^t V^0(Y_s)ds + \sum_{i=1}^{d}\int_0^t V^i(Y_s)dW_s^i + \sum_{i=-1}^{-n}\int_0^t V^i(Y_s)dN_s^{-i}, \tag{4.41}$$

where the N^i are independent standard Poisson processes with parameters λ_i. Solution processes of such equations are often referred to as jump-diffusions in the literature. The shift operators $\tilde{V}_{-i}$ have no mark-dependence; their actions are given by

$$\tilde{V}_{-i} : f(x) \mapsto f(x + V^{-i}(x)) - f(x). \tag{4.42}$$

In computing the action of the deterministic operator $\tilde{V}_0$, we must be aware that equation (4.41) is given in an uncompensated form. We obtain the following by first converting the above equation to an uncompensated equation using the relation (4.6), and then applying the Wagner-Platen expansion.

$$\tilde{V}_0 : f(x) \mapsto \nabla_{V_0}(f(x)) + \frac{1}{2}\sum_{i>0} V_i^j(x)V_i^k(x)\frac{\partial^2 f(x)}{\partial x_j \partial x_k} + \sum_{k\geq 1} \lambda_k \tilde{V}_{-k} \circ f(x). \tag{4.43}$$

It follows that we obtain an separated stochastic Taylor expansion where the $\tilde{V}_w$ are given by the compositions of the above operators, ie $\tilde{V}_w = \tilde{V}_{a_1} \circ \ldots \circ \tilde{V}_{a_n}$. The presence of the shift operators $\tilde{V}_{-k}$ results in a rather different character of the $\tilde{V}_w$ from the familiar composition of vector fields, but it is important to note that the associativity remains, and hence $V_u \circ V_v = V_{uv}$, where uv is the concatenation product of words.

4.3.3 Linear equations

An important class of Lévy-driven equations consists of those for which the vector fields V^i are constant and linear, i.e. they are of the form $V_i(Y) = A^i Y$, where $A^i = [a^i_{jk}]$ are constant $N \times N$ matrices. These are also sometimes known as bilinear equations, and have been studied by Marcus (1978, 1980), for instance. In this case, the Wagner-Platen expansions display the desired separation. Indeed, their form is given in the following result.

Theorem 4.3.2 (Linear Lévy Taylor expansion) *Let φ be the flow map for a bilinear Lévy-driven equation of the form*

$$Y_t = Y_0 + \int_0^t A^0 Y_s ds + \sum_{i=1}^{d} \int_0^t A^i Y_{s-} dW_s^i + \sum_{i=-1}^{-n} \int_0^t A^i Y_{s-} dJ_s^{-i}, \tag{4.44}$$

The Wagner-Platen expansion induces the following expression for the pullback ac-

tion, restricted to the space of linear diffeomorphisms $\mathbb{R}^n \to \mathbb{R}^n$*.*

$$\varphi = \sum_w V_w I_w, \tag{4.45}$$

where $w \in \{-n, \ldots, -1, 0, 1, \ldots, d\}^*$*, the iterated integrals* I_w *are defined in the usual sense, and the vector fields* V_w *compose by the usual matrix multiplication, but in reverse order:*

$$V_{a_1 \ldots a_n}(Y_0) = A^{a_n} \ldots A^{a_1}(Y_0). \tag{4.46}$$

Although this result is simple to derive, we have not found it mentioned anywhere in the literature. The proof is as follows.

Proof: By the hypothesis restricting the pullback action to the space of linear diffeomorphisms, let $f : \mathbb{R}^N \to \mathbb{R}^N$ is linear, of the form $f(x) = Fx$, where $F = [f_{ij}]$ is an $N \times N$ matrix. First, we compute the standard Lie derivative

$$\tilde{V}_i \circ f(x) = \nabla_{V_i}(f(x)) = a^i_{jl} x_l \frac{\partial}{\partial x_j} f_{km} x_m = a^i_{jl} x_l f_{kj} = F A^i x = f(V_i(x)). \tag{4.47}$$

The action of the shift operators $\tilde{V}_{-i}(v)$ on linear functions $f(x)$ is similar:

$$\tilde{V}_{-i}(v) \circ f(x) = f(x + vV_{-i}(x)) - f(x) = vf(V_{-i}(x)). \tag{4.48}$$

Finally, consider the action of the operator $\tilde{V}_0$. Recall the general formulation in the Wagner-Platen expansion:

$$\begin{aligned} \tilde{V}_0 : f(x,t) \quad \mapsto \quad & \frac{\partial f(x,t)}{\partial t} + \nabla_{V_0}(f(x,t)) + \frac{1}{2} \sum_i V_i^j(x) V_i^k(x) \frac{\partial^2 f}{\partial x_j \partial x_k}(x,t) \\ + \quad & \sum_i \int_{\mathbb{R}} \left(\tilde{V}_{-i}(v) \circ f(x,t) - \nabla_{V_{-i}(v)}(f(x,t)) \right) \nu(dv). \end{aligned} \tag{4.49}$$

The first term above vanishes as we are restricting ourselves to autonomous functions $f(x)$. We have already computed the Lie derivative in (4.47), the second term is therefore $f(V_0(x))$. The linearity of f implies that the second order derivatives in the third term vanish. It remains to consider the last term, however note from (4.47) and (4.48) that the two terms under the integration are identical and therefore cancel.

We conclude that the action of $\tilde{V}_0$ reduces in this case to the Lie derivative

$$\tilde{V}_0 \circ f(x) = f(V_0(x)). \tag{4.50}$$

The forms (4.47), (4.48) and (4.50) show that all the operators $\tilde{V}_i$ map linear functions to linear functions. It follows that we may readily deduce the action of $\tilde{V}_w(v)$ on linear functions $f(x)$; for a general word w and multi-dimensional mark $v = (v_1, \ldots, v_{s(w)})$ it is given by

$$\tilde{V}_w(v) \circ f(x) = v_1 \ldots v_{s(w)} F A^{a_n} \ldots A^{a_1} x, \tag{4.51}$$

and the result follows.

The restriction of the pullback action to linear functions is necessary; we require that the mark-dependence is linear to induce the requisite separation of the geometric and stochastic information in the Wagner-Platen expansion. Note however that the pullback flow φ sends linear diffeomorphisms to linear diffeomorphisms; compositions of φ may therefore be computed using the above expansion. This is sufficient to ensure that the above expansion may be encoded in a formal product algebra.

4.3.4 Analytic vector fields

We now demonstrate a general case where a Taylor expansion of the mark-dependent integrands may provide the separation of geometric and stochastic information. Consider the Lévy-driven equation of the form (4.33). We will now show that, provided the vector fields V_{-j} are sufficiently smooth, we may approximate the action of the shift operators $\tilde{V}_{-i}(v)$ by a sum of differential operators. In so doing, we will have constructed a separated stochastic Taylor expansion for the flowmap.

Theorem 4.3.3 *Let φ be the flowmap for a stochastic differential equation of the form (4.33), where the vector fields satisfy the smoothness hypotheses to all orders, and the vector fields of negative index are entire with infinite radius of convergence.*

The flowmap possesses a separated stochastic Taylor expansion, where the operators $\tilde{V}_{i(m)}$ are given by

$$\tilde{V}_{i(m)} = \sum_{i_1+\ldots+i_k=m} \frac{1}{m!} V_{-i}^{i_1} \ldots V_{-i}^{i_k} \frac{\partial^m f^j}{\partial x^{i_1} \ldots \partial x^{i_k}} e_j, \tag{4.52}$$

where $(e_1, \ldots, e_n)$ denotes the standard basis of $\mathbb{R}^n$, and the vector fields $\tilde{V}_i$ arising from non-negative letters i are those appearing in the Wagner-Platen expansion in its full generality. The operators $\tilde{V}_w$ for words with multiple letters are recovered by composition of those corresponding to single letters.

Proof: Let $F(x, v) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be of the form $F(x, v) = f(x)h(v)$, where $f : \mathbb{R}^n \to \mathbb{R}^n$ is entire, that is possesses a Taylor expansion at every point. Moreover, assume that the radius of convergence is everywhere infinite. We may then expand

$$\tilde{V}_{-i}(v) \circ F(x, v) = [f(x + vV_{-i}(x)) - f(x)]h(v) \tag{4.53}$$

in the form

$$\tilde{V}_{-i}(v) \circ F(x, v) = h(v) \sum_{\substack{m \geq 1 \\ i_1+\ldots+i_k=m}} \frac{v^m}{m!} V_{-i}^{i_1}(x) \ldots V_{-i}^{i_k}(x) \frac{\partial^m f^j(x)}{\partial x_{i_1} \ldots \partial x_{i_k}} e_j, \tag{4.54}$$

where $V_{-i} = (V_{-i}^1, \ldots, V_{-i}^n)$, $f = (f^1, \ldots f^n)$ and $(e_1, \ldots, e_n)$ denotes the standard basis of $\mathbb{R}^n$. The result follows.

4.4 Encoding Flowmaps

We demonstrate the encoding of the flowmap of a Stratonovich drift-diffusion equation given in Malham & Wiese (2009), before preceding to its generalizations. We recall that the flowmap φ acts smoothly on the space of diffeomorphisms of $\mathbb{R}^n$ through the pullback $\varphi_*(f) = f(\varphi)$, here this action is given by

$$\varphi_*(f) = \sum_w J_w V_w \circ f, \tag{4.55}$$

where the V_w are interpreted as differential operators arising from the composition of vector fields. Let $\mathbb{V} = \bigoplus_n \mathbb{V}_n$ be the subset of the space of differential operators on smooth functions on $\mathbb{R}^N$ generated by vector fields and their compositions. Letting $\mathbb{J}$ be the ring generated by multiple Stratonovich integrals and the constant random variable 1, under pointwise multiplication and addition, we see that the flowmap φ_* defined in (4.55) lives in the space $\mathbb{J}\langle\mathbb{V}\rangle \cong \bigoplus_{n\geq 0} \mathbb{J} \otimes \mathbb{V}_n$. Define the linear word-to-vector field map $\kappa : \mathbb{K}\langle\mathbb{A}\rangle \to \mathbb{V}$ by $w \mapsto V_w$, and the linear word-to-integral map $\mu : \mathbb{K}\langle\mathbb{A}\rangle \to \mathbb{J}$ by $w \mapsto J_w$. It may be shown that the former is a concatenation homomorphism, and the latter a shuffle homomorphism; the former follows from the associativity of vector field composition, and the latter is Ree's theorem extended to stochastic iterated integrals, see Ree (1958) and Malham & Wiese (2009). It follows that the map $\kappa \otimes \mu$ defines a homomorphism from the product algebra $\mathcal{A} = K\langle\mathbb{A}\rangle_{\mathrm{conc}} \bar{\otimes} K\langle\mathbb{A}\rangle_{⧢}$, where $\bar{\otimes}$ denotes the completion with respect to the A-adic topology (see Reutenauer 1993), to the space of flows. To further our enquiries into integrators for equations driven by Lévy processes, we will discuss how the above perspectives may be adapted to the differing algebraic structure of the flowmap. Indeed, in the previous section we gave a number of expansions for the flowmap of the form

$$\varphi_*(f) = \sum_w I_w \tilde{V}_w \circ f. \tag{4.56}$$

Now, however, the product structure on the I_w arises from the quasi-shuffle algebra of iterated Lévy integrals; moreover, the $\tilde{V}_w$ may arise in a manner different in general from the standard composition of vector fields. In general, there are three cases to consider:

1. Linear Lévy flows, where $\tilde{V}_w = V_w$ is the standard vector field composition
2. Non-linear Itô drift-diffusion flows, where $\tilde{V}_w$ are differential operators in the algebra $\tilde{\mathbb{V}}$, but there is no canonical grading structure on $\tilde{\mathbb{V}}$ analogous to that of $\mathbb{V}$.
3. Non-linear flows for jump equations, including the jump diffusion and analytic vector fields cases. We denote by $\tilde{\mathbb{V}}$ the composition algebra of the operators

$\tilde{V}_w$, but as they include shift operators alongside differential operators $\tilde{\mathbb{V}}$ is no longer a subset of the space of differential operators.

In all cases, the operators $\tilde{V}_w$ arise from an associative composition of operators indexed by words. The word-to-operator map $\tilde{\kappa} : w \mapsto \tilde{V}_w$ corresponding to this indexing thus defines a concatenation homomorphism, though we stress that it is in general distinct from the word-to-vector field map κ considered earlier. Now let $\mathbb{I}$ be the ring of multiple Itô integrals, defined analogously to the above, and define the Itô word-to-integral map $\tilde{\mu} : w \mapsto I_w$. It follows that $\tilde{\kappa} \otimes \tilde{\mu}$ is a homomorphism from the product algebra $\tilde{\mathcal{A}} = K\langle\mathbb{A}\rangle_{\mathrm{co}}\bar{\otimes}K\langle\mathbb{A}\rangle_*$, to the space of Itô flows. Note that in the general case, the series for the flowmap is given by a sum

$$\sum_{w\in A^*\subset\mathbb{A}^*} w \otimes w, \tag{4.57}$$

where $A \subset \mathbb{A}^*$ is a subset not containing letters arising from the completion of the alphabet to account for quadratic covariation processes not in the linear span of the driving processes. Particularly, any letter $i^{(n)}$ corresponding to a Teugels martingale with $n > 1$ is omitted. In this case we must assign zero operators $\tilde{V}_w = 0$ to all words not in A^*; we may then modify the encoding in the product algebra such that the flowmap corresponds to $\sum_{w\in\mathbb{A}^*} w \otimes w$.

4.5 Map-truncate-invert integration schemes

Malham & Wiese (2009) introduced the class of map-truncate-invert schemes for Stratonovich drift-diffusion equations, incorporating the integration scheme of Castell & Gaines, to utilize the algebraic structure of the stochastic Taylor expansion. We here present their natural generalization to Lévy-driven equations, showing how an important subclass of such schemes may be encoded and studied in the quasi-shuffle convolution algebra. We also justify their use with convergence results. Let $f : \mathrm{Diff}(\mathbb{R}^N) \to \mathrm{Diff}(\mathbb{R}^N)$ be an invertible mapping, and suppose the flowmap φ possesses a separated stochastic Taylor expansion. A map-truncate-invert scheme is constructed by, across each timestep, truncating the series $f(\varphi)$ and then composing the resulting mapping with the inverse f^{-1}. We will study such schemes using the algebra structure on $\mathcal{A} = \mathbb{R}\langle A\rangle\bar{\otimes}\mathbb{R}\langle A\rangle$.

Definition 4.5.1 *To a given power series $f = \sum_{k=0}^{\infty} c_k x^k$ and quasi-shuffle algebra, define the convolution power series $f^* : \mathrm{End}(\mathbb{R}\langle A\rangle) \to \mathrm{End}(\mathbb{R}\langle A\rangle)$ by*

$$f^*(X) := \sum_k c_k X^{\star k}, \tag{4.58}$$

where $X^{\star k}$ are the kth powers in the convolution algebra with quasi-shuffle product and deconcatenation coproduct.

We may use convolution power series to encode the effect of applying power series maps to the Taylor expanded flowmap as follows.

Lemma 4.5.2 *Let f be a power series and φ a flowmap possessing separated stochastic Taylor expansion and word-to-operator and word-to-integral maps κ and μ respectively. We have*

$$f(\varphi) = (\kappa \otimes \mu) \circ \psi \circ f^*(\mathrm{id}), \tag{4.59}$$

where $\psi : \mathrm{End}(\mathbb{R}\langle A\rangle) \to \mathcal{A}$ is the embedding given by $f \mapsto \sum_w w \otimes f(w)$, and the quasi-shuffle algebra of the convolution power series is that corresponding to the algebra of iterated integrals.

Proof: This is a direct result of the encoding of the flowmap given in the previous section and Lemma 3.1.12 that the mapping ψ is a homomorphism for the quasi-shuffle convolution product. Indeed, as the the representation of the flowmap in the product algebra $\mathcal{A}$ is given by the series $\sum_{w\in\mathbb{A}^*} w \otimes w$, it follows that is the image in $\mathcal{A}$ under the embedding ψ of the identity endomorphism on words. Powers of the flowmap therefore correspond to convolution powers of the identity under ψ, and the result follows.

The following observation is straightforward but critical.

Lemma 4.5.3 *Suppose the power series $f = \sum c_k x^k$ has an inverse which may be expressed as $f^{-1} = \sum b_k x^k$. Then the compositional inverse of the convolution power series f^* is given by the convolution power series $(f^{-1})^*$.*

Proof: Given two power series $f = \sum f_k x^k$ and $g = \sum_k g_k x^k$, the expansions for $(f \circ g)^*$ and $f^* \circ g^*$ are both given by $\sum_{i+j=k} f_i g_j X^{\star k}$ and hence equal; the result follows immediately as $(f^{-1})^* \circ f^* = (f^{-1} \circ f)^* = \mathrm{id}^* = \mathrm{id}$.

Following the above discussion, we present a formal definition of map-truncate-invert schemes, defined at the algebraic level.

Definition 4.5.4 *Let f be an invertible power series and $\pi_{\leq n}$ a truncation function. The associated map-truncate-invert scheme across a timestep $[t_n, t_{n+1}]$ is given by*

$$\hat{\varphi}_{t_n,t_{n+1}} \coloneqq \left[(\kappa \otimes \mu)_{t_n,t_{n+1}} \circ \psi\right] \circ \left[(f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})\right], \tag{4.60}$$

where $(\kappa \otimes \mu)_{t_n,t_{n+1}}$ is the tensor product of the maps given by the word-to-operator map evaluated at time t_n, and the word-to-integral map evaluated between the limits t_n and t_{n+1}.

More explicitly, let $g \coloneqq (f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})$; the integration scheme described is that given by $\hat{\varphi} = \sum_w V_w I_{g(w)}$. Recall from the introduction that the motivating example of a map-truncate-invert scheme was the Castell-Gaines integration scheme, corresponding to $f = \log$. The logarithm does not possess a power series about the origin, instead we have

$$\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k. \tag{4.61}$$

Let ν be the identity in the convolution algebra, i.e. the composition of the unit and counit of the Hopf algebra structures, that sends non-empty words to 0 and the empty word to itself (see Reutenauer, 1993). It follows that, for any endomorphism X that fixes the empty word, we may define

$$\log^*(X) \coloneqq \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (X - \nu)^{\star k}. \tag{4.62}$$

In particular, we have

$$\log^*(\mathrm{id}) = J - \frac{1}{2} J^{\star 2} + \ldots + \frac{(-1)^{k+1}}{k} J^{\star k} + \ldots, \tag{4.63}$$

where J is the 'augmented ideal projector' given by $J = \mathrm{id} - \nu$. The map J acts as the identity on non-empty words, but sends empty words to 0. We briefly illustrate the distinction between convolution powers of the identity and of the augmented ideal projector in the setting of the shuffle/de-concatenate convolution algebra. In this case, both consist of sums of partitions of the word shuffled together, but in the former case empty words may be included in the partition and in the latter not. Explicitly, where $\sqcup\!\sqcup$ is the shuffle product, we have

$$\mathrm{id}^{\star 2}(ab) = 1 \sqcup\!\sqcup\, ab + a \sqcup\!\sqcup\, b + ab \sqcup\!\sqcup\, 1, \tag{4.64}$$

$$J^{\star 2}(ab) = a \sqcup\!\sqcup\, b. \tag{4.65}$$

The justification for map-truncate-invert schemes is that many such schemes reproduce the stochastic Taylor expansion up to the desired truncation level, but in addition approximate higher order terms. We then have the following.

Theorem 4.5.5 *Let $\hat{\varphi}$ be a map-truncate scheme. Suppose that*

$$(\mathrm{id} \otimes \pi_{\leq n}) \circ \psi \circ \left[(f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})\right] = \psi(\pi_{\leq n}), \tag{4.66}$$

in other words, that the coefficients of all iterated integrals of grade up to the truncation order n are identical to those of the stochastic Taylor series. Suppose further that the expectation of the remainder $\varphi - \hat{\varphi}$ is zero at leading order. Then the scheme $\hat{\varphi}$ converges with the same strong order as the scheme arising from truncating the stochastic Taylor series according to $\pi_{\leq n}$ under the assumption that the latter is convergent.

Proof: Let $\varphi_{n+1} \coloneqq \varphi_{t_n, t_{n+1}}$ be and $\hat{\varphi} \coloneqq \hat{\varphi}_{t_n, t_{n+1}}$ be the exact and approximate flows across the $(n+1)$th timestep respectively. The exact and approximate solutions are given at discretization points by $y_n = (\varphi_n \circ \ldots \circ \varphi_1) y_0$ and $\hat{y}_n = (\hat{\varphi}_n \circ \ldots \circ \hat{\varphi}_1)\hat{y}_0$. The global mean square error is then given by

$$\epsilon^2(t_{n+1}) \coloneqq ||\varphi_{n+1}(y_n) - \hat{\varphi}_{n+1}(\hat{y}_n)||^2_{L^2}. \tag{4.67}$$

The proof given in Bruti-Liberati & Platen (2010) for the convergence of truncated stochastic Taylor expansion schemes relies on the estimate

$$\epsilon^2(t_{n+1}) \leq ||\hat{\varphi}_{n+1}(y_n) - \hat{\varphi}_{n+1}(\hat{y}_n)||^2_{L^2} + ||(\varphi_{n+1} - \hat{\varphi}_{n+1})(y_n)||^2_{L^2}. \tag{4.68}$$

The first term is difficult to estimate, as it is a global error term that involves products of iterated integrals at different times. Milstein & Tretyakov (2004, Theorem 1.1), showed how convergence of integration schemes for drift-diffusion schemes may be obtained from local error bounds using the estimate

$$\epsilon^2(t_{n+1}) \leq ||\varphi_{n+1}(y_n) - \varphi_{n+1}(\hat{y}_n)||^2_{L^2} + ||(\varphi_{n+1} - \hat{\varphi}_{n+1})(\hat{y}_n)||^2_{L^2}, \tag{4.69}$$

under the assumption that the expectation of the local error $E(\varphi_n(x) - \hat{\varphi}_n(x))$ is of higher order than the mean square error, for all x. This result relies on two Lemmas; the first, Lemma 1.3, gives a bound on the first term of (4.69) and generalizes readily to the discontinuous setting by an analogous argument. The generalization of the second, Lemma 1.4, is Lemma 6.6.1 of Bruti-Liberati & Platen (2010). The requisite bound on the local error,

$$||\varphi_n(x) - \hat{\varphi}_n(x)||^2 \leq Ch^m \tag{4.70}$$

for truncated stochastic Taylor expansions, is Lemma 4.5.2 of Bruti-Liberati & Platen (2010), and holds for the map-truncate-invert schemes under the hypotheses of the theorem. Furthermore, the condition on the expectation of the local error holds by the hypotheses. The stated result then follows by an argument entirely analogous to that of Theorem 1.1 of Milstein & Tretyakov (2004).

In our applications, assumption (4.66) is justified through the following result.

Lemma 4.5.6 *Suppose $f(1+x) = \sum_k c_k x^k$ is a power series with zero constant term, $c_0 = 0$, and inverse $f^{-1} = \sum_k b_k x^k$. Suppose further that $g(w) : \tilde{\mathbb{A}}^* \mapsto \mathbb{Z}_+$ is an integer-valued grading function and that the convolution powers f^* are with respect to a quasi-shuffle algebra that preserves the grading. Let $\pi_{\leq n}$ and $\pi_{>n}$ be the* *projections corresponding to the grading g, and define*

$$P \coloneqq \pi_{\leq n} \circ f^*(\mathrm{id}); \quad Q \coloneqq \pi_{>n} \circ f^*(\mathrm{id}). \tag{4.71}$$

Then assumption (4.66) holds, moreover we have the following relation for the leading order local remainder of the associated map-truncate-invert scheme $g \coloneqq (f_{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})$.

$$\pi_{n+1} \circ (\mathrm{id} - g) = b_1 Q. \tag{4.72}$$

Proof: Note that $\mathrm{id} = P + Q$, and hence we have

$$\begin{aligned} \mathrm{id} - g &= (f^{-1})^*(P+Q) - (f^{-1})^*(P) && (4.73)\\ &= \sum_k b_k \left[(P+Q)^{\star k} - P^{\star k}\right] && (4.74)\\ &= b_1 Q + b_2(P \star Q + Q \star P) + O(Q^{\star 2}). && (4.75) \end{aligned}$$

Note that we have assumed $c_0 = 0$, as per the hypotheses. At leading order, we have

$$P \star Q = c_1^2 (J \circ \pi_{\leq n})(J \circ \pi_{>n}). \tag{4.76}$$

The term in the left bracket contains words of grade at least one, and that on the right contains words of grade at least $n+1$, by the hypothesis that the grading is integer-valued. As the quasi-shuffle is graded, all the resulting terms are of grade at least $n+2$, and we observe the same result for the term $Q \star P$. The second result of the theorem then follows. The first statement then holds a fortiori. Note that the assumption that the grading takes integer values may easily be relaxed if required.

4.6 Comparing integration schemes

In the previous section we established a framework for constructing map-truncate-invert schemes through algebraic endomorphisms of the form $(f^*)^{-1} \circ \pi_{\leq n} \circ f^*(\mathrm{id})$. Recall that our aim is to construct integration schemes minimizing the leading order remainder. We now show how to construct an inner product on the space of endomorphisms that encodes the leading order remainder of the associated integration scheme for a class of stochastic differential equations. We conclude with an important minimality result. The material presented here comprises a straightforward generalization of a similar construction given in Ebrahimi–Fard et al. (2012). Recalling that the stochastic information contained in a separated Taylor expanded flowmap is carried by the multiple Itô integrals, we begin by considering algebraic encodings of the expectation.

Definition 4.6.1 (Expectation map and endomorphism) *Given a quasi-shuffle algebra $K\langle\tilde{A}\rangle$ over the alphabet $\tilde{A} = \{\ldots, 0, 1, \ldots, d\}$, we define the expectation map $\bar{E} : K\langle\tilde{A}\rangle \to K[t]$ by*

$$\bar{E}(w) = \begin{cases} \frac{1}{|w|!} t^{|w|} & \text{if } w \in \{0\}^* \\ \\ 0 & \text{otherwise,} \end{cases} \tag{4.77}$$

where $|w|$ denotes the length of the word w, and $K[t]$ is the polynomial ring over a single indeterminate t commuting with K. The composition of the above map with the canonical inclusion $K \to K\langle\tilde{A}\rangle$, sending $k \mapsto k.1$ then induces the expectation endomorphism $E : K\langle\tilde{A}\rangle \to K\langle\tilde{A}\rangle[t]$, explicitly

$$E : w \mapsto \bar{E}(w).1. \tag{4.78}$$

Note that the expectation defined in (4.77) sends a word w to the (probabilistic) expectation of the the integral process that is the image of w under the word-to-integral map $\tilde{\mu}$; this follows as integrals indexed by words not ending in 0 are martingales and hence have zero expectation. More generally, as long as we have at least one nonzero letter we may apply a stochastic Fubini's theorem to show the zero expectation property (but note that, eg, the integral $I_{10}(t)$, whilst having zero expectation, is not a martingale). The derivation of the expectation of the deterministic integrals is immediate as they have no stochastic dependency.

We are now in a position to define an inner product on the space $\mathbb{H} := \mathrm{End}(K\langle\mathbb{A}\rangle)$. For a given quasi-shuffle algebra over the alphabet $\tilde{A}$, set of operators $\tilde{\mathbb{V}}$ and initial data y_0, the map-truncate-invert scheme corresponding to the algebraic endomorphism $X \in \mathbb{H}$ is given by $(\kappa \circ \mu) \circ \psi \circ X$. We accordingly wish to define an inner product on $\mathbb{H}$ such that

$$\langle X, Y \rangle = \mathbb{E}(((\kappa \circ \mu) \circ \psi \circ X)\,((\kappa \circ \mu) \circ \psi \circ Y)). \tag{4.79}$$

This is achieved as follows.

Definition 4.6.2 (Inner product) *For a given set of operators $\tilde{V}_w$, let*

$$\langle X, Y \rangle(y_0) := \sum_{u,v} \bar{\mathbb{E}}(X(u) * Y(v))(\tilde{V}_u \circ \mathrm{id}(y_0))^T(\tilde{V}_v \circ \mathrm{id}(y_0)). \tag{4.80}$$

We then suppress the dependence on the set of operators $\tilde{V}_w$ and the initial condition y_0 by writing $(\tilde{V}_u \circ \mathrm{id}(y_0)^T)(\tilde{V}_v \circ \mathrm{id}(y_0))$ as a set of indeterminates (u, v). Hence

$$\langle X, Y \rangle := \sum_{u,v} \bar{\mathbb{E}}(X(u) * Y(v))(u, v). \tag{4.81}$$

An equivalent characterization is given in Ebrahimi–Fard et al. (2012) in the setting of the shuffle-deconcatenate convolution algebra which shows the positive-definiteness of the inner product. We conclude with the following definitions.

Definition 4.6.3 (Remainder and Pre-remainder) *Given an endomorphism $g = (f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})$, we define the remainder endomorphism to be*

$$R = \mathrm{id} - g. \tag{4.82}$$

The Pre-remainder is defined to be the the difference between the terms removed at the truncation stage, ie

$$Q = f^*(\mathrm{id}) - \pi_{\leq n} \circ f^*(\mathrm{id}). \tag{4.83}$$

Our main comparison result is then the following.

Theorem 4.6.4 *Suppose $g \coloneqq (f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})$ is the algebraic representation of a map-truncate-invert scheme, satisfying*

$$(\mathrm{id} \otimes \pi_{\leq n}) \circ \psi \circ \left[(f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})\right] = \psi(\pi_{\leq n}). \tag{4.84}$$

Let $h \coloneqq \mathrm{id} - g$, and suppose further that at leading order $\langle g, h \rangle = 0$. Then the associated map-truncate-invert scheme of order n, given in Definition 4.5.4, is a efficient integrator in the sense of Chapter 2.7. This holds for all stochastic differential equations driven by independent Lévy processes possessing moments of all orders, and autonomous vector fields satisfying the smoothness hypotheses 4.2.2 to all degrees.

Proof: By Lemma **??**, it suffices to show that the leading order remainder endomorphism $\mathrm{id} \circ \pi_{n+1}$ associated with the truncated stochastic Taylor expansion has a uniformly greater norm than the remainder endomorphism $\pi_{n+1} \circ h$. Now, $\mathrm{id} = g + h$ and $\langle \pi_{n+1} \circ g, \pi_{n+1} \circ h \rangle$, so we have

$$||\pi_{n+1} \circ \mathrm{id}||^2 = ||\pi_{n+1} \circ g||^2 + ||\pi_{n+1} \circ h||^2. \tag{4.85}$$

The result follows immediately.

4.7 Truncations

We have already defined in Section 2.2 the mean-square grading such that truncation of the Wagner-Platen expansion at grade n results in a numerical integration scheme of strong order n. Note that we may instead define the local mean square error grading defined by $g(w) \coloneqq \xi(w) + \frac{1}{2}\zeta(w)$ without the special value for deterministic letters if we consider instead integration schemes of the form

$$\hat{\varphi} = \sum_{g(w) \leq n} V_w I_w + \sum_{g(w) = n + \frac{1}{2}} V_w E(I_w) \tag{4.86}$$

as all non-deterministic words have zero expectation. This is the approach taken in, eg. Malham & Wiese (2009), and has the advantage that the resulting grading is a homomorphism: $K\langle\mathbb{A}\rangle \to (\mathbb{R}_+, +)$. Furthermore, it is the correct form for Stratonovich-based schemes where not all non-deterministic iterated integrals have zero expectation. In realizing numerical schemes derived from stochastic Taylor expansions, the bulk of the computation effort is often associated with the simulation of the iterated integrals I_w and particularly those involving the most distinct non-deterministic letters. For instance, in the case of Milstein schemes - order 1 integrators derived from the stochastic Taylor expansion for drift-diffusion equations, taking the form

$$\tilde{\varphi} = \sum_{a\in\mathbb{A}} V_a I_a + \sum_{i,j\in\mathbb{A}\setminus\{0\}} V_{ij} I_{ij}, \tag{4.87}$$

the dominant computational cost often comes from simulating the iterated integrals I_{ij}. (See, eg Malham & Wiese 2014, Wiktorsson 2001, Kloeden & Platen 1998, and Milstein 1995). Note however, that depending on the method used to simulate the I_{ij}, we may be able to simulate iterated integrals of the form I_{i0} and I_{0i} at minimal additional computational cost. This is the case, for instance, if we employ the method of Kloeden, Platen & Wright (1992) for simulation of iterated integrals based on truncation of the Karhunen-Loève expansion of a Brownian bridge. Indeed, in this case, simulating all the iterated integrals according to the mean-square grading $\{I(w) : g(w) \leq n$ will generate the set of iterated integrals $\{I_w : |w| \leq 2n\}$, grading by word order (see Kloeden & Platen, 1999). This raises the question, if we can generate these extra integrals readily, should we include terms of the form

$$\sum_{\substack{g(w)>n \\ |w|\leq 2n}} V_w I_w \tag{4.88}$$

in the approximation? We will examine this separately for flow maps with iterated Itô integrals and iterated Stratonovich integrals.

4.7.1 Itô integrals

We begin by answering the previous question in the affirmitive.

Theorem 4.7.1 *Let $\tilde{\varphi}_1$, $\tilde{\varphi}_2$ be approximate flowmaps for a Lévy-driven equation, arising from the truncation of the stochastic Taylor expansion using the mean square and word order gradings respectively, and assume that both schemes converge with the same strong order of convergence. The L^2 norm of the local flow remainder associated with the word order approximation $\tilde{\varphi}_2$ is always less than that of the mean square approximation φ_2 of the same order of convergence.*

Proof: The local error associated with the truncation at word order of the stochastic Taylor expansion of the flowmap, or remainder, is given by

$$R_1 = \sum_{|w|>n} \tilde{V}_w I_w. \tag{4.89}$$

On the other hand, the remainder associated with truncation at the mean-square grading is given by

$$R_2 = \sum_{\tilde{g}(w)>\frac{n}{2}} \tilde{V}_w I_w = R_1 + Q, \quad \text{where} \quad Q = \sum_{g(w)>\frac{n}{2};\, |w|\le n} \tilde{V}_w I_w. \tag{4.90}$$

We see that

$$||R_2||^2 = ||R_1||^2 + ||Q||^2 + 2\langle Q, R_1\rangle. \tag{4.91}$$

The leading order term in Q, that is the term in Q with mean-square grade $\frac{n+1}{2}$ is that indexed by the set of words $w : \zeta(w) + 2\xi(w) = n+1; \zeta(w) + \xi(w) \le n$. We see that the leading order terms obey $\xi(w) > 0$, and hence $\zeta(w) < n+1$. In contrast, the leading order terms of R_1 are indexed by words $w : \zeta(w) + 2\xi(w) = n+1; \zeta(w) + \xi(w) > n$. Combining the equations, we obtain $2\xi(w) \le \xi(w)$, as $\xi(w)$ is a non-negative integer we deduce that it is zero. We conclude that $\zeta(w) = n+1$. Given a word w, define $k(w)$ to be the subword obtained by deleting the zero letters. Now the expectation of the product of any two iterated integrals $I_w, I_{w'}$ such that $k(w), k(w')$ do not obey $k(w) = k(w')$, is zero. We see this by noting that $E(I_w I_{w'}) = E(I_{w*w'}) = E(I_{z(w*w')})$, where $z(w)$ is the identity map on deterministic words, but maps all non-deterministic words to zero (the latter relation follows from the properties of the expectation map). Particularly, if w, w' are such that

$\zeta(w) \neq \zeta(w')$, it follows that $E(I_w I_{w'}) = 0$. Therefore Q and R_1 are orthogonal at leading order, and we have

$$||R_2||^2 = ||R_1||^2 + ||Q||^2. \tag{4.92}$$

In particular we obtain $||R_2||^2 \geq ||R_1||^2$, and it follows that the word order truncation yields an integration scheme that is universally more accurate than the mean square grading truncation.

4.7.2 Stratonovich integrals

We will show here that approximate flows based on truncating Stratonovich flowmaps at word order are not necessarily more accurate at leading order than those based on truncating at the variance grading. Specifically, we compare the Milstein approximate flow map $\hat{\varphi}$ given by (4.87) with the word order 2 approximation

$$\tilde{\varphi} = \hat{\varphi} + \sum_{i \in \mathbb{A}\backslash\{0\}} V_{i0} J_{i0} + V_{0i} J_{0i}. \tag{4.93}$$

Define R_1 to be the leading order remainder of the word-order scheme and R_2 that of the Milstein scheme. Recalling that we must subtract the expectations of leading order terms in the integrators, we have

$$R_1 = \sum V_{ijk} J_{ijk} + V_{ijk}(\delta_{ij} + \delta_{jk})\frac{t^2}{4}, \tag{4.94}$$

$$R_2 = R_1 + \sum V_{i0} J_{i0} + V_{0i} J_{0i} := R_1 + Q. \tag{4.95}$$

It follows that $||R_2||^2 = ||R_1||^2 + ||Q||^2 + 2\langle Q, R_1\rangle$. We therefore have efficiency, here meaning that the mean-square error associated with the word-order scheme, R_1, is always smaller than that of the Milstein scheme, R_2, if and only if $||Q||^2 + 2\langle Q, R_1\rangle$ is a positive semidefinite quadratic form.

In order to assist with our calculation, we will convert all the above Stratonovich integrals into Itô integrals. Using the method described in Kloeden & Platen (1998, pp. 175–6) we find that $J_{i0} = I_{i0}$, $J_{0i} = I_{0i}$, and that $J_{ijk} = I_{ijk} + \frac{1}{2}(\delta_{ij} I_0 k + \delta_{jk} I_{i0})$.

Lemma 4.7.2 *Let I_{ij} be iterated Itô integrals with respect to independent Wiener processes and a determinstic drift term, indexed by $\{0, 1, \ldots, d\}$. The following identities describing the expectation of their products hold:*

$$E(I_{i0} I_{j0}) = E(I_{0i} I_{0j}) = 2\delta_{ij} \frac{t^3}{6}, \tag{4.96}$$

$$E(I_{i0} I_{0j}) = E(I_{0i} I_{j0}) = \delta_{ij} \frac{t^3}{6}. \tag{4.97}$$

These may be verified by computing the quasi-shuffles of the two words and using the algebraic expectation map. Computing

$$||Q||^2 = E((V_{i0} I_{i0} + V_{0i} I_{0i})(V_{i0} I_{i0} + V_{0i} I_{0i})), \tag{4.98}$$

we obtain

$$||Q||^2 = \frac{t^3}{6} \begin{pmatrix} V_{10} \\ V_{20} \\ V_{01} \\ V_{02} \end{pmatrix}^T \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} V_{10} \\ V_{01} \\ V_{20} \\ V_{02} \end{pmatrix}. \tag{4.99}$$

Similarly, we calculate

$$2\langle Q, R_1 \rangle = 2E\left((V_{ijk} \frac{1}{2} (\delta_{jk} I_{i0} + \delta ij I_{0k}))(V_{i0} I_{i0} + V_{0i} I_{0i}) \right), \tag{4.100}$$

where we have omitted the determinstic terms in R_1 as the expectations $E(tI_{ij})$ are all zero. We find

$$2\langle Q, R_1\rangle = \frac{t^3}{12}\begin{pmatrix} V_{111} \\ V_{122} \\ V_{211} \\ V_{222} \\ V_{112} \\ V_{221} \\ V_{10} \\ V_{20} \\ V_{01} \\ V_{02} \end{pmatrix}^T \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} V_{111} \\ V_{122} \\ V_{211} \\ V_{222} \\ V_{112} \\ V_{221} \\ V_{10} \\ V_{20} \\ V_{01} \\ V_{02} \end{pmatrix}. \tag{4.101}$$

Explicitly it may be shown that

$$R_3 = \frac{t^3}{12}\begin{pmatrix} V_{111} \\ V_{122} \\ V_{211} \\ V_{222} \\ V_{112} \\ V_{221} \\ V_{10} \\ V_{20} \\ V_{01} \\ V_{02} \end{pmatrix}^T \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 & 0 & 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 2 & 3 & 1 & 0 & 0 & 4 & 0 & 2 \\ 3 & 1 & 0 & 0 & 0 & 2 & 2 & 0 & 4 & 0 \\ 0 & 0 & 1 & 3 & 2 & 0 & 0 & 2 & 0 & 4 \end{pmatrix} \begin{pmatrix} V_{111} \\ V_{122} \\ V_{211} \\ V_{222} \\ V_{112} \\ V_{221} \\ V_{10} \\ V_{20} \\ V_{01} \\ V_{02} \end{pmatrix}. \tag{4.102}$$

The matrix has positive and negative eigenvalues. We conclude that, in the Stratonovich case, truncation at word order does not necessarily result in a scheme more efficient than truncating using the mean-square grading.

Chapter 5

Map-truncate-invert Schemes

Having laid the algebraic groundwork, we are now in a position to compare integration schemes of the form $\hat{\varphi} = f^{-1} \circ \pi \circ f(\varphi)$, which we associate with endomorphisms $(f^*)^{-1} \circ \pi \circ f^* \in \mathbb{H}$. We aim to construct integration schemes such that their associated remainder endomorphisms minimize within a certain class of schemes the norm induced by the inner product defined on the space of endomorphisms. We begin by surveying the results of Ebrahimi–Fard et al. (2012) which gave such a scheme for the case of Stratonovich flows for drift-diffusion equations. The results of this paper are accordingly derived using the shuffle/de-concatenate Hopf algebra associated with Stratonovich integration, in particular the definitions of expectations and inner products of endomorphisms differ from those given in Definitions 4.6.1 and 4.6.2, although they are perfectly analogous. We conclude the chapter by examining the proof of the above and highlighting obstacles to its immediate generalization to equations driven by a broader class of driving process. Particularly, the proofs rely on the shuffle (and not quasi-shuffle) algebraic structure and hence issues arise for driving processeses for which the Stratonovich formalism does not yield a shuffle algebra. We will examine resolutions of these issues in the chapters following this.

5.1 Stratonovich drift-diffusion equations

In this section we will consider exclusively Stratonovich drift-diffusion equations. Recall that the power series map $f(1+x) = \text{sinhlog}(1+x) = x + \frac{1}{2}\sum_{k=2}^{\infty}(-1)^{k-1}x^k$

induces a mapping of endomorphisms $f^*(X) = \frac{1}{2}(X - X^{*(-1)})$. The approximation of the flowmap $\hat{\varphi} = \text{sinhlog}^{-1} \circ \pi \circ \text{sinhlog}(\varphi)$, corresponding to the algebraic endomorphism $(\text{sinhlog}^*)^{-1} \circ \pi \circ \text{sinhlog}^*(\text{id})$ is an encoding of the sinhlog integrator considered in Malham & Wiese (2009). We note the identity $\text{sinhlog}^*(\text{id}) = \frac{1}{2}(\text{id} - S)$, where S is the shuffle convolution algebra antipode, we have similarly $\text{coshlog}^*(\text{id}) = \frac{1}{2}(\text{id} + S)$. Malham & Wiese (2009) established that, in the absence of drift, the approximate flowmap associated to the endomorphism $\text{sinhlog}^*(\text{id})$ possesses the desired efficiency of minimizing the coefficient of the leading order remainder term in comparison with the truncated stochastic Taylor approximation. The truncation utilized was that corresponding to the word-order grading. This was generalized to include drift in Ebrahimi–Fard et al. (2012). We follow the exposition of the latter paper in showing these results. The following result was critical in Ebrahimi–Fard et al. (2012), it is essentially a reformulation of Lemma 4.3 of Malham & Wiese (2009).

Lemma 5.1.1 *For all endomorphisms X, Y, we have $\langle X, Y\rangle = \langle R \circ X, R \circ Y\rangle$, where R is the reversal map.*

Before giving a proof of the above statement, we note that in the shuffle setting, the expectation map acquires a somewhat different form than that given in the previous section. This is because Stratonovich integrals with respect to martingales are not in general martingales. Indeed, we have

$$\hat{E} : w \mapsto \begin{cases} \frac{t^{n(w)}}{2^{d(w)} n(w)!}, & w \in \mathbb{D}^*, \\ 0, & w \in \mathbb{A}^* \setminus \mathbb{D}^* \end{cases} \tag{5.1}$$

where $\mathbb{D}^* \in \mathbb{A}^*$ free monoid on $\mathbb{D} = \{0, 11, \ldots, dd\}$, $d(w)$ is the number of consecutive pairs of non-zero letters in w, and $n(w)$ is length of w when viewed as an element of $\mathbb{D}^*$, see Ebrahimi–Fard et al (2012).

Proof: It is possible to prove the result directly through the formula for the expectation of iterated integrals with respect to Wiener processes given in Milstein (1995) and Kloeden & Platen (1999), as was done in Malham & Wiese (2009). However, we provide an alternative proof as we believe it is simpler, and is readily adaptable to the general discontinuous Itô setting, as we will show in Chapter 7. We begin by noting that $\hat{E}(R(w)) = \hat{E}(w)$. We see this as R preserves the spaces $\mathbb{D}^*$ and $\mathbb{A}^* \setminus \mathbb{D}^*$, and within these spaces $\hat{E}$ is evaluated by counting letters. The number of occurences of a letter in a given word is invariant under R. Note that the identity $E(R(w)) = E(w)$ also holds for E, the expectation associated with multiple Itô integrals. Note further that R is a shuffle homomorphism, as stated in Ebrahimi–Fard et al (2012). This suffices to give the result, as

$$\sum_{u,v} E\{R(u) \sqcup\!\sqcup R(v)\}(u,v) = \sum_{u,v} E\{R(u \sqcup\!\sqcup v)\}(u,v) = \sum_{u,v} E\{u \sqcup\!\sqcup v\}(u,v). \tag{5.2}$$

The following result of Ebrahimi–Fard et al. (2012) is a key step in establishing the main result of this section.

Lemma 5.1.2 *Let π_n be the canonical projection from $\mathbb{R}\langle\mathbb{A}\rangle$ onto its nth homogeneous component $\mathbb{R}\langle\mathbb{A}\rangle_n$ with respect to the grading $g(w) = |w|$. We then have*

1. $\langle S \circ \pi_n, S \circ \pi_n \rangle = \langle \mathrm{id} \circ \pi_n, \mathrm{id} \circ \pi_n \rangle$,
2. $\langle \mathbb{E} \circ S \circ \pi_n, \mathbb{E} \circ S \circ \pi_n \rangle = \langle \mathbb{E} \circ \mathrm{id} \circ \pi_n, \mathbb{E} \circ \mathrm{id} \circ \pi_n \rangle$,
3. $\langle \mathrm{sinhlog}^*(\mathrm{id}) \circ \pi_n, \mathrm{coshlog}^*(\mathrm{id}) \circ \pi_n \rangle = 0$,
4. $\langle \mathbb{E} \circ \mathrm{sinhlog}^*(\mathrm{id}) \circ \pi_n, \mathbb{E} \circ \mathrm{coshlog}^*(\mathrm{id}) \circ \pi_n \rangle = 0$.

These follow readily from the preceding Lemma as $S(u) \sqcup\!\sqcup S(v) = R(u) \sqcup\!\sqcup R(v)$ where u, v are of the same length (and $\mathbb{E} \circ \mathbb{E} = \mathbb{E}$). The main result of Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012) is the following.

Theorem 5.1.3 *The map-truncate-invert integration scheme induced by the sinhlog power series is efficient, in the sense that its leading order error norm is less than that obtained by truncating a Stochastic Taylor expansion, where we grade by word order on the shuffle-convolution algebra. This holds for all Stratonovich drift-diffusion equations where the driving vector fields are smooth in the sense that they have bounded derivatives of all orders.*

We comment that the method of proof utilized in Ebrahimi–Fard et al. (2012) can be realized as an implementation of Theorem 4.6.4 of this thesis. Indeed, there exists an explicit expression for the compositional inverse of sinhlog, given by

$$\mathrm{sinhlog}^{-1}(X) = X + (X^{\star 2} + \nu)^{\star(\frac{1}{2})} = \nu + X + \frac{1}{2}X^{\star 2} - \frac{1}{8}X^{\star 4} + \dots \tag{5.3}$$

It follows that the assumptions of Lemma 4.5.6 are satisfied, and in particular we obtain that the preremainder and remainder are equal at leading order. This result, in combination with Lemma 5.1.2 shows that the hypotheses of Theorem 4.6.4 are satisfied. The result follows immediately. Ebrahimi–Fard et al. (2012) obtained further results for related integration schemes by similar methods. For instance, it was shown that any integrator of the form $\gamma = A\mathrm{sinhlog}^*(\mathrm{id}) + B\mathrm{coshlog}^*(\mathrm{id})$ is efficient in this sense provided $A \neq 0$ and $|\frac{B}{A}| < 1$, indeed we obtain

$$||\mathrm{id} - R(\gamma)||^2 = \left(1 - \frac{B}{A}\right)^2 ||\mathrm{coshlog}^*(\mathrm{id})||^2, \tag{5.4}$$

and hence

$$||\mathrm{id}||^2 = ||R||^2 + \left[1 - \left(\frac{B}{A}\right)\right] ||\mathrm{coshlog}^*(\mathrm{id})||^2. \tag{5.5}$$

Ebrahimi–Fard et al. (2012) then noted that the sinhlog integrator ($A = 1, B = 0$) is therefore optimal in this class. Further optimality results concerning perturbations of the defining power series at leading order are given in Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012), we will not pursue this here.

5.2 Generalization

We wish to investigate the applicability of the previous results to numerical approximations of quasi-shuffle flowmaps of more general SDEs, given in Itô form. We assume we are able to give a quasi-shuffle product algebra representation of the flowmap. Consider the integration scheme $(f^*)^{-1} \circ \pi \circ f^* \in \mathbb{H}$ corresponding to the modified sinhlog series $f = \frac{1}{2}(\mathrm{id} - \hat{S})$, where $\hat{S}$ is the quasi-shuffle antipode given by (3.1.22). We have already noted that the first of the above lemmas continues to hold, therefore we might hope that such integrators will retain their efficiency. This will turn out not to be true in general. We begin by showing that Lemma 5.1.2 does not hold in the new setting. Consider the Itô drift-diffusion flowmap on the alphabet $\{0, 1\}$, i.e. corresponding to an SDE of the form

$$dY_t = V_0(Y_t)dt + V_1(Y_t)dW_t. \tag{5.6}$$

First consider the 'true' grading and its projectors π_n, we have $\pi_2(\mathbb{A}^*) = 11$, and $S(11) = 11 + 0$. It follows that

$$\langle S, S\rangle \circ \pi_2 = E\{S(11) * S(11)\}(11, 11) = \frac{3}{2}t^2(11, 11). \tag{5.7}$$

However, we have

$$\langle R, R\rangle \circ \pi_2 = R\{11 * 11\}(11, 11) = \frac{1}{2}t^2(11, 11) \neq \langle S, S\rangle \circ \pi_2. \tag{5.8}$$

A similar result holds where we use the word order truncation, we will obtain $\pi_2(\mathbb{A}^*) = \{11, 10, 01, 11\}$. Then $S = R$ on all words in this set except 11, and we obtain

$$\langle S, S\rangle \circ \pi_2 = \langle R, R\rangle \circ \pi_2 + E\{0*0\}(11, 11) + E\{0*00\}(11, 00) + E\{00*0\}(00, 11). \tag{5.9}$$

This precludes the possibility of an orthogonal decomposition of remainders

$$||\pi_n||^2 = ||\pi_n \circ \operatorname{sinhlog}^*(\mathrm{id})||^2 + ||\pi_n \circ \operatorname{coshlog}^*(\mathrm{id})||^2 \tag{5.10}$$

in all but exceptional cases. The problem lies in the differing form of the quasi-shuffle antipode. Before continuing our pursuit of efficient integrators from an algebraic perspective, we will provide a more detailed analysis of power-series based integrators and low-order schemes for drift-diffusion equations.

Chapter 6

Power series integrators

In this chapter we examine integration schemes induced by power series for Itô drift-diffusion equations of the form

$$dY_t = V_0(Y_t)dt + \sum_{i=1}^{d} V_i(Y_t)dW_t. \tag{6.1}$$

The integration schemes are map-truncate-invert schemes induced by convolution power series of the form $f^* = J + \alpha J^{\star 2} + \beta J^{\star 3} + \ldots$. We write out in full the quadratic form in the vector fields giving the difference between the norms of the leading order remainders of such integration schemes and the truncated stochastic Taylor scheme with the same order of convergence. This is achieved under the assumption that the pre-remainder of the power series schemes equals the remainder at leading order. The justification for this assumption is Lemma 4.5.6. We will show that for integration schemes of order half, any power series with α taking a value between -1 and 0 induces an integrator with error norm uniformly less than that of the stochastic Taylor integrator. The optimal value is $\alpha = -\frac{1}{2}$, corresponding to the sinhlog and Castell-Gaines integrators, which are equivalent at order half. In constrast, we will show that for integrators of strong order one, no such efficient power series integrator exists.

6.1 Order half integrators

The order half stochastic Taylor integrator is given by

$$\varphi^{st} = \sum_{i \geq 0} V_i I_i. \tag{6.2}$$

For ease of notation, we have dropped the tildes from the V_i. This means that, for the rest of this chapter, V_0 will denote the second order differential operator $\tilde{V}_0$. The power series integrator induced by $f^* = J + \alpha J^{\star 2} + \ldots$ is then given at leading order by

$$\varphi^{ps} = V_0 I_0 + \sum_{i>0} V_i I_i + \alpha V_{ii} I_0, \tag{6.3}$$

assuming that the pre-remainder equals the remainder at leading order. We have here employed the mean-square grading in the truncations. We therefore obtain leading order remainders

$$R^{st} = \sum_{i,j>0} V_{ij} I_{ij} \tag{6.4}$$

and

$$R^{ps} = \sum_{i,j>0} V_{ij} I_{ij} + \alpha \sum_{i,j>0} V_{ij}(I_{ij} + I_{ji}) = R^{st} + \tilde{R}. \tag{6.5}$$

As all the sums in the above are over $i, j > 0$ we will henceforth employ summation convention. We have $R^{st} = R^{ps} - \tilde{R}$, and hence

$$||R^{st}||^2 = ||R^{ps}||^2 - ||\tilde{R}||^2 - 2\langle R^{st}, \tilde{R}\rangle. \tag{6.6}$$

The following is a straightforward consequence of the formula for the expectation of the product of iterated Wiener integrals given in Milstein (1995) and Kloeden & Platen (1999), and is recorded here to simplify the following calculations.

Lemma 6.1.1 *Suppose $i, j, k, l > 0$. The following formula holds for the expectation of the product of iterated integrals of word order two.*

$$\mathbb{E}\{I_{ij} I_{kl}\} = \delta_{ik}\delta_{jl}\frac{t^2}{2}. \tag{6.7}$$

It follows that

$$
\begin{aligned}
||\tilde{R}||^2 &= \alpha^2 \langle V_{ij}(I_{ij} + I_{ji}), V_{kl}(I_{kl} + I_{lk}) \rangle && (6.8)\\
&= \alpha^2 V_{ij}^T V_{kl} \left[\langle I_{ij}, I_{kl} \rangle + \langle I_{ij}, I_{lk} \rangle + \langle I_{ji}, I_{kl} \rangle + \langle I_{ji}, I_{lk} \rangle \right] && (6.9)\\
&= \frac{t^2\alpha^2}{2} V_{ij}^T V_{kl} \left[\delta_{ij}\delta_{jl} + \delta_{il}\delta_{jk} + \delta_{jk}\delta_{il} + \delta_{jl}\delta_{ik} \right] && (6.10)\\
&= t^2\alpha^2 V_{ij}^T V_{kl} \left[\delta_{ij}\delta_{jl} + \delta_{il}\delta_{jk} \right] && (6.11)\\
&= t^2\alpha^2 (V_{ij}^T V_{ij} + V_{ij}^T V_{ji}). && (6.12)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\langle R^{st}, \tilde{R} \rangle &= \alpha \langle V_{ij} I_{ij}, V_{kl}(I_{kl} + I_{lk}) \rangle && (6.13)\\
&= \frac{\alpha t^2}{2} V_{ij}^T V_{kl} (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) && (6.14)\\
&= \frac{\alpha t^2}{2} \left[V_{ij}^T V_{ij} + V_{ij}^T V_{ji} \right]. && (6.15)
\end{aligned}
$$

Combining the above, we obtain

$$
||R^{st}||^2 = ||R^{ps}||^2 - \alpha(\alpha+1) \left[V_{ij}^T V_{ij} + V_{ij}^T V_{ji} \right]. \tag{6.16}
$$

We therefore wish to study the quadratic form $-\alpha(\alpha+1)\left[V_{ij}^T V_{ij} + V_{ij}^T V_{ji}\right]$. We may represent the above as $-\alpha(\alpha+1)V^T A V$, where V is a vector comprising all the vector fields V_{ij}, $i, j > 0$, and A a matrix. It remains to find the eigenvalues of the matrix A such that $V^T A V = V_{ij}^T V_{ij} + V_{ij}^T V_{ji}$.

Lemma 6.1.2 *The matrix A such that $V^T A V = V_{ij}^T V_{ij} + V_{ij}^T V_{ji}$ has eigenvalues of 2, with multiplicity $n + \frac{1}{2}n(n-1) = \frac{1}{2}n(n+1)$, and 0 with multiplicity $\frac{1}{2}n(n-1)$.*

Proof: We may decompose the matrix A as follows:

$$
V^T A V = \sum_{i>1} 2V_{ii}^T V_{ii} + \sum_{i,j>1} \begin{pmatrix} V_{ij} \\ V_{ji} \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} V_{ij} \\ V_{ji} \end{pmatrix}. \tag{6.17}
$$

The 2×2 matrix where every entry is one has eigenvalues of 2 and 0. The statement follows by counting the number of terms in each sum.

Theorem 6.1.3 *Let Y_t be the solution of a stochastic differential equation driven by independent Wiener processes and autonomous vector fields obeying the smoothness hypotheses to all orders. The order half power series integrator that minimizes the mean square remainder compared with that of the Euler scheme is that given by $f = J - \frac{1}{2}J^{\star 2} + \dots$.*

Proof: By the preceding Lemma, the matrix A is positive semidefinite. It therefore suffices to choose α to maximise $-\alpha(\alpha+1)$. This is positive for α between -1 and 0, attaining its maximum at $\alpha = -\frac{1}{2}$.

6.2 Order one integrators

Here we consider integrators induced by $f = J + \alpha J^{\star 2} + \beta J^{\star 3}$. The stochastic Taylor integrator is:

$$\hat{\varphi} = \sum_{g(w)=1} I_w V_w, \tag{6.18}$$

and the power series integrator is given by inverting

$$\hat{\psi} = \sum_{g(w)=1} I_w V_w + \sum_{u,v \in \mathbb{A}} \alpha V_{uv} I_u I_v, \tag{6.19}$$

where $\mathbb{A}$ is the set of all non-deterministic letters $\{1, \dots, d\}$. Again we assume that the pre-remainder and remainder are equal at leading order, following Lemma 4.5.6. We therefore obtain the following expressions for the leading order remainders of the above schemes.

$$R^{st} = \sum_{g(w)=1.5} I_w V_w, \tag{6.20}$$

$$\begin{aligned} R^{ps} \quad &= \sum_{g(w)=1.5} I_w V_w + \sum_{u \in \mathbb{A}} \alpha (V_{u0} + V_{0u}) I_{u \sqcup\!\sqcup 0} && (6.21)\\ &+ \sum_{u,v,w \in \mathbb{A}} V_{uvw} \{ (\alpha + \beta) \left[I_{u \sqcup\!\sqcup v \sqcup\!\sqcup w} + \delta_{uv} I_{w \sqcup\!\sqcup 0} + \delta_{uw} I_{v \sqcup\!\sqcup 0} + \delta_{vw} I_{u \sqcup\!\sqcup 0} \right] && (6.22)\\ &- \alpha \left[I_{wvu} - I_{uvw} + \delta_{uv} I_{w0} + \delta_{vw} I_{0u} \right] \}, && (6.23) \end{aligned}$$

where we have, for instance, used the relation

$$uv \sqcup\!\sqcup w + u \sqcup\!\sqcup vw = u \sqcup\!\sqcup v \sqcup\!\sqcup w - (wvu - uvw) \tag{6.24}$$

to relate the terms $\alpha V_{uvw}(I_{uv \sqcup\!\sqcup w} + I_{u \sqcup\!\sqcup vw})$ arising from $\alpha J^{\star 2}$ with the terms $\beta V_{uvw}(I_{u \sqcup\!\sqcup v \sqcup\!\sqcup w})$ arising from $\beta J^{\star 3}$. We have similarly related the quasi-shuffle terms of these products. Recall that we may write $R^{ps} = R^{st} + \tilde{R}$, from which we obtain

$$||R^{st}||^2 = ||R^{ps}||^2 - ||\tilde{R}||^2 - 2\langle R^{st}, \tilde{R}\rangle. \tag{6.25}$$

Our aim is therefore to establish an explicit form for the difference $-||\tilde{R}||^2 - 2\langle R^{st}, \tilde{R}\rangle$ derived above. The coming calculations are simplified by noting the following simple consequences of the formula for the expectations of iterated Wiener interals found in Milstein (1995) and Kloeden & Platen (1999).

Lemma 6.2.1 *For $u, v \in \mathbb{A}_+$, we have the following formulae for the expectations of products of iterated integrals of word order two.*

$$\mathbb{E}\{I_{u \sqcup\!\sqcup 0} I_{v0}\} = \delta_{uv}\frac{t^3}{2}, \tag{6.26}$$

$$\mathbb{E}\{I_{u \sqcup\!\sqcup 0} I_{0v}\} = \delta_{uv}\frac{t^3}{2}, \tag{6.27}$$

$$\mathbb{E}\{I_{u \sqcup\!\sqcup 0} I_{v \sqcup\!\sqcup 0}\} = \delta_{uv} t^3. \tag{6.28}$$

Lemma 6.2.2 *Suppose that $u, v, w \in \mathbb{A}_+$. The following formulae for the expectations of products of iterated integrals of word order three hold.*

$$\mathbb{E}\{I_{u \sqcup\!\sqcup v \sqcup\!\sqcup w} I_{p \sqcup\!\sqcup q \sqcup\!\sqcup r}\} = \Gamma t^3, \tag{6.29}$$

$$\mathbb{E}\{I_{uvw} I_{p \sqcup\!\sqcup q \sqcup\!\sqcup r}\} = \Gamma(u, v, w, p, q, r)\frac{t^3}{6}, \tag{6.30}$$

$$\mathbb{E}\{I_{uvw} I_{pqr}\} = \delta_{up}\delta_{vq}\delta_{wr}\frac{t^3}{6}. \tag{6.31}$$

where Γ is defined such that $\Gamma(u, v, w, p, q, r) = 1$ if there exists a permutation σ of p, q, r for which $\sigma(p) = u, \sigma(q) = v, \sigma(r) = w$, and $\Gamma(u, v, w, p, q, r) = 0$ otherwise.

We treat first the term $-||\tilde{R}||^2$. Using the formulae for the leading order remainders R^{st} and R^{ps} previously derived, we obtain

$$\tilde{R} = \sum_{u,k\in\mathbb{A}} I_{u\sqcup 0}\left[\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup kk}\right] - \alpha\left[V_{ukk}I_{0u} + V_{kku}I_{u0}\right] \tag{6.32}$$

$$+ \sum_{u,v,w\in\mathbb{A}} V_{uvw}\left[(\alpha+\beta)I_{u\sqcup v\sqcup w} + \alpha(I_{uvw} - I_{wvu})\right]. \tag{6.33}$$

Write $\tilde{R} = P + Q$, where P is the first sum above and Q the second. We note that P and Q are orthogonal with respect to the inner product, so we have

$$||\tilde{R}||^2 = ||P||^2 + ||Q||^2. \tag{6.34}$$

We now treat separately the terms P and Q. We begin by splitting the sum up further into $Q = q_1 + q_2$, where q_1 is the term on the left in $\alpha+\beta$ and q_2 the term on the right in α. It follows that

$$||Q||^2 = ||q_1||^2 + ||q_2||^2 + 2\langle q_1, q_2\rangle. \tag{6.35}$$

Furthermore, we may calculate

$$||q_1||^2 = \sum_{u,v,w;\sigma\in S_3} (\alpha+\beta)^2 t^3 \left[V_{uvw}^T V_{\sigma(uvw)}\right] \tag{6.36}$$

and

$$||q_2||^2 = \sum_{u,v,w} \frac{\alpha^2 t^3}{3}\left[V_{uvw}^T V_{uvw} - V_{uvw}^T V_{wvu}\right]. \tag{6.37}$$

Note that by the second relation of Lemma 6.2.2, we have

$$\mathbb{E}\{I_{uvw} I_{p\sqcup q\sqcup r}\} = \mathbb{E}\{I_{wvu} I_{p\sqcup q\sqcup r}\}. \tag{6.38}$$

It follows that $\langle q_1, q_2\rangle = 0$. We again write $P = p_1 + p_2$ where p_1 is the term in $I_{u\sqcup 0}$ and p_2 the term in I_{u0} and I_{0u}, although note that this decomposition will not be orthogonal as above. Explicitly, we have

$$p_1 = \sum_{u,k\in\mathbb{A}} I_{u\sqcup 0}\left[\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup kk}\right] \tag{6.39}$$

and

$$p_2 = \sum_{u,k\in\mathbb{A}} -\alpha\left[V_{ukk}I_{0u} + V_{kku}I_{u0}\right]. \tag{6.40}$$

First we compute $||p_2||^2$. We have

$$||p_2||^2 = \sum_{u,k,k'} \frac{\alpha^2 t^3}{6}\left[2V^T_{ukk}V_{uk'k'} + V^T_{kku}V_{uk'k'} + V^T_{ukk}V_{k'k'u} + 2V^T_{kku}V_{k'k'u}\right]. \tag{6.41}$$

We follow by computing $||p_1||^2$; we obtain

$$||p_1||^2 = \sum_{u,k,k'} t^3[\alpha V^T_{u\sqcup 0} + (\alpha+\beta)V^T_{u\sqcup kk}][\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup k'k'}]. \tag{6.42}$$

The inner product is calculated to be the following

$$\langle p_1, p_2\rangle = \sum_{u,k,k'} -\frac{\alpha t^3}{2}V^T_{u\sqcup k'k'}[\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup kk}]. \tag{6.43}$$

We obtain the following coefficients of t^3 in the sum for $||P^2||$:

$$\begin{aligned}
V^T_{u\sqcup 0}V_{u\sqcup kk} &: \alpha(\alpha+\beta) - \frac{1}{2}\alpha^2 = \alpha(\frac{1}{2}\alpha+\beta), && (6.44)\\
V^T_{u\sqcup 0}V_{u\sqcup 0} &: \alpha^2, && (6.45)\\
V^T_{ukk}V_{uk'k'}\ ,\ V^T_{kku}V_{k'k'u} &: \beta(\alpha+\beta) + \frac{1}{3}\alpha^2, && (6.46)\\
V^T_{ukk}V_{k'k'u}\ ,\ V^T_{kku}V_{uk'k'} &: \beta(\alpha+\beta) + \frac{1}{6}\alpha^2. && (6.47)
\end{aligned}$$

The following coefficients contain contributions from both $||P||^2$ and $||Q||^2$:

$$\begin{aligned}
V^T_{ukk}V_{ukk}\ ,\ V^T_{kku}V_{kku} &: (\alpha+\beta)(2\alpha+3\beta) + \frac{2}{3}\alpha^2, && (6.48)\\
V^T_{ukk}V_{kku}\ ,\ V^T_{kku}V_{ukk} &: (\alpha+\beta)(2\alpha+3\beta) - \frac{1}{6}\alpha^2, && (6.49)
\end{aligned}$$

this follows as there are two relevant permutations in the sum of $||q_1||^2$ each adding an $(\alpha+\beta)^2$ term, and we add or subtract a term in $\frac{\alpha^2}{3}$ from $||q_2||^2$ according to which of the above entries we are in.

We now calculate the second term in the difference of $||R^{st}||^2$ and $||R^{ps}||^2$; this is the term $-2\langle R^{st},\tilde{R}\rangle$. Recalling the identity

$$R^{st} = \sum_{i,j,k} V_{ijk}I_{ijk} + \sum_{u} V_{u0}I_{u0} + V_{0u}I_{0u}, \tag{6.50}$$

we call the first sum above r_1 and the second r_2. It is clear then that

$$\langle R^{st},\tilde{R}\rangle = \langle r_1, Q\rangle + \langle r_2, P\rangle, \tag{6.51}$$

where we recall that P and Q were given by the following expressions.

$$\begin{aligned} P &= \sum_{u,k\in\mathbb{A}} I_{u\sqcup 0}\left[\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup kk}\right] - \alpha\left[V_{ukk}I_{0u} + V_{kku}I_{u0}\right] && (6.52)\\ &:= p_1 + p_2. && (6.53)\\ Q &= \sum_{u,v,w\in\mathbb{A}} V_{uvw}\left[(\alpha+\beta)I_{u\sqcup v\sqcup w} + \alpha(I_{uvw} - I_{wvu})\right]. && (6.54)\end{aligned}$$

We obtain

$$\langle r_1, Q\rangle = \sum_{u,v,w;\sigma} V^T_{uvw}\left[(\alpha+\beta)V_{\sigma(uvw)} + \alpha(V_{uvw} - V_{wvu})\right]\frac{t^3}{6}. \tag{6.55}$$

We then calculate $\langle r_2, P\rangle$ using the decomposition $P = p_1 + p_2$. The contribution from $\langle r_2, p_2\rangle$ is

$$[V^T_{u0}(V_{ukk} + 2V_{kku}) + V^T_{0u}(2V_{ukk} + V_{kku})]\frac{-\alpha t^3}{6}. \tag{6.56}$$

and that from $\langle r_2, p_1\rangle$ is

$$V^T_{u\sqcup 0}(\alpha V_{u\sqcup 0} + (\alpha+\beta)V_{u\sqcup kk})\frac{t^3}{2}. \tag{6.57}$$

We obtain the following coefficients of t^3 in the expansion for $\langle r_2, P\rangle$:

$$\begin{aligned} V^T_{u\sqcup 0}V_{u\sqcup 0} &: \frac{\alpha}{2}, & (6.58)\\ V^T_{u0}V_{ukk}\ ,\ V^T_{0u}V_{kku} &: \frac{1}{3}\alpha+\frac{1}{2}\beta, & (6.59)\\ V^T_{u0}V_{kku}\ ,\ V^T_{0u}V_{ukk} &: \frac{1}{6}\alpha+\frac{1}{2}\beta. & (6.60)\end{aligned}$$

Gathering and summing all the previous results, we obtain the following coefficients of t^3 in the sum.

$$\begin{aligned} V^T_{u\sqcup 0}V_{u\sqcup 0} &: \alpha(\alpha+1), & (6.61)\\ V^T_{u0}V_{ukk}\ ,\ V^T_{0u}V_{kku} &: \frac{1}{3}\alpha+\frac{1}{2}\beta+\alpha\beta+\frac{1}{2}\alpha^2, & (6.62)\\ V^T_{u0}V_{kku}\ ,\ V^T_{0u}V_{ukk} &: \frac{1}{6}\alpha+\frac{1}{2}\beta+\alpha\beta+\frac{1}{2}\alpha^2, & (6.63)\\ V^T_{ukk}V_{uk'k'}\ ,\ V^T_{kku}V_{k'k'u} &: \beta(\alpha+\beta)+\frac{1}{3}\alpha^2, & (6.64)\\ V^T_{ukk}V_{k'k'u}\ ,\ V^T_{kku}V_{uk'k'} &: \beta(\alpha+\beta)+\frac{1}{6}\alpha^2, & (6.65)\\ V^T_{uvw}V_{\sigma(uvw)} &: (\alpha+\beta)(\alpha+\beta+\frac{1}{3}), & (6.66)\\ V^T_{uvw}V_{uvw} &: \frac{1}{3}\alpha(\alpha+1), & (6.67)\\ V^T_{uvw}V_{wvu} &: -\frac{1}{3}\alpha(\alpha+1), & (6.68)\\ V^T_{ukk}V_{ukk}\ ,\ V^T_{kku}V_{kku} &: (\alpha+\beta)(2\alpha+3\beta+\frac{2}{3})+\frac{1}{3}\alpha(2\alpha+1), & (6.69)\\ V^T_{ukk}V_{kku}\ ,\ V^T_{kku}V_{ukk} &: (\alpha+\beta)(2\alpha+3\beta+\frac{2}{3})-\frac{1}{6}\alpha(\alpha+2), & (6.70)\\ V^T_{kuk}V_{ukk}\ ,\ V^T_{kuk}V_{kku}\ ,\ V^T_{kuk}V^T_{kuk} &: 2(\alpha+\beta)(\alpha+\beta+\frac{1}{3}). & (6.71)\end{aligned}$$

We must also consider carefully the coefficients of terms in V_{uuu}. We expect that we obtain the new expressions by summing existing ones, but doing so by hand is a useful check. Note that the relevant terms in $\tilde{R}$ are

$$V_{uuu}[(\alpha+2\beta)I_{u\sqcup 0}+6(\alpha+\beta)I_{uuu}]. \qquad (6.72)$$

We compute first

$$2\langle R^{st},\tilde{R}\rangle = V^T_{uuu}\{2(\alpha+\beta)t^3V_{uuu}+\frac{1}{2}(\alpha+2\beta)t^3V_{u\sqcup 0}\}+\ldots, \qquad (6.73)$$

where the ... signify that we have omitted a term in $V_{u\sqcup 0}^T V_{uuu}$ whose coefficient is clear from the symmetry of the matrix. We obtain the following coefficients of t^3 in the expansion for $||\tilde{R}||^2$:

$$V_{uuu}^T V_{u\sqcup 0} \quad : \quad \alpha(\alpha+2\beta) + \frac{\alpha+2\beta}{2} = (\alpha+\frac{1}{2})(\alpha+2\beta), \tag{6.74}$$

$$V_{uuu}^T V_{uuu} \quad : \quad 2(\alpha+\beta) + 6(\alpha+\beta)^2 + (\alpha+2\beta)^2, \tag{6.75}$$

$$: \quad 6(\alpha+\beta)(\alpha+\beta+\frac{1}{3}) + (\alpha+2\beta)^2. \tag{6.76}$$

These are consistent with the previous coefficients; the latter coming from 6 terms in $(\alpha+\beta)(\alpha+\beta+\frac{1}{3})$ and

$$(\alpha+2\beta)^2 = 4\beta(\alpha+\beta) + \frac{2}{3}\alpha(2\alpha+1) - \frac{1}{3}\alpha(\alpha+2). \tag{6.77}$$

Note that we can't sum up the values for $V_{ukk}^T V_{ukk}$, $V_{kku}^T V_{kku}$, $V_{ukk}^T V_{kku}$, $V_{kku} V_{ukk}$ here as we will end up counting 8 permutations of uuu, not 6. The final value (coefficient of t^3) is

$$V_{uuu}^T V_{u\sqcup kk} : (\alpha+2\beta)(\alpha+\beta) - \frac{\alpha}{2}(\alpha+2\beta) = \frac{1}{2}\alpha^2 + 2\beta(\alpha+\beta). \tag{6.78}$$

The last value is as we expect.

6.2.1 Error form analysis

As in the order half case, we may write the quadratic form $-||\tilde{R}||^2 - 2\langle R^{st}, \tilde{R}\rangle$ in the form $V^T A V$, where V is a vector of indeterminates $(V_{10}, V_{01}, \ldots)$ corresponding to the vector fields, and A is a matrix. To obtain an integration scheme with mean square error less than that associated with the stochastic Taylor scheme, we would then require that A be positive semidefinite. As before, this form will split into a sum $\sum_i V_i^T A_i V_i$ where the V_i are disjoint. The eigenvalues will then be the totality of eigenvalues of the A_i. As iterated integrals of the form I_{uvw}, $u, v, w \in \mathbb{A}_+$ are orthogonal to those of the form I_{u0}, $u, \in \mathbb{A}_+$, the decomposition will be twofold. We will require both quadratic forms to be positive definite. The first part comprises a sum over all vectors of indeterminates

$$(V_{uvw}, V_{uwv}, V_{vuw}, V_{wuv}, V_{vwu}, V_{wvu}) \tag{6.79}$$

corresponding to the threefold composition of vector fields with distinct indices $u, v, w \geq 1$. The matrices for each distinct set of indices u, v, w is the same. We obtain

$$\begin{pmatrix} V_{uvw} \\ V_{uwv} \\ V_{vuw} \\ V_{wuv} \\ V_{vwu} \\ V_{wvu} \end{pmatrix}^T (\gamma A + \delta B) \begin{pmatrix} V_{uvw} \\ V_{uwv} \\ V_{vuw} \\ V_{wuv} \\ V_{vwu} \\ V_{wvu} \end{pmatrix}, \tag{6.80}$$

where $\gamma = \frac{1}{3}\alpha(\alpha+1)$, $\delta = (\alpha+\beta)(\alpha+\beta+\frac{1}{3})$, B is the 6x6 matrix where every entry is a 1, and

$$A = \begin{pmatrix} 1 & & & & & -1 \\ & 1 & & & -1 & \\ & & 1 & -1 & & \\ & & -1 & 1 & & \\ & -1 & & & 1 & \\ -1 & & & & & 1 \end{pmatrix}. \tag{6.81}$$

There are values of α and β for which this expression is negative semidefinite, and hence its negation is positive semidefinite as we wish. Both the sinhlog and log integrators possess this property. We hence focus on the other sum in the decomposition. The second part of the decomposition comprises vectors of indeterminates corresponding to compositions of vector fields with non-distinct indices or indices including a zero. Again, there is a sum $\sum V_i^T A_i V_t$ here, but the matrices A_i are the same. They are given by

$$\begin{pmatrix} V_{u0} \\ V_{0u} \\ V_{ukk} \\ V_{kku} \\ V_{kuk} \\ V_{uk'k'} \\ V_{k'k'u} \\ V_{k'uk'} \\ V_{uuu} \end{pmatrix}^T \begin{pmatrix} \kappa & \kappa & \gamma & \delta & 0 & \gamma & \delta & 0 & \zeta \\ \kappa & \kappa & \delta & \gamma & 0 & \delta & \gamma & 0 & \zeta \\ \gamma & \delta & \nu & \mu & \phi & \psi & \xi & 0 & \eta \\ \delta & \gamma & \mu & \nu & \phi & \xi & \psi & 0 & \eta \\ 0 & 0 & \phi & \phi & \phi & 0 & 0 & 0 & 0 \\ \gamma & \delta & \psi & \xi & 0 & \nu & \mu & \phi & \eta \\ \delta & \gamma & \xi & \psi & 0 & \mu & \nu & \phi & \eta \\ 0 & 0 & 0 & 0 & 0 & \phi & \phi & \phi & 0 \\ \zeta & \zeta & \eta & \eta & 0 & \eta & \eta & 0 & \omega \end{pmatrix} \begin{pmatrix} V_{u0} \\ V_{0u} \\ V_{ukk} \\ V_{kku} \\ V_{kuk} \\ V_{uk'k'} \\ V_{k'k'u} \\ V_{k'uk'} \\ V_{uuu} \end{pmatrix} t^3, \tag{6.82}$$

where the coefficients of the matrix are

$$\kappa = \alpha(\alpha+1), \tag{6.83}$$
$$\phi = 2(\alpha+\beta)(\alpha+\beta+\frac{1}{3}), \tag{6.84}$$
$$\delta = \frac{1}{6}\alpha + \frac{1}{2}\beta + \alpha\beta + \frac{1}{2}\alpha^2, \tag{6.85}$$
$$\gamma = \frac{1}{3}\alpha + \frac{1}{2}\beta + \alpha\beta + \frac{1}{2}\alpha^2, \tag{6.86}$$
$$\psi = \beta(\alpha+\beta) + \frac{1}{3}\alpha^2, \tag{6.87}$$
$$\xi = \beta(\alpha+\beta) + \frac{1}{6}\alpha^2, \tag{6.88}$$
$$\zeta = \frac{\alpha}{2} + \beta + \alpha\beta + \alpha^2, \tag{6.89}$$
$$\eta = 2\beta(\alpha+\beta) + \frac{\alpha^2}{2}, \tag{6.90}$$
$$\nu = \frac{8}{3}\alpha^2 + 5\alpha\beta + 3\beta^2 + \alpha + \frac{2}{3}\beta, \tag{6.91}$$
$$\mu = \frac{11}{6}\alpha^2 + 5\alpha\beta + 3\beta^2 + \frac{\alpha}{3} + \frac{2}{3}\beta, \tag{6.92}$$
$$\omega = 7\alpha^2 + 16\alpha\beta + 10\beta^2 + 2(\alpha+\beta). \tag{6.93}$$

This matrix possesses at least one negative eigenvalue for all values of α and β except $\alpha = \beta = 0$. The eigenvalues were computed numerically. We present two visualizations of the smallest eigenvalue of the above matrix for general values of α and β. The first is a contour plot, and the second is a surface in three dimensions.

In particular, we see that there are no values of α and β for which the matrix A is positive semidefinite. It follows that we may not obtain an efficient integrator in the sense of Section 2.7 of this thesis of order 1 using power series methods.

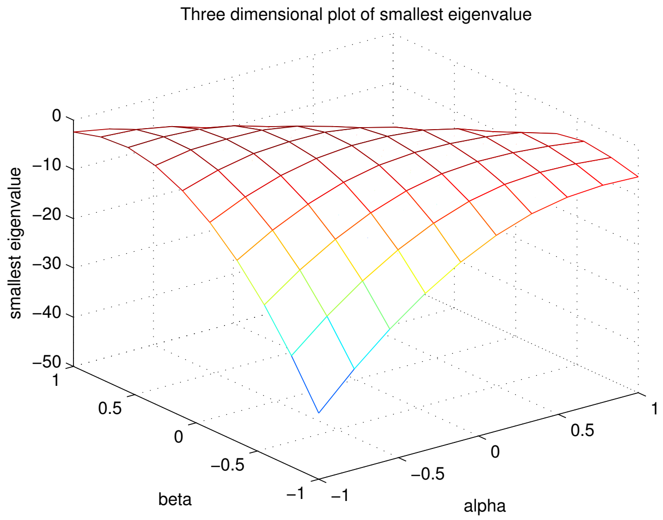

Figure 6.1: Contours of smallest eigenvalue of the difference of remainder forms between general power series integrator and stochastic Taylor integrator

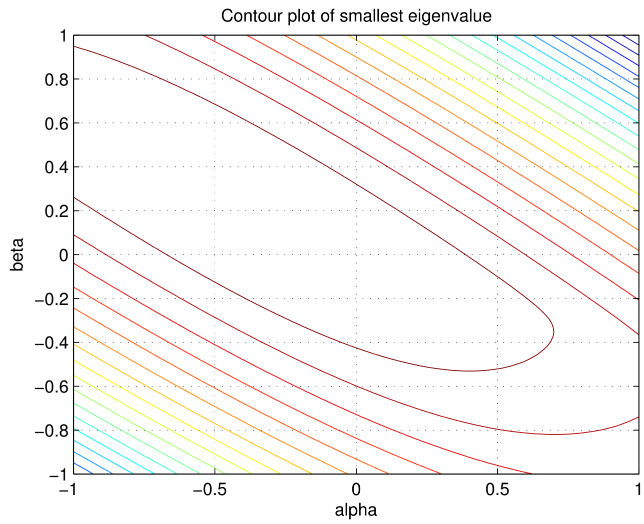

Figure 6.2: 3D plot of smallest eigenvalue of the difference of remainder forms between general power series integrator and stochastic Taylor integrator

Chapter 7

One-step Schemes and Sign-reverse Integrator

In this chapter, we introduce a new method of generalizing the sinhlog integrator of Malham & Wiese (2009) designed to preserve its efficiency properties for broader classes of driving processes. In doing so, we will examine a new method for realizing map-truncate-invert schemes in a single step, removing many of the difficulties inherent in inverting such schemes. We begin with a discussion of the algebraic encoding of power-series schemes, and show how the algebraic structure naturally gives rise to a broader class of scheme. This provides the stimulus for introducing the new sign-reverse integrator, a scheme arising from the consideration of the sinhlog integrator. In the main result of this chapter and indeed the thesis, to be published in Curry et al. (2014b), we show that this integration scheme is efficient for Lévy-driven equations with sufficiently smooth vector fields, in the same sense as the sinhlog integrator of Malham & Wiese.

Present knowledge, however, does not provide a method of computing the inversion step in this map-truncate-invert scheme. Accordingly, in the next section we introduce a new perspective for analysing map-truncate-invert schemes. We show that the terms up to the desired strong order of convergence coincide with the schemes arising from a truncation of the stochastic Taylor expansion. They differ, however, in that they possess terms of higher order than the truncation. We show how these are related to the remainder endomorphisms discussed previously, and show how they may be generated from lower order terms utilizing the algebraic structure. In doing so we resolve the difficulties inherent in inverting such schemes. We then return to our discussion of the sign-reverse integrator, showing how it fits into this framework. We conclude with a brief discussions of stochastic differential equations where the stochastic and geometric information contained in the flowmap may not be separated, and comment on the extent to which our integration schemes may be applied there.

7.1 The sign-reverse integrator

We adopt the following perspective. We have viewed integration schemes under two lenses, first as approximations of the flowmap of the form $\hat{\varphi} = f^{-1} \circ \varphi \circ f(\varphi)$ for smooth, invertible f; and second as approximations induced by the algebraic maps $(f^*)^{-1} \circ \pi \circ f^* \in \mathbb{H}$. We showed in the section on encoding integration schemes that the homomorphic embedding $\psi : \mathbb{H} \to \bar{\mathcal{A}}$, where $\bar{\mathcal{A}} = K\langle \mathbb{A} \rangle \otimes K\langle \mathbb{A} \rangle_*$ provides a relation between f and f^* above where f is a power series. Composition powers of the identity do not however exhaust the space $\mathrm{End}(K\langle \mathbb{A} \rangle)$, the reversal map R is one example of a map not expressible as a power series. We will therefore expand our search to include approximations to the flowmap in the wider image space $\psi(\mathbb{H})$. It is worth pausing a moment to reflect on the role of the convolution product in the broader setting. Note that the only place in the computation of $\varphi = \mathrm{sinhlog}^{-1} \circ \pi \circ \mathrm{sinhlog}^*(\mathrm{id})$ where we use the convolution structure is in the expression of the algebraic map $\mathrm{sinhlog}^{-1}$; all the products are standard compositions. Indeed, the power of the convolution formalism is that it allows us to compute composition inverses through Lemma 4.5.3.

Definition 7.1.1 *Let $f(1+x) = \mathrm{sinhlog}(1+x)$. The sign-reverse integration scheme is defined to be the map-truncate-invert scheme*

$$\hat{\varphi}_{t_n,t_{n+1}} := \left[(\kappa \otimes \mu)_{t_n,t_{n+1}} \circ \psi\right] \circ \left[(f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})\right], \qquad (7.1)$$

where the convolution powers are taken in the shuffle/deconcatenate convolution algebra, regardless of the quasi-shuffle algebra structure of the iterated integrals.

The sign-reverse takes its name from the relation $f^*(\mathrm{id}) = \frac{1}{2}(\mathrm{id} + (-1)^{|w|+1}R(w))$. We can not in general give an expression for these maps as convolution power series, except where the convolution algebra is a shuffle algebra. To see this, compare the action by $S(w) := (-1)^{|w|}R(w)$ and a map $f := \alpha J + \beta J^{\star 2}$ on a general word of length two ; we have

$$f(ab) = (\alpha + \beta)ab + \beta ba + \beta[a, b]. \tag{7.2}$$

Comparing this to the identity $S(ab) = ba$, we see that $-\alpha = \beta = 1$, and that $[a, b] = 0$ for all a, b. This implies that the operation $[,]$ is trivial and the quasi-shuffle algebra is therefore the shuffle algebra.

Lemma 7.1.2 *For all endomorphisms X, Y, we have $\langle X, Y\rangle = \langle R \circ X, R \circ Y\rangle$, where R is the reversal map.*

Proof: As $E(w) = 0$ unless $w \in \{0\}^*$, we have $E(w) = E(R(w))$. Moreover, by Proposition 4.2 of Hoffman & Ihara (2011), R is a homomorphism for any quasi-shuffle algebra to itself. The result then follows as

$$\sum_{u,v} E\{R(u) * R(v)\}(u, v) = \sum_{u,v} E\{R(u * v)\}(u, v) = \sum_{u,v} E\{u * v\}(u, v). \tag{7.3}$$

Let π_n be the projector of $\mathbb{R}\langle\mathbb{A}\rangle$ onto its nth homogeneous component in the local mean square grading $g(w) = \frac{1}{2}(|w| + \xi(w))$; for this section we will use this form of the mean-square grading, remembering to add the expectation of leading order terms. Define $\mathbb{A}_g = \mathrm{Im}(\pi_n)$. We now note the following.

Lemma 7.1.3 (Orthogonal decomposition) *The subspaces $\mathbb{A}_g$ decompose orthogonally (with respect to the inner product) into word-homogenous subspaces $\oplus_n \mathbb{A}_g^n$.*

Proof: Eliminating $\xi(w)$ from the simultaneous equations

$$\begin{aligned} \frac{1}{2}\zeta(w) + \xi(w) &= g(w) && (7.4) \\ \zeta(w) + \xi(w) &= |w|, && (7.5) \end{aligned}$$

we obtain $|w| = g(w) + \frac{1}{2}\zeta(w)$. Recall that if the polynomial $w * w'$ is to have non-zero expectation, the words w, w' must have the same number of non-zero letters, ie $\zeta(w) = \zeta(w')$. If we have $w, w' \in \mathbb{A}_g$, some g, then by definition $g(w) = g(w')$. It follows that $|w| = |w'|$, as required.

We have the following.

Corollary 7.1.4 *Let π_n be the projector of $\mathbb{R}\langle\mathbb{A}\rangle$ onto its nth homogeneous component in the grading $g(w) = \frac{1}{2}(|w| + \xi(w))$, and let $\langle , \rangle$ be the inner product on the quasi-shuffle algebra. We then have:*

1. $\langle S \circ \pi_n, S \circ \pi_n \rangle = \langle \mathrm{id} \circ \pi_n, \mathrm{id} \circ \pi_n \rangle$,
2. $\langle \mathbb{E} \circ S \circ \pi_n, \mathbb{E} \circ S \circ \pi_n \rangle = \langle \mathbb{E} \circ \mathrm{id} \circ \pi_n, \mathbb{E} \circ \mathrm{id} \circ \pi_n \rangle$,
3. $\langle \mathrm{sinhlog}^*(\mathrm{id}) \circ \pi_n, \mathrm{coshlog}^*(\mathrm{id}) \circ \pi_n \rangle = 0$,
4. $\langle \mathbb{E} \circ \mathrm{sinhlog}^*(\mathrm{id}) \circ \pi_n, \mathbb{E} \circ \mathrm{coshlog}^*(\mathrm{id}) \circ \pi_n \rangle = 0$,

where $\mathrm{sinhlog}^*(\mathrm{id})$ *and* $\mathrm{coshlog}^*(\mathrm{id})$ *are given by convolution power series in the shuffle/deconcatenate convolution algebra.*

Proof: The first point holds as, by the orthogonal decomposition lemma, all contributing terms in the expectations $E(S(u) * S(v))$ are expectations of products of words of the same length. We therefore have $E((-1)^{|u+v|}R(u) * R(v)) = E(R(u) * R(v))$, and the result follows from Lemma 7.1.2. The rest are immediate consequences of the above, noting the identities $\mathrm{sinhlog}^*(\mathrm{id}) = \frac{1}{2}(\mathrm{id} - S)$ and $\mathrm{coshlog}^*(\mathrm{id}) = \frac{1}{2}(\mathrm{id} + S)$.

If we adopt the word order grading, the proof of the analogous result is similar, indeed we do not need the orthogonal decomposition lemma.

Lemma 7.1.5 *Let π_n be the projector onto the nth homogeneous component according to the word-order grading. With the usual inner product, we obtain*

1. $\langle S \circ \pi_n, S \circ \pi_n \rangle = \langle \mathrm{id} \circ \pi_n, \mathrm{id} \circ \pi_n \rangle$,
2. $\langle \mathrm{shl} \circ \pi_n, \mathrm{chl} \circ \pi_n \rangle = 0$.

Proof: As above, all words u, v in the image of π_n are of the same length. It follows that $(-1)^u(-1)^v = 1$, and we obtain the result using Lemma 7.1.2.

We are now in a position to state our main result.

Theorem 7.1.6 *Let φ be the flowmap of a stochastic differential equation driven by independent Lévy processes with moments of all orders. Assume that φ possesses a separated stochastic Taylor expansion, and that the driving vector fields satisfy the smoothness Hypotheses 4.2.2 to all orders. The sign-reverse approximation of a given order n is then 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, independent of the initial conditions. This holds for truncations at both the mean square grading and the word order grading.*

Proof: The main ingredient in the proof is Theorem 4.6.4. Let $g := (f^{-1})^* \circ \pi_{\leq n} \circ f^*(\mathrm{id})$, and define $h := \mathrm{id} - g$. The result follows if we can show that $\langle g, h \rangle = 0$. Note that sinhlog possesses a power series inverse, given by

$$\mathrm{sinhlog}^{-1}(x) = 1 + x + \frac{1}{2}x^2 + \dots, \tag{7.6}$$

see Ebrahimi–Fard et al. (2012). Furthermore, the sinhlog series is an expansion in convolution powers of the augmented ideal projector with zero constant term. The hypotheses of Lemma 4.5.6 are then satisfied, as the shuffle product in the convolutions preserves all grading functions. It follows that $g = \mathrm{sinhlog}^*(\mathrm{id})$ at leading order, and $h = \mathrm{coshlog}^*(\mathrm{id})$. The preceding lemmas then guarantee that the hypotheses of Theorem 4.6.4 hold, and the result therefore follows.

The one caveat is that we may now only compute the inverse at an algebraic level, and not with operations on the flowmap itself. In what follows, we explore a method of resolving this issue.

7.2 One-step realizations

We begin by noting that any endomorphism $X = f^{-1} \circ \pi_n \circ f$ inducing an integrator of strong order n is necessarily the identity on $\mathbb{A}_{\leq n}$, where $\mathbb{A}_{\leq n} = \pi_{\leq n}(\mathbb{A}^*)$ is the projection operator $\pi_{\leq n}$ according to the mean square grading, this is essentially Theorem 4.5.5. Furthermore, any higher order terms in $X = f^{-1} \circ \pi_n \circ f$ may be expressed in terms of functions of elements in $\mathbb{A}_{\leq n}$. Consider the strong order 1 sinhlog scheme for a Stratonovich drift-diffusion equation. It is possible to work out an explicit expression for the endomorphism $X = \text{sinhlog}^{-1} \circ \pi_{\leq 1} \circ \text{sinhlog}^*(\text{id}) : \mathbb{R}\langle \mathbb{A} \rangle \to \mathbb{R}\langle \mathbb{A} \rangle$. We will simplify matters by computing only the leading order terms, i.e. in this case $\pi_{\leq \frac{3}{2}} \circ X$. At leading order, we have $\text{sinhlog}^{-1}(X) = X + \frac{1}{2} X^{\star 2}$ and $\pi_{\leq 1} \circ \text{sinhlog}^*(\text{id}) = \pi_{\leq 1} - \frac{1}{2} J^{\star 2} \circ \pi_{\leq 1}$. Note the following identities:

$$\begin{aligned}
\pi_{\leq \frac{3}{2}} \circ (\pi_{\leq 1} * \pi_{\leq 1}) &= J^{\star 2} \circ \pi_{\leq \frac{3}{2}}, && (7.7)\\
\pi_{\leq \frac{3}{2}} \circ ((J^{\star 2} \circ \pi_{\leq 1}) * \pi_{\leq 1}) &= J^{\star 3} \circ \pi_{\frac{3}{2}}, && (7.8)\\
\pi_{\leq \frac{3}{2}} \circ (J^{\star 2} \circ \pi_{\leq 1})^{\star 2} &= 0. && (7.9)
\end{aligned}$$

We may now proceed:

$$\pi_{\leq \frac{3}{2}} \circ \text{sinhlog}^{-1} \circ \pi_{\leq 1} \circ \text{sinhlog}^*(\text{id})$$

$$\begin{aligned}
&= \pi_{\leq \frac{3}{2}} \circ (\pi_{\leq 1} - \frac{1}{2} J^{\star 2} \circ \pi_{\leq 1} + \frac{1}{2}(\pi_{\leq 1} - \frac{1}{2} J^{\star 2} \circ \pi_{\leq 1})^{\star 2}),\\
&= \pi_{\leq \frac{3}{2}} \circ (\pi_{\leq 1} - \frac{1}{2}(J^{\star 2} \circ \pi_{\leq 1} + \pi_{\leq 1}^{\star 2} - (J^{\star 2} \circ \pi_{\leq 1}) \circ \pi_{\leq 1} + \frac{1}{4}(J^{\star 2} \circ \pi_{\leq 1})^{\star 2})),\\
&= \pi_{\leq 1} + \frac{1}{2}(J^{\star 2} - J^{\star 3}) \circ \pi_{\frac{3}{2}}. && (7.10)
\end{aligned}$$

It follows that the integration scheme given by

$$\psi(X) = \sum_{g(w)\leq 1} V_w I_w + \frac{1}{2} \sum_{g(w)=\frac{3}{2}} V_w I_{(J^{\star 2} - J^{\star 3})} \tag{7.11}$$

is equivalent, at leading order, to that constructed by $\tilde{X} = \text{sinhlog}^{-1} \circ \pi_{\leq 1} \circ \text{sinhlog}^*(\text{id})$. We may establish this more simply however, from Lemma 4.5.6; if we have an expression for the remainder R at leading order (ie, $n + \frac{1}{2}$), the identity $\text{id} - R = X \circ \pi_{n+\frac{1}{2}}$ holds. In the sinhlog case, we have $R = Q$ where Q is the pre-remainder $\text{sinhlog}^*(\text{id}) \circ \pi_{\geq n+\frac{1}{2}}$. It follows that we may extend (7.11) to all orders immediately; we have

$$\psi(X) = \sum_{g(w)\leq n} V_w I_w + \frac{1}{2} \sum_{g(w)=n+\frac{1}{2}} V_w I_{(J^{\star 2} - J^{\star 3} + \ldots)(w)} \tag{7.12}$$

Note that a similar derivation allows us to perform a leading-order equivalent version of the Castell-Gaines scheme; $R = Q$ also holds in this setting. We also note the expression $\frac{1}{2}(J^{\star 2} - J^{\star 3} + \ldots)(w) = \text{coshlog}^*(\text{id})(w)$. This gives us an alternative method of proof of Theorem 7.1.6 by immediate application of the sinhlog/coshlog orthogonality. It also gives a different insight into the result of Ebrahimi–Fard et al. (2012) that $f = \text{coshlog}^*(\text{id})$ fails to induce a satisfactory integrator: the integrator includes a term in $\text{sinhlog}^*(\text{id})(w)$, containing in turn a term in $J(w)$ and hence may not be generated from lower order terms. Indeed, the computational power of (7.12) is that the terms of grading $> n$ are all of the form $J^{\star n}(w)$, that is to say they are all shuffles of lower order words. In practice, this means that we need only simulate the integrals I_w up to order n, and the remaining terms may be generated by products of these without incurring significant computational costs.

7.3 Signed-reverse integrator revisited

Note that the derivation of (7.12) is purely a computation of the algebraic endomorphism $\varphi_n = \text{sinhlog}^{-1} \circ \pi \circ \text{sinhlog}^*(\text{id})$. At no point does it invoke the result that the embedding $\mathbb{H} \to K\langle\mathbb{A}\rangle \bar{\otimes} K\langle\mathbb{A}\rangle$ is an algebra homomorphism, nor that the quasi-shuffle products of words correspond to real products of stochastic integrals. Recall that the integrator for more general Itô equations given in Section 7.1 corresponds to the identical endomorphism $\varphi_n \in \mathbb{H}$. It follows that we may derive the same algebraic expression

$$X = \mathrm{id} \circ \pi_{\leq n} + \mathrm{coshlog}^*(\mathrm{id}) \circ \pi_{n+1}, \tag{7.13}$$

and hence obtain the integrator

$$\psi(X) = \sum_{g(w) \leq n} V_w I_w + \frac{1}{2} \sum_{g(w)=n+1} V_w I_{J^{\star 2} - J^{\star 3} + \ldots (w)}, \tag{7.14}$$

where the convolutions $J^{\star 2}$ etc are in the shuffle convolution algebra, and not the more natural quasi-shuffle convolution. We will refer to the extra terms in such an integrator compared with the stochastic Taylor scheme as the remainder terms. We must note, however, that the last paragraph of the previous section no longer applies exactly; the integrals $I_{u \sqcup\!\sqcup v}$ are no longer equal to their real products $I_u I_v$. We can at least say that the components of greatest word order of $u \sqcup\!\sqcup v$ and $u * v$ are the same. For this reason we will always truncate according to the word order grading. We may therefore derive expressions of the form

$$I_{u \sqcup\!\sqcup v} = I_u I_v + \sum_{g(w) \in \mathbb{A}_{\leq n}} I_w. \tag{7.15}$$

Indeed, this observation is behind the result of Hoffman (1999) that the Lyndon words form a basis for the quasi-shuffle algebra. We now give a brief derivation of the remainder terms

$$X(\varphi) = \frac{1}{2} \sum_{g(w)=n+1} V_w I_{\mathrm{chl}(w)} \tag{7.16}$$

in terms of lower order iterated integrals (under the word order grading $g(w)$). This amounts to computing the difference between $I_{(J^{\star 2} - J^{\star 3} + \ldots)(w)}$ where the convolutions are quasi-shuffles, and the same term where the convolutions are shuffles. This is accomplished most readily by noting that in both cases $(J^{\star 2} - J^{\star 3} + \ldots)(w) = \frac{1}{2}(\mathrm{id} + S)(w)$, where S is the relevant antipode. It follows that this difference is

$\frac{1}{2}(S-\hat{S})$, where $\hat{S}$ is the quasi-shuffle antipode and S the shuffle antipode. Explicitly, we have

$$(S-\hat{S})(ijk) = [k,j]i + k[i,j] + [i,j,k]. \tag{7.17}$$

For instance, we see that the order 1 sign-reverse approximation of the drift-diffusion flowmap has remainder term:

$$X(\varphi) = \frac{1}{2} \sum_{ijk \geq 1:\, i \neq k} V_{ijk}(I_{ij}I_k + I_iI_{jk} - I_iI_jI_k + \delta_{ij}I_{k0} + \delta_{jk}I_{0i}). \tag{7.18}$$

Note that the expression

$$(J^{\sqcup\!\sqcup 2} - J^{\sqcup\!\sqcup 3})(ijk) = ij \sqcup\!\sqcup k - i \sqcup\!\sqcup kj = ijk - kji, \tag{7.19}$$

where $J^{\sqcup\!\sqcup n}$ denotes the n-fold convolution power in the shuffle/de-concatenate convolution algebra, implies that the terms in (7.18) where $i = k$ are zero, hence the restricted summation given. More generally, a similar expression may be derived from (7.17) in the various quasi-shuffle settings.

Chapter 8

Practical Implementation

We here give a brief overview of some of the practical issues pertaining to strong numerical schemes for Lévy-driven equations, before presenting the results of our numerical experiments.

8.1 Simulation

We give a brief sketch of certain issues concerning practical implementation of integrators for SDEs driven by Lévy processes. There are two main problems to surmount. First, simulating multiple integrals, already a difficult task for drift-diffusion SDEs, may be prohibitively difficult for Lévy driven equations. Second, even simulating the increments of a Lévy process is difficult in many cases. Knowledge of the canonical decomposition $Z = \alpha t + \sigma W_t + J_t$, where J_t is characterized through its random jump measure, may not provide a method of calculating the law of increments $\mathcal{L}(Z_{t+h} - Z_t)$ in the case that J_t is an infinite activity process. We also briefly discuss the consequences of the infinite alphabet appearing in the linearized flow.

The most important class of Lévy processes we can readily simulate consists of jump diffusions. A Lévy process is a jump diffusion if its Lévy measure is finite, i.e. $\nu(\mathbb{R}) = \lambda < \infty$. In the absence of drift and Brownian components, such a process is a compound Poisson process (see Bertoin 1996, p. 16), that is of the form

$$Z_t = \sum_{i=1}^{N_t} Y^i, \tag{8.1}$$

where N_t is a standard Poisson process, and the marks Y^i are independent, identically distributed random variables with a known distribution. In this instance, we may simulate Z_t by simulating the Poisson process $N = N_t$ to give the number of jumps in the interval considered. As the jump times are uniformly distributed on this interval, we sample N independent uniformly distributed jump times, and then sample Y^i according to the known law (see Cont & Tankov 2004, p.174). More challenging is the computation of the iterated integrals featuring a Wiener process and a compound Poisson process. In our simulations, we have computed

$$I_{12}(t) = \int_0^t W_s dZ_s = \sum_{i=1}^{N_t} (W(t_i) - W(t_{i-1}))Y^i, \tag{8.2}$$

where the $\{t_i\}$ are the jump times of Z, by simulating the increments of the Wiener process W_t at all the jump times. Note that this is computationally expensive where the intensity of the compound Poisson process is large. Furthermore, it is close in spirit to the jump-adapted schemes described in Bruti-Liberati & Platen (2010), an interesting question is how our analysis translates into this framework?

More generally, it may be shown that the integrability conditions on Lévy measures require $\nu(A) < \infty$ where $A \subset \mathbb{R}$ is bounded away from the origin. We may therefore hope to approximate Lévy processes of infinite intensity with jump diffusions by ignoring or approximating jumps below a given small magnitude ϵ. The simplest such approximation is given by

$$Z_t \approx Z_t^\epsilon = \alpha t + \beta W_t + \int_t \int_{|x|>\epsilon} x[Q(dx, ds) - s\nu(dx)]. \tag{8.3}$$

Let $R_t^\epsilon := Z_t - Z_t^\epsilon$. As we are working with compensated jump measures, we have $\mathbb{E}(R_t^\epsilon) = 0$. Defining $\sigma^2(\epsilon) := \int_{|x|<\epsilon} x^2 \nu(dx)$, we find that $\mathrm{Var}(R_t^\epsilon) = t\sigma^2(\epsilon)$. The accuracy of this approximation depends on the rate at which $\sigma(\epsilon)$ converges to zero. Various results concerning weak convergence of this approximation are given in Cont & Tankov (2004). Integration schemes based on performing such an approximation and applying an Euler method have been studied at length by Protter & Talay (1997), Jacod et al. (2005), but only in the weak sense. We can often improve the above in approximating the small jumps by a Brownian motion. Critical is the following result, due to Asmussen & Rosinski, (2001):

Theorem 8.1.1 *$\sigma(\epsilon)^{-1}R^\epsilon \to W$ in distribution as $\epsilon \to 0$, where W a standard Brownian motion, if and only if, $\forall k > 0$,*

$$\frac{\sigma(k\sigma(\epsilon) \wedge \epsilon)}{\sigma(\epsilon)} \to 1, \quad \text{as} \quad \epsilon \to 0. \tag{8.4}$$

A sufficient condition for the above to hold is that $\epsilon^{-1}\sigma(\epsilon) \to \infty$ as $\epsilon \to 0$, necessary if ν has no atoms in a neighbourhood of 0 as will often be the case.

Weak convergence of integration schemes based on solving approximate SDEs driven by $\hat{Z}_t^\epsilon = Z_t^\epsilon + \sigma(\epsilon)W_t$ has been investigated by Jacod et al (2005), for instance. A recent paper of Fournier (2011) showed strong convergence in the case of a single driving Lévy process. The difficulty is that we need convergence of the above approximation in the Wasserstein metric, not just in distribution. Fournier was able to use a recent result of Rio (2009) on convergence of the central limit in the Wasserstein metric. Rio's result only applies for convergence to a one-dimensional Gaussian random variable, hence the restriction to a single driving process, although it is conjectured that this restriction is not necessary. Furthermore Rosinski (2001) showed a class of alternative methods for approximating Lévy processes by jump diffusions based on summing a random number of terms of an infinite series. Interesting questions remain concerning the application of such approximations to strong numerical schemes.

We remark that, in the case of the linearized flow, the free monoid $\tilde{\mathbb{A}}^*$ is not locally finite with respect to the word order or variance grading, that is to say for each graded homogenous component there may be an infinite number of terms. To obtain an implementable numerical integration scheme from its truncation, we would therefore have to truncate the Taylor expansion employed in the linearization after a fixed number of terms, say N. For such a scheme to be practical, the Taylor series would have to exhibit rapid convergence. At present, such expansions seem to be of greater theoretical than practical interest.

8.2 Numerical experiments

We here present a number of numerical experiments concerning the results given above. First, consider the drift-diffusion equation driven by linear vector fields

$$dY_t = V_0 Y_t dY_t + V_1 Y_t dW_t^1 + V_2 Y_t dW_t^2, \tag{8.5}$$

where the W^i are independent, standard Wiener processes, $Y_t \in \mathbb{R}^2$ with initial condition $Y_0 = (1, 0.5)^T$ and the linear fields are given by the matrices

$$V_0 = \begin{pmatrix} -0.0721 & -0.2173 \\ -0.1719 & -0.9581 \end{pmatrix}, \tag{8.6}$$

$$V_1 = \begin{pmatrix} 0.0800 & 0.5769 \\ -0.5961 & -0.9619 \end{pmatrix}, \quad \text{and} \quad V_2 = \begin{pmatrix} -0.9438 & 0.5520 \\ -0.4684 & 0.1591 \end{pmatrix}. \tag{8.7}$$

We compare the global mean square error $\mathbb{E}(\sup_{0\le t\le T} |Y_t - \hat{Y}_t|^2)^{1/2}$, estimated by sampling 10000 paths, for three different numerical schemes: (i) the Milstein scheme

$$\hat{Y}_{t_n} = \left(\Delta_{t_n} V_0 + \sum_{i=1,2} \Delta W^i(t_n) V_i + \sum_{i,j=1,2} \Delta I_{ij}(t_n) V_j V_i \right) \hat{Y}_{t_{n-1}}, \tag{8.8}$$

where $\Delta_{t_n} = t_n - t_{n-1} = h$ (we will consider only schemes with a fixed, uniform discretization grid), $\Delta W^i(t_n) = W^i(t_n) - W^i(t_{n-1})$, and $\Delta I_{ij}(t_n) = \int_{t_{n-1}}^{t_n} W^i dW^i$;

(ii) The order 1 word-order truncated stochastic Taylor expansion:

$$\hat{Y}_{t_n} = \left(\Delta_{t_n} V_0 + \sum_{i=1,2} \Delta W^i(t_n) V_i + \sum_{i,j=0,1,2} \Delta I_{ij}(t_n) V_j V_i \right) \hat{Y}_{t_{n-1}}, \tag{8.9}$$

where I_{ij} is defined as above, using the convention $W^0(t) = t$, and (iii) the Sign-reverse (word order) scheme

$$\begin{aligned} \hat{Y}_{t_n} &= \Bigg(\Delta_{t_n} V_0 + \sum_{i=1,2} \Delta W^i(t_n) V_i \sum_{i,j=1,2} \Delta I_{ij}(t_n) V_j V_i \\ &+ \sum_{i,j,k=1,2} (\Delta I_{ij} \Delta W^k(t_n) + \Delta W^i \Delta I_{jk}(t_n) - \Delta W^i \Delta W^j \Delta W^k(t_n)) V_k V_j V_i \\ &+ \sum_{i,j,k=1,2} (\delta_{ij} \Delta I_{k0}(t_n) + \delta_{jk} \Delta I_{0i}(t_n)) V_k V_j V_i \Bigg) \hat{Y}_{t_{n-1}}. \end{aligned} \tag{8.10}$$

In figure 8.1 we have plotted the global mean square error of the three numerical

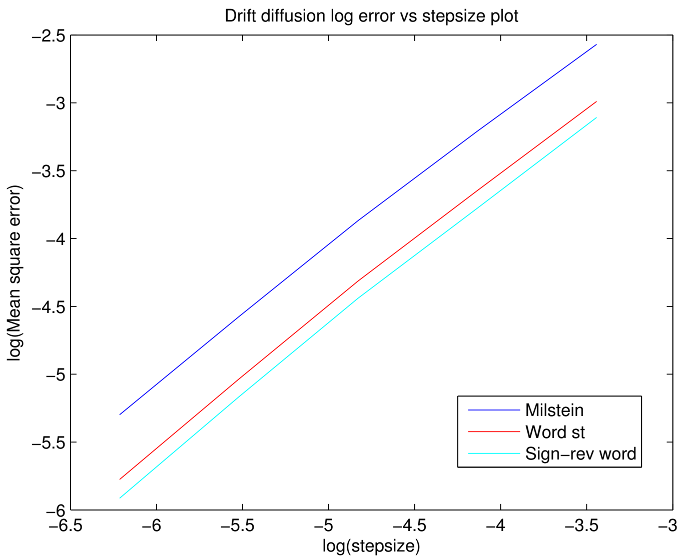

Figure 8.1: log-log plot of mean square error against stepsize for equation (8.5) with parameters as given, derived through numerical means

approximations against the discretization grid stepsize (the exact solution being unavailable, we have computed an approximate exact solution using an order 1 integrator with a smaller stepsize and accumulated the increments of the stochastic processes when computing the approximations of a higher stepsize). We observe strong convergence of order 1 in all three methods, as predicted. Furthermore,

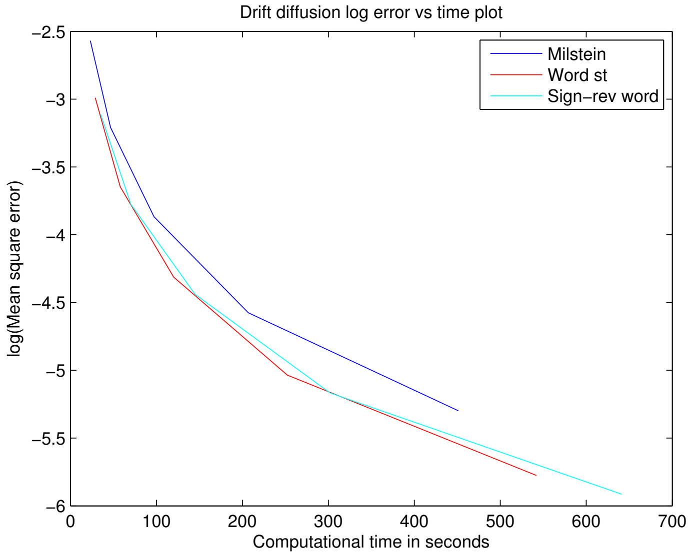

Figure 8.2: Plot of log(mean square error) against computation time for equation (8.5) with parameters as given, derived through numerical means

in line with our analytic results, we notice that 'sign-reverse' integrator gives a significant improvement on the word-order stochastic Taylor scheme, which is in turn significantly more accurate than the Milstein scheme. Explicit integration schemes for the equation (8.5) show less tendency to blowup when eigenvalues of the driving fields are of less than unit magnitude, and have been chosen thus. Note, however, that the theoretical improvement in mean-square error of the sign-reverse scheme over the word order Taylor scheme is measured in terms of the form

$$\frac{1}{3}V_kV_jV_i(V_kV_jV_i - V_iV_jV_k).y_n. \tag{8.11}$$

Heuristically, it follows that this will be small (compared with the improvement from Milstein to word-order Taylor) unless the commutators $[V_i, (V_j)^2]$ have eigenvalues of comparable magnitude to the vector fields. Thus, generically for linear drift-diffusion equations with driving matrices with spectra contained in the unit disc, the improvement demonstrated by the sign-reverse integrator is likely to be more modest than that shown in figure 8.1. Of greater practical interest is the question of whether the improvement in accuracy resulting from the extra terms in the word order and sign reverse integrators. In figure 8.2 we have plotted the global mean square error of the approximations against the total computational time used by each integration scheme in the same simulation as above. We see that, in this case, both the word order stochastic Taylor and sign-reverse outperform the Milstein scheme according to this criterion. There is little difference between the word order Taylor and sign-reverse schemes however, the extra computations required by the latter in this case roughly offset the improved accuracy.

Consider the equation

$$dY_t = V_0 Y_t dY_t + V_1 Y_t dW_t + V_2 Y_t dJ_t, \tag{8.12}$$

where W_t is a standard Wiener process and J_t is the compound Poisson process $J_t = \sum_{i=1}^{N_t} Y_i$, where N_t is a (standard) Poisson process with intensity λ, and the independent, identically distributed marks $Y_i \sim \mathcal{N}(0, \lambda)$. $Y_t \in \mathbb{R}^2$ and the initial condition is again $Y_0 = (1, 0.5)^T$. We compare the three order 1 schemes: (i) Milstein, (ii) word-order stochastic Taylor (iii) sign-reverse integrator, where the first two are as given before (employing the convention $W^1(t) = W(t); W^2(t) = J(t)$), and the sign-reverse integrator is given by:

$$\begin{aligned}
\hat{Y}_{t_n} &= \Bigg(\Delta_{t_n} V_0 + \sum_{i=1,2} \Delta W^i(t_n) V_i + \sum_{i,j=1,2} \Delta I_{ij}(t_n) V_j V_i \\
&+ \sum_{i,j,k=1,2} (\Delta I_{ij} \Delta W^k(t_n) + \Delta W^i \Delta I_{jk}(t_n) - \Delta W^i \Delta W^j \Delta W^k(t_n)) V_k V_j V_i \\
&+ \sum_{i,j,k=1,2} (\mathbb{I}_{i=j=1} \Delta I_{k0}(t_n) + \mathbb{I}_{j=k=1} \Delta I_{0i}(t_n)) V_k V_j V_i \\
&+ \sum_{i,j,k=1,2} (\mathbb{I}_{i=j=2} \Delta I_{k2^{(2)}}(t_n) + \mathbb{I}_{j=k=2} \Delta I_{2^{(2)}i}(t_n)) V_k V_j V_i \Bigg) \hat{Y}_{t_{n-1}},
\end{aligned} \tag{8.13}$$

where $I_{2^{(2)}i}(t_n) = \int_{t_{n-1}}^{t_n} [J]_s dW_s^i$ and $I_{i2^{(2)}}(t_n) = \int_{t_{n-1}}^{t_n} W_s^i d[J]_s$, $[J]$ being the quadratic variation process $[J]_t = \sum_0^{N_t} Y_i^2$. We have approximated the iterated integrals using the methods outlined earlier. For our simulations, we have taken intensity parameter $\lambda = 43.40$, and matrices

$$V_0 = \begin{pmatrix} -0.6841 & 0.5588 \\ 0.9335 & -0.0273 \end{pmatrix}, \tag{8.14}$$

$$V_1 = \begin{pmatrix} 0.9530 & -0.6357 \\ 0.9556 & -0.3250 \end{pmatrix}, \quad \text{and} \quad V_2 = \begin{pmatrix} -0.1763 & -0.9088 \\ 0.3382 & 0.8925 \end{pmatrix}. \tag{8.15}$$

As before, we observe in figure 8.3 strong convergence of order 1, with the sign-

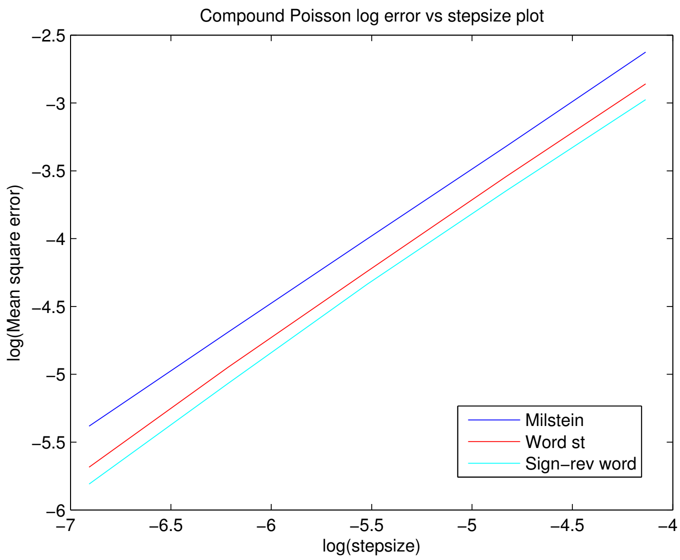

Figure 8.3: log-log plot of mean square error against stepsize for equation (8.12) with parameters as given, derived through numerical means

reverse integrator more accurate than the word-order stochastic Taylor integrator, which in turn is more accurate the Milstein scheme. Note that the previous remarks on generic behaviour of the integration schemes and the particular vector fields chosen apply equally here. In this instance, however, the Milstein scheme obtains the better results when we factor in the additional computational time taken by the other schemes. This is because, unlike in the previous case, we require extra computational effort to simulate the additional iterated integrals such as I_{i0}.

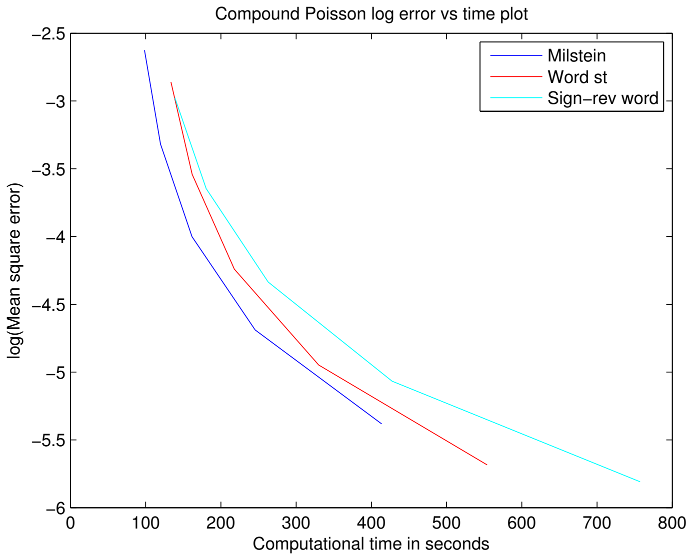

Figure 8.4: plot of log(mean square error) against computational time for equation (8.12) with parameters as given, derived through numerical means

Chapter 9

Conclusion and Future Work

We have shown how to express the flowmap of a stochastic differential equation driven by independent Lévy processes, with moments of all orders and autonomous vector fields that are uniformly bounded with uniformly bounded derivatives of all orders, in the form

$$\varphi = \sum_{w} \tilde{V}_w I_w, \tag{9.1}$$

where the $\tilde{V}_w$ are operators constructed from the driving vector fields V_i and the I_w are the multiple iterated integrals with respect to the driving processes. From this, we have constructed an integration scheme based on the approximation of the above flowmap in the form

$$\hat{\varphi} = \sum_{|w|\leq n} \tilde{V}_w I_w + \frac{1}{2} \sum_{|w|=n+1} \tilde{V}_w(I_{(\mathrm{id}+s)(w)}), \tag{9.2}$$

utilising the fact that $(\mathrm{id} + S)(w)$ may be expressed as linear combinations in the quasi-shuffle algebra of words of length smallar than $n + 1$. It was shown that the above scheme is efficient in the sense that the coefficient of its leading order remainder is always less than or equal to that of the integration schemee

$$\hat{\varphi} = \sum_{|w|\leq n} V_w I_w, \tag{9.3}$$

obtained by simple truncation of the stochastic Taylor series. As the above integrator is not a power series integrator as discussed in Chapter 6, however, we can no longer show the optimality of this scheme with respect to certain perturbations, as was shown in Malham & Wiese (2009) and Ebrahimi–Fard et al. (2012). The following question is then natural. Let $V \subset \mathbb{H}$ be the subspace of endomorphisms f of $K\langle\mathbb{A}\rangle$ such that $f(w)$ is computable from linear combinations of quasi-shuffles of words of lower order. Any power series map of the form $f^* = c_2 J^{\star 2} + c_3 J^{\star 3} + \ldots$ is in this space, for instance, as is $(\mathrm{id} + S)$. We have endowed $\mathbb{H}$ with the structure of an inner product space. If we can show that this is a Hilbert space, and V is as closed subspace, then there exists an orthogonal projection $P_V : \mathbb{H} \to V$. Under our formalism, this would show the optimality of the integration scheme

$$\hat{\varphi} = \sum_{|w| \leq n} V_w I_w + \sum_{|w|=n+1} V_w I_{P_V(\mathrm{id})(w)}. \tag{9.4}$$

Further questions arise concerning the applicability of our results. We have assumed that the vector fields are sufficiently smooth for the stochastic Taylor expansions to exist to all orders. This is somewhat restrictive, and perhaps unnecessarily so. The convergence of integration schemes arising from the truncation of the stochastic Taylor series only requires the existence of uniformly bounded derivatives up to the order of the remainder, and it seems likely that such conditions would be sufficient for the schemes considered here. To establish this through algebraic methods would, however, require the algebraic encodings to be modified to allow for truncated series representations of approximate flows.

As was noted in the section on practical experiments, the improved accuracy obtained by the sign-reverse integration scheme comes at the expense of increased computational effort. This is to be expected, indeed this was discussed in detail in Newton (1986), and the order one sign-reverse integration scheme for drift-diffusion equations driven by a single Wiener process is exactly Newton's order one asymptotically efficient integrator. Newton (1986) considered the problem of when the increased accuracy could be considered worth the extra computational effort. A detailed appraisal of the same considerations applied to our integration scheme in some generality would be valuable. In the simulations presented earlier, for the compound Poisson case the extra computational effort was found to dominate. This is largely due to the inefficient method utilized to compute the iterated integrals appearing in the expansions. It was commented that the methods employed were similar in spirit to the jump-adapted schemes described in Bruti-Liberati & Platen (2010). There the discretization grid is generated by superimposing the jump times of the compound Poisson processes on the deterministic grid of fixed step size. In such a situation, any iterated integrals involving the compound Poisson processes vanish. In view of the detrimental effect on performance arising from the simulation of these iterated integrals, it would be extremely useful to examine whether the integration schemes derived here translate to the jump-adapted framework.

In many cases, the computation of the operators $\tilde{V}_w$, containing derivatives of the driving vector fields is complicated and may even dominate the computational cost of the numerical scheme. For such situations it is advisable to consider the implementation of derivative-free schemes, where the composition of operators $\tilde{V}_w$ are replaced with finite differences. It would be useful to examine whether our results our applicable to such schemes. At the least, Newton (1991) derived an asymptotically efficient, derivative free integration scheme of order one for drift-diffusion equations driven by a single Wiener process.

It was commented in Chapter 8 that for drift-diffusion equations, the improvement in accuracy attained by the sign-reverse integrator compared with the truncated Taylor methods is greatest where the magnitudes of the diffusion coefficients dominate those of the drift coefficient. Such equations are liable to blow up, and are therefore best tackled using implicit schemes (see Milstein & Tretyakov, 2004). The derivation of an analogous implicit integration scheme using algebraic methods would therefore be of substantial practical and theoretical interest. Indeed, fully implicit schemes are required for the derivation of symplectic structure-preserving integration schemes for stochastic differential equations on manifolds, see Milstein, Repin & Tretyakov (2002). The generalization of our methods to equations on manifolds in general would be of interest, see Malham & Wiese (2008).

We conclude by commenting that the establishment of a quasi-shuffle algebra structure for iterated Lévy integrals has potential implications beyond the derivation of numerical integration schemes. The algebra of iterated stochastic integrals has applications to, for instance, chaotic expansions of representation of martingales, see Jamshidian (2011). Whether the quasi-shuffle algebra structure may be applied in such situations remains a topic for future exploration.

Bibliography

[1] Abe, E. 1980 *Hopf algebras.*, Cambridge University Press.

[2] Appelbaum, D., 2004 *Lévy Processes and Stochastic Calculus*, Cambridge studies in advanced mathematics 93, Cambridge University Press, Cambridge.

[3] Asmussen, S., Rosiński, J. 2001 Approximations of small jumps of Lévy processes with a view towards simulation, *J. Appl. Probab.,* **38** 482–493

[4] Azencott, R. 1982 Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman., *Séminaire de Probabilités XVI*, Volume 921 of *Lecture Notes in Mathematics*, 237–285. Springer-Verlag, Berlin.

[5] Bardou, F., Bouchard, J., Aspect, A., Cohen-Tannoudji, C. 2002 *Lévy Statistics and Laser Cooling*, Cambridge University Press

[6] Bates, D. 1996 Jumps and stochastic volatility: the exchange rate processes implicit in Deutschemark options, *Rev. Fin. Studies* **9** 69–107

[7] Baudoin, F. 2004 *An introduction to the geometry of stochastic flows.*, London, UK: Imperial College Press.

[8] Barndorff-Nielsen, O. E., 1997 Normal inverse Gaussian distributions and stochastic volatility modelling, *Scand. J. Statist.,* **24** 1–13.

[9] Ben Arous, G. 1989 Flots et series de Taylor stochastiques., *Probab. Theory Related Fields* **81**, 29–77.

[10] Bertoin, J. 1996 *Lévy Processes*, Cambridge: Cambridge University Press.

[11] Black, F., Scholes, M., 1973 The Pricing of Options and Corporate Liabilities., *Journal of Political Economy* **81**(3), 637–654

[12] Boyarchenko, S., Levendorskii, S., 2002 Perpetual American options under Lévy processes, *SIAM J. Control Optim.* **40**, 1663–1696.

[13] Boyarchenko, S., Levendorskii, S., 2011 Optimal stopping in Lévy models, for non-monotone discontinuous payoffs, *SIAM J. Control Optim.* **49**(5), 2062–2082.

[14] Bruti-Liberati, N., Platen, E. 2005 Strong approximations of stochastic differential equations with jumps, *Journal of Computational and Applied Mathematics* **205** 982–1001

[15] Bruti-Liberati, N., Platen, E. 2010 *Numerical Solution of Stochastic Differential Equations with Jumps in Finance.*, Springer.

[16] Buckwar, E., Malham, S.J.A., Wiese, A., 2012 An introduction to SDE simulation., Preprint. Heriot-Watt University.

[17] Cartier, P., 1972 On the structure of free baxter algebras, *Advances in Mathematics* **9**(2), 253–265.

[18] 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.

[19] Castell, F., Gaines, J., 1996 The ordinary differential equation approach to asymptotically efficient schmes for solution of stochastic differential equations., *Ann. Inst. H. Poincaré Probab. Statist.* **32**, 231–250.

[20] Chen, K.-T. 1957 Integrations of paths, geometric invariants and a generalized Baker–Hausdorff formula., *Annals of Mathematics*, **65**, 163–78.

[21] Chen, 1961 Formal differential equations, *Annals of Mathematics*, **73**, 110–133.

[22] Chen, K.-T. 1968 Algebraic paths., *Journal of Algebra*, **10**, 8–36.

[23] 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**, Springer-Verlag, Berlin, New York

[24] Connes, A., Kreimer, D. 2000 Renormalization in quantum field theory and the Riemann-Hilbert problem. I. The Hopf algebra structure of graphs and the main theorem., *Comm. Math. Phys.*, 210(1) 249–273.

[25] Cont, R., Tankov, P. 2004 *Financial Modelling with Jump Processes*, Chapman & Hall/CRC

[26] Courrège, P. 1963 Intégrale stochastique par rapport à une martingale de carré intégrable, in *Séminaire de Théorie du Potentiel*, Faculté des Sciences de Paris.

[27] Curry, C., Ebrahimi–Fard, K., Malham, S.J.A., Wiese, A. 2014 Lévy processes and quasi-shuffle algebras, *Stochastics* **86**(4), 632–642.

[28] Curry, C., Ebrahimi–Fard, K., Malham, S.J.A., Wiese, A. 2014b Efficient integrators for systems driven by Lévy processes., In preparation.

[29] Davis, M.H.A. 2005 Martingale Representation and All That, in *Advances in Control, Communication Networks, and Transportation Systems, ed. Abed, E.H.*, Bosteon, Birkaüser, 57–68.

[30] Doléans–Dade, C., Meyer, P.A., 1970 Intégrales stochastiques par rapport aux martingales locales, in *Séminaire de Probabilités IV*, Volume 124 of *Lecture Notes in Mathematics*, 77–107 Springer-Verlag, Berlin.

[31] Doob, J.L. 1984 *Classical Potential Theory and Its Probabilistic Counterpart* Springer-Verlag, Berlin, Heidelberg, New York

[32] 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* doi:10.1098/rspa.2012.0024.

[33] Ebrahimi–Fard, K., Patras, F. 2013. *Rota-Baxter Algebra. The Combinatorial Structure of Integral Calculus*, arXiv:1304.1204v1 [math.RA].

[34] Fliess, M. 1981 Fonctionelles causales non linéaires et indéterminées non-commutatives, *Bull. Société Mathémat. France* **109**, 3–40.

[35] Fournier, N. 2011 Simulation and Approximation of Lévy-driven Stochastic Differential Equations, *ESAIM: Prob. Stat.* **15** 233–248

[36] Gaines, J. 1994 The algebra of iterated stochastic integrals., *Stochast. Stochast. Rep.* **49**, 169–179

[37] Gray W.S., Duffaut Espinosa, L.A. 2011 Faà di Bruno Hopf algebra for a group of Fliess operators with applications to feedback, *Systems Control Lett.*, **60** 441449.

[38] Gray, W.S., Duffaut Espinosa, L.A., Ebrahimi–Fard, K., 2014 Recursive Algorithm for the Antipode in the SISO Feedback Product *Proc. 21st International Symposium on the Mathematical Theory of Networks and Systems*, Groningen, The Netherlands, to appear.

[39] Grossman, R., Larson, R.G. 1989 Hopf-algebraic structure of families of trees, *J. Algebra* **126**(1), 184–210.

[40] Hairer, M., 2014 A theory of regularity structures., *Invent. Math.* doi:10.1007/s00222-014-0505-4.

[41] Hairer, M., Kelly, D., 2014 Geometric vs non-geometric rough paths., To appear in *Annals of IHP*.

[42] Hoffman, M.E. 2000 Quasi-shuffle products., *Journal of Algebraic Combinatorics* **11**, 49–68.

[43] Hoffman, M.E., Ihara, K. 2012 Quasi-shuffle products revisited., Preprint. Max-Planck-Institut für Mathematik Bonn.

[44] Ikeda, N., Watanabe, S., 1989 *Stochastic Differential Equations and Diffusion Processes (2nd ed.)*., North-Holland/Kodansha, Tokyo, Amsterdam, New York.

[45] Itô, K. 1951 On stochastic differential equations., *Mem. Amer. Math. Soc.* **4**

[46] Jacod, J., Kurtz, T.G., Méléard, S., Protter, P. 2005 The approximate Euler method for Lévy driven stochastic differential equations, *Ann. institut H. Poincaré (B) Prob. et Stat.* **41**(3) 523–558.

[47] Jacod, N., Shiryaev, A.N., 1987 *Limit Theorems for Stochastic Processes.*, Springer.

[48] Jamshidian, F. 2005 Chaotic expansion of powers and martingale representation. Preprint.

[49] Jamshidian, F. 2011 On the combinatorics of iterated stochastic integrals., *Stochastics* **83**, 1–15.

[50] Kawski, M. 2001 The combinatorics of nonlinear controllability and noncommuting flows, Lectures given at the Summer School on Mathematical Control Theory, Trieste.

[51] Kloeden, P.E., Platen, E. 1999 *Numerical Solution of Stochastic Differential Equations* (3rd Printing), Berlin etc: Springer.

[52] Kloeden, P.E., Platen, E., Wright, I. 1992 The approximation of multiple stochastic integrals., *Stochastic Anal. Appl.* **10**(4), 431–441.

[53] Kou, S. 2002 A jump–diffusion model for option pricing, *Management Science* **48** 1086–1101.

[54] Kunita, H., Watanabe, S., 1967 On square integrable martingales., *Nagoya Math. J.* **30** 209–245.

[55] 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.

[56] Loday, J.-L. 2007 On the algebra of quasi-shuffles, *manuscripta mathematica* **123**(1), 79–93.

[57] Lord, G., Malham, S.J.A., Wiese, A., 2008 Efficient Strong Integrators for Linear Stochastic Systems, *SIAM J. Numer. Anal.* **46** 2892–2919.

[58] Lothaire, M., 1983, *Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **17**, Addison-Wesley.

[59] Lyons, T., Victoir, N. 2004 Cubature on Wiener space., *Proc. R. Soc. Lond. A* **460** 169–198.

[60] Lyons, T.J., 1998 Differential equations driven by rough signals., *Rev. Mat. Iberoamericana*, **14**(2), 215–310.

[61] Lyons, T., Caruana, M., Lévy, T., 2007 Differential equations Driven by Rough Paths, in *École d'été des probabilités de Saint-Flour XXXIV – 2004, ed. J. Picard*, volume 1908 of *Lecture Notes in Mathematics*, Berlin, Springer.

[62] Malham, S.J.A., Wiese, A. 2008 Stochastic Lie gorup integrators., *SIAM J. Sci. Comput.* **30**, 597–617.

[63] 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.

[64] Malham, S.J.A., Wiese, A. 2014 Efficient almost-exact Lévy area sampling, *Preprint, available at http://arxiv.org/pdf/1107.0151.pdf*

[65] Madan, D., Carr, P., Chang, E., 1998 The variance gamma process and option pricing, *Euorpean Finance Review* **2** 79–105.

[66] Marcus, S. 1978 Modeling and analysis of stochastic differential equations driven by point processes., *IEEE Transactions on Information Theory*, IT-24, No. 2 164–172.

[67] Marcus, S. 1981 Modeling and analysis of stochastic differential equations driven by semimartingales., *Stochastics* **4**, 223–45.

[68] Manchon, D. 2008 Hopf algebras and renormalisation, in *Handbook of algebra*, Vol. 5 (Ed. M. Hazewinkel), 365–427. Elsevier/North-Holland.

[69] Manchon, D. 2011 A short survey on pre-Lie algebras, *E. Schrödinger Institut Lectures in Math. Phys., "Noncommutative Geometry and Physics: Renormalisation, Motives, Index Thoery", Eur. Math. Soc., A. Carey Ed.*

[70] Mandelbrot, B.B., 1963 The variation of certain speculative prices, *Journal of Business* **36**, 392–417

[71] McCulloch, J., 1996 Financial applications of stable distributions, in *Statistical Methods in Finance, eds. Maddala, G. and Rao, C.*, Volume 14 of *Handbook of Statistics*, Elsevier.

[72] Melançon, G., Reutenauer, C., 1989 Lyndon words, free algebras and shuffles, *Can. J. Math.* **41**(4), 577–591.

[73] Merton, R. 1973 Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science* **4**(1), 141–183.

[74] Merton, R. 1976 Option pricing when underlying stock returns are discontinuous *J. Financial Economics*, **3**, 125–144.

[75] Meyer, P.A. 1967 Intégrales stochastiques I–IV, in *Séminaire de Probabilités I*, Volume 39 of *Lecture Notes in Mathematics*, 95–117. Springer-Verlag, Berlin.

[76] Milstein, G.N. 1995 *Numerical integration of stochastic differential equations*, Kluwer Academic Publishers, 1994.

[77] Milstein, G.N., Repin, M. Yu., Tretyakov, M.V. 2002 Numerical methods for stochastic systems preserving symplectic structure, *SIAM J. Numer. Anal.*, **40**(4), 1583–1604.

[78] Milstein, G.N., Tretyakov M.V. 2004 *Stochastic Numerics for Mathematical Physics*, Springer.

[79] Munthe-Kaas, H.Z., Wright, W.M. 2007 On the Hopf algebraic structure of Lie group integrators, *Foundations of Computational Math.* **8**(2), 227–257.

[80] Murua, A. 2006 The Hopf algebra of rooted trees, free Lie algebras, and Lie series, *Foundations of Computational Math.* **6**(4), 387–426.

[81] Newton, N.J. 1986 An asymptotically efficient difference formula for solving stochastic differential equations, *Stochastics*, **19**, 175–206

[82] Newton, N.J. 1991 Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations, *SIAM J. Appl. Math.* **51**, 542–567.

[83] Nualart, D., Schoutens, W. 2000 Chaotic and predictable representations for Lévy processes, *Stochastic Processes and Their Applications* **90**, 109–122.

[84] Papoulis, D. 1991 *Probability, Random Variables, and Stochastic Processes (3rd ed.)*, New York, etc: McGraw–Hill.

[85] Parzen, E. 1962 *Stochastic Processes*, San Francisco:Holden–Day.

[86] 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.

[87] Platen, E. 1982 A generalized Taylor formula for solutions of stochastic differential equations., *SANKHYA A* **44**(2), 163–172

[88] Platen, E., Wagner, W. 1982 On a Taylor formula for a class of Itô processes., *Probab. Math. Statist.* **3**(1), 37–51.

[89] Protter, P. 2003 *Stochastic Integration and Differential Equations (2nd ed.)*, Berlin, etc: Springer.

[90] Protter, P., Talay D. 1997 The Euler Scheme for Lévy driven Stochastic Differential Equations, *Annals of Probability* **25** No. 1, 393–423

[91] Radford, D.E. 1979 A natural ring basis for the shuffle algebra and an application to group schemes., *Journal of Algebra* **58**, 432–54

[92] Rao, K.M. 1969 Quasimartingales., *Math. Scand.* **24** 79–92.

[93] Ree, R., 1958 Lie elements and an algebra associated with shuffles., *Annals of Mathematics*, **68**, 210–20.

[94] Reutenauer, C. 1993 *Free Lie Algebras*, London Mathematical Society Monographs, Clarendon Press, Oxford.

[95] Rio, E. 2009 Upper bounds for minimal distances in the central limit theorem, *Annales de l'institut Henri Poincaré - Probabilités et Statistiques* **45** No. 3, 802–817

[96] Rosiński, J. 2001 *Series representations of Lévy processes from the perspective of point processes*, in Lévy processes - Theory and Applications, Barndorff-Nielsen, O., Mikosch, T., Resnick, S., eds., Birkhäuser: Boston.

[97] Sato, K.-I., 2013 *Lévy Processes and Infinitely Divisible Distributions (2nd ed.)*, Cambridge Studies in Advanced Mathematics, CUP.

[98] 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. Alg'ebre et théorie des nombres* **12**(1), 1–23.

[99] Strichartz, R.S. 1987 The Campbell–Baker–Hausdorff–Dynkin formula and solutions of differential equations., *J. Funct. Anal.* **72**, 320–345.

[100] Sweedler, M.E., 1969 *Hopf Algebras*, Mathematics lecture notes, W. Benjamin.

[101] Wiktorsson, M. 2001 Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions, *Ann. Appl. Probab.* **11**(2) 470–487.

[102] Zudilin, V.V., 2003 Algebraic relations for multiple zeta values, *Russian Math. Surveys* **58**(1) 3–32.